SOUTHERN AFRICAN INSTITUTE FOR
INDUSTRIAL
ENGINEERING

steering the 4th INDUSTRIAL REVOLUTION

29th SAIIE ANNUAL CONFERENCE

Spier, Stellenbosch
24 - 26 October 2018

PROCEEDINGS

[Intentionally left blank]

Steering the 4th INDUSTRIAL REVOLUTION

29th SAIIE ANNUAL CONFERENCE

24-26 October 2018 Spier, Stellenbosch www.conference.saiie.co.za

Proceedings of the 29th Annual Southern African Institute for Industrial Engineering Conference

24 - 26 October 2018
Spier
Stellenbosch, South Africa

ISSN: 2308-8265
ECSA CPD No: **SAIIE / CPD / I / 03-18**

Editor:
Prof CSL Schutte, Stellenbosch University, South Africa

Contents available on the conference website:www.saiie.co.za/SAIIE29
Published by: Southern African Institute for Industrial Engineering (SAIIE)
- see www.saiie.co.za

[Intentionally left blank]

PREFACE

The Southern African Institute for Industrial Engineering's (SAIIE) annual conference has become a popular choice for industry, academics and researchers in Industrial Engineering and related disciplines. I am excited and honoured to be the editor again for this proceedings.

This year's editorial process was administrated by a very capable individual: Stephan Snyman, a postgraduate student from Stellenbosch University, who acted as track director and technical editor. Thank you for your professional hard work to put the proceedings together and making the editorial process so effortless!

For this conference, prospective speakers were offered the following submission options:

- Presentations submitted for the *"**Industry Presentations**"* or *"**Academic Presentation only**"* track were approved on submission of the abstract only, checking for suitability.
- Submissions for the "***Academic Papers"*** track were provisionally approved on the basis of an abstract, where-after the authors were invited to submit a full-length academic paper, which was reviewed by a double-blind peer review refereeing process, described below.

204 submissions were received of which 90 peer-reviewed papers made it through the review process, with another 4 non-peer-reviewed submissions, and 4 keynotes.

The review process was managed through an on-line conference system, allowing referees to provide on-line feedback, and to ensure that a record exists for all editorial decisions taken during the process. Papers were allocated at least two reviewers, often teaming up an experienced academic, with a less experienced author, so as to follow a true **peer**-review process and also to serve as a learning experience for the less experienced reviewer, without sacrificing the credibility of the peer review process.

Only papers that passed the peer reviewed process are published in the conference proceedings. In addition, each reviewer's feedback was considered, and a rating calculated. The papers with the best ratings (and also after checking for suitability) were selected for the special edition of the South African Journal for Industrial Engineering. 19 Papers were selected for the special edition, and as a consequence, withdrawn from the Proceedings.

This conference has therefore three outputs:

- The printed ***Conference Programme*** includes an abstract of each Peer Reviewed paper, as well as all the other non-peer-reviewed submissions (Presentations, Tutorials, and Invited Presentations), to enable the delegates to plan which sessions to attend.
- The ***Conference Proceedings*** (this document) is an electronic document distributed to all delegates, and contains full-length papers that were submitted, reviewed and approved for the Peer Reviewed Tracks. Its purpose is to give full access to the complete conference material for many years after the conference is over. The proceedings are also available on-line, on a conference website hosted and archived by Stellenbosch University, to ensure that it remains accessible and indexed by scholarly search engines.
- The **Special Edition of the South African Journal of Industrial Engineering** that will appear in early November, honouring the best work submitted to this conference. The Special Edition also contains submissions from other related conferences.

We trust that you will enjoy the 29th Annual SAIIE Conference, and that this publication will serve as a first step for exposing the work of our authors to the world!

Prof Corné Schutte
Editor
October 2018

[Intentionally left blank]

ORGANISING TEAM

ORGANISING COMMITTEE

- **(Chair) Jacques Fauré**, Fortna and SAIIE Council
- **Corné Schutte,** Stellenbosch University, Stellenbosch
- **Lynette Pieterse,** SAIIE and SAJIE Administrator
- **Thereza Botha,** TechnoScene (Pty) Ltd (conference organiser)

EDITORIAL COMMITTEE

- **Corné Schutte,** Editor, Stellenbosch University, South Africa
- **Stephan Snyman,** Track Director and Technical Editor. Stellenbosch University, South Africa

EDITORIAL REVIEW PANEL

The Editorial Committee would like to acknowledge the following reviewers who all contributed to the reviewing process.

- **Stephen Adeyemi**, Vaal University of Technology, South Africa
- **Dr Sanat Agrawal**, Jaypee University of Engineering and Technology, India
- **Prof Alfayo Anyika Alugongo**, Vaal University of Technology, South Africa
- **Adetoye Ayokunle Aribisala**, Federal University, Nigeria
- **Ridge McGibbon Armstrong**, Stellenbosch University, South Africa
- **Wouter Bam**, Stellenbosch University, South Africa
- **Louzanne Bam**, Stellenbosch University, South Africa
- **Toussaint Victor Bamporiki**, Stellenbosch university, South Africa
- **Jean-Claude Munyaka Baraka**, Durban University of Technology, South Africa
- **Prof James Bekker**, Stellenbosch University, South Africa
- **Tshegofatso Bele**, Vaal University of Technology, South Africa
- **Blaise Ebouele Benga Bolan**, Vaal University of Technology, South Africa
- **Henda Besselaar**, Stellenbosch University, South Africa
- **Vivek Bhagwan**, Stellenbosch University, South Africa
- **Walter Booysen**, North-West University, South Africa
- **Prof Stephen Bosman**, Cape Peninsula University of Technology, South Africa
- **Jessica Botes**, Stellenbosch University, South Africa
- **Jozine Botha**, University of Pretoria, South Africa
- **Dr Anthon Botha**, TechnoScene (Pty) Ltd, South Africa
- **Dr Hendrik Gideon Brand**, North-West University, South Africa
- **Jana Breedt**, Eskom Holdings SOC Ltd, South Africa
- **Prof Alan Brent**, Stellenbosch University & Victoria University of Wellington, New Zealand
- **Rhett Desmond Butler**, Stellenbosch University, South Africa
- **Kristy Campbell**, North-West University, South Africa
- **Sylvester Tatenda Chatukuta**, Stellenbosch Univesrity, South Africa
- **Kudzanayi Chiteka**, Harare Institute of Technology, Zimbabwe

- **Charl Cilliers**, North-West University, South Africa
- **Jacques Coetzee**, University of Pretoria, South Africa
- **Rojanette Coetzee**, North-West University, South Africa
- **Pieter Conradie**, Fourier Approach, South Africa
- **Pieter Conradie**, Stellenbosch University, South Africa
- **Dr Oliver Damm**, Stellenbosch University, South Africa
- **Hasan Darwish**, North-West University, South Africa
- **Dr Edward Davies**, Nelson Mandela University, South Africa
- **Imke de Kock**, Stellenbosch University, South Africa
- **Jeanne-Mari de Villiers**, Stellenbosch University, South Africa
- **Dr Marné De Vries**, University of Pretoria, South Africa
- **Mendon Dewa**, Durban University of Technology, South Africa
- **Wesley Emile Dharmalingum**, University of Kwazulu-Natal, South Africa
- **Theuns Dirkse van Schalkwyk**, Stellenbosch University, South Africa
- **Marno Du Plessis**, Stellenbosch University, South Africa
- **Anli Du Preez**, Stellenbosch University, South Africa
- **Willie Du Preez**, Council for Scientific and Industrial Research, South Africa
- **Dr Jacques du Toit**, Spatialedge (Pty) Ltd, South Africa
- **Jarred Durbach**, University of Witwatersrand, South Africa
- **Zandaline Els**, Stellenbosch University, South Africa
- **Prof Neels Fourie**, Stellenbosch University, South Africa
- **Lyndall Fourie**, University of Pretoria, South Africa
- **Matt Simon Futcher**, University of Witwatersrand, South Africa
- **Pieter Goosen**, North-West University, South Africa
- **Dr Andries Gustav Stephanus Gous**, North-West University, South Africa
- **Dr Sara S (Saartjie) Grobbelaar**, Stellenbosch University, South Africa
- **Jacomine Grobler**, University of Pretoria, South Africa
- **Dr Hendrik Johannes Groenewald**, North-West University, South Africa
- **Devon Hagedorn-Hansen**, Stellenbosch University, South Africa
- **Waldt Hamer**, North-West University, South Africa
- **Li Hao**, Northwestern Polytechnical University, China
- **Dieter Hartmann**, Hazaa! Consulting, South Africa
- **Teresa Hattingh**, University of the Witwatersrand, South Africa
- **Mieke Henning**, Stellenbosch University, South Africa
- **Matthew Ray Higgo**, Stellenbosch University, South Africa
- **Biancé Huysamen**, Stellenbosch University, South Africa
- **Dr Christianah Olakitan Ijagbemi**, Tshwane University of Technology, South Africa
- **John Mosoke Ikome**, Vaal University of Technology, South Africa
- **Dr Wyhan Jooste**, Stellenbosch University, South Africa
- **Bronwyn Joubert**, Stellenbosch University, South Africa
- **Lélani Joubert**, Stellenbosch University, South Africa
- **Dr André Joubert**, Vaal University of Technology, South Africa
- **Denzil Kennon**, Stellenbosch University, South Africa
- **Sumarie Koetsier**, Council for Scientific and Industrial Research, South Africa
- **Dr Eduan Kotze**, University of the Free State, South Africa
- **Gerrit Kotze**, Sasol Mining, South Africa

- **Prof Paul Kruger**, University of Pretoria, South Africa
- **Wouter Jacobus Kühn**, Stellenbosch University, South Africa
- **Dr Daniel Petrus Lotter**, Stellenbosch University, South Africa
- **Dr Ann S lourens**, Nelson Mandela University, South Africa
- **Dr Louis Louw**, Stellenbosch University, South Africa
- **Ruan Luies**, North-West University, South Africa
- **Kate Perry Lusiba**, Durban University of Technology, South Africa
- **Prof Eric Lutters**, University of Twente, Netherlands
- **Gideon Jacobus Maasz**, Master Drilling Group, South Africa
- **Whisper Maisiri**, North-West University, South Africa
- **Vusumuzi Malele**, Tshwane University of Technology, South Africa
- **Johannes Bekane Mapokgole**, Vaal University of Technology, South Africa
- **Newton Mapowo**, Stellenbosch University., South Africa
- **John Jostinus Mashala**, South African Production and Inventory Control Society, South Africa
- **Lodrina Masiyazi**, Harare Institute of Technology, Zimbabwe
- **Dr Stephen Matope**, Stellenbosch University, South Africa
- **Dr Andrew Meyer**, Logistic & Mechanical Services, South Africa
- **Samson Mhlanga**, National University of Science and Technology, Zimbabwe
- **Prof Mellet Moll**, Cape Peninsula University of Technology, South Africa
- **Kenneth Moodley**, Unilever, South Africa
- **Mphegolle Ephraim Moshidi**, North-West University, South Africa
- **Zanele Promise Mpanza**, University of Johannesburg, South Africa
- **Prof Khumbulani Mpofu**, Tshwane University of Technology, South Africa
- **Dr Josephine Musango**, Stellenbosch University, South Africa
- **Dr Michael Mutingi**, Namibia University of Science & Technology, Namibia
- **Andrew Kisten Naicker**, Durban University of Technology, South Africa
- **Coenraad Josephus Nel**, Stellenbosch University, South Africa
- **Diaan Nell**, North-West University, South Africa
- **Francis Nturanabo**, Vaal University of Technology, South Africa
- **Dr Hamieda Parker**, University of Cape Town, South Africa
- **Pieter FH Peach**, North-West University, South Africa
- **Dr Jimoh Pedro**, University of the Witwatersrand, South Africa
- **Nalini Sooknanan Pillay**, Eskom Holdings SOC Ltd, South Africa
- **Prof Leon Pretorius**, University of Pretoria, South Africa
- **Jan-Harm Pretorius**, University of Johannesburg, South Africa
- **Jaco Prinsloo**, North-West University, South Africa
- **Prof Kemlall Ramsaroop Ramdass**, UNISA, South Africa
- **Colin Arthur Saffy**, University of the Witwatersrand, South Africa
- **Monique Liana Schoombie**, Council for Scientific and Industrial Research, South Africa
- **Prof Corné Schutte**, Stellenbosch University, South Africa
- **Vishal Shrivastava**, Operations Research Society of South Africa, South Africa
- **Vennan Sibanda**, Tshwane University of Technology, South Africa
- **Kgomotso Simango**, North-West University, South Africa
- **Akshay Pradeep Singh**, University of KwaZulu Natal, South Africa
- **Stephan Snyman**, Stellenbosch University, South Africa

- **Henrietta Steenkamp**, University of Johannesburg, South Africa
- **Prof Herman Steyn**, University of Pretoria, South Africa
- **Prof Arnesh Telukdarie**, University of Johannesburg, South Africa
- **Prof Thomas Bobga Tengen**, Vaal University of technology, South Africa
- **Prof Fanie Terblanche**, North-West University, South Africa
- **Victoria Thomas**, Stellenbosch University, South Africa
- **Prof Andre Francois Van Der Merwe**, Stellenbosch University, South Africa
- **Dr Karl Robert van der Merwe**, Nelson Mandela University, South Africa
- **Johannes Hendrik van der Merwe**, University of the Witwatersrand, South Africa
- **Prof Liezl Van Dyk**, North-West University, South Africa
- **Prof Wynand Van Dyk**, Stellenbosch University, South Africa
- **Dr Joubert van Eeden**, Stellenbosch University, South Africa
- **Quintin Van Heerden**, Council for Scientific and Industrial Research, South Africa
- **Dr Stephan van Jaarsveld**, North-West University, South Africa
- **Nina Louise van Rooyen**, Pragma, South Africa
- **Dr Chris Van Schoor**, Institute for Business Efficiency, South Africa
- **Prof Jan H van Vuuren**, Stellenbosch University, South Africa
- **Dr Brian John Van Vuuren**, Stellenbosch University, South Africa
- **Maria Van Zyl**, North-West University, South Africa
- **Ilse Van Zyl**, Stellenbosch University, South Africa
- **Nadia Millis Viljoen**, Council for Scientific and Industrial Research, South Africa
- **Prof Jacobus Krige Visser**, University of Pretoria, South Africa
- **Prof PJ Vlok**, Stellenbosch University, South Africa
- **Konrad von Leipzig**, Stellenbosch University, South Africa
- **Marisa Walters**, Stellenbosch University, South Africa
- **Dr Dafeng Xu**, University of Pretoria, South Africa
- **Prof Sarma Venkata Seshachala Yadavalli**, University of Pretoria, South Africa
- **Philani Nduna Zincume**, Stellenbosch University, South Africa

ACKNOWLEDGEMENTS

SAIIE29 would like to thank the following sponsors, partners and exhibitors for their generous support

ACADEMIC PARTNER	Universiteit iYunivesithi Stellenbosch University 100 1918-2018; Engineering Ezobunjineli Ingenieurswese
PLATINUM SPONSOR	Universiteit van Pretoria University of Pretoria Yunibesithi ya Pretoria; Graduate School of Technology Management
GOLD SPONSOR	MMI HOLDINGS
BRONZE SPONSOR	University of the Witwatersrand, Johannesburg
BAG SPONSOR	UNISA college of science, engineering and technology
EXHIBITORS	BLUESTALLION technologies; pps for professionals since 1941

CORPORATE PARTNERS	DSV Global Transport and Logistics MMI HOLDINGS PiLog Tshwane University of Technology *We empower people*
ASSOCIATION PARTNER	SAPICS your supply chain community

Table of Contents

[1] The Reference Number is a unique paper reference that is used throughout the conference to identify the paper. This is also used as a page number prefix in the proceedings, and papers are sorted according to this number in the proceedings.

Alphabetic Author Index

[2] The Reference Number is a unique paper reference that is used throughout the conference to identify the paper. This is also used as a page number prefix in the proceedings, and papers are sorted according to this number in the proceedings.

[Intentionally left blank]

DISASTER RELIEF CHAINS ANALYSIS IN THE SADC REGION

J.C.M Baraka[1*], S.S. Yadavalli[2*] & M. Dewa[3*]

[1]Department of Industrial Engineering
Durban University of Technology, Durban, South Africa
20618160@dut4life.ac.za

[2]Department of Industrial & System Engineering
University of Pretoria, Pretoria, South Africa
Sarma.yadavalli@up.ac.za

[3]Department of Industrial Engineering
Durban University of Technology, Durban, South Africa
mendond@dut.ac.za

ABSTRACT

SADC region has been ravaged, for decades by severe man-made and natural disasters killing over 120 thousand people while affecting millions other. Disaster relief chains represent, among many others, the biggest challenge for humanitarian agencies in the region. This paper intends to present the outcome of a survey developed, targeting disaster management agencies operating in SADC. The detailed analysis associated with a transportation problem approach assisted the research investigates the region disaster relief chains readiness. The survey focused on the supply chain planning, the transportation modes decision approach, the natural of relief`s supports as well as the coordination between relief agencies. The outcome showcased the region lacks and potentials in disaster emergency response and lives saving operations.

[1] The author was enrolled for an D Eng (Doctorate in Engineering) degree in the Department of Industrial Engineering, Durban University of Technology (DUT)
[2] The author is a Professor(Doctorate Engineering) at the Department of Industrial and System Engineering, University of Pretoria
*Corresponding author

1. INTRODUCTION

'The challenge for relief agencies responding to global emergencies such as volcanic eruptions, earthquakes, floods or wars is how to prepare and manage relief activities especially in an unpredictable environment' [7] . Although science has made significant advances in predicting disasters, without preparedness, any humanitarian organization risks underachieving their goals.

Natural and man-made disasters are documented for causing loss of life and environmental damage throughout SADC. According to the Centre of research on the Epidemiology of Disasters (CRED) [8], 'earthquakes, floods, landslides, and extreme weather conditions made 2010 the deadliest year for disaster in the past two decades. Their report suggests that some 373 natural disasters killed almost 300 thousand people in 2010, affecting nearly 207 million others, and costing nearly US$ 10 billion in economic losses' ([17]; [9]).

Problems deriving from the disaster relief operations differ depending on various factors [4]. Among the key factors are the type of disasters, the disaster impacts on the region, and location of the disaster, or infrastructure availabilities in the affected areas, etc. [14] claim a lack of coordination between organizations that plan, design, construct, operate, maintain, and manage complex infrastructures and humanitarian organizations. "*Both, structure-based (hardware, structures, facilities) and human-based (institutions, organizations, culture and language) infrastructures have the potential to disrupt a bigger part of the humanitarian effort system when attacked or hit by a natural disaster*" [2]. Although disaster relief chains problems are documented by several studies conducted on the topic [4], insufficient available structures and resources for supply chain management activities and techniques in many SADC countries have hindered collective relief solutions (Bilal, 2010). For instance, logisticians in rescued organizations have to deal with various types of difficult and complex tasks that produce delays and ultimately loss of lives and resources (Bilal, 2010). One such situation is the high risk and low reliability infrastructures that have a direct effects on deliveries, while impacting human lives and health' [16].

The survey conducted in this study aim at investigating SADC relief supply operations readiness, using appropriate transportation mode, with suitable coordination between relief agencies and their use of all relevant parties. The survey targeted relief organizations operating in the region. The paper will interpret the survey responses, and then present detail outcomes. Questions are grouped into categories and the feedback from each category of questions is balanced with the survey outcome and the peer reviewed data available in order to identify the region readiness to each of the categories.

2. LITERATURE REVIEW

[28] defines disaster relief chains as 'a process and a system involved in mobilizing people, resources, skills and knowledge to help vulnerable people affected by natural disasters and complex emergencies". The function of disaster relief chain embraces all range of activities, including preparedness, planning, procurement, transport, warehousing, tracking and tracing, customs and clearance' [28].

According to [7], regardless of the relief agencies operations, 'each shares the common critical objective of rapidly delivering the correct amount of goods, people and monetary resources to locations in need. However, two pillars for supply chain and logistic are capable of directly affecting the success of a relief effort, those pillars are availability of infrastructures and transportation.

After the Lisbon atrocity, Jean-Jacques Rousseau (Genevan philosopher, writer, and composer of the 18th century) wrote in a letter to Voltaire in 1756 stating that natural disaster does not just happen, but is a result to bad planning and lack of infrastructure (as cited by [12], from Masters and Kelly, 1990). [25] further stated that high-income areas such as Europe, North America, or Japan are as highly exposed to natural disasters as are low-income areas in Africa and Asia. While the effects of earthquakes and floods in low income areas are often magnified, due to poor housing situations, roads, insufficient hospitals or inadequate construction requirements [21]. A real economical breakthrough is unlikely to the achieved until significant investments on infrastructure are made, allowing the region to effectively implement the low cost disaster relief plans.

Transportation is considered the cornerstone of the last mile distribution [4], 2008) and the second largest overhead cost to humanitarian organization after personnel [32], [11]. Relief transportation is defined as the movement of goods from origins to destinations using one or more transportation modes such as air, ocean, rail, and road [6]. With both natural and man-made disasters being forecast to increase five-fold over the next 50 years [29], there is therefore need to develop an efficient and effective transportation system to assist rescue operations. To access affected people rapidly and cost-effectively, humanitarian organization such as Fritz Institute begun the initial steps of developing models that improve the efficiency and effectiveness of humanitarian supply chains [24]. **[19]** first publish a transportation problem modeling a large amounts involving many different commodities, such as food, clothing, medicine supplies, machinery and personnel.

[15] then argued above the need for transportation and infrastructure, the need for a prepositioned are for disaster supply chain rather than dealing with logistical response. For such argument to benefit the SADC region,

[33] dealing the importance of collaboration between humanitarian organizations and disaster operation parties in order to alleviate the administrative and logistical bottlenecks. Coordination enhances disaster relief chains capacity especially during the delivery and distribution of relief cargoes to the needy in conflict zones [33].

3. RESEARCH METHODOLOGY

The research intends to incorporate the survey feedback; the peer reviews documentation as well as the available declassified information on disaster agencies operating in the region. The survey was drafted with aim of filling the existing information gaps in regards to the current relief organization operation in the region. Then, the study opted for quantitative approach for obtaining a fist hand knowledge on the region disaster relief decision making in the area of preparedness, dealing with infrastructure quality, dealing with available transportation system, transportation modes decision, criteria used in transportation modes selection, etc. As stated by [5], additional information in humanitarian efforts are useful in order to design of an optimal routing plan for first aid material, food, equipment, and rescue personnel from supply points to a large number of destination nodes from geographically scattered areas.

3.1. Target Population

The study has targeted all relief organization with representative operating in the region. SADC region constitutes of 15 countries with the population surrounding 300 million people [23] in which nearly half (144 million) has been affected by natural disaster and 120 thousands by man-made disaster between1980-2016. According to the UN-Habitat publication of cities highly vulnerable to man-made and natural disasters [31], disasters accounted for over US$667 billion in material loss between 1995 and 2005.

Three population's types are targeted in the survey namely: Governmental Organizations, Non-Governmental Organizations and Intergovernmental Organizations. All three population groups are represented by belonging disasters agencies and related entities responsible for it data collection from the SADC countries.

3.2. Data Analysis

3.2.1. Disaster Relief Chains framework

Disaster relief chains involve different but equally vital stages and simultaneous activities. Figure 1 shows the relationship of each of those activities and the relief process from the supplier to destination (consumer). The framework below also illustrates the SADC process this research is intending to investigate and recommend an improvement.

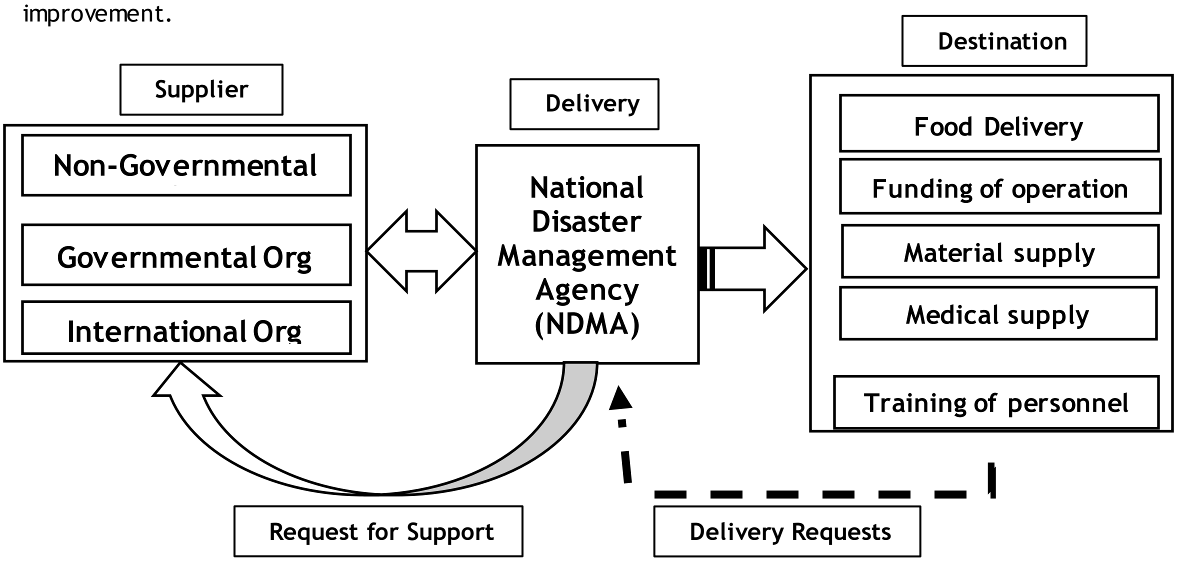


Figure 1: Disaster Relief Chains framework

3.2.2. Analysis of the questionnaire

The research questions are utilised as a primary source of data collection, databases from investigated relief organizations and reviewed literatures served the research for clarification, validation, investigation and elicitation.

Four categories of questions are used in this study and each of the categories intends to determine the impacts disaster relief chains have on the overhaul readiness and people livelihoods. Figure 2 illustrates the sequence of questioning used from category A to D.

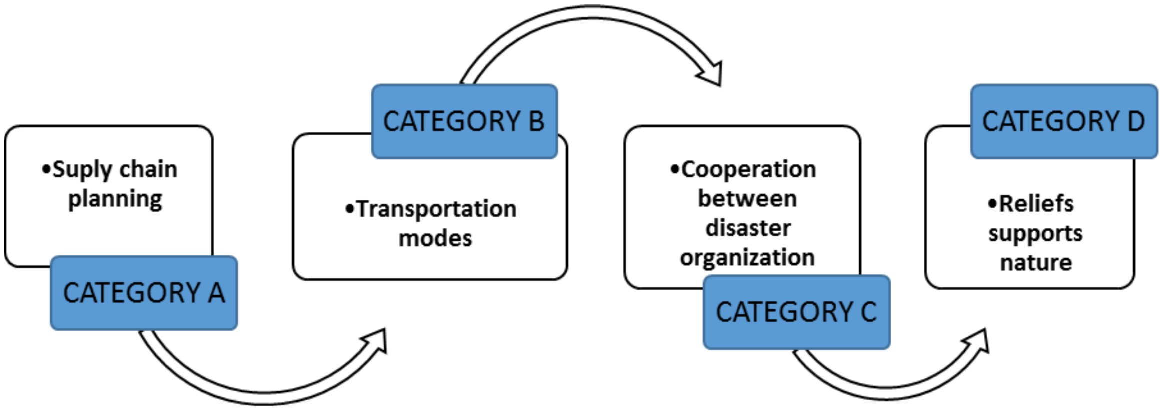


Figure 2: categories of questions

Category A: questions from this category aim at bringing clarity on the type of supplies, available facilities for delivery of reliefs. Questions from this category also deal with systems to be implemented to facilitate the existing processes.

Category B: This category looks at the humanitarian relief chain current decision making practices on transportation modes, and investigates ways of improvement.
Category C: This category intends to investigate the cooperation and coordination between disaster organization and other relevant entities.

Category D: This category of questions tends to illicit all factors needed for an effective and safe humanitarian transportation activity.

3.2.2.1. Sample Analysis
In total, 60 questionnaires were despatched, 48 were returned (80% response rate). The data collected from the survey responses were analysed using with SPSS version 21.0. The results present the descriptive statistics in the form of graphs, cross tabulations and other figures for the quantitative data that was collected.
Reliability is computed by taking several measurements on the same subjects. A reliability coefficient of 0.70 or higher is generally considered as "acceptable". The research reliability score exceeds the recommended value of 0.700. This indicates a high (overall) degree of "acceptable", consistent scoring for question of the research.

3.3. Transportation Problems

On top of the statistical analysis described above, the study reviews a suitable transportation problem. According to [34], transportation problems require the determination of a minimum cost transportation plan for a single commodity from a number of sources to a number of destinations. [22] expressed the transportation problem in mathematical language. The mathematical language of the transportation problem is as follows:

$$\min z_q = \sum_{i=1}^{m}\sum_{j=1}^{n} C_{ij}^{q} X_{ij}, q = 1,2 \tag{1}$$

Subject to:

$$\sum_{j=1}^{n} X_{ij} = a_i, i = 1,2,3,4,5,6,7, \dots, m \tag{2}$$

$$\sum_{i=1}^{m} X_{ij} = b_j,\ j = 1,2,3,4,5,6,7, \dots, n \tag{3}$$

$$X_{ij} \geq 0, \qquad \forall i,j \tag{4}$$

With C_{ij}(t) as the unit cost of transportation from supply point m $i(i = 1,2,\dots,m)$ to demand point n $j(j = 1,2,\dots,n)$ when the duration allowed for the transportation is t units and the destination q units. Now, setting X_{ij} to be the variable denoting the amount transported from supply point i to demand point j, let a_i be the availability at source points i and b_j the demand at demand point j.

4. RESULTS

This section presents the results and discusses the findings from the survey and the reviewed of literature. The feedback from the 48 returned questions (80% response rate) are first presented through summarised percentages for the variables that constitute each category. Then further analyses are done for statistical purposes. Each category responses is summarised in a table, then a graphs analysis are followed.

4.1. Supply Chain Planning (Category A)

Table 1 summarizes the responses from category A questions.

Table 1: Category A summarised responses

Organizational roles	Supply Chain Challenges	Supply Chain Encounter	Supply Chain Strategy	Relief Quantity
• Funding Operations • Transportation of Relief supplies	• Safety • Road Conditions	• Poor condition of Infrastructure • Long wait at the bordergate	• smaller vehicles • Prepositing location	• Population living in the area • Transport as much as possible

4.1.1. Organizational roles

Figure 3 is the feedback from the following question: *"In what sections does your organization plays a humanitarian role?"* 45% of respondent's answers revealed that their activities involve the transportation of relief supplies, 30% fund local agencies for operations. The recent episode of drought in SADC has shown the importance of partners in assisting financially starved areas in the region. According to the Regional Situation Update on the El Niño-Induced Drought (2016), 74% of the required Regional Humanitarian Appeal of $2.9 billion is yet to be raised as only $757 million has been raised by governments ($222m) and partners ($535m).

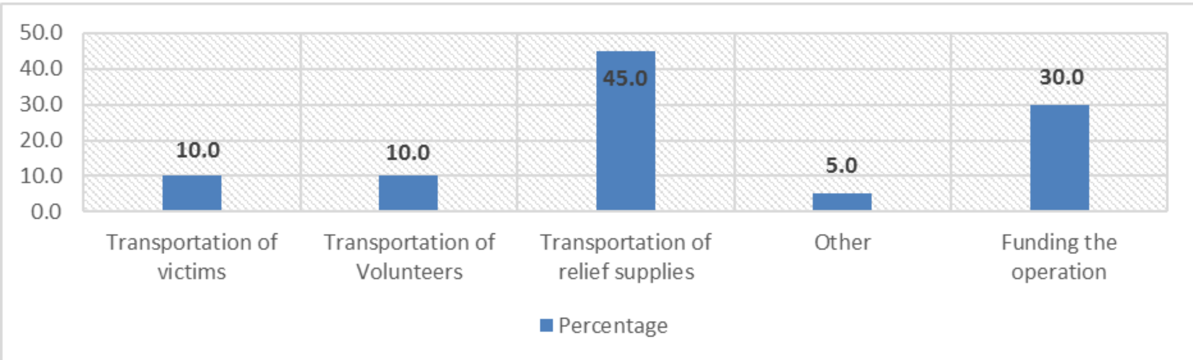


Figure 3: Organizational role during disaster relief operation

4.1.2. Supply Chain challenges

Figure 4 is the feedback from the enquiry on the *"Challenges faced by humanitarian organizations during supply chain."* Among the 48 respondents, 25 % sees safety of the supplies to destinations as the main concern during a humanitarian supply chain especially in the aftermath of an earthquake, flood or a cyclone. While another 25% consider instead the road condition as the biggest challenge especially during last mile distributions. Although roads in SADC are generally in good condition, infrastructures are often affected during violent natural disaster such as volcanoes or earthquakes.

Another relevant aspect of humanitarian organizations revealed in this enquiry is the importance of "cost" during transportation in the region. Although transportation costs in SADC are cheaper than in other African region (with air transportation costing between US$ 4000 and US$ 6000, road transportation between US$ 120 and US$ 280 and train between US$100 and US320), it remains high in comparison to the world standard. Transportation costs in the SADC are hugely affected by the delays at border crossings, weighbridges, ports, and lengthy customs processes. Simulation done in the SADC led to the suggestion that reduction on the border delays could reduce transport costs to over US$ 200 a day per trip, which represent a loss of US$ 120 million per year in the region, given the current traffic volumes.

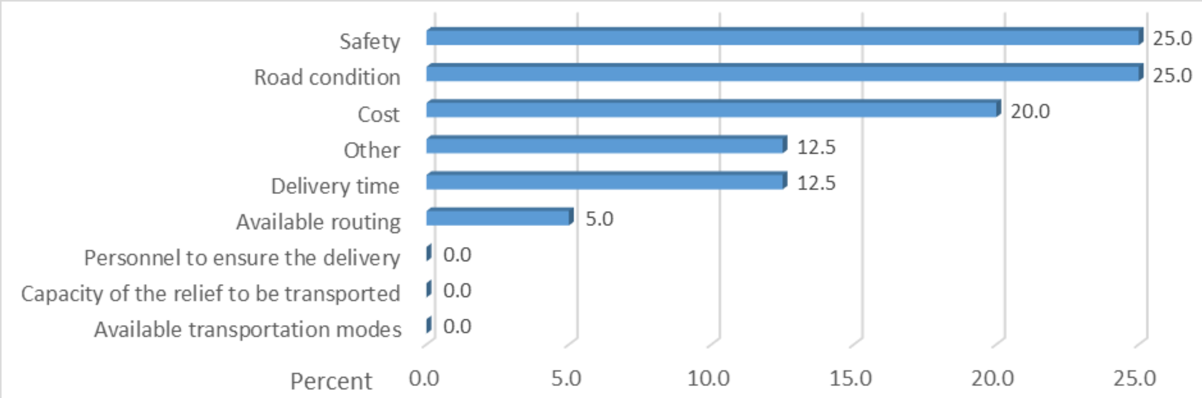


Figure 4: Supply Chain challenges

4.1.3. Supply Chain Encounter

Figure 5 is the feedback from the following question: *"What type of difficulties that disaster organization encounters during the supply of the relief?"* From this question responses, as many as 40 % sees poor condition of infrastructures the main difficulty affecting the supply of reliefs. Except South Africa and few other SADC countries, the region has struggled with deteriorating railways, connecting roads, also limited airports access to affected zones. Many regional countries also suffer from lack or poor health care facilities which force organization to transfer critical victims to the nearest medical facilities. Although in the last decades the region has seen large investment into infrastructures, significant work is still needed to close the gap with developed countries.

Another supply chain difficulty raised from the responses is the long wait at the border gate. Close look at two active borders in SADC namely Beit Bridge (a Border between South Africa and Zimbabwe) and Chirundu (a border post between Zambia and Zimbabwe), Beit Bridge handles as many as 500 trucks a day; delays for northbound traffic are 34 hours and for southbound traffic 11 hours [10] . Estimation for Chirundu border indicates that it takes northbound traffic approximately 39 hours to cross the border and Southbound traffic 14 hours. According to [10], "the northbound delays are significantly more than the southbound delays at the Chirundu border because the Zambian Revenue Authority takes on average 17.4 hours to process documents and inspect loads. Idle time for northbound loads is approximately 10.9 hours per trip due to delays in documents handling upon truck arrival, border clearance, and driving".

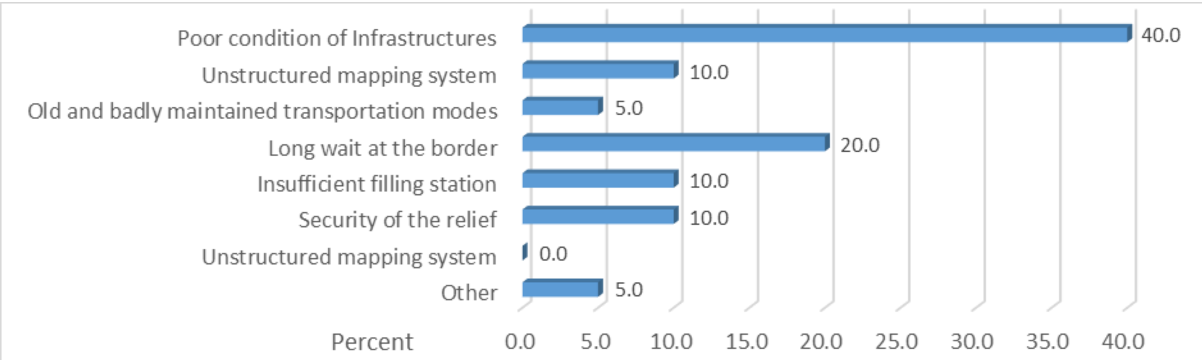


Figure 5: Supply Chain Encounter

4.1.4. Supply Chain Strategy

Figure 7 is the feedback from the following question: *"What strategy do you use to ensure that reliefs are supplied quicker and cost effectively?"* Respondent's answer show that 40% rely on smaller vehicles such as (4x4 or mini buses) instead of trucks as strategy to supply quickly to needy areas especially in last miles distribution. 4x4 offers the organizers practical advantages due to its flexibility especially manoeuvring around poor road conditions. Although with 4x4, more trip are needed in order for the supply chain objectives to be achieved. 30% believes preposition reliefs to strategic locations ensure that goods are available immediately when need arise.

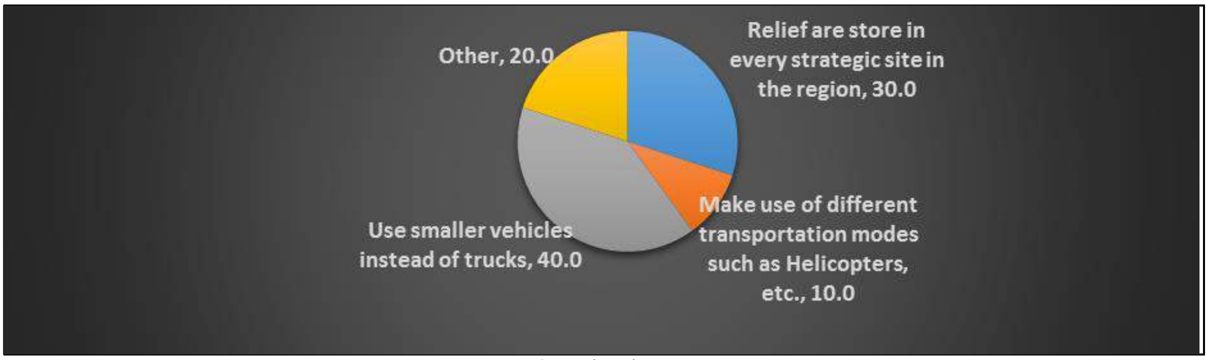


Figure 6: Supply chain strategy

4.1.5. Relief Quantity Estimation

Figure 7 is the feedback from the following question: *"During a disaster, how do you determine the total number affected by disaster in relation to the number of reliefs?"* From responses collected, 30% of respondents determine the total affected numbers by estimating the population size of the affected zones. 27 % of respondents prefer transporting available quantity. Only 18 % investigates first with accuracy the number of victims, then send the relief accordingly. 25 % of organization that found this question no applicable represents to their operations. Their mission could instead be financing or coordinating operations.

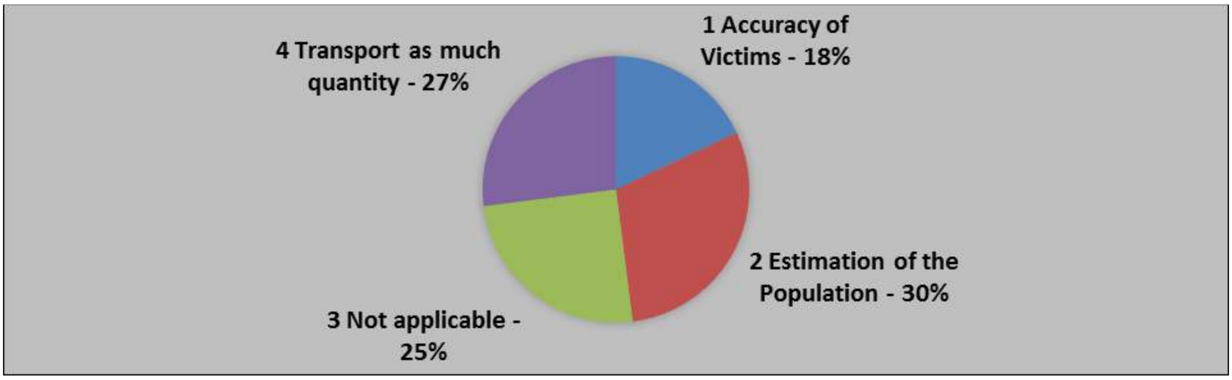


Figure 7: Relief Quantity Estimation

4.2. Transportation Modes (Category B)

The Table below (Table 2) summarizes the responses from category B questions.

Table 2: Category B summarised responses

Change of Transportation Modes	Choice of Transportation modes	Transportation of Volunteers/staffs	Transportation modes used	Transportation of Evacuees
•Yes	•Trucks •Helicopters	•Yes	•4x4 or 4x2 Vehicles •Mini Buses	•Buses •Military Vehicles

4.2.1. Change of transportation modes

Figure 8 is the feedback from the following question: *"During an entire trip, do you ever transfer the supplies from one to another (e.g. Ship to box truck, train to truck, etc.)?"* According to the Graph 4.6, most disaster organizations change two or more transportation modes during the transportation of foods, materials, construction equipment or other from point origins to destinations. For SADC island (Madagascar, Seychelles and Mauritius), it is logistically necessary to use at least two modes of transportations in order to reach the affected zones from other SADC countries. Similarly, landlocked countries need either to use the air transportation from the country airport or utilize the closest ports to the affected zones. Trucks, Mini buses, Vehicles, or helicopters are often used for last miles distribution while train, cargo airplanes as well as cargo ships are usually needed for intermodal transportation.

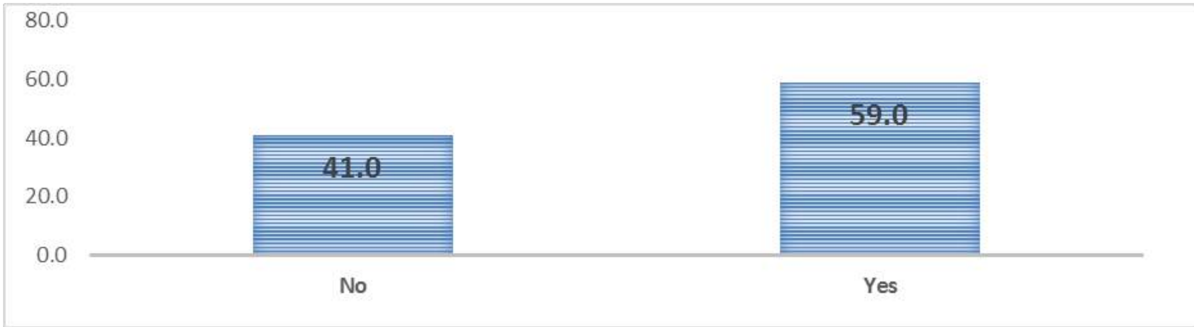


Figure 8: Change of transportation modes

4.2.2. Choice of transportation modes

Figure 9 shows the feedback from the following: *"When delivering supplies, how often does your organization use each of the following transportation modes?"* According to the feedback received, the most common mode of transportation in SADC is trucks (62.5%). According to [27], the Southern African Corridors has the highest volume per kilometre of goods than other regions on the African Continent. Findings from the World Bank Africa region Sustainable Development Unit confirm that the road freight tariff is at 5 cents per tonne-km in SADC, the lowest on the continent; goods however, move faster in SADC than any other areas in the continent due to its effective velocity of 18.6 kilometres per hour. From Durban to Lusaka for instance, the price and performance levels approach those of developed countries. Other transportation modes mentioned include Helicopters (37.0%) and Buses (37.5%) victims and personnel's.

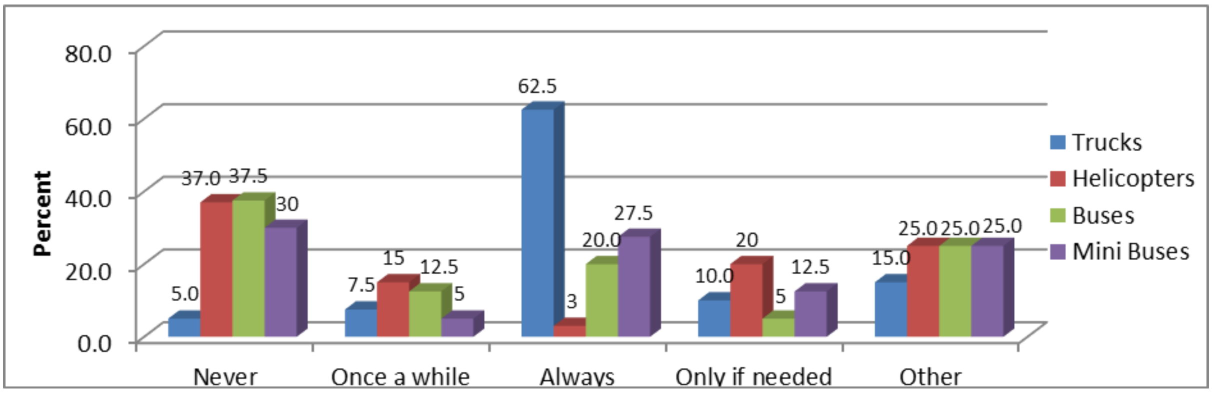


Figure 9: Choice of transportation modes

4.2.3. Transportation of Volunteers/staffs

Figure 10 shows the feedback from the following: *"Does your organization transport volunteers/staff to the affected areas?"* More humanitarian organizations in the region transport volunteers/staffs to and from the affected areas.

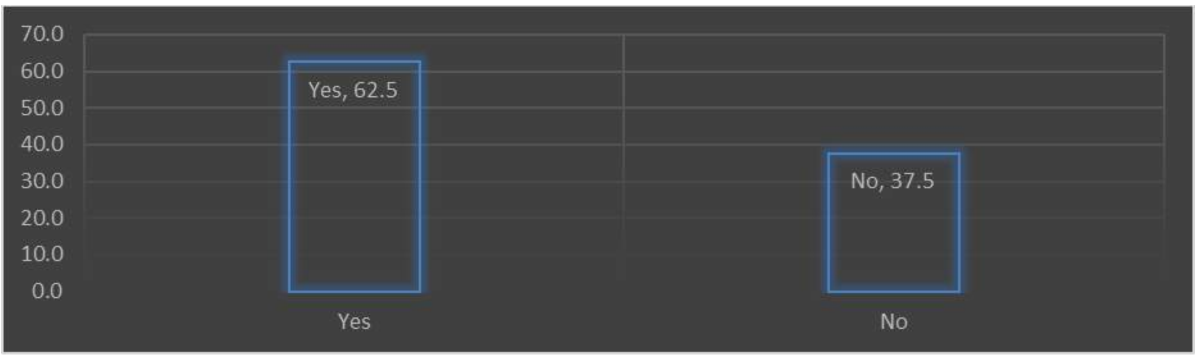


Figure 10: Transportation of Volunteers

4.2.4. Transportation modes decisions

Figure 11 shows the feedback from the following: "*When transporting volunteers/staffs to the affected areas, what transportation modes do you organizations use?*" For transporting volunteers/staffs to the affected areas in the regions, the majority of disaster organizations use 4x4 or 4x2 vehicles due to their flexibility and accessibility even in the poorly maintain roads. Mini buses and buses are used on accessible roads.

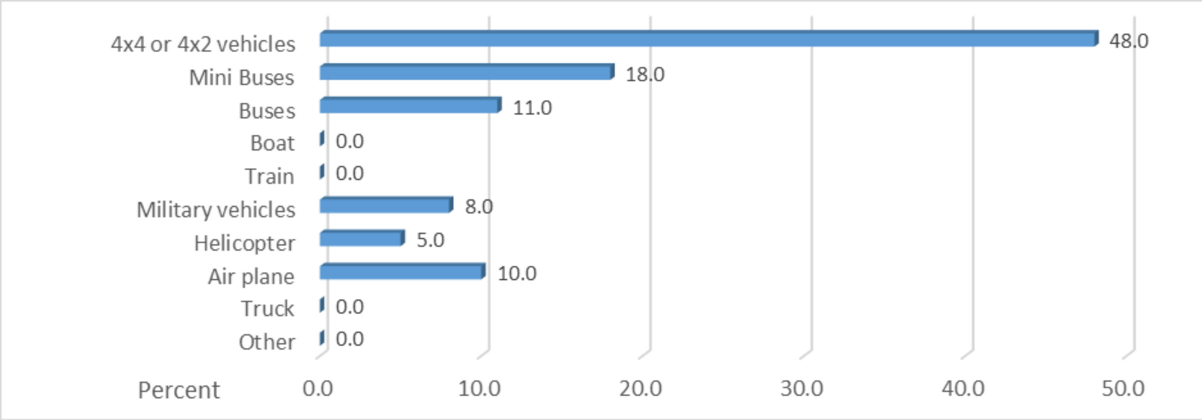


Figure 11: Transportation modes decisions

4.2.5. Transportation modes for Evacuees

Figure 12 shows the feedback from the following question: *"When evacuating people out of the affected areas, what transportation modes do you utilize often?"* The majority of respondents use the mini buses option for the evacuation of people out of the affected areas followed by military vehicles and helicopters. Depending on road condition and on the type of disasters, military vehicles and helicopters were selected as necessary in order to maximise lifesaving.

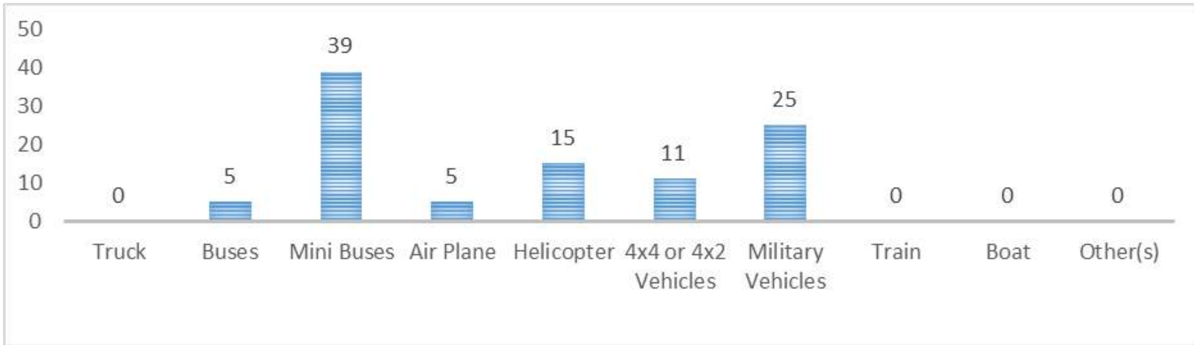


Figure 12: Transportation modes for Evacuees

4.3. Cooperation between Relief Organizations (Category C)

The Table below (Table 3) summarizes the responses from category C questions.

Table 3: Category C summarised responses

Type of Organization	Sharing of Loads
•Non-Governmental Organization •International Organization •Governmental organization	•Never •Only if needed

4.3.1. Organization involved

Figure13 shows the feedback from the following question: *"What kind of organization would you classify your organization to be?"* 56% of organizations involve in this study are non-governmental organization (56%), followed by 31% of Governmental agencies (31%) and International Organization representing 13%.

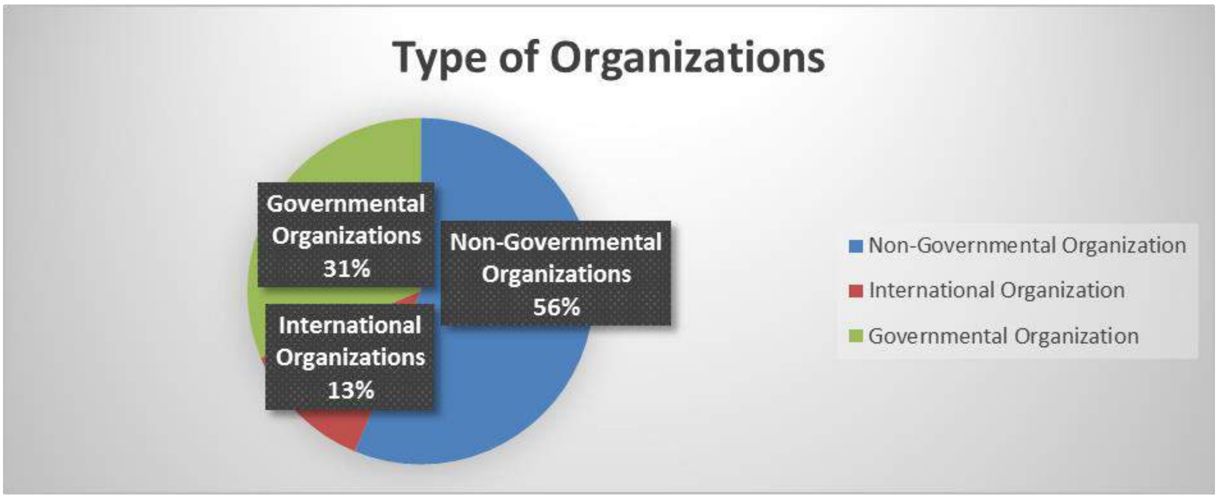


Figure13: Type of Organizations

4.3.2. Sharing of Loads

Figure 14 shows the feedback from the following question: *"How often does your organization share loads when cooperating with other organizations during the transportation of reliefs?"* The majority humanitarian organizations never share load with other organizations during the transportation of reliefs. While a small portion only share load if it is needed. Among participants, 20 % do not consider this question as relevant to their activities, 30 % of the participants had no information to share. Nevertheless, 50 % suggest the importance of policies to encourage disaster organizations coordinate their efforts in order to maximize the delivery of re lief and save lives. According to [21], collaboration brings both benefits and numerous challenges to organizations. The authors suggest developing an effectiveness paradigm that focuses on value creation, knowledge sharing and collaboration. Instead of competing for donors, improving relationships enables humanitarian organizations manage complexities and ambiguities, favouring dialogue, coordination and learning [21]. For [3]; [18], coordination resolves the geographical dispersion and the insufficient or inaccurate communication between the field and the head offices of most agencies, also between different stakeholders. Sharing loads could help organizations avoid exceeding their authorities and act in a controlling or domineering manner [13]. For functionally structured and decentralized aids organizations, a formal networks need to be create in order to improve their regional responses [20]. Finally, sharing loads raise the urgency to a structured logistical resource allocations instead of organization headquarters being more concerned with relationships with donors [8].

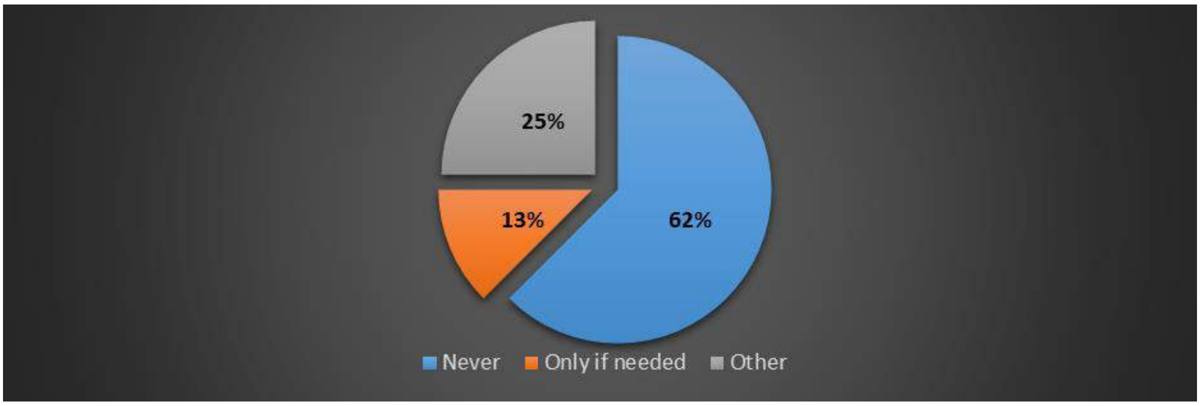


Figure14: Sharing of Loads.

4.4. Relief support nature (Category D)

The Table below (Table 4) summarizes the responses from category D questions.

Table 4: Category C summarised responses

Assistance offers	Decision making factors	Training of Volunteers/Staffs	Type of training provided
•Foods •Materials	•Cost •Infrastructure	•Distribution of reliefs •Inventory control	•Inventory control •Evacuation of Victims

4.4.1. Assistance offers to affected communities

Figure 15 shows the feedback from the following questions: "What type of help that your organization offers?" The majority of targeted disaster organizations offer foods to the victims (35 %), 23 % of the respondents offer materials such as tents, clothing's, blankets, plates, etc., while 17% prefer offering financial assistance to local communities in the aftermath of a disaster. A good number of those organizations offer training to volunteers, personnel and officials or logistical supports.

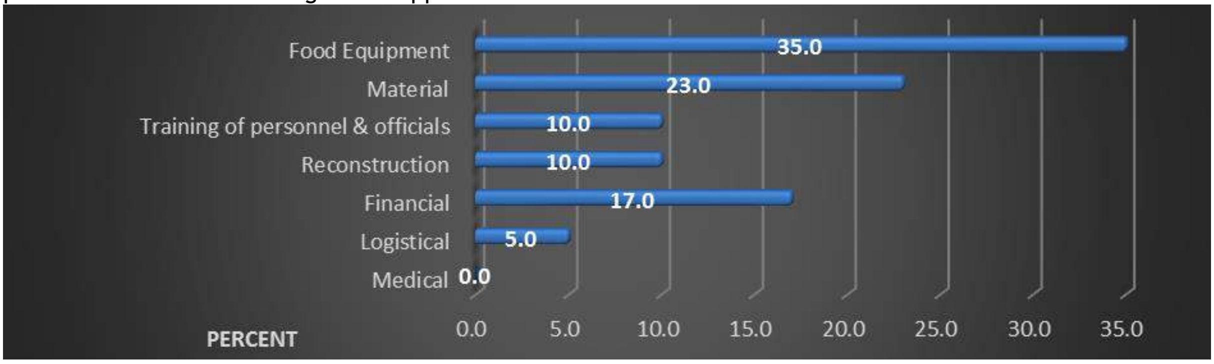


Figure15: Assistance offers to affected communities

4.4.2. Decision making factors

Figure 16 shows the feedback from the following question: *"In transferring supplies from one destination to another, what factors are generally included in your decisions?"* Cost is the major decision factor for disaster organizations during the transfer of relief from one destination to another (25%). Although SADC transportation modes and infrastructures are generally in good condition, it is up to the decision makers to opt for either the cost effective transportation modes with increased delivery time or the one that minimize the time while using the fastest way to reach destination. Another cost factor worth more emphasis is the border crossing cost. Although already stated at point 4.1.3, the journey from Lusaka to Durban takes over eight days to complete, with almost four days spent at border crossings. The cost of delays for an eight-axle interlink truck has been estimated to be around US$ 300 per day; given traffic volumes, this representsa loss of more than US$ 50 million annually' ([26]; [10] ; [1]). Another decision making factors include security, infrastructures (Hospital, road, communication, railways, Airways, Buildings, etc.) and distance are also important factors in decision making during humanitarian operations.

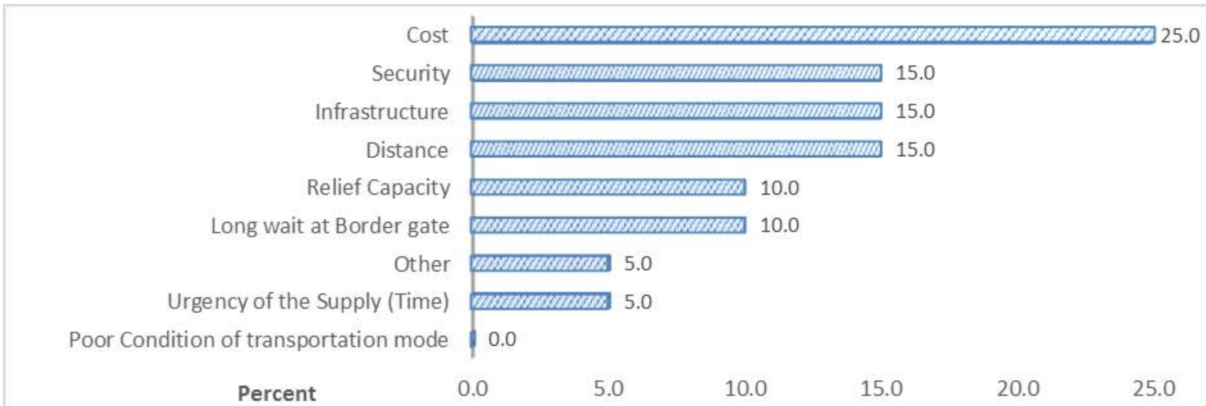


Figure 16: Decisions making factors

4.4.3. Training of staff and volunteers

Figure 17 shows the feedback from the following: *"Does your organization train volunteers/staff to the disaster prevention and response?"* Half of the organizations in the region trained the volunteers for disaster preventions and responses. Communities around SADC are often not prepared to respond effectively to disaster cases. [30] conducted a survey determining the logistic skill level for humanitarian aid organizations, the feedback revealed that the majority of about 300 logisticians surveyed indicated that they typically get their education and training on the job by co-workers. The inexperienced and unskilled logisticians in the field represent a threat to an effective disaster response.

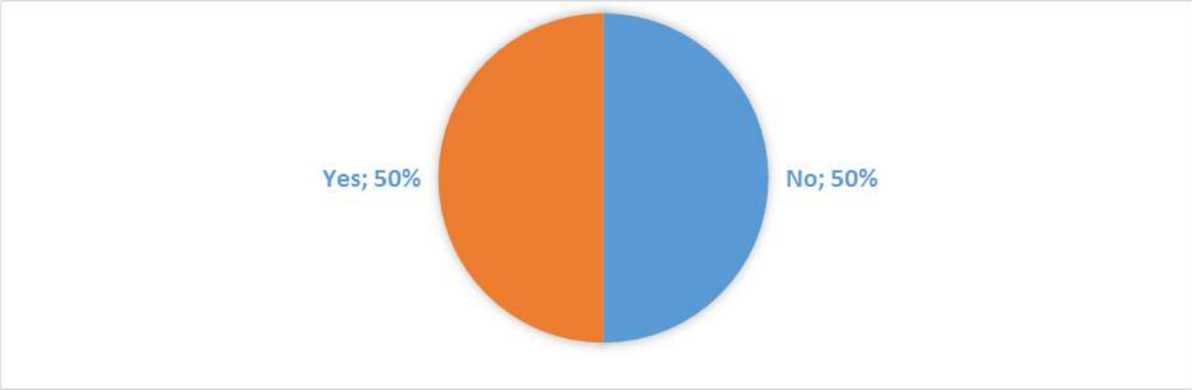


Figure 17: Training of Volunteers/staff

4.4.4. Type of Training Provided

Figure 18 shows the feedback from the enquiry about the *"Type of training provided"*. Most aid organizations view evacuation of victims as the main training priority during a humanitarian disaster. This is followed by the distribution of reliefs (medical supply, foods, equipment, etc.) to affected population.

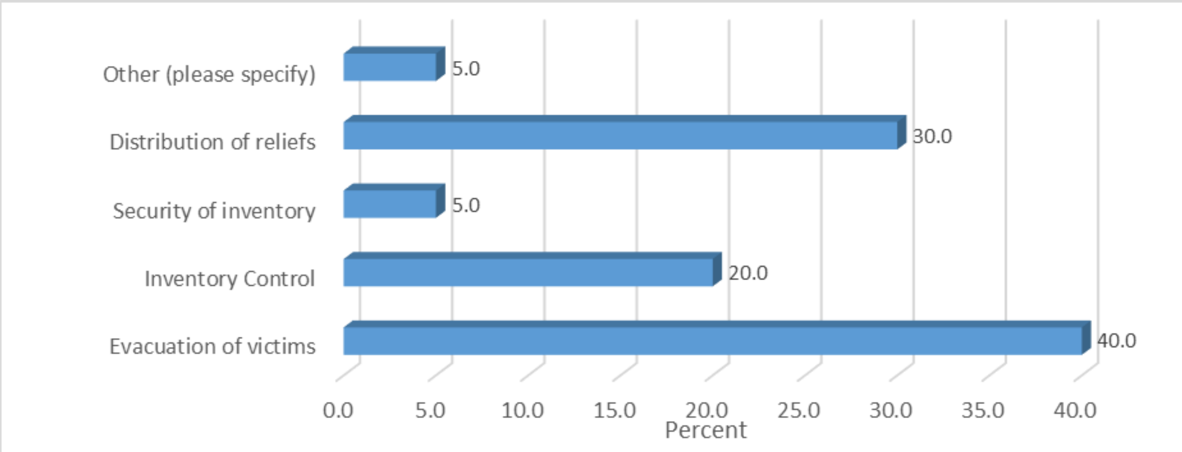


Figure 18: Type of Training provided

4. CONCLUSION

With the growing threat of natural and man-made disasters, it is imperative for disaster organizations operating the region to continually share their knowledge and experiences in order to be more effective in their humanitarian operation quests. This study exposed the existing gaps in SADC disaster relief chain operation and offered agencies key decision making options. To investigate the current humanitarian regional affairs, a survey was developed, the questionnaires was divided into four categories of questions (enquiries). And the responses highlighted critical issues affecting the current disaster relief chains situation in the region.

Among the research findings, an overwhelming number of respondents agreed that disaster agencies in SADC needs an ongoing operation and information coordination mechanism to alleviate the cost, the time as well as the needed resources to saves lives. There is also an urgent need to resume or perhaps accelerate investment into infrastructure in the region. Such development not only improves the region connectivity and livelihood but also reduces the time and the cost of disaster operations. A strategic investment into infrastructure and communication will not only boast the region economy but also minimize the disaster costs.

REFERENCES

[1] **Africa Infrastructure Country Diagnostic (AICD)**. 2013. World Bank website. http://data.worldbank.org/data-catalog/africa-infrastructure.

[2] Altay, N., and Green III, W. G. 2006. OR/MS research in disaster operations management. *European Journal of Operational Research*, *175*(1), pp. 475-493.

[3] Auf der Heide, E. 1989. Disaster Response: Principles of Preparation and Co-ordination. CV Mosby, Toronto.
[4] Balcik, B., Beamon, B. M., and Smilowitz, K. 2008. Last mile distribution in humanitarian relief. *Journal of Intelligent Transportation Systems*, *12*(2), pp. 51-63.
[5] Barbarosoğlu, G., Özdamar, L., and Cevik, A. 2002. An interactive approach for hierarchical analysis of helicopter logistics in disaster relief operations. *European Journal of* Operational Research, 140(1), pp. 118-133.
[6] Barnhart, C, Donald, H.R. 1993. Modelling intermodal routing. *Journal of Busines Logistics*, 14(1), pp.205–223.
[7] Beamon, B.M. and Kotleba, S.A. 2006. Inventory modelling for complex emergencies in humanitarian relief operations. International *Journal of Logistics Research and Applications*, 9(1), pp. 1-18. http://dx.doi.org/10.1080/13675560500453667.
[8] Byman, D., Lesser, I.O., Pirnie, B.R., Bernard, C. and Waxman, M. 2000. Strengthening the Partnership: Improving Military Coordination with Relief Agencies and Allies in Humanitarian Operations, Rand Corporation, Santa Monica, CA, available at: www. rand.org/pubs/monograph reports/MR1185/ (accessed April 29, 2007).
[9] Centre of research on the Epidemiology of Disasters Website (EM-DAT CRED) 2013. *Southern Africa disasters Database*. 26/09/2013. http://www.em-dat.net.
[10] Curtis, B. 2009. The Chirundu Border Post: detailed monitoring of transit times. *Sub-Saharan Africa Transport Policy Program Discussion Paper 10*.
[11] Disparte, D. 2007. The postman's parallel. *Car Nation*, 2 (1) pp. 22-27.
[12] Dynes, R. R. 1999. The dialogue between Voltaire and Rousseau on the Lisbon earthquake: the emergence of a social science view. http://udspace.udel.edu/handle/19716/435.
[13] Dynes, R.R. 1994. Community emergency planning: false assumptions and inappropriate analogies. *International Journal of Mass Emergencies and Disasters*, 12(1), pp. 141-158.
[14] Haimes. Y.Y and Longstaff, T., 2002. The role of risk analysis in the protection of critical infrastructures against terrorism. *Risk Analysis*, 2002; 22(3), pp. 439–444.
[15] Hale, T., and Moberg, C. R. 2005. Improving supply chain disaster preparedness: a decision process for secure site location. *International Journal of Physical Distribution and Logistics Management*, 35(3), pp. 195-207.
[16] Hamedi, M., Haghani, A., and Yang, S. 2012. Reliable Transportation of Humanitarian Supplies in Disaster Response: Model and Heuristic. *Procedia-Social and Behavioral Sciences*, *54(1)*, pp. 1205-1219.
[17] Lancet, T., 2010. Growth of aid and the decline of humanitarianism. *The Lancet*, 375 (1), pp. 3-11.
[18] Mileti, D. 1999. Disasters by Design: A Reassessment of Natural Hazards in the United States, Joseph Henry Press, Washington, DC.
[19] Oh, SC. and Haghani, A, 1996. Formulation and Solution of multi-commodity, multi-model network flow model for disdater relief operations. Transportation research Part A: Pollicy and Practice, 1996 - Elsevier.
[20] Oloruntoba, R, Gray, R, 2006. Humanitarian aid: an agile supply chain? Supply Chain Management. An International Journal of Physical distribution and Logistics, 11(2), pp. 115–120.
[21] Pateman, H, Hughes, K, Cahoon, S. 2013. Humanizing humanitarian supply chains: A synthesis of key challenges. *The Asian journal of shipping and logistics*, 29 (1) pp. 081-102.
[22] Prasad, V. R., Nair, K. P. K., and Aneja, Y. P. 1993. A generalized time-cost trade-off transportation problem. *Journal of the Operational Research Society*, pp. 1243-1248.
[23] Southern African Development Community 2013. Members and states representations; 2009; 2013/08/12. http://www.sadc.int/member-states/.
[24] Strawderman, L., and Eksioglu, B. 2009. The role of intermodal transportation in humanitarian supply chains final report.
[25] Strömberg, D. 2007. Natural disasters, economic development, and humanitarian aid. *The Journal of Economic Perspectives*, pp. 199-222.
[26] Ranganathan, R. and Foster, V., (2011). The SADC"s infrastructure: A regional perspective. Policy Research. World Bank, Africa Region.
[27] Teravaninthorn, S., and Raballand, G. 2009. Transport prices and costs in Africa: a review of the main international corridors. *World Bank Publications*.
[28] Thomas, A. 2004. Elevating humanitarian logistics. *International Aid and Trade Review, Special Edition*, pp. 102-106.
[29] Thomas, A. and Kopczak, L. 2005. From Logistics to Supply Chain Management – The Path Forward to the Humanitarian Sector. *Fritz Institute*.
[30] Thomas, A., and Mizushima, M. 2005. Logistics training: Necessity or Luxury? *Forced Migration Review*, 22(22), pp. 60-61.
[31] United Nations habitat 2013. Global Report on Human Settlements; 2009; 2013/06/12. http://www.unhabitat.org/documents/GRHS09/FS5 pdf.
[32] Van Wassenhove, L. N., and Pedraza Martinez, A. J. 2012. Using OR to adapt supply chain management best practices to humanitarian logistics. *International Transactions in Operational Research*, 19(1-2), pp. 307-322.
[33] Van Wassenhove, L.N. and Samii, R. 2003. The United Nations Joint Logistics Centre (UNJLC): The Genesis of a Humanitarian Relief Coordination Platform, *INSEAD, Fontainebleau*.

[34] Vignaux, G. A., and Michalewicz, Z. 1991. A genetic algorithm for the linear transportation problem. Systems, Man and Cybernetics. *IEEE Transactions* on, 21(2), pp. 445-452.

A CASE STUDY - MANAGING OPERATIONAL AND SUPPLY CHAIN EFFICIENCY: STOCHASTIC MODELLING OF A NEW PRODUCTION FACILITY

J.I. de Wet[1*] & J. Pretorius[2]

**[12]University of Johannesburg,
South Africa**
[1]jacobusdewet26@gmail.com, [2]jhcpretorius@uj.ac.za

ABSTRACT

An industrial company has embarked on an expansion project to create more capacity to meet the growing demand for one of its high-end product ranges. The final step in the production process constitutes of solidification, packaging, local storage and distribution. A stochastic simulation model in Arena Simulation Software was developed to represent the packaging activities, transportation between intermediate storage facilities and the product outbound supply chain. The aim of this study is to analyse the impact of the additional production volumes on the high-end product range packaging facility and the outbound supply chain capabilities and constraints. The study determined the required fleet and staging capacity at the production facility. A compelling business case is justified to expand the staging capacity as it holds significant opportunity cost.

*Corresponding author

1. INTRODUCTION

The company has invested in a production expansion project to create more capacity to meet the growing demand for one of its high-end product ranges. The production expansion project is in its final stages, and a new/expansion plant is expected to come online in the coming months. With the increased production, it is required to determine capacity capabilities in the downstream distribution of the increased volumes.

The objective is to evaluate the packaging and solidification throughput capacity by means of a stochastic simulation. The simulation model will take into account possible bottlenecks, buffer staging capacities and transport requirements due to increased production volumes.

The objectives for the model are:
- Determine the impact of the additional volumes on the outbound supply chain;
- Identify plant staging capacity constraints;
- Determine transportation requirements between the Production Facility and the Intermediate Storage Facility.

The effectiveness of a production facility is imperative to the enterprise's competiveness. Critical factors that are taken into account that affect the efficiency of manufacturing systems and processes [1] are:
- Operational times;
- Buffer capacities;
- Setup times;
- Batch sizes.

Figure 1 depicts the operational and supply chain network that is modelled in the Arena Simulation Software Package:

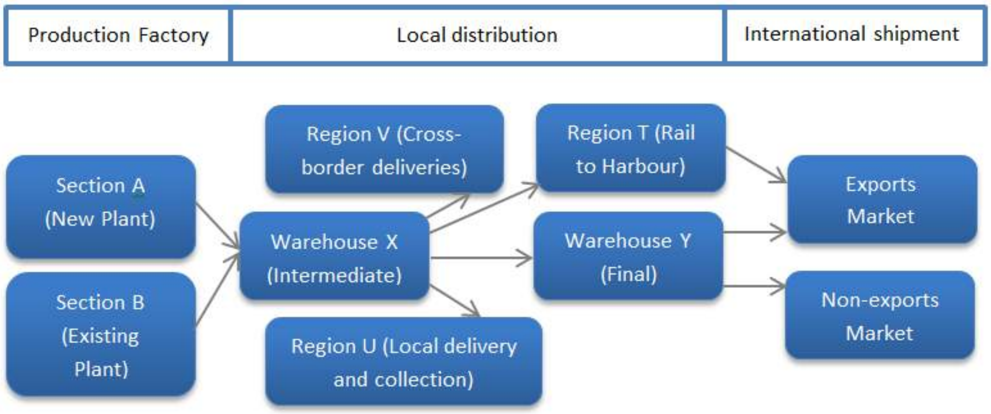


Figure 1: Operational and supply chain network

Supply chain simulation includes the behaviour of an integrated production and logistics network that represents the complex interactions and characteristics of intra- and inter-facilities, resources and tangible entities. In general, simulation is applied to study the stochastic variability and impact caused by [2]:
- End customer demand;
- Transportation modes;
- Packaging/container changes;
- Production schedules;
- Supply base changes.

With the newly commissioned plant, the simulation model analyses the effect of the additional product volumes on the loading staging area's capacity and outbound supply chain requirements.

2. LITERATURE REVIEW

Excellence in operational execution in the production and logistics of an organisation's supply chain is dependent on the effective and timely translation of customer demand into material handling and control decisions across the supply chain [2]. The challenge of translating these customer requirements and demands into material control decisions is complicated by the factors [3] as set out in Figure 2:

Figure 2: Challenges in achieving operational excellence

In order for products to be manufactured at the lowest possible cost and delivered in the shortest amount of time to customers, supply chain operational planning is vital to determine how execution should be done. Numerous organisations use a supply chain planning (SCP) tool whereby historical production data and supply chain performance together with customer demand information are taken to create material and distribution requirements planning (DRP) from an analytical perspective. Deterministic planning approaches such as DRP and material requirements planning (MRP) contain inherent limitations which are only manifested when it comes to the implementation of these planning techniques [3]:

- Real life scenario experimentation is rarely cost-effective, often disruptive and in most cases impossible (especially in large commercial supply chains)
- Accurate prediction of a dynamic and ever-developing system over time is nearly impossible.
- Incorporating random effects that occur is difficult to accommodate.
- Operational performance assessment is dependent on real historical information of the system. Influencing parameters and variables cannot be changed, altered or adjusted.

These shortcomings can be addressed by discrete event simulation (DES). Banks et al. [2] emphasises that simulation has been used for a long time to address issues relating to the management of the supply chain within the factory boundaries – through verifying the effect of different system configurations on how execution should be done. Grabau [4] raises the concern that averages kill. Banks et al. [2] note that this is a concern beyond the factory boundaries. In contrast to the factory floor, there is great potential in supply chains for many industries to optimise throughput and reduce lead times. Small changes in a supply chain where large quantities of stock are involved may have a great impact in absolute terms.

Over time the development of high-performance and robust distributed simulation has enabled the undertaking of concerns such as:
- Data shielding;
- Local maintenance;
- Scalability;
- Geographical distribution of supply chains.

2.1 Simulation Approaches

Implementation approaches for distributed supply chain simulations contingent on the issues to be tackled can be applied. Figure 3 depicts the two different approaches.

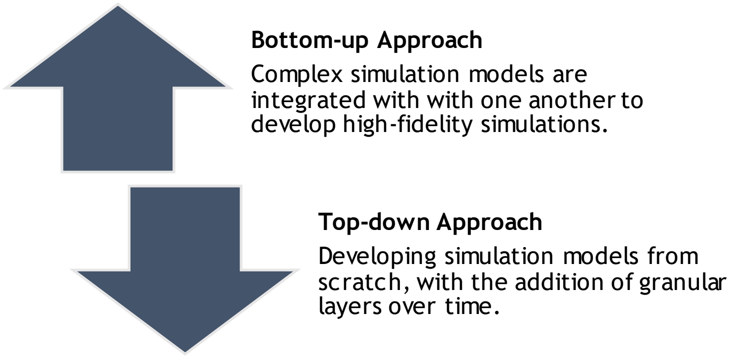


Figure 3: Simulation approaches [2]

Simulation models are particularly useful to understand the complex interactions between several components within a supply chain. It is cautioned [2] that the models generally include an unsophisticated abstraction of the associated planning process. Simulations are typically fed with material releases. However, in today's pull driven environment with high product demand variability as well as frequent phasing-in and out of new and old products, it remains a challenge to generate these input releases. Therefore the process or system that the simulation model represents is ultimately driven by customer demand scenarios, i.e. demand fulfilment management.

Further, DES permits the assessment of a supply chain's dynamic behaviour in a non-steady state [3]. A common example is the breakdown of a critical resource in a process and how the system will recover after the incident. DES is exceptionally beneficial in environments and industries that are subjected to the following characteristics, as in the case of the industrial sector:

- The need for the optimisation of capacity utilisation and sequencing in production is high and consequently the flexibility of capacity adjustments and adaptions are low, usually due to the high capital cost associated with these amendments.
- Organisations that produce in large quantities with many customers are subjected to great variability and stochastic uncertainties across the whole supply chain.
- Numerous complex interdependencies between customers and suppliers allow for significant prospective optimisation and foster opportunities for collaborative performance, enhancement and improvement.
- Value added operations from labour are low.
- Value added operations from logistics are significant.

2.2 Decision Support in Dynamic Markets

Growing competition in the marketplace compels industrial corporations to continuously improve their business performance in order to remain competitive and relevant [5]. Simulation and modelling technologies are gaining acceptance in industrial industries to provide effective decision support [6].

Advanced experimentation techniques are required to test and analyse system capabilities, especially in highly capitalised industries where real system experimentation is too disruptive, costly and time consuming [7]. Experimental techniques, such as stochastic simulation models take into account the specific demand-driven/pull production environments where operational planning is the result of an intricate balance between frequently changing customer requirements and material flow within the production area at predetermined times [8]. However, it remains a challenge to maintain and support the models in these fast changing business environments [7].

Lendermann et al. [9] outlines that today's demand driven and production pull-environments lead to high process variability and frequent new product introductions. In the case of this study, the company's Solidification Plant experience high process variability due to the pull-driven market demand which causes frequent product change-over and results in regular system flushing that severely impacts product losses.

Simulation based architectures for operational planning are developed for responding to production discrepancies and disturbances. Ramakrishnana and Wysk [10] provide four advantages of simulation decision support:

- The ability to predict future production behaviour, given the current state of operations.
- Improving the control logic of the production facility by analysing various scenarios.
- Capacity planners to effectively allocate resources by introducing resource scheduling constraints into the simulation model to improve decision-making.
- Simulation offers schedulers and planners the ability to evaluate production strategies to increase production throughput.

The importance of simulation models is evident in cases where there are many constraints and decision variables with complex and non-linear interactions between them, as well as conflicting and multiple objectives [11]. These constraints, variables and objectives can only be sufficiently understood through a simulation analysis [12, 13]. Longo [13] points out, in these contexts, one does not get a sound understanding of the system and its intricacies through mathematical models alone. These models are usually restricted with multiple constraints and assumptions of which the outcome does not represent reality and making the model inapplicable. Simulation models permits [13, 14, 15]:

- Artificially real complex systems;
- Time compressed analysis;
- Multiple performance measurements monitored; and

- Delineated system's behaviour.

Several studies [16, 17] claim that literature lacks studies, with its basis in real cases, where simulation models are used to analyse the complex behaviour of systems in field services. Further, the presence of studies investigating how simulation models can aide and support in the design of systems are limited, especially where systems evolve over time and need to cope with expansion and global growth. In the aerospace industry [18] there is a compelling need to assess whether field system delivery will comply with industry service level agreements.

This study aims to contribute to literature to show how stochastic simulation can provide effective decision support in an industrial production environment where efficiency assessment is challenging.

2.3 Synthesis

A top-down approach to simulation modelling is adopted to add granularity for the requested scenarios to be tested. This approach further allows for the complexities of the process at each stage of the outbound supply chain to be augmented. Key opportunities for stochastic simulation modelling lay within the company's high-end product range division as customer demand dictates regular transportation and packaging changes. Especially in this study, the impact of adding additional infrastructure will be scrutinised by the simulation model.

The high-end product range division of the company clearly illustrates today's demand driven and production pull-environment which lead to high process variability. Stochastic simulation modelling has the ability to take variability into account to aid and improve decision-making.

3. METHODOLOGY

A structured and rigorous method is followed whereby secondary quantitative data is sourced and statistically analysed with discrete histogram probabilities to model variability. The Arena [20] based simulation model is built, verified and validated, where after scenarios are tested and outcomes presented to enable improved decision-making. Figure 4 illustrates the approached followed:

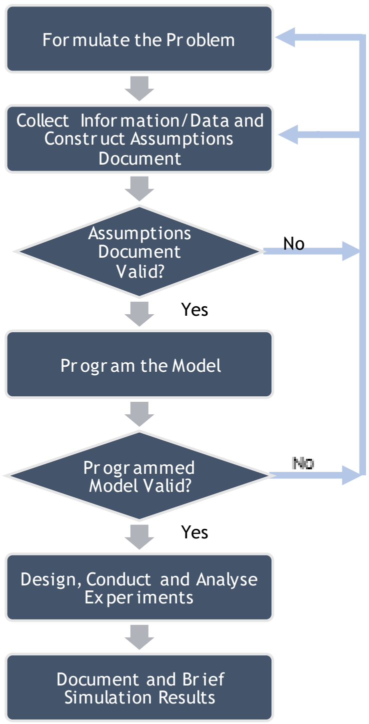


Figure 4: Research Approach

The simulation model is used to model the throughput and provide recommendations to the plant's staging buffer capacity requirements as well as the transportation requirements for the ramped-up production rates. The research will be done for two solidification and packaging plants: the old plant, which existed before the expansion project, and the newly commissioned plant that exists after completion of the expansion project. The model will include the total outbound supply chain of the combined production throughput of the old and new plant. This includes intermediate storage facilities, product movement via road and rail, and final product allocation (both local and export markets).

4. ANALYSIS OF SYSTEM FACTORS AND ASSUMPTIONS

For each source of randomness in the stochastic model, probability histograms or distributions were calculated. Figure 5 illustrates the scope on a high-level and the factors that were taken into account in the stochastic model:

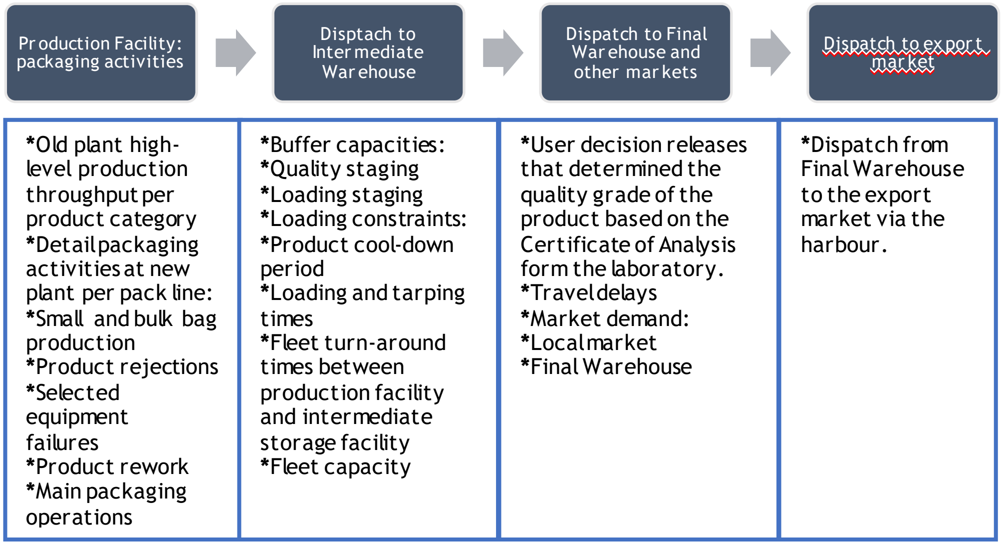


Figure 5: Scope and Factors account for in model

Data was collected from various sources for the different sections in the simulation model:

Section in model	Measure	Data source	Assumption or Analysis
Pack lines	Production rate (number of pallets per day)	Daily production report maintained by Production personnel.	Probability distributions
Pack lines	Days of zero production (percentage of days per year)	Daily production report maintained by Production personnel.	Fixed percentage
Pack lines	Production design capacity (bags per hour)	Equipment manual	Assumption
Plant	Plant availability	Operating standards	Assumption
Product	Product weight (kg per bag)	Operating standards	Assumption
Palletising and staging	Buffer capacities (pallets per plant)	Design layout	Assumption
Palletising and staging	Product cool-off period (hours per day)	Quality standards	Assumption
Palletising and staging	Lifting operations and idling time (seconds)	Time study	Probability distributions
Palletising and staging	Facility layout (distances, product flow)	Visual inspection	Assumption
Palletising and staging	Conveyor belt speeds (meter per minute)	Operating standards	Assumption
Palletising and staging	Pallet rejection rates (percentage of pallets rejected per day)	Visual inspection	Assumption
Palletising and staging	Mean time between equipment failures (days)	Failure records	MTBF calculation
Transportation	Truck capacity (pallets per truck)	Operating standards	Assumption
Transportation	Loading times (hours per day)	Operating standards	Assumption
Transportation	Warehouse capacity (pallets per facility)	Design layout	Assumption

Transportation	Travel times (min per meter). Including turnaround, idling and inspection times.	Vehicle tracking data collection system	Probability distributions
Transportation	Number of deliveries per day	Warehouse data reports	Probability distributions

The results were refined nu applying calculated factors to the probability distributions to account for the input data discrepancies. In cases where the calculated probability distribution produced consistently higher or lower values, factors were applied to the distributions to achieve a more representative result.

The model was run for a year period, as it was considered representative of all key events that the model need to capture. Each run had 20 repetitions, equating to 10 minutes running time.

Production throughput and transport turnaround times were considered as top priority constraints and was tested according to this priority setting.

5. BASE CASE VALIDATION

Verification allows confirming that the model behaves in the way it was intended to behave, whereas validation determines whether the model is able to accurately reflect reality [19]. Verification is a detailed step in the simulation study and is simulation software orientated. The main action in verification is debugging of the model; verifying that the computerised model works correctly. The model was verified through:

- Checking model logic;
- Structured walk-through of the model logic with modelling expert;
- Correcting syntax errors;
- Run-time errors corrected that manifested during program execution.

In consultation with a subject matter expert, a structured walk-through of the model logic and assumptions were done. Animation was also used to verify the model's execution. In addition, the model was executed by creating and following a single entity (pallet) through the model to test model logic.

Conceptual validation was done by interacting with production and supply chain logistic managers on a regular basis to ensure that the model represents the real world adequately. Operational validation was done by comparing the average model output with the actual data. The delta column, in the next four tables, indicates the percentage difference between the model average and the actual average. The company considered the validation acceptable if the delta values were within a 5% range. Given the small delta between the model results and the actual results, it can be concluded that the model is an accurate representation of reality.

Table 1 indicates the model average vs. the actual data average for the small and bulk bag production and pallet rejections per pack line at the new plant. It further indicates the production throughput of the old plant per product category as well as the total production per plant and product rework.

Table 1: Base Case Validation — Production

Description	Units	Model Average	Actual Average	Delta
Small bags created at new plant	bags/day	5 542	5 537	0.1%
Number of small-bag pallets at new plant	pallets/day	153	158	-3.4%
Bulk bags created at new plant	bags/day	97.9	97.7	0.2%
Rejected pallets derived from small bags at new plant	pallets/day	7.7	7.7	-0.5%
Number of pallets created at old plant	pallets/day	280	273	2.6%
Total production	t/d	345	348	-0.8%
Old Plant	t/d	180	175	2.7%
New Plant	t/d	165	173	-4.3%

The average inventory statistic is showed in Table 2. The average staging buffers are well below its capacity of 284 pallets for the new plant and 460 pallets for the old plant. The intermediate warehouse's average inventory of 5127t is below its capacity of 7700t and the content of the quarantine buffer of 2002 pallets is below the capacity of 2500 pallets. It is expected that the average is to be lower than capacity. From Table 2, it is evident that on average the inventory levels at the intermediate warehouse are near its capacity.

Table 2: Average inventory

Average Inventory Statistics		
Old Plant Staging Buffer	pallets	98
New Plant Staging Buffer	pallets	92
Intermediate Warehouse	tons	5 127
Quarantine at Intermediate Warehouse	pallets	2 002
Final Warehouse	tons	4 890

Table 3 indicates operational statistics of the number of trucks rotating between the intermediate warehouse and the Production Facility. It also indicates the product tonnage received at the intermediate warehouse. With total production of 345t/day sent to the intermediate warehouse, 344t/day is received. This indicates that there is not a build-up of entities and the model logic is working.

Table 3: Operational Statistics

Description	Unit	Model Average	Actual Average	Delta
Number of trucks per day between intermediate warehouse and Production Facility	Truck/ day	11.9	11.83	**0%**
Total mass received at Intermediate Warehouse	t/d	344	348	**1.0%**
Intermediate Warehouse demand	t/d	349	348	**0.2%**

Table 4 indicates the demand generated and fulfilled from the intermediate warehouse to the final warehouse. The demand generated between the average model output and actual data is within a 5% range and the demand fulfilment within 95% to 100%. The demand fulfilment indicates the percentage of the demand generated that was fulfilled. For the export market, the demand generated from the Final Warehouse is on average 290t/d. On average the Final Warehouse produces 280t/d. Demand fulfilment of the Final Warehouse is 96%.

Table 4: Final Warehouse demand generation and fulfilment

Final warehouse export demand generated	t/d	290	290	0%
Final Warehouse export demand fulfilled	t/d	280		
Final Warehouse export demand fulfilment	%	96%		
Final Warehouse Non-export demand generated	**t/d**	**3.6**	**3.4**	**4%**
Final Warehouse Non-export demand fulfilled	t/d	4		
Final Warehouse Non-export demand fulfilment	%	99%		

The credibility of the model was affirmed by the end-user having confidence in the model's results. End-users also confirmed that the discrete input histograms adequately represent the variability of the real system. No statistical technique was employed to account for stochastic validation between the real and simulated system. The end-users were comfortable that the observed and real values were close enough and that the model can be used to simulate scenarios.

6. RESULTS

Simulation models are not optimisers, but are used to answer "what if" questions. The findings outline the baseline and increased production capability, staging and storage capacity requirements as well as the transportation need between the Production Facility and the intermediate warehouse.

Table 5: Model Results

Scenarios	Production (t/d)	Staging Requirement (pallets)	Fleet Requirements (number of trucks)	Inventory Overrun at intermediate warehouse
Base line production	165	350	-	14% at 7700t 1.9% at 10400t
Rated production	188	401	4	14.63% at 7700t 2% at 10400t
Maximum sustained production	220	500	5	15% at 7700t 2% at 10400t

7. CONCLUSION

The number of trucks rotating between the Production Facility and the intermediate warehouse has a limited effect on clearing the daily production due to the product cool-down requirement period of 12 hours and that pick-up loads are restricted to specific times (07:00-16:00).

For the increased volumes at the new plant, a buffer staging capacity of at least 401 pallets is required. A fleet size of four trucks is required to clear the daily production. An additional truck (5th truck) is not necessary as it would have no benefit in reducing the loading staging buffer size.

Therefore a fleet of eight trucks is required to clear the old and new plants' daily production for the rated production case. The maximum sustained production of 220t/d at the new plant requires a loading staging buffer of 500 pallets with a fleet of five trucks. A total fleet requirement of nine trucks is required for the old and new plants' daily production to be cleared. To reduce the staging buffer to 451 pallets a 6th truck for the new plant is required.

The intermediate warehouse with a capacity of 10400 ton is recommended and should be able to manage the 2% inventory overrun.

Should the staging buffer capacity of 284 pallets not increase to 500 pallets, the industrial company may need to cut back production by 234 pallets per day which holds a significant revenue loss that presents a compelling business case to pursue the expansion of the staging area at the new plant.

REFERENCES

[1] S. Kłos and J. Patalas-Maliszewska, "Throughput Analysis of Automatic Production Lines Based on Simulation Methods", Intelligent Data Engineering and Automated Learning – IDEAL 2015, pp. 181-190, 2015.

[2] J. Banks, S. Buckley, S. Jain, P. Lendermann and M. Manivannan, "Opportunities for simulation in supply chain management", in Winter Simulation Conference, E. Yucesan, C.-H. Chen, J. L. Snowdon, and J.M Charnes, eds., 2002.

[3] P. Lendermann, B. Julka and L. McGinnis, "Framework for Distributed Supply Chain Simulation: Technical Feasibility and Implementation Approaches for the Semiconductor Industry", in International Conference on Modelling and Analysis of Semiconductor Manufacturing (MASM), 2002.

[4] M. Grabau, "Averages kill (or how to sell business process simulation)", in 2001 Winter Simulation Conference, Institute of Electrical and Electronic Engineers, 2001, pp. 1262-1265.

[5] S. Ramakrishnan, "A simulation-based framework for integrating operational models with business process models in a global business environement", Ph.D, Industrial and Systems Engineering in the Graduate School of Binghamton University State University of New York, 2008.

[6] P. Lendermann, "About the Need for Distributed Simulation Technology for the Resolution of Real-world Manufacturing and Logistics Problems", in 2006 Winter Simulation Conference, 2006, pp. 1119-1128.

[7] P. Lendermann, S. Turner, M. Low, B. Gan, N. Julka, L. Chan, W. Cai, L. Lee, E. Chew, S. Teng and L. McGinnis, "An integrated and adaptive decision-support framework for high-tech manufacturing and service networks", Journal of Simulation, vol. 1, no. 2, pp. 69-79, 2007.

[8] P. Lendermann, B. Gan and L. McGinnis, "Distributed simulation with incorporated APS procedures for high-fidelity supply chain optimization", in Winter Simulation Conference, Washington DC, 2001, pp. 1138-1145.

[9] P. Lendermann, N. Julka, L. Chan and B. Gan, "Integration of Discrete Event Simulation Models with Framework-Based Applications", in 2003 Winter Simulation Conference, New Orleans, LA, 2003, pp. 1797-1804.

[10] S. Ramakrishnan and R. Wysk, "A Real Time Simulation-based Control Architecture for Supply chain Interactions", in Industrial Engineering Research Conference, 2002.

[11] F. Visintin, I. Porcelli and A. Ghini, "Applying discrete event simulation to the design of a service delivery system in the aerospace industry: a case study", Journal of Intelligent Manufacturing, vol. 25, no. 5, pp. 1135-1152, 2013.

[12] P. Lendermann, A. Shirrmann, B. Gan, H. Fromm and N. Nulka, "Unlocking value from equipment exchange contracts in aviation using simulation-based optimization", in 2010 Winter Simulation Conference, 2010, pp. 2034-2045.

[13] F. Longo, "Advances of modeling and simulation in supply chain and industry", SIMULATION, vol. 87, no. 8, pp. 651-656, 2011.

[14] J. Gagliardi, J. Renaud and A. Ruiz, "A simulation modeling framework for multiple-aisle automated storage and retrieval systems", Journal of Intelligent Manufacturing, vol. 25, no. 1, pp. 193-207, 2012.

[15] D. Zhang and A. Anosike, "Modelling and simulation of dynamically integrated manufacturing systems", Journal of Intelligent Manufacturing, vol. 23, no. 6, pp. 2367-2382, 2012.

[16] M. Jahangirian, T. Eldabi, A. Naseer, L. Stergioulas and T. Young, "Simulation in manufacturing and business: A review", European Journal of Operations Research, no. 2031, pp. 1-13, 2010.

[17] S. Agnihothri and A. Mishra, "Cross-training Decisions in Field Services with Three Job Types and Server-Job Mismatch*", Decision Sciences, vol. 35, no. 2, pp. 239-257, 2004.

[18] M. MacDonnell and B. Clegg, "Designing a support system for aerospace maintenance supply chains", Journal of Manufacturing Technology Management, vol. 18, no. 2, pp. 139-152, 2007.

[19] K. Kendall and J. Kendall, Systems analysis and design. Boston, Mass: Pearson, 2014.

[20] "Discrete Event Simulation Software | Manufacturing, Supply Chain & Healthcare Simulation Software | Arena Simulation", Arenasimulation.com, 2017. [Online]. Available: https://www.arenasimulation.com/. [Accessed: 11- Jul- 2017].

COMPETITIVE ADVANTAGE DURING INDUSTRY 4.0: THE CASE FOR SOUTH AFRICAN MANUFACTURING SMES

M.S. Futcher[1]* & B.P. Sunjka[2]

[1]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
mattfutcher@gmail.com

[2]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
Bernadette.Sunjka@wits.ac.za

ABSTRACT

With the expected disruption of industry 4.0 and the current challenges that SMEs face in South Africa, there is an increasing threat that SMEs will lose any competitive advantage they currently have. This exploratory study investigates how South African manufacturing SMEs can remain competitive during the fourth industrial revolution. Data, in the form of current literature, was analysed using thematic content analysis. From the analysis process, 8 emergent themes were used to organise the results of the study. Notable findings towards generating competitive advantage included: the location of SMEs within clusters, collaboration with disruption leaders, the sharing of outcomes across the value chain, the shift of business models towards a service and software orientation, the use of data-driven insights to find and capture high margin markets and the increased effectiveness of labour through technology use. The study also found that the use of the internet of things and cloud computing can significantly reduce infrastructure requirements and promote a competitive advantage.

[1] The author was enrolled for an MSC Eng 50/50 (Engineering Management) degree in the Department of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand
*Corresponding author

1. INTRODUCTION

One of the key drivers of knowledge capital and economic prosperity in a country is manufacturing [1]. According to Kaldor [2], an increase in manufacturing output leads to a direct overall increase in gross domestic product and an increase in productivity from sectors outside of the manufacturing sector. Manufacturing is one of the few sectors that exhibits this characteristic, where a net positive output leads to growth both within and outside of the sector. It is, therefore, one of South Africa's most important sectors for overall economic growth.

Small to medium enterprises (SMEs) contribute 40% of South Africa's manufacturing gross domestic product output [3]. SMEs are a vital component of the South African economy and provide a large portion of South Africa's much-needed jobs [4]. SMEs in South Africa have traditionally operated in a challenging environment, facing issues such as limited access to skills, funding and developed infrastructure as well as crime and strict regulation [3,4].

With the dawn of the next industrial revolution (Industry 4.0), manufacturing as a whole is expected to undergo significant change. Industry 4.0 differentiates itself from the third industrial revolution by having technology link operations and make localised decisions through the internet. This will enable "intelligence" to be built into operations with production items such as machines making decisions [5]. Industry 4.0 is linked to technologies and concepts such as cloud computing, artificial intelligence, 3D printing, cyber-physical systems and smart factories [6]. Through industry 4.0, manufacturing processes and resource allocations will significantly change giving manufacturing SMEs the opportunity to develop and provide downstream services [6,7].

With the expected change and disruption that industry 4.0 will bring, along with the challenges that SMEs face in South Africa, there exists a very real threat for the small to medium manufacturing sector in South Africa. It is therefore important to understand how industry 4.0 will affect manufacturing SMEs and what can be done to retain or grow their competitive advantage. This study explored how South African manufacturing SMEs can remain competitive during the fourth industrial revolution. This was achieved by investigating possible sources of competitive advantage during the fourth industrial revolution and determining the actions and means by which manufacturing SMEs can generate competitive advantage during the fourth industrial revolution. A framework for decision-makers in SMEs was finally developed that will assist the creation of competitive advantage during the fourth industrial revolution.

2. LITERATURE REVIEW

2.1 Small to Medium Enterprises

SMEs in South Africa are defined as businesses that have no more than 200 employees and generate a maximum revenue of R40 million per annum [8]. Some of the challenges SMEs face in South Africa include; access to finance [9], labour law [10], crime [10], access to necessary resources [11], access to markets [12], research and development [3], regulation and policy [13] and unskilled labour [3].

2.1.1 Growth and Competitive Advantage

The International Trade Centre [14] released a competitiveness outlook for SMEs in 50 different countries. The SMEs were examined from both an internal and external point of view. Internally, SMEs were scored on their ability to connect, compete and change. An SMEs ability to `connect' was determined by examining its website and the degree to which the company uses e-mails for communication. An SMEs ability to `compete' was determined by the presence of an international quality certificates, capacity utilisation, managerial experience and the use of banking. Lastly an SMEs ability to `change' was determined by the the presence of audited financial statements, the degree of investment financed by banks, the presence of a formal training program and whether or not the SME makes use of foreign technology licenses. The competitiveness profile for South African SMEs, Figure 1, shows that there is a strong positive correlation between SME size and competitive performance. Figure 1 also shows that it is only large SMEs that score above 70% in their ability to connect. Both small and medium size SMEs score below 50%.

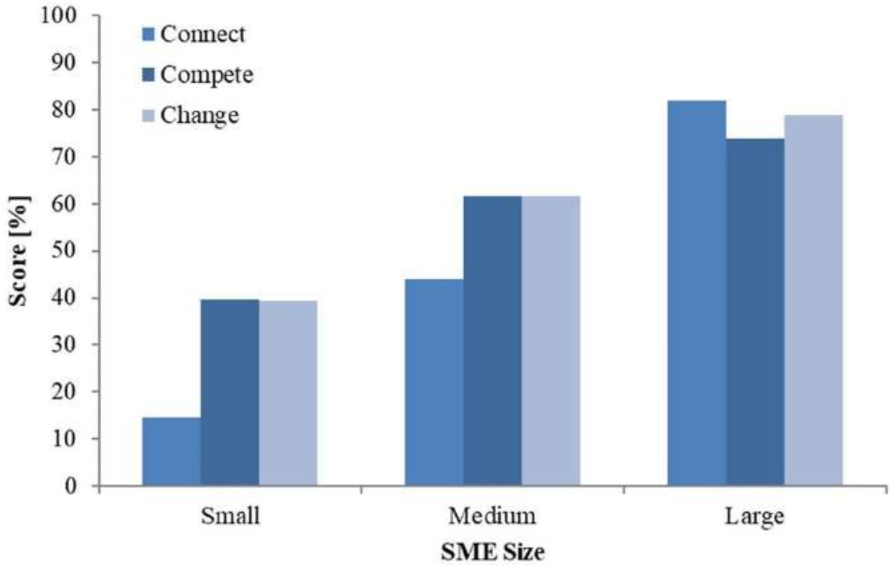


Figure 1 - SME competitiveness outlook profile South Africa [14]

2.1.2 The potential impact of industry 4.0

It cannot be said for certain that industry 4.0 will have a negative or positive effect for South African SMEs especially those within the manufacturing sector. There are very real threats that industry 4.0 poses such as increased competition from the global market, loss of unskilled jobs and an increase in financial commitment. There is evidence to suggest that technology has had a positive impact on SMEs and that cloud computing could actually alleviate many of the challenges SMEs currently face [15,16].

2.1.3 Adoption of industry 4.0

Industry 4.0 adoption by SMEs in South Africa and other BRICS countries is relatively low, with the focus of industry efforts being put towards skills development and preparation of the workforce for industry 4.0 [17]. Although South African SMEs face a number of barriers to industry 4.0 adoption, many international SMEs face similar challenges, even those in industry 4.0 leaders such as Germany [18]. One of the biggest challenges facing all SMEs is the considerable amount of investment which is required to adopt industry 4.0 [18].

As this study explored how South African manufacturing SMEs can remain competitive during the fourth industrial revolution, the concepts of competitive advantage and industry 4.0 will be elaborated on through the literature.

2.2 Competitive Advantage

Competitive advantage is an area of knowledge with many theoretical views and theories. Two academic definitions of competitive advantage are:

> *"Competitive advantage is obtained when an organisation develops or acquires a set of attributes (or executes actions) that allow it to outperform its competitors"* [19].

> *"The degree to which a firm has reduced costs, exploited opportunities and neutralized threats"* [20].

The following list serves as a brief summary of each theoretical view covered in the study with Figure 2 outlining a conceptual framework of competitive advantage theories.

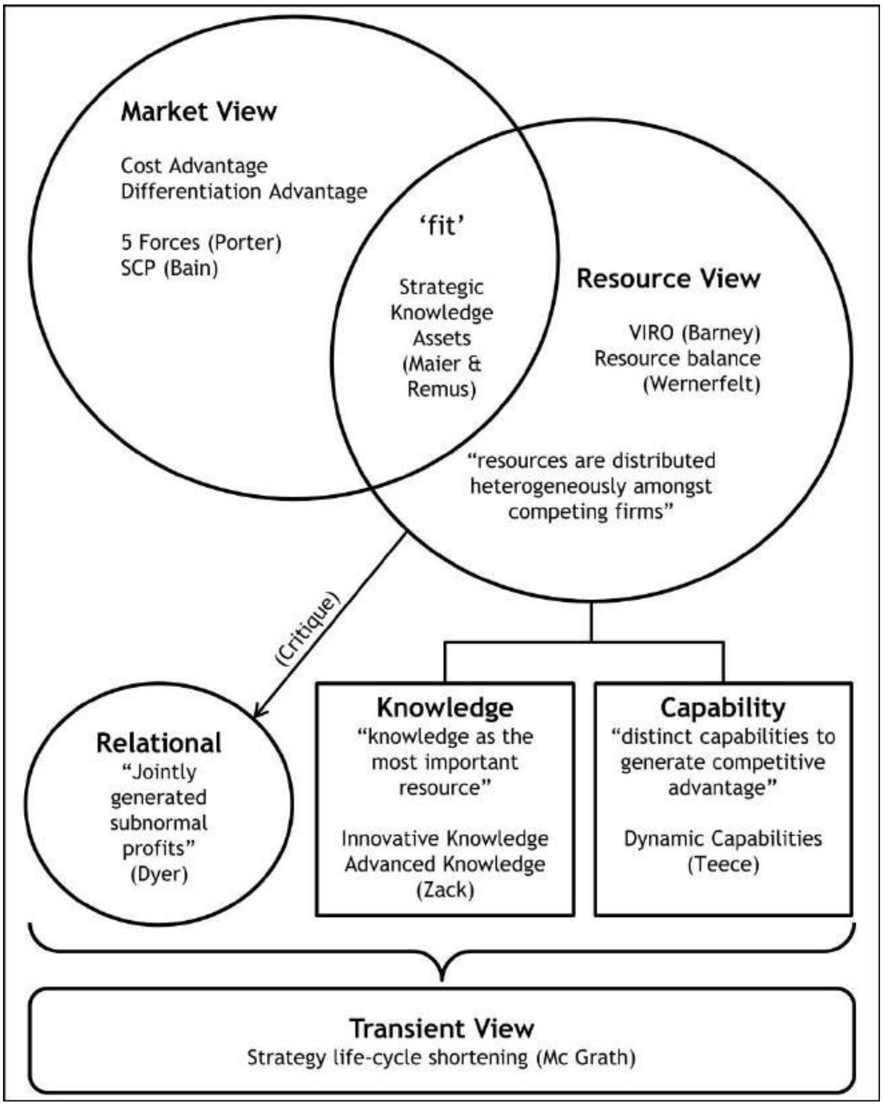


Figure 2 - Conceptual framework of competitive advantage

- Market-Based View: External factors and a firms positioning in the market predominantly effect its performance [21].
- Resource-Based View: Internal process and management of resources affect the performance of a firm [22].
- Knowledge-Based View: Knowledge is the most important resource and separate from other resources [23].
- Capability-Based View: The use of resources through distinct capabilities governs competitive advantage [24].
- MBV and RBV 'fit': The necessity of both internal (resources-based view) and external (market-based view) perspectives [25].
- Relational View: A firm's relation within the network leads to competitive advantage [26].
- Transient View: Opportunities for competitive advantage are transient and long-term perspectives can no longer achieve sustainableed competitive advantage regardless of whether or not an internal or external view is taken [27].

2.3 Design Principles of industry 4.0 and key components

There is no generally accepted working definition of industry 4.0 [28]. There has however been a set of industry 4.0 design Principles outlined by Hermann, Pentek and Otto [28], which are useful for understanding the fundamentals of Industry 4.0. The design Principles are:

2.3.1 Design Principle 1: Interconnectedness

The connection of people, machines, devices and sensors to create a networked ecosystem is fundamental to industry 4.0. Without interconnection, the sharing of information in real-time cannot happen and many of the efficiency benefits gained through industry 4.0 will not be realised. Interconnection also forms the backbone of collaboration between businesses and supply-chains, allowing them to share information (design principle 2) and achieve data-driven decision making [28].

Two important technologies that relate to the design Principle of interconnectedness are cyber-physical systems and the internet of things. Cyber-physical systems can be defined as "a new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities" [29]. These systems take into account the interaction and interdependency between what would have traditionally been seen as a physical system and a cyber-system [30]. They can be roughly thought of as the merger between the physical environment and the internet.

The internet of things is "a network of internet-connected objects able to collect and exchange data using embedded sensors" [31]. Chebudie, Rotondi and Minerva [32] found that there are certain fundamentals to the internet of things. These are:

- The connection of objects – For the network to function the connection of the "things" is necessary (design principle 1). Each object must also be uniquely identifiable and have sensing/actuation capabilities often making it a smart device.
- Ubiquity – The network is available anywhere or anytime and there is a continual link between the network and the internet.
- Embedded intelligence – smart devices exhibit their own emergent behaviour.
- Self-configurability – Each object connected the network manages its own configuration, software requirements, resources and energy consumption. Allowing the network to scale to a large size.
- Programmability – Refers to devices in the network that can execute different tasks without having to change physical form.

2.3.2 Design Principle 2: Information transparency

Industry 4.0 requires almost unrestricted data sharing, with information transparency increasing tenfold relative to industry 3.0. A true industry 4.0 state will exist when hundreds of sensors are capturing data and feeding that information into plant and supply chain models, creating a virtual copy of the physical world. These models will be filled with "context-aware" information, allowing participants in the network to make hyper-informed decisions in real-time [28]. Two key components linked to information transparency are Big Data and Cyber Security.

Big Data is defined as data sets (or streams) that are "not only big, but also high in variety and velocity, which makes them difficult to handle using traditional tools and techniques" [33]. Due to the large number of sensors and information streams in industry 4.0, significant data analytic capabilities will be necessary for businesses to make correct decisions [33].

Cyber-crime will be one of the greatest threats for business during industry 4.0 [34]. Although companies will share large amounts of internal information with the network, there will be data that is sensitive and important to that companies success. With all the connection points into the network, companies will need to strengthen their cybersecurity in order to protect data that leads to competitive advantage [34].

2.3.3 Design Principle 3: Decentralised decision making

Once interconnection (design principle 1) and information transparency (design principle 2) have been established, decentralised decisions can be made. "Decentralised" refers to an object in the network having a form of local intelligence to operate and make independent decisions [28]. Decentralised decision making will require artificial intelligence and the in many cases the use of cloud computing.

Artificial intelligence is simply machines (or virtual machines such as bots) exhibiting human-like cognition [35]. It is an extremely broad field but a critical component to industry 4.0 and one of the distinguishing factors from industry 3.0.

The access and "sharing of web infrastructure for resources, software and information over a network" is known as the Cloud or Cloud Computing [36]. It is simply the pooling of IT infrastructure, storage and computing power. It will enable businesses to access key digital capabilities from almost any geographical location in a manner that will allow scalability. The Cloud in many instances runs off a pay-per-use business model which will allow businesses of all sizes access to very important industry 4.0 capabilities such as data analytics and artificial intelligence [36].

2.3.4 Design Principle 4: Technical assistance

The role of people will begin to shift from operators to strategic decision makers as industry 4.0 matures [28]. With the vast amount of connections and data streams, people will rely on systems to support and assist them, presenting information in usable formats for problem-solving purposes [37]. Robots will also carry out difficult, unsafe and monotonous tasks on behalf of people [28].

Industry 4.0 is expected to change the nature of work, with a shift towards human-machine collaboration. Jobs that require repetitive non-skilled tasks are at greatest threat of being made redundant by robots [38]. Many jobs will be lost to industry 4.0 advancement, but a net positive increase in jobs is expected. New forms of work will be created in the areas of IT and data analytics [38]. The challenge from companies will be re-skilling staff to make the shift towards industry 4.0 [38].

3. RESEARCH DESIGN AND METHODOLOGY

The nature of this study is substantially forward looking from a time perspective. The specific research area, SMEs in the context of Industry 4.0, is unexplored with little existing work on the topic. It is for these reasons that the research is exploratory in nature. Exploratory research is primarily used when there is little known about the topic under study. It aims to better define an area of study and provide initial findings that can be further explored in more detail [39]. Exploratory research can normally also be classified as inductive research because it moves from the general towards the specific [40]. An inductive research approach was used for this study with data being qualitative in nature. Qualitative data is suited towards an exploratory study because it is rich in information and can be used to cover a large scope of research [39].

It was believed that using semi-structured interviews would be the most useful due to their balance between participant freedom and interviewer control. Semi structured interviews involve a standard set of questions being posed to a participant but additional probing questions can be asked throughout the interview [40].Through investigation into this method it became clear that there existed a very broad range of opinion on industry 4.0 with not many true subject matter experts. It was determined that companies within South Africa are in the initial stages of investigation on the potential of industry 4.0. For many companies, adopting an industry 4.0 state right now would be too expensive. Once some of the barriers to entry are removed such as the high cost of technology and the low levels of infrastructure, companies may begin to focus their research and knowledge on it. This would lead to more subject matter experts existing in South Africa.

At the moment internal knowledge of industry 4.0 is at a surface level and semi-structured interviews would be useful for understanding the general consensus on industry 4.0 but not how it can be used for competitive advantage. Through informal discussions with potential subject matter experts, it was determined that most of their knowledge was sourced through reading literature and not necessarily through experience in the research topic. It was therefore decided that collecting data directly from literature may improve the accuracy of the research and reduce potential bias by gathering information directly from the source.

3.1 Literature search and selection

Literature was found through keyword searches and snowball sampling [41]. Keywords were used in multiple combinations and included the terms; industry 4.0, Industrie 4.0, digitisation, manufacturing, competitive, advantage, industrial, revolution, smart factory and future. Google Scholar was used along with databases such as E.I Compendex, Inspec, National Technical Information Services and Scopus. Literature was then selected on the basis that it was relevant to the research objectives in part or in full and two additional criteria [42]. The first criterion was concerned with the target audience. Literature that was written for business leaders and managers was favoured to ensure that the research would be targeted at decision makers within SMEs and offer practical value. The second criteria related to the diversity of thought. Literature that covered a broad range of the research topic was favoured in order to ensure that a holistic outlook on the subject was met and that research remained exploratory [43]. The search and selection process was done in parallel to the analysis of the research so that the selection could stop once information saturation had been reached [44].

3.2 Analysis

The research data was analysed using thematic content analysis. There is no single method for conducting a thematic analysis but rather used by research as necessary [45]. This type of analysis has been criticised for being open to reliability errors due to researcher interpretation and understanding [46]. It is however well suited to large qualitative data sets [46]. The literature analysis process involved reading through the research several times for familiarisation. Direct quotes were then extracted from the literature that related to competitive advantage under any of the theoretical views examined through the literature survey. These quotes were then classified into a set of initial themes. These initial themes were refined through a co-occurrence analysis and a set of final themes were selected.

4. RESULTS

A total of 36 emergent themes were found in the literature. A co-occurrence analysis was done on the quotes to understand which themes frequently occurred together in the same quote. Themes that had a 60% or higher co-occurrence rate another theme were considered to be sub-themes. After refinement 8 final themes were used toward the final results. Table 1 below shows the theme and its ranking by frequency of occurrence.

Table 1 - Theme ranking

Theme Ranking	Ranking
Collaboration	1
Digital Capabilities	2
Business Model	3
Data	4
Decision Making	5
People	6
Organisational change	7
Operations	8

The results of the final themes are presented in Table 2.

Table 2 - Results by theme

Theme	Findings
Collaboration	• Employee exchange as a means of bringing new skills into the business. • Sharing of outcomes across the value chain. • Collaborating with companies who are involved in disruptive trends as a means to survival. • Building relationships outside of the traditional value chain.
Digital Capabilities	• Use the on-demand economy for acquiring capabilities. • Only develop digital capabilities that are necessary and align with the business model. • Be proactive and aggressive towards developing digital capabilities.
Business Model	• All companies need to become “technology” companies. • The business model must support partnerships and platforms. • Understand and forecast disruption by constantly checking for new revenue models outside of the current industry that are transferable. Expect integration between digital and physical. • Maintain the core business, inspire innovation at the edge aligned with disruptive trends but focus on solving customer problems.
Data	• Data is a source of competitive advantage and a business asset. • Data security is linked to digital trust and important for partnerships. • Data should support employee decision making and create a holistic understanding of the business.
Decision Making	• Should be based on feedback learning instead of past data extrapolation due to the fast pace of change. • Must be done faster and automated where possible • Companies need to remain informed to remain competitive in a dynamic landscape. • Decision making is dispersed amongst empowered employees.
People	• Employees made more effective through technology. • Leaders must become digitally fluent and have an intuitive sense of how to use partnerships for success. • Success depends on leaders’ ability to transform their organisations, and employees’ ability to implement digital initiatives.
Organisational change	• Ability to change becomes fundamental to competitive advantage. • Implement short-term initiatives immediately and medium/long term initiatives that focus on transformation, not augmentation.

	• Smaller companies are more agile and can change business models faster.
Operations	• Big data, the internet of things and cloud technology offer significant opportunities to reduce costs and infrastructure requirements. • Operations must be view holistically and beyond a company's own boundaries. • Industry 4.0 is underpinned by lean manufacturing which needs to be embraced to operate with the speed and flexibility of the ecosystem.

4.1 Framework

The results were refined into a framework that can be used by decision-makers in SMEs which is shown in Figure 3.

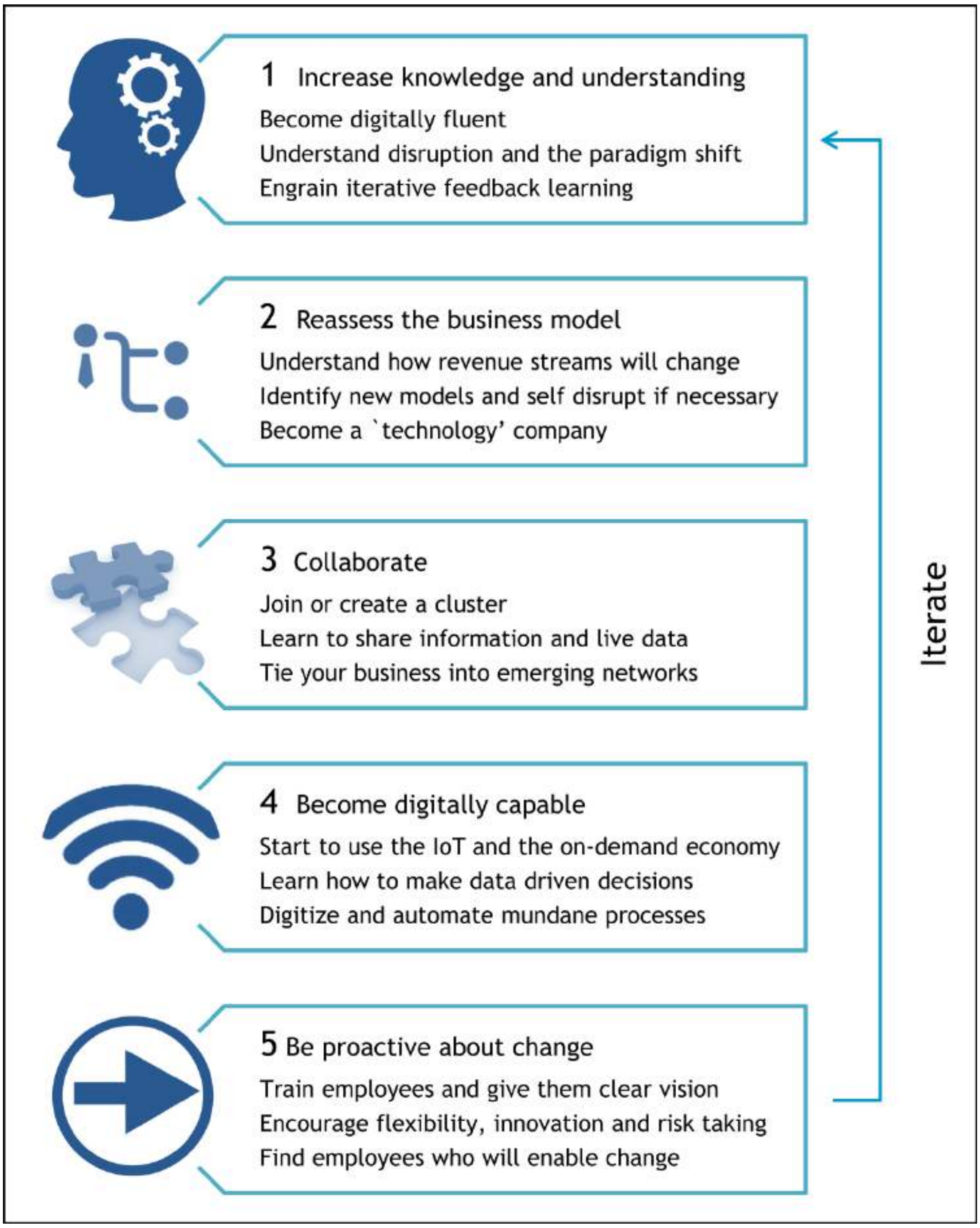


Figure 3 - Industry 4.0 framework

5. DISCUSSION

The research findings presented in this paper offer ideas and thoughts towards maintaining or growing a competitive advantage from the perspective of manufacturing SMEs in South Africa. The findings presented are by no means exhaustive or applicable in every case, they merely offer a starting point to further investigation and should be used as such. The research thus has a number of limitations that should be considered:

- The research is purely based on literature, which limits its practical application. Although there are few true subject matter experts with knowledge of both manufacturing SMEs and industry 4.0, there are people who have expertise in overlapping or similar fields. Their contribution would offer an ever further diversity of findings and possible make the research more practically applicable.
- The research is positioned as an introduction to further work and therefore lacks a degree of detail to make it practically sound. This is especially true for the framework presented in Figure 3. The framework is primarily used towards simplifying the main findings and translating them towards high-level intentions a manufacturing SME can take. The specific details of application are not included in this study and should be investigated further as developments within industry 4.0 unfold and true adoption occurs in international locations.

Related to the limitations of the research, there are also a number of limitations and constraints for SMEs in taking steps towards industry 4.0. These limitations are mainly governed by the current state of SMEs in SA and include access to finance, physical resources, knowledge and skilled labour [3,9,11]. These factors constrain the adoption of industry 4.0 and make implementing any findings from the research challenging. Despite these limitations, there will be some degree of change that SMEs can begin to make. The strategic level actions and intentions presented in Figure 3 will offer guidance to this change and provide a fixed point that can be worked towards while Industry 4.0 develops further.

The most prominent themes were collaboration and digital capabilities. The idea that collaboration is fundamental to competitive advantage during industry 4.0 is not surprising since industry 4.0 will essentially be the development of a giant network or ecosystem [47]. A businesses ability to tie into the ecosystem will be crucial, where each new connection (or relationship) will be an opportunity to offer value. As the number of routes to market increase and become more flexible, traditional linear supply chains will be less common. A single business may be a node for many different routes to market and thus sharing information much further up and down the supply chain becomes important. Digitally integrating operations with partners up and down the supply chain will ensure a business remains relevant.

The idea of collaboration goes past just business-to-business relationships. Collaboration between humans and the cyber-physical environment will also be an important element of survival. Business who understand how to use their current workforce in partnership with technologies such as artificial intelligence and robotic will outcompete firms who choose to ignore these advancements. It is difficult to determine when the most economical point in time is to make an investment into these kinds of technologies but business should prepare themselves for integration and ensure their knowledge and understanding of these technologies remain current. Clusters (partnerships between universities, industry and government) where resources are shared and new technology can be tried and tested will be a good way of ensuring a business remains informed [48].

The last major source of collaboration will be through outsourced capabilities. The on-demand economy and cloud services offer small business the opportunity to scale and pick-up necessary skills as and when they need them. This leads directly to the theme of digital capabilities. Companies will ultimately remain competitive by developing digital capabilities that will allow them to remain relevant keep up with the manufacturing efficiency and quality of the world market.

Digital capabilities speak directly to the resource-based view of competitive advantage, where above normal opportunities are produced through the management of internal resources such as knowledge and machinery. For small businesses, obtaining all digital capabilities is not necessary. Rather capabilities which align to the business model should be actively sought after and developed [47]. First mover advantage will be important. A business that starts to acquire digital capabilities early on will build up internal knowledge on these capabilities much faster than a company that delays. With the rate at which industry 4.0 will bring about change, there may be a critical point where learning and development needs to happen faster than what a business can actually manage. Business should develop digital capabilities as soon as possible to avoid being left behind.

6. CONCLUSIONS AND RECOMMENDATIONS

Two of the biggest challenges that face the South African manufacturing sector is the lack of skills necessary to implement industry 4.0 along with a lack of exposure to industry 4.0. South African industry 4.0 learning factories, which simulate an industry 4.0 environment as far as possible, could potentially address these challenges and should be explored by relevant stakeholders. These factories should be set up within clusters so that knowledge and understanding is shared across academia, business and government.

Cloud computing also has the potential to alleviate some of the challenges faced by SMEs. [16] It can potentially be used by the government to offer mass support to SMEs (within all industries) through the development of free software packages that assist in core business functions such as accounting, finical management, legal and labour requirements.

7. REFERENCES

[1] **Hausmann, R.,** 2011. The Atlas of Economic Complexity, *The Atlas of Economic Complexity,* USA, pp 1-91.

[2] **Kaldor, N.,** 1968. Causes of the Slow Rate of Economic Growth of the United Kingdom: an Inaugural Lecture, *Louvain Economic Review,* 34(2).

[3] **Stellenbosch University**, 2016. *The Small, medium and micro enterprise sector of South Africa,* Stellenbosch, Bureau for Economic Research.

[4] **OECD**, 2017. OECD Economic Surveys: South Africa, *OECD Economic Surveys, South Africa*, pp 7-64.

[5] **Drath, R.,** 2014. Industrie 4.0 - Hit or Hype?, *INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS Industrial Electronics Magazine*.

[6] **Macdougall, W.,** 2014. Industrie 4.0: Smart Manufacturing for the Future, *Germany Trade & Invest*.

[7] **Weng, J.,** 2015. How Data Analytics would Transform High Tech Semiconductor Manufacturing in the Era of Big Data, *Semicon South East Asia*

[8] **Republic of South Africa**, 1996. *National Small Business Act*.

[9] **T.T.G of the Policy Board for Financial Services and Regulation**, 2007. *SME's Access to Finance in South Africa: A Supply-side Regulatory Review*.

[10] **OECD**, 2015. *OECD Economic Surveys: South Africa*.

[11] **Singer, S., Amoros, J. E. & Moska, D.,** 2015. *Global Entrepreneurship Monitor - 2014 Global Report*.

[12] **Ladzani, M. & Fulufhelo, N.,** 2009. Support for rural small businesses in Limpopo Province, South Africa, *Development Southern Africa,* 26(2), pp 225 - 239.

[13] **Schwab, K.,** 2014. *The Global Competitiveness Report 2014-2015,* World Economic Forum.

[14] **Wiklund, J. et al.,** 2009. Building an integrative model of small business growth, *Small Business Economics*, 32(4), pp 351-374.

[15] **Kholopane, P.,** 2016. Boosting competitive advantages of small and medium manufacturers in South Africa by applying continuous improvement and operational strategies, *Portland International Conference on Management of Engineering and Technology: Technology Management For Social Innovation, Proceedings,* pp 1442-1447.

[16] **Kumalo, N. & Van Der Poll**, J. A., 2016. The role of cloud computing in addressing SME challenges in South Africa, *International Conference on Cloud Computing Research and Innovation*.

[17] **Aulbur, W., CJ, A. & Bigghe, R.,** 2016. *Skill Development for Industry 4.0*.

[18] **Andulkar, M., Le, D. T. & Berger, U.,** 2018. A multi-case study on Industry 4.0 for SME's in Brandenburg, Germany, *Proceedings of the 51st Hawaii International Conference on System Sciences,* pp 4544-4553.

[19] **Wang, H.-L.,** 2014. Being Practical with Theory A window into Business Research, *Theories for competitive advantage,* pp 33-43.

[20] **Newbert, S.,** 2008. Value, rareness, competitive advantage, and performance: A conceptual-level empirical investigation of the resource-based view of the firm, *Strategic Management Journal,* 29(7), pp 745-768.

[21] **Bergen, M. & Peteraf, M.,** 2003. Scanning dynamic competitive landscapes: A market-based and resource-based framework, *Strategic Management Journal,* 24(10), pp 1027-1041.

[22] **Karnani, A. & Wernerfelt, B.**, 1987. Competitive Strategy Under Uncertainty, *Strategic Management Journal,* 8(2), pp 187 - 194.

[23] **Zack, M.,** 1999. Developing a Knowledge Strategy, *California Management Review,* 41(3), pp 125-145.

[24] **Grant, R.**, 1991. The resource-based theory of competitive advantage, *California Management Review,* 33(3), pp 114-135.

[25] **Maier, R. & Remus, U.**, 2002. Defining process-orientated knowledge management strategies, *Knowledge and Process Management*, 9(2), p 103.

[26] **Singh, H. & Dryer, J.**, 1998. The relational view: Cooperative strategy and sources of interorganizational competitive advantage, *Academy of Management Review,* 23(4), pp 660-679.

[27] **McGrath, R. G.,** 2013. Transient Advantage, *Harvard Business Review,* 91(6), pp 1-10.

[28] **Herman, M., Pentek, T. & Otto, B.,** 2016. Design Principles for Industrie 4.0 Scenarios. *Proceedings of the Annual Hawaii International Conference on System Sciences,* pp 3928-3937.

[29] **Samad, T. & Annaswamy, A.,** 2011. the Impact of Control Technology, *Institute Of Electrical and Electronics Engineers Control Systems Society 1,* p 246.

[30] **Zhang, Y. et al**, 2013. High Fidelity Visualization of Cyber-Physical Systems, *International Journal of Modeling Simulation and Scientific Computing,* 4(2).

[31] **Meola, A.,** 2016. What is the Internet of Things (IoT)?, *Business Insider*.

[32] **Chebudie, A., Rotondi, D. & Minerva, R.,** 2015. Towards a definition of the Internet of Things (IoT), *Institute of Electrical and Electronics Engineers Internet of Things, 1.*

[33] **Elgendy, N. & Elragal, A.,** 2014. Big Data Analytics: A Literature Review Paper, *Advances in Data Mining Applications and Theoretical Aspects,* pp 214-227.

[34] **Symantec**, 2016. *Smarter Security for Manufacturing in the Industry 4.0 Era,* Mountain View.

[35] **Bughin, J. et al.**, 2017. *Artificial Intelligence - The Next Digital Frontier?,* London, Mckinsey Global Institute.

[36] **Srivastava, A.,** 2014. A Detailed Literature Review on Cloud Computing, *Asian Journal of Technology & Management Research,* 4(2).

[37] **Gorecky, D., Schmitt, M., Loskyll, M. & Zuhlke, D.,** 2014. Human-machine-interaction in the industry 4.0 era. *Institute of Electrical and Electronics Engineers.*

[38] **Lorenz, M. et al.,** 2015. *Man and Machine in Industry 4.0.,* Boston Consulting Group.

[39] **Andulkar, M. et al.,** 2018. A multi-case study on Industry 4.0 for SME's in Brandenburg Germany, *Proceedings of the 51st Hawaii International Conference on System Sciences*, pp 4544-4553.

[40] **Aulbur, W. & Bigghe, R.,** 2016. *Skill Develoment for Industry 4.0.*

[41] **Wohlin, C.,** 2014. Guidliness for snowballing in systematic litertaure studies and a replication in software engineering, *Proceedings of the 18th International Conference on Evaluation and Assesmnet in Software Engineering*, pp 1-10.

[42] **Braun, V. & Clarke, V.,** 2006. Using thematic analysis in psychology, *Qualitative Research in Psychology*, 3(2), pp 77-101.

[43] **Kumar, R.,** 2011. *Research Methedology: A step by step guide for begineers,* 109(1), pp 5-10.

[44] **Fugard, A. & Potts, H.,** 2015. Supporting thinking on sample sizes for thematic analyses: a quantitative tool, *International Journal of Social Research Methedology,* 18(6), pp 669-684.

[45] **Braun, V. & Clarke, V.,** 2006. Using thematic analysis in psychology, *Qualitative Research in Psychology,* 3(2), pp 77-101.

[46] **MacQueen, K., Guest, G. & Namey, E.,** 2012. Introduction to applied thematic analysis, *Applied Thematic Analysis*. Sage, pp 3-20.

[47] **Geissbauer, R., Vedso, J. & Schrauf, S.,** 2016. *Industry 4.0: Building the digital enterprise.*

[48] **KPMG**, 2016. *The Factory of the Future.*

PUBLIC-PRIVATE SUPPLY CHAIN INTEGRATION AS A POSSIBLE MEANS TO IMPROVE PUBLIC HEALTH SUPPLY CHAINS

J. Botes[1], W. Bam[2] & I.H. de Kock[3]

[123]Department of Industrial Engineering
Stellenbosch University
[1]16496698@sun.ac.za, [2]wouterb@sun.ac.za, [3]imkedk@sun.ac.za

ABSTRACT

Struggling health supply chains and poor health outcomes in developing countries, have highlighted the need to improve these supply chains. A number of different methods have been used to improve health supply chains. However, it has been argued that the results are not sufficient and sustainable, neither do they aim to resolve impending challenges. In this paper, we put forward that public-private supply chain integration may be an important tool towards improving public health supply chains. However, further research is required to establish tools that support the determination to improve the supply chains and it. Moreover, research is required to determine whether improvements can be accomplished, as well as what the impact would be on the health supply chains in specific circumstances.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University.
*Corresponding author

1. INTRODUCTION

Supply chains are an essential component of the health system because they are responsible for providing a continuous supply of quality, affordable products to locations that are accessible to the country's population [1]. In order to effectively prevent, identify and respond to diseases, health supply chains must be able to respond rapidly and reliably when sourcing and distributing medicines, equipment and other health commodities [2]. Health supply chains also provide information regarding the supply and demand of products to planners and policymakers, who in turn ensure that the system has adequate supply and resources [1].

Health supply systems, including procurement, distribution and warehousing, must be able to meet the dynamic public health and security needs and demands [2]. According to Barillas [3], the success of a health system is determined by the availability of pharmaceuticals. Pharmaceuticals are critical to the health system because the expiration of medicines incurs extra costs, while the failure to deliver medicines, delayed deliveries or insufficient stock costs people their lives [4].

However, in many developing countries the problem is not finding methods to prevent or cure the high burden of disease, but to deliver quality health commodities at healthcare facilities on time, in the right quantities, at a reasonable cost to patients who need it [5]. A number of solutions have been applied to improve public health supply chains in developing countries. However, various authors have pointed out that these solutions may not be sustainable, proactive and sufficient [4, 6-8] and that there is a need for innovative solutions [9], not only for current health supply chain problems but also for impending supply chain problems [7, 10].

This paper explores the problems that public health supply chains are currently facing in developing countries, the methods that are being used to improve public health supply chains and some of the key shortcomings of these methods. The paper then suggests a research agenda that may further support the continuous improvement of public health supply chains.

2. LITERATURE REVIEW

Poor functioning health supply chains can severely damage the health system and hamper health outcomes [1] and result in higher costs, expiration and wastage of products, redundant efforts, stockouts and subsequently poorer health outcomes [11]. Unfortunately, this is common amongst the majority of developing countries [6]. Many countries lack the required resources to ensure the availability of health products [6]. This includes a poor functioning or absent logistics information system and the lack of transport and storage resources [6].

Reliable and complete data is required for the effective management of health supply chains [4]. This data is used to quantify how much of each product is needed by the population and, therefore ensures that there is adequate stock available at each health facility [4]. However, in many developing countries the lack of, or inadequate data for forecasting and supply planning [1, 4, 11], results in unresponsive supply chains [11], stock-outs and supply chain inefficiencies [4].

In many developing countries there is also a lack of trained staff [6]. Often, supply chain tasks are given to workers who are either not trained to perform these tasks, or who do not have the necessary qualifications [4, 6]. Consequently, at the lower levels of the health supply chain, clinical staff have to spend valuable time on supply chain tasks instead of serving and helping patients [6].

Furthermore, health supply chains are strained due to the increasing demand for health services and increasing volumes of products that need to be provided by the public health sector [2, 6, 7]. According to Allain et al. [6], the demand for increased health services in turn increases the complexity of health supply chains, resulting in higher costs. Funding from donors has risen, which further pushes up the volume of products that flow through the supply chains [7].
While health supply chains are trying to meet these needs, consumers, partners and civil society demand greater cost effectiveness and improved performance [7]. Meeting the higher demand for health commodities and services requires more robust regulatory oversight, so that the quality of the increasing number of products can be ensured. Thus, additional pressure is placed on both the public and private sector [2].

Most public health systems have multiple vertical, programme-specific supply chains, each of which receive different levels of attention and funds from donors [6]. Some ministries of health develop these separate vertical supply chains specifically for each health programme. Thus allowing for investments, information and supply chain activities, such as procurement, forecasting and product flow, to be better aligned with specific program outcomes and objectives [6]. In addition, some investors have focused on developing alternative supply chains to compensate for poor performing supply systems, or to respond to bigger health sector reforms [7]. However, only a handful of public health supply chains are able to accommodate various parallel supply initiatives [6].

Vertical supply chains are often criticised by some health reformists to be counterproductive and inefficient in strengthening the public health system [12]. Moreover, these supply chains are criticised for promoting fragmentation [13]. This fragmenting results in the duplication of resources and effort [14], and increasing the complexity of the supply system [6]. International agencies, such as the World Health Organization, PATH [15], VillageReach [14] and USAID [8], have been involved in projects that aim to merge some, or all, of the vertical programme supply chains. The expectation is that redundancies will be reduced and that performance and efficiencies will improve [6]. Other factors that have been increasing the complexity of public health supply chains are: (i) The fundamental operational changes that some governments are making to the health system [6]; (ii) an increase in the number of decision makers, financing options and stakeholders who have contributed to the decentralisation of government services [7]; and (iii) the poor design of health supply chains, which are often established based on assumptions that are outdated [6].

Other challenges faced by public health supply chains include a lack of adequate transport, limited geographic reach, malfunctioning cold chains, insufficient warehousing and distribution, and frequent stock outs [1, 6, 11]. In summary, public health supply chains do not have the capacity and resources required to supply an ever increasing volume and complexity of products [1].

Numerous international initiatives and donors have saved many lives [10] by providing much-needed funds [6] and facilitating the distribution and purchasing of various medicines, such as HIV, tuberculosis and malaria medication [10]. However, the funding is often uneven, usually allocated for specific diseases or programmes, and resources that are distributed by health programmes may not always be shared effectively. Thus resulting in underutilised resources [6]. The International Finance Corporation [10] argues that it is unclear whether donor funds are a sustainable and sufficient solution in addressing future health problems. Other initiatives have dedicated a lot of effort and investment to addressing these health supply chain challenges, in the form of supply chain strengthening and capacity building [4, 7].

Capacity building is defined as “the development of the ability of individuals and organizations or organizational units to perform functions effectively, efficiently and sustainably” [4, 16]. There have been some successful cases where capacity building and systems strengthening have improved the efficiency [4]. Moreover, the performance of pharmaceutical supply chains also improved, which resulted in an increased availability of health commodities [7]. Examples include countries such as Tanzania, where the Medical Stores Department operates at a level that is better than expected. In addition, Zambia and Rwanda’s pharmaceutical supply chains have also been reported to run efficiently [4]. Nevertheless, capacity-building efforts, aimed at improving supply chain efficiency, have been to a large extent unsuccessful [4]. For example, the Malawian Central Medical Store continuously encounters, problems despite receiving fulltime technical assistance from the Global Fund over a period of two years [4]. According to Bornbusch et al. [7], the improvements from capacity building and supply strengthening are “tenuous”[1] and suggest that public health leaders should ask whether current solutions are working as well as they need to, and whether current solutions will be able to solve future challenges. Similarly, [4] argues that the capacity building approach needs to change due to the poor effectiveness of these projects. However, [4] does not state how the approach should change, or how it can be improved.

On the other hand, Bornbusch et al. argue that the core competency of the government is not the operation of supply chains because numerous governments lack the expertise required to operate an efficient supply chain [7]. Moreover, there is are no career structures to facilitate high performance in supply chain workers, and thereby enable professionalism [7]. Instead, the government should play a stewardship role where it is responsible for providing the necessary guidance, oversight and vision to ensure that health supply chains achieve results [7]. Within a stewardship role, it is not mandatory for the government to directly control facilities and services. However, it is the government’s responsibility to engage and coordinate various actors in order to collectively achieve common development goals [7].

Currently, health reform seems to favour the implementation of supply chain integration in order to improve the overall health system, as well as the efficiency of health supply chains [12]. In the health sector community, the term ‘supply chain integration’ can refer to two types of integration [17]. The first involves the merging of programme or disease-specific supply chains, which are also referred to as “product integration” [17], or “horizontal integration” [18]. The second involves the “integration of information flows, physical flows, and financial flows between a firm and its supply chain partners” [19]. When speaking about supply chain integration, it generally refers to the second type of integration. Integrated supply chains have six characteristics, namely: (i) Clarity of roles and responsibilities; (ii) agility; (iii) streamlined processes; (iv) visibility of information; (v) trust and collaboration amongst actors within the supply chain; and (vi) alignment of objectives [20, 21]. While product integration does reduce redundancy, improve efficiency and reduce complexity in the public health system, it does not improve product availability [8, 20]. Supply chain integration improves supply chain performance, reduces costs and improves customer service [8]. However, according to the USAID, the implementation of either product integration or supply chain integration is not sufficient, rather, both and other approaches should be implemented [8]. On the other hand van Olmen et al. [22] argue that although theory may

[1] Tenuous definition: Very weak or slight; very slender or fine; insubstantial (Oxford dictionary)

suggest that merging vertical supply chains into a single, central supply chain increases the efficiency, it also creates weak links within the chain in areas such as haphazard ordering systems, stock management and slow distribution. According to van Olmen et al. [22] these weak links could result in the weakening of the entire chain.

Although many health supply chains have been substantially improved through both product and supply chain integration, as illustrated in case studies provided by [20, 21, 23], some questions remain. For example, questions regarding whether supply chain integration will be able to solve future supply chain challenges, considering factors such as the increasing number and volume of products [2, 6, 7], the increasing health service demand [6], demographic changes such as population growth [9] and increases in the burden of disease [9]. Looking at the documents that supply chain integration advocates, such as USAID, WHO, JSI and PATH, provide, there is no indication as to what resources are required for such integration endeavours. The resources required will vary from country to country [21]. However, the case studies by [20, 21, 23] support the notion that substantial financial and physical resources are required to implement supply chain integration. The questions surrounding the resources include: What resources are required and how do developing countries, that already lack financial and physical resources, obtain the necessary resources to implement these approaches? Do they rely on donors and NGOs? If developing countries do manage to integrate their supply chains, how can these supply chains further improve to accommodate increased volumes, populations, diseases?

According to Bornbusch and Bates [12], private sector supply chain research and application suggest that multiplicity in supply systems is the preferred method for improving efficiencies. Multiplicity involves "structuring a supply system to take advantage of multiple supply chains or segments to reduce risk and maintain supply" [12]. Multiplicity has been partially applied in public health supply chains in the sense that the public sector may take advantage of the private sector's strengths to improve supply chains, usually through public-private initiatives [1, 11]. However, from the existing literature it is clear that the public-private initiatives have not been applied to take advantage of entire supply chains as suggested by [12]. In contrast, public-private initiatives are applied to specific problematic or underperforming supply chain areas [24, 25]. In addition, most public-private initiatives predominantly occur in disease-specific programs [26], due to the increased external funding for vertical disease-specific supply chains and programs. This funding is opposed to the increase in available resources that are at the public sector's disposal [26]. This means that improvements through public-private initiatives are targeted to specific supply chains and rely on external funding.

In order for the public health system to ensure the availability of health products, while the burden of disease increases and health supply chain face the above mentioned challenges, more efficient supply strategies need to be found [9]. This means that decision makers will need to adopt new frameworks and models [6], as well as innovative delivery systems [9]. According to [9], it is crucial to investigate the private sector's role in providing increased service levels.

3. ALTERNATIVE RESEARCH AGENDA

Given the current context of health supply chains in developing countries, it can be argued that the health sector should take a proactive approach to addressing the supply chain challenges and improving health outcomes. We suggest that an alternative approach may be used to overcome these problems. However, very little research has been conducted with regards to this approach. We argue that the health sector, and academics, investigate an approach involving horizontal supply chain integration and horizontal supply chain collaboration - in public and private health supply chains. This approach may be referred to as 'public-private supply chain integration'. The meaning of horizontal supply chain integration is briefly discussed, followed by a discussion of horizontal supply chain collaboration along with the benefits of these approaches. Next, a definition of public-private integration is provided, and the reasons for public-private integration are included. Lastly, the final research agenda is presented together with the challenges that this approach may present and how they may be overcome.

3.1 Horizontal Supply Chain Integration

There are two types of horizontal integration, namely forward and backward integration [27]. Forward horizontal integration involves the collaboration with other organisations that provide substitute options, for example an organisation can provide road transport in addition to rail transport [27]. On the other hand, backward horizontal integration involves the integration of a company with other similar companies, for example a retailer can work with a second retailer or own a second retailer [27]. However, usually horizontal integration refers to the consolidation of companies [28], either by merging with or acquiring a competitor [27], that is in the same supply chain tier (i.e. manufacturer and manufacturer or distributer and distributer) [28]. Before continuing with horizontal integration, a decision needs to be made regarding what exactly will be integrated and how [18]. There are two options when deciding what should be integrated, namely products or processes [18]. Supply chain functions include procurement, forecasting, transport, information systems, orders, storage and transport [18]. Product integration involves the last two functions, storage and transport, whereas process integration involves the other functions that consist of supply chain management processes [18].

In terms of integration, supply chains can either be fully integrated or integrated via segmentation [18]. Full integration occurs when the multiple supply chains are essentially merged into one supply chain [18]. Segmented integration involves grouping the products according to specific characteristics and delivering according to these characteristics [18]. For instance, products that need to deliver to a specific location may be integrated. The benefits of horizontal supply chain integration include economies of scale, increased flexibility and adaptability, improved efficiency and improved performance [18]. However, it should be noted that these benefits and methods were achieved by integrating parallel public supply chains and not supply chains from stakeholders who are competitors, as we are suggesting.

3.2 Horizontal Supply Chain Collaboration

Horizontal supply chain collaboration occurs when “unrelated or competing organisations, producing similar products or different components of a product, form a cooperative association to share resources such as warehouse space and manufacturing capacity” [29, 30]. Many companies optimise and improve their supply chains to a point where further improvement efforts do not yield significant improvements or cost savings [31]. However, when companies participate in horizontal supply chain collaboration far greater improvements are achieved in efficiency and sustainability [31].

Horizontal supply chain collaboration is a relatively new research field [31] and so far the focus has been predominantly on horizontal collaboration in transportation and logistics management [32]. Vanovermeire et al. [31] investigated the viability of implementing horizontal logistics collaborations in supply chains, through the analysis of a case study. In the case study three companies currently deliver their own products. However, 57% of the orders are delivered to a customer that is common to two or all of the companies. Vanovermeire et al. [31] examined the costs of: (i) each company delivering their products individually; (ii) each company delivering the products after internal optimisation has been carried out; and (iii) the cost of the three companies collaborating with one another. It was found that internal optimisation achieved a cost saving of 13.65%, whereas collaboration between the three companies saved up to 25.83% in costs, and the number of trips decreased by 26.58% [31].

In the logistics area of supply chain, horizontal collaboration can improve the efficiency between 10 and 30% [31]. Horizontal supply chain collaboration could lead to the following outcomes [27, 31, 33, 34]: (i) Economies of scale when delivering to customers and decreased logistics costs; (ii) increased service levels, which result in more frequent deliveries and increased throughput; (iii) increased market share, which provides shared opportunities for new customers; (iv) shared investments; (v) sustainable logistics due to the efficient use of transport; and (vi) the exchange of best practices and innovation. An increasing amount of companies are forming horizontal collaborations in the logistics area of supply chains, where orders are consolidated into a common transportation channel [31]. Some companies take collaborations a step further by sharing assets, such as warehouses, collaborating on supply chain decisions according to a shared strategy, and harnessing the additional bargaining power to achieve economies of scale [31]. Ultimately these companies create one large supply chain [31].

According to Björnfot and Torjussen [35], supply chains need to incorporate structural flexibility in order to overcome and manage increased demand and uncertainty in markets. In addition, Björnfot and Torjussen argue that horizontal collaboration improves a supply chain’s structural flexibility and stabilises the market. Shared resources and capabilities enable this flexibility [33].

Horizontal collaboration presents its own challenges [31]. These include the risk of divulging information, cultivating a relationship based on trust, dividing gains amongst partners and absence of IT support as well as case studies [31]. One of the biggest risks of horizontal collaboration, according to academics and experts, is determining how the gains from the collaboration will be shared amongst partners [31]. According to a literature review by Cruijssen et al. [35], barriers to horizontal collaboration include selecting the right partner, negotiations and coordination, and the adoption of information and communication technology. Horizontal collaboration is long-term in nature and requires commitment and trust from all partners [31].

3.3 Public-private Supply Chain Integration

The following definition of supply chain integration was created by combining elements from horizontal supply chain integration and collaboration [11, 36–38]: Supply chain integration is defined as two or more autonomous supply chains (in this case public and private pharmaceutical supply chains) that work together. The integration is supposed to (i) improve their collective efficiency and effectiveness; (ii) find synergistic combinations of resources; and (iii) find solutions to problems, that each supply chain may not achieve on its own, by constructively exploring their differences and combining expertise from different organisations within the supply chains.

According to Donato et al. [2], countries that can take advantage of the public and private sector strengths have supply chains that are more adaptable, effective and more immune to disease outbreaks and epidemics. For example, the private health sector in countries of the Organisation for Economic Cooperation and Development (OECD) are leveraged to a much greater extent to accomplish increased effectiveness. On the other hand, the private sector involvement is significantly less in health supply chains in developing countries [1].

In contrast to health supply chains in developing countries, health supply chains in OECD countries largely depend on the private sector to provide services such as distribution, supply and other auxiliary services to complement a predominantly public health sector [1]. These supply systems effectively provide a continuous supply of pharmaceutical products to health facilities [1]. This led to the deliberation of how greater private sector involvement and leveraging, may improve public health supply chains in low- and middle-income countries. Many authors, such as Nishtar, Tennyson, Prybil et al., Kula and Fryatt, have argued that the public and private health sectors cannot address current and emerging health problems individually, but that the two sectors should work together [39–43]. Public-private collaboration is therefore imperative and unavoidable [39].

The private sector needs to be part of the solution if health supply chains are to react responsively to the current dynamic environment [44]. Engaging with private sector providers can improve the effectiveness of supply chains and consequently improve the health system [26]. Private sector initiatives can increase the efficiency and reliability of supply chains among all sectors and improve disadvantaged communities' access to products. Thus, contributing to public supply chain challenges [1]. Private sector initiatives can enhance the efficiency and effectiveness of supply chains, lead to the adoption of private sector best practices and expand the private sector's reach [1].

According to Bornbusch et al. [7], the government needs to recognise that public health supply chains encompass numerous other supply chains that comprise of a multitude of actors, from facilities to distributors of the public and private sector, NGOs and faith-based organisations. This collection of supply chains and actors may be a complex system. However, if the system is adequately understood and managed, Bornbusch et al. [7] argue that the supply chains can be "woven into a rationally integrated system". As a result, governments may have the flexibility and option to reduce or even eliminate distributors, funders, suppliers, procurement arrangements and quality assurance since all actors will be working together to improve health outcomes [7].

The supply chain functions that will remain the government's responsibility include the regulation of pharmaceuticals, policymaking, developing the overall supply system strategy and vision, managing the public sector's expenditures and supervising the health system [7]. This type of supply chain integration is similar to the concept of multiplicity suggested by Bornbusch and Bates [12], who identify the need for research to determine how multiplicity can be built into public health supply chains in order to optimise the performance, cost and risk management. The role of multiplicity in public health supply chains is crucial as it will assist future systems with the ability to handle increasing volumes of products [12]. However, public-private integration does not only have to benefit the public sector. In countries such as Ghana, Tanzania and Kenya, to name a few, the private for-profit sector has saturated their target market of high-income earners [45]. Subsequently, the private sector now aims to provide products and services to lower income groups [45]. Integration can expand the private sector's channels to lower income groups, while simultaneously improving access to health products [1]. It has been established that a number of private sector initiatives generate a profit, which indicates that the private sector has the opportunity to improve the health system while making a profit by engaging with the public sector [1]. However, the government will need to effectively regulate the private sector in order to manage its profit motive [1].

3.4 Research Agenda

The public sector has been involved in numerous engagements with the private sector. However, currently most of the engagements have been created in disease-specific programmes or with the focus of solving specific problems within health supply chains. We argue that the integration of health supply chains, as defined in section 3.3, may be a possible solution to proactively address supply chain issues and improve health outcomes.

Health supply chains from the various sectors may be integrated and the strengths of each sector harnessed to address current and future health and supply chain problems. This potential solution will likely prove to be very complex to implement, for the reasons discussed in section 3.3. Very little research has been done on the integration of end-to-end supply chains, as well as horizontal supply chain collaboration in the health sector. Therefore, we identify crucial areas for further research framed around the following questions:

- Can the public and private sector health supply chains be integrated (as defined in section 3.3) to address supply chain issues?
- Is it a feasible solution for developing countries that lack resources?
- What requirement will be needed to achieve public-private integration?
- What is the implication of public-private supply chain integration?

The suggested research agenda may be ambitious, yet it will undoubtedly be worthwhile investigating if it means that the access, availability and affordability of health products will improve as a result of better functioning supply chains. it may not be feasible from the outset for the supply chains to integrate at all points, there is a need to identify the costs, benefits and risks for integrating different aspects. Starting with limited sections of the supply chain can reduce the complexity somewhat and may provide information about how to go about the integration process in future as well as where these supply chains can or should integrate, including the impacts that integration has for the two supply chains.

4. CONCLUSION

The lack of resources along with an increase in the burden of disease, population and volume of health products in developing countries have led to poorly performing health supply chains and low health outcomes. Some solutions have not been as successful as hoped while others focus on disease specific supply chains or specific problems. Although improvements have been made, various authors argue that the solutions are not good, sustainable or proactive enough. There is a need to not only solve current supply chain problems, but also adapt supply chains for future supply challenges that may occur. This paper has suggested that public-private supply chain integration may be a potential solution to the problem. This paper also highlighted the need for further research to determine the feasibility of the solution, the implications of the solution and how it may be implemented. Further research may enable developing countries to deliver much needed improvements in the efficiency and effectiveness of health supply chains and thereby improve health outcomes.

REFERENCES

[1] **DGDA and MIT-Z ILP**., 2008. Private Sector Role in Health Supply Chains: Review of the Role and Potential for Private Sector Engagement in Developing Country Health Supply Chains, 13.

[2] **Donato, S.,Roth, S.,and Parry, J.**, 2016. Strong supply chains transform public health, *Asian Development Bank Briefs*, (72), pp 1–8.

[3] **Barillas, E.**, 2005. Pharmaceutical System Performance within the context of Health Sector Reform, *Rational Pharmaceutical Management Plus Program*. Management Sciences for Health, Arlington, VA.

[4] **Matowe, L.**, 2015. Improving Pharmaceutical Supply Chain Management Systems in Resource-Limited Countries: Time to Chage Approaches to Capacity Building, *Journal of Pharmacy and Pharmaceutical Sciences*, 4(2), pp 56–59.

[5] **WHO**., 2007. Everybody's Business: Strengthening Health Systems to Improve Health Outcomes. WHO's Framework for Action. World Health Organization Press, Geneva.

[6] **Allain, L.,Goentzel, J.,Bates, J.,and Durgavich, J.**, 2010. Reengineering Public Health Supply Chains for Improved Performance: Guide for Applying Supply Chain Segmentation Framework, *USAID | DELIVER PROJECT, Task Order 1*. Arlington, VA.

[7] **Bornbusch, A.,Dickens, T.,Hart, C.,and Wright, C.**, 2014. A stewardship approach to shaping the future of public health supply chain systems, *Global Health: Science and Practice*, 2(4), pp 403–409.

[8] **USAID**., 2009. Putting Integration into Perspective: Proven Practices to Strengthen Public Health Supply Chains. .

[9] **World Health Organization**., 2011. *The World Medicines Situation 2011*, 3rd Edition. World Health Organization Press.

[10] **International Finance Corporation**., 2017. Private Sector Pharmaceutical Distribution and Retailing in Emerging Markets: Making the Case for Investment. pp 1–15.

[11] **United Nations Commission on Life-Saving Commodities Technical Refrence Team**., 2015. Private Sector Engagement: A Guidance Document for Public Health Supply Chains. .

[12] **Bornbusch, A. and Bates, J.**, 2013. Multiplicity in public health supply systems: a learning agenda, *Global Health: Science and Practice*, 1(2), pp 154–159.

[13] **O'Hanlon, B. *et al.***, 2017. Exploring Partnership Opportunities to Achieve Universal Health Access: 2016 Uganda Private Sector Assessment in Health. .

[14] **Village Reach**., 2015. Exploring New Distribution Models for Vaccines and other Health Commodities Adapted to the on the Ground Realities of the Equateur and Tshuapa Provinces , Democratic Republic of Congo.

[15] **WHO and PATH**., 2013. Integration of vaccine supply chains with other health product supply chains: a framework for decision-making. Seattle, WA.

[16] **Kaplan, A.**, 2000. Capacity Building: Shifting the Pradigms of Practice, *Development in Practice*, 10, No. 3(10th Anniversary Issue), pp 517–526.

[17] **USAID**., 2011. Supply Chain Integration: Seamlessly Linking the Pieces. .

[18] **PATH**., 2013. Integrating the supply chains of vaccines and other health commodities.

[19] **Rai, A.,Patnayakuni, R.,and Seth, N.**, 2006. Firm Performance Impacts of Digitally Enabled Supply Chain

Integration Capabilities, *MIS Quarterly*, 30(2), pp 225–246.

[20] **John Snow Inc.**, 2012. Getting Products to People: The JSI Framework for Integrated Supply Chain Management in Public Health. .

[21] **USAID.**, 2011. Supply Chain Integration: Case Studies from Nicaragua, Ethiopia, and Tanzania. .

[22] **van Olmen, J. *et al*.**, 2010. *Analysing health systems to make them stronger*, 27. .

[23] **World Health Organization and PATH.**, 2013. Optimize: Senegal Report, Seattle.

[24] **USAID.**, 2010. Emerging Trends in Supply Chain Management: Outsourcing Public Health Logistics in Developing Countries, Arlington, VA.

[25] **UNICEF.**, 2016. A Process Guide and Toolkit for Strengthening Public Health Supply Chains through Capacity Development. .

[26] **Kaboru, B. B.**, 2012. Uncovering the potential of private providers' involvement in health to strengthen comprehensive health systems: a discussion paper, *Perspectives in Public Health*, 132(5), pp 245–252.

[27] **Aboutalebi, R.**, 2015. Strategies for Effective Worldwide Supply Chains, in *Handbook of Research on Global Supply Chain Management*, IGI Global, pp 1–15.

[28] **Huemer, L. and Furlan, A.**, 2011. Re-Conceptualizing Integration Strategies and Positioning Choices: Beyond the Upstream-Downstream Dimension, in *27th IMP Conference*.

[29] **Simatupang, T. M. and Sridharan, R.**, 2002. The Collaborative Supply Chain : A Scheme for Information Sharing and Incentive Alignment, *The International Journal of Logistics Management*, (February), pp 1–32.

[30] **Soosay, C.**, 2010. Assessing Cold Chain Integration, Innovation and Information Sharing, in *Pioneering Solution in Supply Chain Management: A Comprehensive Insight Into Current Management Approaches*, Schmidt Erich Verlag, pp 225–240.

[31] **Vanovermeire, C.,Sörensen, K.,Van Breedam, A.,Vannieuwenhuyse, B.,and Verstrepen, S.**, 2014. Horizontal logistics collaboration: Decreasing costs through flexibility and an adequate cost allocation strategy, *International Journal of Logistics Research and Applications*, 17(4), pp 339–355.

[32] **Soosay, C. and Hyland, P.**, 2015. A decade of supply chain collaboration and directions for future research, *Supply Chain Forum: An International Journal*, 20(6), pp 613–630.

[33] **Sanchez Rodrigues, V.,Harris, I.,and Mason, R.**, 2015. Horizontal logistics collaboration for enhanced supply chain performance: an international retail perspective, *Supply Chain Management: An International Journal*, 20(6), pp 631–647.

[34] **Cruijssen, F.,Dullaert, W.,and Fleuren, H.**, 2007. Horizontal Cooperation in Transport and Logistics: A Literature Review, *Transportation Journal*, 46(3), pp 22–39.

[35] **Björnfot, A. and Torjussen, L.**, 2012. Extent and Effect of Horizontal Supply Chain Collaboration Among Construction SME, *Journal of Engineering, Project, and Production Management*, 2(1), pp 47–55.

[36] **Axelsson, R. and Axelsson, S. B.**, 2006. Integration and collaboration in public health: A conceptual framework, *International Journal of Health Planning and Management*, 21(1), pp 75–88.

[37] **Foreman, S. E. and Roberts, R. D.**, 1991. The Power of Health Care Value-Adding Partnerships: Meeting Competition through Cooperation, *Hospital and Health Services Administration*, 36(2), pp 175–190.

[38] **National Research Council,Committee on Supply Chain Integration,Commission on Engineering and Technical Systems,Division on Engineering and Physical Sciences,National Academy of Sciences,and Board on Manufacturing and Engineering Design.**, 2000. Executive Summary, in *Surviving Supply Chain Integration: Strategies for Small Manufacturers*, National Academies Press.

[39] **Nishtar, S.**, 2004. Public – private "partnerships" in health – a global call to action, *Health Research Policy and Systems*, 2(1), pp 5.

[40] **Tennyson, R.**, 2011. *The partnering toolbook: An essential guide to cross-sector partnering*. .

[41] **IFC World Bank.**, 2011. *Healthy Partnerships: How Governments Can Engage the Private Sector to Improve Health in Africa*. World Bank Publications.

[42] **Prybil, L.,Jarris, P.,and Montero, J.**, 2015. A Perspective on Public-Private Collaboration in the Health Sector, *National Academy of Medicine*. pp 1–13.

[43] **Kula, N. and Fryatt, R. J.**, 2014. Public-private interactions on health in South Africa: Opportunities for scaling up, *Health Policy and Planning*, 29(5), pp 560–569.

[44] **John Snow Inc.**, 2016. Getting Products to People: How Private Sector Solutions Can Strengthen Supply Chains for Public Health. .

[45] **O'Hanlon, B. and Jeffers, J. B.**, 2013. Reference Guide for Development of Public Private Partnerships in the Health Sector for Countries in the SADC Region. .

TRAIN DRIVER AUTOMATION STRATEGIES TO MITIGATE SIGNALS PASSED AT DANGER ON SOUTH AFRICAN RAILWAYS

J.H. van der Merwe[1*] & B.P. Sunjka[2]

[12]School of Mechanical, Aeronautical and Industrial Engineering
University of the Witwatersrand, South Africa
[1]jhvdmerwe@gmail.com, [2]Bernadette.sunjka@wits.ac.za

ABSTRACT

Train derailments or collisions have the potential to result in catastrophic loss of life and/or destruction of property. Ever higher demands for train density (i.e. trains per hour for a given section of track) as well as the catastrophic results when accidents do occur have given rise to the development of railway signalling systems as mitigation measures.

Signals Passed At Danger (SPADs) refers to when a train driver passes a stop signal without authority and is one of the typical causes of such accidents resulting in significant damages reported within Transnet Freight Rail (TFR) in recent years. Studies have shown human train driver error and violation of signals to be a significant cause of SPAD events.

This study investigated the application of train driver automation as a mitigation measure against SPADs within the South African railway environment in general and TFR in particular. The study was qualitative in nature, following a model development methodology and used in-depth, semi-structured interviews with railway signalling engineers for data collection. The primary goal was defined to be the development of a train driver function automation method that could be considered the most appropriate within the TFR operational environment.

The study determined the most appropriate method to be that of having a human driver with technical supervision. In this arrangement, the human driver could remain in his conventional role of driving the train but with a technical supervision system superimposed that automatically intervenes if a train driver exceeds his movement authority (e.g. Automatic Train Protection or ATP). This approach mitigates many of the costs imposed by human failure associated with SPAD events, yet retains the value of human flexibility which is especially useful under abnormal circumstances.

*Corresponding author

1. INTRODUCTION

Since its construction in the 1860's, the South African railway network has grown to roughly 23 273 route kilometers [19, 26, 28]. Of the major South African railway operators, Transnet Freight Rail (TFR), Metrorail and Gautrain, TFR is the most prominent, representing 20 953 route kilometers of infrastructure ownership [26] and 60.5% of train traffic – 73.2 million train kilometers registered on average, anually, for the period 2013 to 2016 [20,21]. TFR railway infrastructure is located in both urban and rural areas and its operational control systems facilitate both freight and passenger trains [18], causing the TFR environment to best represent the South African railway context as a whole.

Railway transportation concentrates large quantities of people or goods on networks between various destinations. Inherent in this method of transportation also lies its great weakness - train derailments or collisions which have the potential to result in catastrophic loss of life and/or destruction of property. The 4490 operational safety incidents reported over the period 2013 to 2016 resulted in 1845 injuries, 461 fatalities and R469 964 458 in direct economic losses (annual averages) [21]. Ever higher demands for train density (i.e. trains per hour for a given section of track) as well as the catastrophic results when accidents do occur, have given rise to the development of railway signalling systems as mitigation measures [22, 25].

While the South African Railway Safety Regulator regulatory framework distinquishes between several categories of operational incidents [21], Signals Passed At Danger (SPADs) is the subject of this particular study. Railway signals, most often taking the form of traffic-lights located next to the railway line, are used to coordinate train traffic and protect train movements by communicating to train drivers their current movement authorities. In Figure 1, train B has received a proceed signal (Sig Y) to move onto the top line while trains A and C are ordered to stop by signals X and Z. These signals, and the drivers obeying them, are the only things protecting train B from a possible collision.

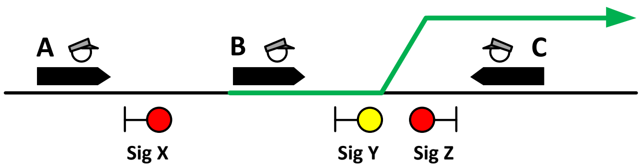


Figure 1 Signalling train movements

Signals X and Z are said to be "at danger" (meaning stop) and an insident where a driver passes such a signal is called a Signal Passed at Danger or SPAD [7]. SPADs are considered the most serious precursors to railway accidents ascribable to operator error [10]. The average number of unauthorized movements, including SPADs, reported on South African railways for the period 2013 to 2016 is 102.7 (annual average) or 1.4 SPADs per million train kilometers traveled (normalization parameter) [20, 21]. This compares quite favourably to Australia at 1.5 [16] but less favourably to Great Britain at 0.5 [13, 14, 15] for the same period, illustrating that South African SPAD levels are not especially high. Yet, of the top five most costly incidents reported for 2014/15, three were related to train handling (associated damages R93 million) with the most costly incident directly attributed to a SPAD (Maputo train-on-train collision – R56 million) [20]. These numbers indicate that while SPADs are not the most common incidents, the associated costs when they do occur can be relatively high, presenting a significant risk.

The Independent Transport Safety Regulator (ITSR) body within the New South Wales Government in Australia, concluded that the causes of SPADs generally include technical deficiencies associated with rolling stock or infrastructure but also driver error and violation of authorities [7]. Driver errors and violations typically include unsafe actions by the train crew such as the misreading of signal aspects, disregard for cautionary signals, incorrect braking technique, failure to communicate correctly and a range of external and internal distractions. The ITSR also includes organizational factors, such as poor safety culture, poorly designed procedures and inadequate monitoring or supervision, as contributors to such violations.

The study, therefore, proposed to investigate the application of train driver automation as mitigation to SPAD incidents in the South African railway environment in general and TFR in particular.

2. LITERATURE REVIEW

2.1 Strategies and countermeasures in SPAD risk management

Theeg & Vlasenko [25] describe the functions of train protection and control systems, as falling into three categories i.e.

- cab signalling functions that constitute driver warnings and information displays located in the train cab and that typically include audible signal warnings, visual repetition of track side signals and dynamic speed information;
- supervision functions that constitute the monitoring of train conditions and driver behavior that typically include checks in driver ability, attentiveness and compliance with speed limits; and
- intervention functions that constitute automatic interventions that typically include forced-response driver warnings, disabling of traction power and the application of train service or emergency brakes, to slow the train or bring it to a complete standstill.

Based on a study performed in Australia and New Zealand, Naweed et al. [12] defined the two general categories of formal countermeasures and informal strategies to mitigate SPAD risk in train driving. Formal countermeasures typically include forced-response devices that promote driver awareness to SPAD risk (Automatic Warning System - AWS) as well as automatic intervention systems (Automatic Train Protection - ATP). Informal strategies focus on the driver's intention to drive safely. These strategies are created by the drivers themselves to aid them in safely executing their duties – an exercise that can in and of itself be considered safety awareness or safety culture. According to Naweed et al. [12], these strategies became more important as systems like the AWS and ATP are increasingly not used as intended. Some drivers reported that they sometimes turn down the volume of the AWS system, considering it a distraction, even though such an action was against policy. Others reported using the ATP system as a general method of monitoring speed and not as a safety warning device.

The strategies identified, which were grouped as specifically applying to the context of service delivery or the context of the driver-signal dynamic were,

- **Assessment strategies** that focus on the driver's own assessment of the train he is driving and the driver monitoring of his own fitness to drive safely [12]. The first entails the personal assessment by the driver of the state of the train he is driving. This would include knowledge of potential mechanical issues such as braking dynamics that may affect train handling. The second strategy entails the driver's honest self-monitoring of his levels of emotional distraction and fatigue and reporting concerns about his fitness to drive prior to starting his shift.
- **Task prioritization strategies** that include the conscious decision to focus on tasks considered imperative to safe driving and the postponement of competing tasks [12]. A typical example would be ignoring an incoming call from the controller whilst navigating a difficult section in the route. When confronted with time table pressure, experienced drivers would consciously prioritize safety.
- **Automatic-attention strategies** that are built around cognitive processes and awareness strategies to assist drivers in automatically registering signals and changes in signals [12]. A typical example would be when a driver observes a signal switching from clear to caution, where that driver would then "turn on a switch" [12] in his head helping him to focus on the signal state and changes in that state. This becomes much more important in demanding situations.
- **Behavioral strategies** that include token actions that drivers devise as aids to help them overcome sighting restrictions, retain signal awareness and act as reminders during station dwells [12]. Such actions include consciously setting the direction switch to the neutral position during station dwells, or physically standing up in the cab when a caution signal is encountered.

To apply operator automation as a mitigation strategy in SPAD incidents in a South African railway context, the following areas of literature pertaining to train driver automation are explored:

2.2 Fundamental challenges to train driver automation

There are fundamental challenges or practical limitations to the concept of operator automation that should also be considered. These include certain ironic circumstances that have to be confronted when considering operator (train driver) automation as well the additional roles that train drivers fulfil above that of train control.

2.2.1 The ironies of automation

Bainbridge [1], a seminal author on the topic of operator automation, discusses the ironies faced by the system designer aiming to eliminate the operator from a control system due to his perceived unreliability or inefficiency. Bainbridge [1] considered it ironic that this same designer cannot control the errors that he himself introduces into the design of the automatic control system (designer unreliability). The operator then tends to end up with an arbitrary set of tasks that could not be easily or practically automated but that are also not really suited to the abilities of the operator (designer inefficiency). These tasks can generally be reduced to the categories of monitoring of the automatic system for correct operation and the intervention or "taking control" in the case of actual system failure. Bainbridge [1] suggested that if an operator is required to take control of a process and stabilize it, a certain measure of manual control skill is required. The same applies to fault diagnosis underpinning judgements on process shut down or recovery actions. Such judgements and actions require a certain measure of cognitive skill.

The manual control skills required to control a complex process can take a long time to master and deteriorate fairly quickly if not practiced regularly. This is evidenced by the fact that experienced operators tend to make step changes or transitions more smoothly, quickly and with fewer actions than less experienced operators [1]. More experienced operators tend to demonstrate an insight or intuition into what the effects of their actions will be. Contrary to this, inexperienced operators rely more on feedback and correction – a process that can be slow and result in significant oscillation or "back and forth" action. At the same time it will be more difficult for an inexperienced operator to be able to judge if the feedback received is the result of poor driving or system failure. It may require a very skilled operator to take over to cope without relying on the system. Train driving can be considered a good example of such a complex process. The driver has to adjust and control the speed of the train – a vehicle that has a lot of inertia and is, therefore, relatively unresponsive to his control actions. At the same time, the driver has to adjust for changing speed restrictions, keep to a strict time table, drive economically and avoid hazards [1].

The cognitive skills required, can be categorized into categories of "long-term knowledge" and "working storage". Long-term knowledge represents the methods and strategies developed over time, while engaged in the control of a process [1]. These methods and strategies are developed during the long run performance of control actions, considering the feedback and retaining what works and works well while rejecting those methods that work less well or not well at all. Unfortunately, this knowledge cannot be acquired without practice and deteriorates if not practiced [1]. Working storage represents knowledge about the current state and behavior of the process on a given day. A train driver may have to adjust to driving a different kind of train or a train with a different makeup affecting the handling. There may be workmen on the track this week or line of sight may be affected by the weather resulting in the driver adjusting his approach to driving. This kind of knowledge will not be available to the operator when control has to be taken in the moment of need [1].

Bainbridge [1] also suggests that before relegating the role of the operator to that of only monitoring an automated process, it should be remembered that the operator was to be replaced by the automatic system because he was perceived to be inferior at the control task to that very same automatic system. Now the operator will be required to only monitor and judge if the automatic system is working effectively. The irony of the situation is again evident when considering the monitor has to remain vigilant for an abnormality or failure event that rarely occurs. Owing to the fact that human operators are not always competent at this, they tend to rely on the alarm system to highlight abnormalities. This prompts us to ask: Is the alarm system working correctly and how should the monitor detect alarm system failure? This seems to be an impossible task [1]. It is also important to consider the impact the job of monitoring has on the operator's attitude. It can easily seem like a job requiring very little skill but a lot of responsibility. It has been shown that such circumstances are conducive to low job satisfaction, high stress, poor health and can also lead to increased error rates.

Even after being reviewed by Baxter et al. [2] thirty years after first publication, examples of Bainbridge's ironies [1] were found to still be prevalent and are expected to persist for some time to come.

2.2.2 The hidden roles of the train driver

In a study by Karvonen et al. [9] on the challenges presented in the full automation of the Helsinki Metro, the point was made that there is more to the train driver than basic train control. The case is presented that when the Helsinki Metro operation is considered as a whole, the driver is responsible for operating the train, taking care of passengers, observing events outside the train and acting positively in exceptional situations.

The driver's position in the train cab provides him with a direct view of the track, stations, platforms and passengers while operating the train. This places him in a unique position to anticipate, observe, interpret and react to events in that environment. Some of these events may require a speedy reaction such as when an obstacle is observed on the track ahead or when intruders or vandals are detected inside the security fence – observations and reactions that may prove difficult if performed from a remote location [9].

There are also certain tasks that are fairly trivial for a human being to perform but that can be extremely difficult for an automatic system to perform [9]. When considering the design of safety critical systems, it should be remembered that computers excel only at repetitive, basic tasks, and not at complex problem solving. A typical example of this limitation is passenger care. Passengers are independent actors whose actions may be unpredictable. A human train driver, therefore has a natural interface with human passengers. In the case of emergency, be it a medical condition of a passenger, train accident, or a fire, the presence of a human driver on site to calm other passengers, assist with orderly and speedy evacuation and to coordinate with emergency services is incomparable. In general, the train driver can act as the on-site representative of the railway company for passengers, clients or emergency services when the need arises [9].

Lastly, the train driver also acts as an important link with other actors in the metro system such as the traffic controllers, security guards and maintenance personnel [9]. His actions may include reporting of potential hazards, security risks and technical failures. The driver is also near at hand to fix small technical faults such as train door faults that can bring the whole operation to a stand-still if an unmanned model was pursued [9].

2.3 Theoretical Automation Framework

The theoretical automation framework of Theeg & Vlasenko [25] was selected for the study. It is considered preferable because it presents a more linear progression from total manual control to total automatic control while also resisting the impulse for what can be referred to as the "proliferation of gadgets" – the simple and unthoughtful addition of a multitude of buttons, sirens, flashing lights and other countermeasures to the train cab. Such measures often prove more disruptive than helpful to the driver in the execution of his duties. This framework presents a simple and theoretically elegant approach to train driver automation design in general, and is based on discrete steps of increased automation, as discussed below, in reverse order of automation:

2.3.1 Level 4 – Full automation

The train is normally driven automatically with no supervision from a human driver. In some cases a person that is normally charged with other responsibilities like ticket collecting, may be available to take control if needed [25]. At first glance, full automation seems to deliver on all of the perceived advantages of train driver automation: increased safety (driver induced SPADs) [12], increased cost effectiveness (personnel reductions and efficient driving profiles) and operational flexibility [9]. Yet, difficulties are soon revealed when the train control function is considered within the greater context of the railway operation, and operating in the real world. This perspective can be considered as the systems level or system engineer's view. It may not be that difficult to automate train control, but to automate the train driver may not prove as simple when considering the following:

When the inevitable technical failure does occur, what fall-back modes and recovery mechanisms can be put in place to maintain safety and not bring the whole operation to a stand-still? In such cases a ready and able human train driver is inevitably required to monitor the system for correct operation and take control when needed [1]. Increased automation may also not lead to a reduction in staff as a reduction in train drivers due to automation may only lead to increased requirements in support and supervisory staff and more sophisticated levels of training [9]. Then, there are certain functions that may be quite trivial for a human train driver to perform but almost impossible to automate. These include the ability of a human train driver to observe potential hazards and react to those hazards, passenger care under normal and emergency circumstances, debugging of small technical failures such as door problems and a whole range of interactions with operations and maintenance staff [9]. Automation Level 4 was therefore not considered feasible for this study.

2.3.2 Level 3 – Automatic driving with human supervision

The train is normally driven automatically with the driver observing and intervening in case of danger or technical failure [25]. Automatic driving with human supervision effectively counters most of the defects introduced with full automation but it is confronted by another kind of problem – that of having a fully capable human driver at hand to monitor and intervene when necessary. Bainbridge [1] asserted that, the manual control skills and cognitive skills required for train control will soon deteriorate in a driver if not practiced regularly.

A driver whose responsibility has been relegated to only monitoring the driving performance of a machine whose driving skills are perceived to be superior to his own may not be the right person for the job. He may also not have enough insight into the decision making process to judge when things are going wrong and require intervention. A driver assigned to monitor an automatic system may very well feel that his job requires little skill, that he has very little insight and control of what is happening, yet is held responsible for the outcomes. These working conditions have shown to lead to high levels of stress and unhappiness and increased error rates. Automation Level 3 was therefore not considered feasible for this study.

2.3.3 Level 2 – Partially automatic operation

Train driving tasks and responsibilities are divided between the train protection system and the driver. The train protection system is fully responsible for some tasks and the driver is fully responsible for the others [25]. Level 2 automation does not seem to counter the criticisms of Level 3 and was therefore not considered feasible for this study.

2.3.4 Level 1 – Manual driving with technical supervision

This involves a train protection system supervising the driver and enforcing safety in case of driver error [25]. Manual driving with technical supervision seems to be the most practical option because the driver is maintained in his traditional role mitigating most of the fundamental criticisms to automatic drivers faced by levels two, three and four.

In addition, the technical supervision system can be scaled to cover only the functions critical to safety and can therefore be much less sophisticated and less expensive than a fully functional automatic driver. Lastly, the technical supervision system can simply be superimposed over the current manual driving model and the implementation will therefore result in minimum disruption to ongoing operations.

2.3.5 Level 0: Manual driving without automation

The driver is fully responsible for driving and there is no train protection systems present [25].

2.3.6 Recommendations

While Level 1 automation (manual driving with technical supervision) is the recommended automation model to be implemented, it should be noted that, to improve the practicability of the new model in the TFR railway environment, the model should be based on the current signalling and operational practices within TFR.

Additionally, the formal countermeasures and informal strategies presented by Naweed et al. [12] align reasonably well with the two elements of probability and criticality that make up the European Committee for Electrotechnical Standardization (CENELEC) concept of risk [4]. The informal strategies and vigilance improvement systems attempt to reduce driver error rates while intervention systems minimise the effects if driver errors were made. Although it is desirable to try and reduce the probability of train driver error, the method seems to be somewhat open-ended as it cannot enforce safety. It does not mean that such methods are ineffective and should not be investigated or implemented or that intervention systems are not subject to failure [6, 23 cited in 12] but it can be argued that vigilance improvement systems do not present a solid foundation for safety control [10, 25]. Reduction in adverse consequences (mitigation) seems to be the most appropriate and accessible approach in this case as it closes the loop in terms of safety enforcement (the result is forced) and is based on technologies that are well established and well understood. These systems (Automatic Train Protection for example) are based on the monitoring of the human driver and the enforcement of safety by applying the brakes if the driver significantly exceeds the required speed restrictions or movement authorities [25]. In many ways such an approach constitutes having all of the advantages of a human train driver, which has been shown to be considerable [1, 9], whilst mitigating the disadvantages. It is therefore recommended that the new model approach be along similar lines to ATP systems.

Finally, it may be worthwhile to consider adding functions that aid the driver in better driving but such functions should not form the basis of safety.

3. RESEARCH METHODOLOGY

The study followed a qualitative approach, utilizing a model development methodology and semi-structured verification interviews with signalling engineers. Qualitative research is interpretative and aims to provide a depth of understanding [5]. Such models provide symbolic, textual or graphic answers to research questions, typically based on logic, theory or verbal descriptions [3]. This study opted for the development of qualitative models, descriptive of the structure and behavior of the system by which train movements are and should be authorized and effected within TFR.
System structure was represented using a simple block diagram scheme, illustrating the different components and role players within the system and how they relate to one another. System behavior was represented using the Enhanced Functional Flow Block Diagram (EFFBD) scheme. Conventional FFBD's are made of functional blocks, each representing a definite, finite, discrete action to be accomplished. The behavior model is developed using a series of diagrams to show the functional decomposition and to display the functions in their logical, sequential relationship. Diagrams are laid out so that flow direction is generally from left to right. Lines and arrows connecting functions indicate function flow and not lapsed time or intermediate activity. Common logical operations (i.e. concurrency, selection, iteration, repetition and loops) can also be implemented [11].

3.1 Data Collection

3.1.1 Pre-modelling concept development and review

The pre-modelling conceptual framework of South African signalling practices is based on the researcher's own knowledge and experiences as a signalling engineer. Reference was made to written records and technical literature where possible. To ensure that the framework was accurate, the resulting write-up was submitted to an experienced signalling engineer for review and comment.

3.1.2 Model development interview sample size and selection

The interviewee sample consisted of three railway signalling engineers (designated ENG1, ENG2 and ENG3), the the technical discipline specifically responsible for safety in railway movements or the creation of safe railway capacity. They were selected for their extensive experience (30+ years) as signalling engineers within the TFR railway operations, signalling and projects environment.

3.1.3 Model development interview planning, preparation and execution

Interviews were in-depth, semi-structured and consisted of the development and review of the current practice as well as proposed ideal models by expert signalling engineers. Three separate interviews were arranged – one with each engineer. One or two simple modelling examples were discussed to familiarize the interviewees with the modelling language. The background, motivations and recommendations upon which the models were to be based were also explained and discussed with the interviewees. Thereafter the engineers were required to develop proposed models and analyze them to confirm that the models met all the needs set out by the critical research question. The engineers, aided by the interviewer, were free to rework or modify their models until completely satisfied and before signing-off. Any qualifying statements or comments were also recorded in the process. Interviews were not audio recorded and transcribed verbatim. Interview notes, however, took the form of drawings, diagrams, descriptions and qualifying statements.

3.2 Data analysis

Data analysis took the form of the harmonization or consolidation of the interview models into a generic or representative model. This was possible due to the large degree of commonality between the models. Where the models did diverge, the diverging features were highlighted and discussed as possible options or customizations that could be applied on top of the generic model.

3.3 Study validity, reliability and ethics considerations

Various measures were put in place to asure the validity, reliability and ethics of the study. Validity measures included specifically interviewing experienced signalling engineers (experts in the subject of study) and utilizing a graphical modelling language instead of text based descritptions to increase clarity. Reliability measures included following an accepted research method (model development), arranging interviews at convienient times and places as well as interviewee anonymity. Interviewees were properly briefed beforehand and the whole process was accepted by the univerity ethics committee.

4. RESULTS AND DISCUSSION

4.1 TFR conceptualization

The purpose of this conceptualization was to develop an understanding of how railway movements are effected within TFR. Specific emphasis was given to the roles and functions signalling system actors (e.g. the train driver and Train Control Officer) and signalling system components (e.g. interlocking and field elements) play in those movements. This section presents a brief introduction to the railway signalling practices in TFR while highlighting the implications for SPAD incidents and causes.

4.1.1 Functional Structure: Typical TFR Signalling and Operational Control System

A typical TFR Signalling and Control System can best be visualized in layers stacked on top of each other connecting the Train Control Officer (TCO), who represents railway operational objectives such as the train movement schedule, at the very top to the trains, and signalling field elements at the very bottom. TCO controls are translated to the relevant station via the remote control system, is filtered through a safety interlocking layer after which it affects the desired change to the field elements (e.g. throwing a points set or changing a signal aspect) through field element control units. Field element status information (e.g. lay of a points set or train positions) is reported back to the TCO in similar fashion.

Figure 2 has been adapted from Trinckauf's [27] *Functional structure of the railway control system* to be more representative of TFR signalling and control.

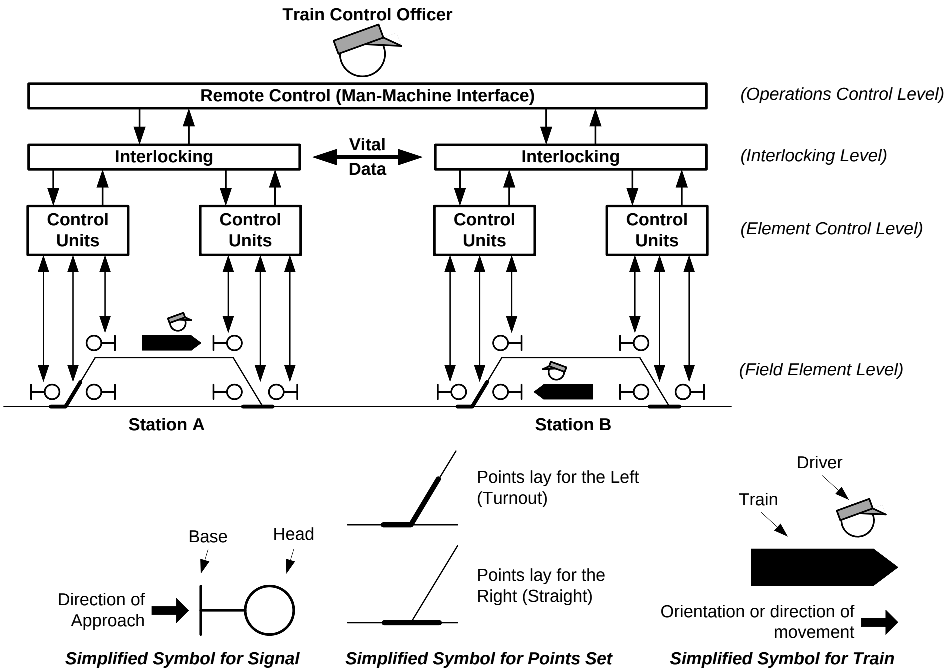


Figure 2 Typical TFR Signalling & Operational Control System

4.1.2 Principles of Train Separation

The most common train separation methods include Signalled Fixed Block Operation and Cab Signal Operation of which the Fixed Block is the dominant method on TFR signalled lines. On un-signalled TFR lines, constituting roughly 60% of all TFR lines, operational control is governed by a Track Warrant System (TWS) with movement authorities managed within the in-house developed CS90 remote control and Visual Display Unit (VDU) system.

4.1.3 Signals and Signal Aspects

On TFR signalled lines the dominant practice is that of color-light signalling with fixed block operation. Signals are used to guide a train though a railway network by communicating compulsory movement authorities to the train driver who is then expected to follow them to the letter. These communications include the following:

- Communication of movement authority into a block section including the nature of that authority (eg. permission to proceed and speed restrictions).
- Inform the train driver as to the upcoming features of the track ahead (eg. upcoming turnouts, entry into yards and sidings).

As many different possible meanings have to be conveyed to the driver, the signal has been equipped with several different lamps of various colors and arranged in several different configurations. These lamps can then be activated in predetermined combinations to indicate specific meanings called signal aspects. Figure 3 is a representation of the physical signal pole of a typical TFR signal that a driver would encounter when approaching a station [8]. The signal pole is populated with a green lamp, red lamp, white lamp, two yellow lamps, a signal identification plate and a shunt lamp-set (two small white lamps arranged diagonally). If all these different meanings or signal aspects, including a “dark” signal (no lamps burning), are summated, a signal with this physical configuration can be used to display up to thirteen different aspects [8].

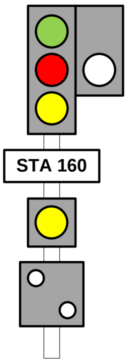


Figure 3 Color light signal (physical representation)

The consequences of misreading or even completely missing a signal can vary from the benign to the severe. For example, if a flashing yellow is misread as a solid yellow, the train may apply brakes and slow down too soon, simply resulting in an operational inefficiency. If the reverse happened (i.e. reading a solid yellow as a flashing yellow), the train will not slow down and prepare to stop at the stop signal ahead, possibly resulting in a train collision. The worst case would obviously be to miss a stop signal or to come upon a stop signal unprepared which may result in a train collision while traveling at line speed.

4.1.4 Interlocking Principles

TFR interlocking systems typically implement the "Protected route and Overlap" method. TFR interlocked routes can also be manually cancelled by the TCO, but if the train has already entered the route or occupied the approach track to the entry signal, the signal is put back to Danger (stop/red aspect) but the route remains reserved and locked for a predetermined time before normalizing [8]. This prevents endangering the train and allows the driver to bring the train under control and to a safe stop without fear of a collision or derailment due to points sets being thrown under the train.

4.1.5 Signalling a Layout (Crossing Station)

One of the simplest illustrative examples of signalled layouts within TFR is that of the crossing station. The purpose of the crossing station is to increase bi-directional traffic density over a single line, albeit at the added cost of additional operational complexity. Figure 4 illustrates a typical signalling layout for a crossing station with functional descriptions for the different types of signals below.

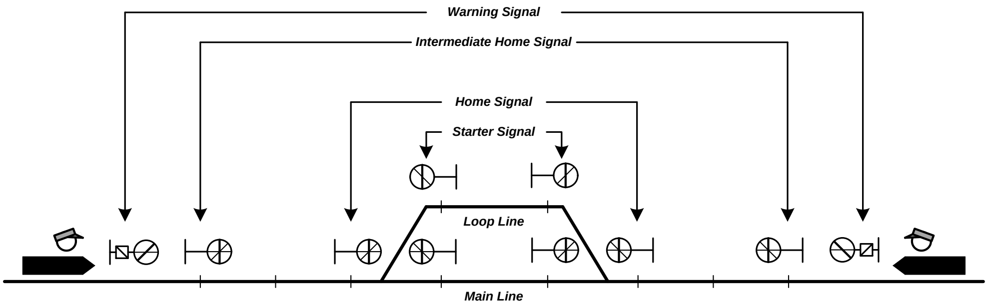


Figure 4 Signalled Layout (Crossing Station)

The functions of the indicated signals in this layout are as follows [24]:

> Warning Signal: Warns the driver that he is approaching a stop signal. Note this signal does not have a Stop aspect. This signal is sometimes erroneously referred to as a distant signal.
> Intermediate Home Signal: First stop signal when approaching the station and acts to protect the overlap of a route setup from the opposite side of the station.
> Home Signal: Stop signal before station entry. This signal is to regulate access to the station.
> Starter Signal: This signal has a dual purpose – it can either be used as a destination signal for a train entering the station (Danger Aspect) or as a departure signal for a train exiting the station (Proceed Aspect).

4.1.6 Dispatching Principles

Centralised Traffic Control (CTC) operation is the predominant model on TFR Signalled lines and is illustrated in Figure 5. In CTC operation, all points and signals inside the controlled area are directly controlled by the CTC

TCO. All train movements are governed by signal indications. The local interlockings are remote-controlled without local staff [17].

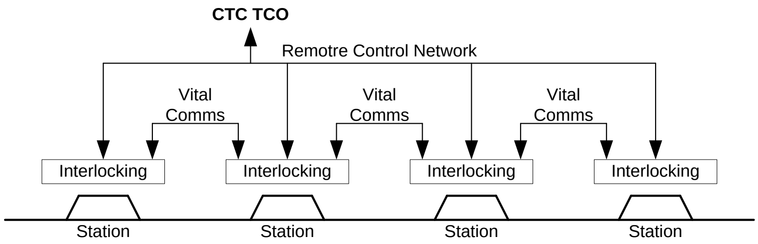


Figure 5 Centralized Traffic Control (CTC) [17, p. 60]

In a TFR Track Warrant System, movement autorities are communicated to train drivers directly via a radio or cellular network (Figure 6).

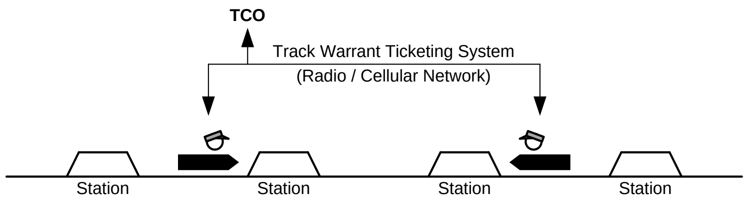


Figure 6 Dispatching by Track Warrant

4.2 Current Practice and Proposed Automation models (Consolidated)

Due to the high degree of commonality between the Current Practice models and the Proposed Automation models produced by ENG1 and ENG2, the interview format for ENG3 was rather directed towards the confirmation or rejection of those models. The consolidated or generic model was therefore generated and presented to ENG3 for endorsement. All three engineers agreed with the recommendations that flowed from the literature review. Both ENG1 and ENG2 based their automation models on Automatic Train Protection systems that intervene when the driver exceeds the required speed profiles while also promoting efficient driving patterns. These views were also endorsed by ENG3.

The first difference between the models generated by ENG1 and ENG2 was that of the level of detail. ENG2 preferred to define the system on component level where ENG1 preferred to stay on the system level. The consolidation process also preferred the system level as the component level may become too prescriptive about the specific technological implementation and move away from a neutral description of system behavior.

The second difference between the models spoke to the operational environment. ENG2 defined models for both the Signalled and Track Warrant operating environments while ENG1 felt that it was not necessary to consider train driver automation in areas where conventional signalling systems were not considered or warranted in the first place. The two models generated by ENG2 were very similar, only emphasizing the safety integrity ratings of the systems and equipment used in implementation. In this case as well, when consolidating to the system level, the models for the two operating environments merged into one.

The models as well as the consolidation process were discussed with ENG3 and the resulting model was accepted and endosed by ENG3. The resulting generic model consists of both a structural component (Figure 7) and a behavioural component (Figure 9). As the proposed automation model was based on the current practice, the current practice model elements are shown in grey while the modifications or improvements are indicated in blue – see Figure 9.

It should be noted that the actual models generated in the interview process did not always strictly adhere to the proper notation conventions associated with the EFFBD modelling language. As these models were intended to be primarily descriptive, these deviations were not considered problematic as long as the models were readable, understandable, logical and unambiguous. Per implication, these models may require a measure of reformulation if they are to be ported into software tools for simulation purposes.

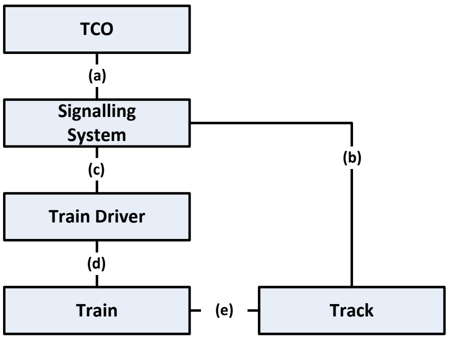


Figure 7 Generic model - structure

Interface listing:

a)	TCO – Signalling System	Computer based control terminal
b)	Signalling System – Track	Train detection device (axle counter / track circuit)
c)	Sig. System – Train Driver	Signal lamp
d)	Train Driver – Train	Driver instrument panel
e)	Train – Track	Wheel on track

The behavior model was based on the basic train movement on a signalled line - depicted in Figure 8.

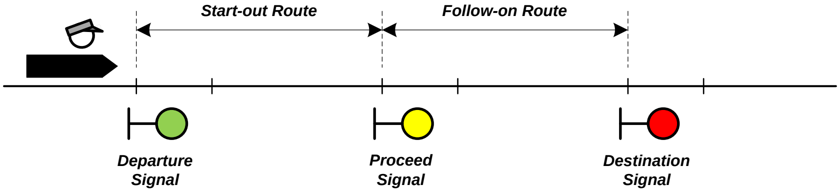


Figure 8 Train Movement

A train route was setup and cleared from the departure signal to the destination signal. The Solid Green departure signal is indicating to the driver that the train may proceed at line speed. The follow on signal displays a Solid Yellow aspect, indicating to the driver that the train may proceed but that the driver must be prepared to stop at the next signal (Red / Danger).

The behavior model (Figure 9) can be logically devided into the phases of Movement Setup, Route Traversal and End of Authority.

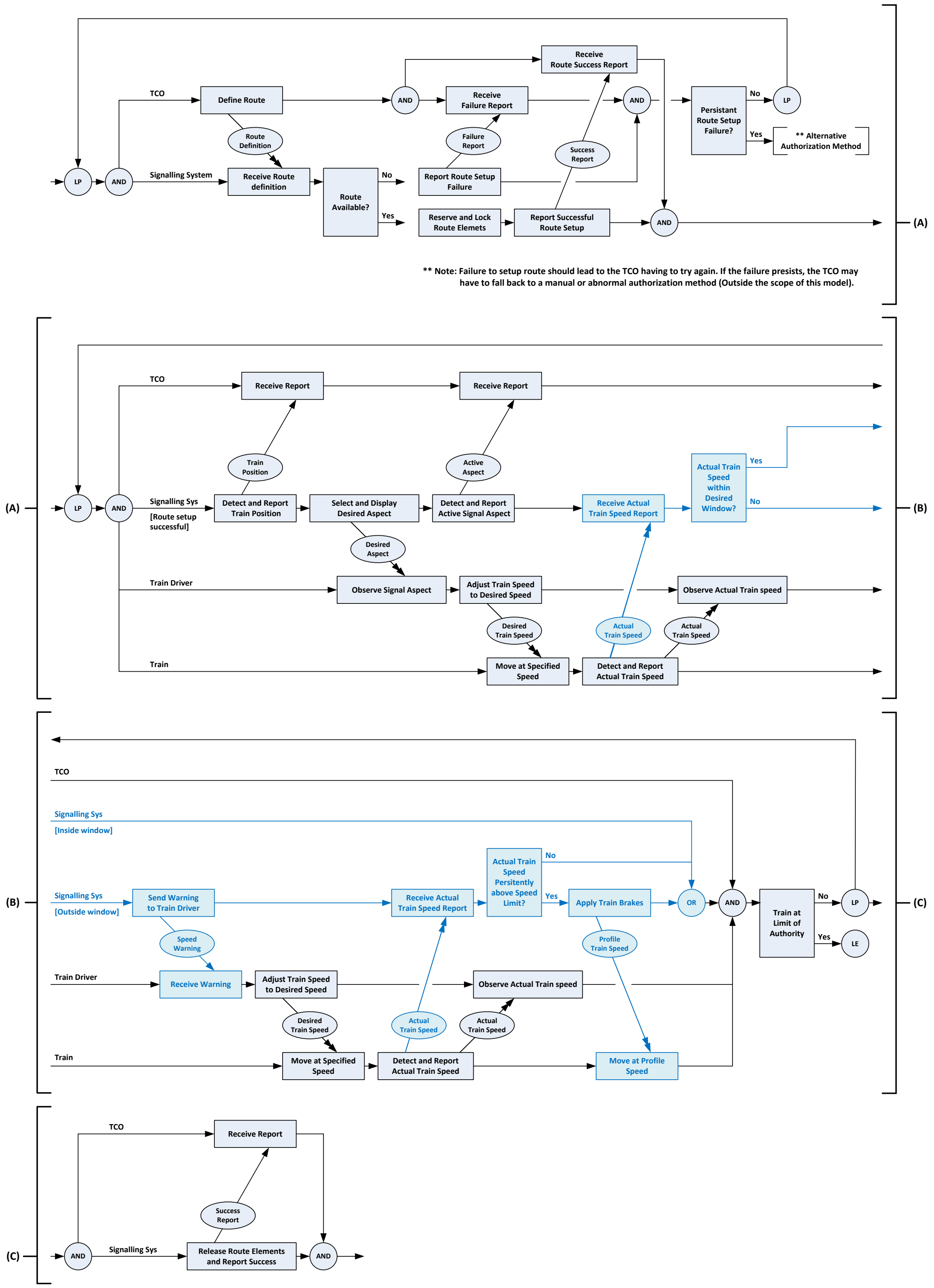


Figure 9 Generic model - behavior

Ultimately, all the engineers explicitly agreed that the resulting models successfully answered the following questions:

- Do you agree with the recommendations from the literature review and were those recommendations faithfully implemented in the proposed automation model?
- Would the resulting automation model be effective in mitigating SPADs and would you consider it to be the most appropriate train driver automation method for the TFR environment?

All of the interviewed signalling engineers answered in the affirmative.

5. CONCLUSION

Operator automation is a field that presents enormous potential advantages when it comes to the technological augmentation of human capabilities, yet it comes with significant engineering challenges as well. As has been evident in this study, this statement also holds true when considering the automation of the train driver within the TFR environment. At first glance, the completely automatic control of a train seems to be a simple process to implement, yet when train control is holistically considered within the greater context of railway operation (the systems engineer's view), there are significant challenges that become apparent.

A thorough review of the literature has shown that human train driver errors and violations significantly contribute as causes of SPADs (a fundamental failure in driving safety) yet it has also been shown that the additional roles and functions fulfilled by a human train driver (i.e. intervention and support under abnormal circumstances) are not as easily automated. The most practical and readily available train driver automation model, mitigating most of the costs imposed by human failure associated with SPAD events yet retaining the value of human flexibility, was demonstrated to be that of retaining the human driver yet with technical supervision. In this arrangement, the human driver would remain in his conventional role of driving the train but with a technical supervision system superimposed that automatically intervenes if a train driver exceeds his movement authority. In addition, such a system could be tailored to also guide the driver towards optimal driving profiles.

The recommendations from the literature as well as the resulting generic model were endorsed by the signalling engineers interviewed, affirming it met the requirements set out by the research question: That of being the most appropriate method of train driver function automation to mitigate SPADs in the TFR railway environment.

5.1 Futher study

The financial costs typically attached to safety systems and technologies such as Automatic Train Protection (ATP) can be significant and often prohibitive for large scale implementation and support in third world countries such as South Africa. In many cases (e.g. rural lines that carry predominantly freight traffic) the traffic densities and potential risk levels may not justify the costs associated with equipment carrying the Safety Integrity Level (SIL) ratings of SIL3 or SIL4 [4], typically associated with the high traffic density on metro lines. In such cases, the application of SIL1 or SIL2 systems may be sufficient if combined with soft strategies like the informal driver strategies presented by [12]. The formalization and scientific evaluation of the effects on risk reduction of such strategies and initiatives, may prove a fruitful field for further study.

REFERENCES

[1] Bainbridge, L. 1983. Ironies of Automation, *Automatica,* 19.6, 775-779.

[2] Baxter, G., Rooksby, J., Wang, Y., khajeh-Hosseini, A. 2012. The ironies of automation ... still going strong at 30?, *Proceedings of ECCE 2012 Conference,* pp. 65-71.

[3] Buede, D.M. 2009. *The Engineering Design of Systems: Models and Methods,* 2nd Edition, John Wiley and sons.

[4] British Standards Institution. 1999. *BS EN 50126-1:1999 Railway applications – The specification of Reliability, Availability, Maintainability and safety (RAMS) – Part1: Basic requirements and generic process,* BSI.

[5] Dudovskiy, J. 2011. Research Methodology, [Online] Available from: http://research-methodology.net/ [Accessed 16th November 2017].

[6] Edkins, G.D., Pollock, C.M. 1997. The influence of sustained attention on railway accidents, *Accident Analysis and Prevention,* vol. 29, No. 4, pp. 533-539.

[7] Independent Transport Safety Regulator (N.S.W.). 2011. Managing signals passed at danger, Independent Transport Safety Regulator.

[8] Jooste, J.P. 2013. Required Operational Capability for Transnet Freight Rail Signal Interlocking Systems, Transnet Freight Rail Technology Management.

[9] Karvonen, H. et al. 2011. Hidden Roles of the Train Driver: A Challenge For Metro Automation, *Interacting with Computers,* 23.4, 289-298.

[10] Kyriakidis, M., Hirsch, R., Majumdar, A. 2012. Metro railway safety: An analysis of accident precursors, *Safety Science,* vol. 50, no. 7, pp. 1535-1548.

[11] National Aeronautics and Space Administration (NASA). 2007. Systems Engineering Handbook, Rev.1, National Aeronautics and Space Administration.

[12] Naweed, A., Rainbird, S., Chapman, J. 2015. Investigating The Formal Countermeasures And Informal Strategies Used To Mitigate SPAD Risk In Train Driving, *Ergonomics,* 58.6, 883-896.

[13] Office of Road and Rail. 2017. Freight Train Kilometres by operator – Table 13.25, [Online] Available from: http://dataportal.orr.gov.uk/browsereports/13 [Accessed 16th November 2017].

[14] Office of Road and Rail. 2017. Passenger Train Kilometres by operator – Table 12.13, [Online] Available from: http://dataportal.orr.gov.uk/browsereports/12 [Accessed 16th November 2017].

[15] Office of Road and Rail. 2017. SPADs on NR controlled infrastructure – Table 5.25, [Online] Available from: http://dataportal.orr.gov.uk/browsereports/5 [Accessed 16th November 2017].

[16] Office of the National Railway Safety Regulator. 2016. Rail safety report 2015 – 2016, [Online] Available from: https://www.onrsr.com.au/publications/rail-safety-report [Accessed 16th November 2017].

[17] Pachl, J. 2009. Railway Operation Processes, *In:* Theeg, G., Vlasenko, S., *Railway Signalling & Interlocking,* Eurailpress, 2009, p. 39-60.

[18] Passenger Rail Agency of South Africa. 2013. Shosholoza Meryl home page, [Online] Available from: http://www.shosholozameyl.co.za [Accessed 16th November 2017].

[19] Passenger Rail Agency of South Africa. 2015. Annual Report 2014/15, [Online] Available from: http://www.prasa.com/Annual Reports [Accessed 16th November 2017]

[20] Railway Safety Regulator. 2015. State of safety report 2014/15, [Online] Available from: http://www.rsr.org.za/infohub/knowledgehub [Accessed 16th November 2017].

[21] Railway Safety Regulator. 2016. State of safety report 2015/16, [Online] Available from: http://www.rsr.org.za/infohub/knowledgehub [Accessed 16th November 2017].

[22] Rolt, L.T.C. 2009. *Red for Danger*, The History Press.

[23] Scott, A., Gibson, H. 2012. Red means stop, doesn't it? A human factors trial of a UK train safety system driver-machine interface, *In:* Dadashi, N., Clarke, T., Wilson, J., Rajan, J., Mills, A., *Rail Human Factors around the World,* CRC Press, 2012, pp. 457-464

[24] Spoornet. 2002. Train Working Rules, Spoornet.

[25] Theeg, G., Vlasenko, S. 2009. Train Protection, *In:* Theeg, G., Vlasenko, S., *Railway Signalling & Interlocking,* Eurailpress, 2009, p. 208-251.

[26] Transnet Freight Rail. 2017. Transnet Annual Results 2017: Freight Rail, [Online] Available from: http://www.transnet.net/InvestorRelations/Pages/Annual-Results-2017 [Accessed 16th November 2017].

[27] Trinckauf, J. 2009. Basic Characteristics of Railway Systems and the Requirements for Signalling, *In:* Theeg, G., Vlasenko, S., *Railway Signalling & Interlocking,* Eurailpress, 2009, p. 17-23.

[28] Van der Merwe, J. 2008. Gautrain from dream to reality, *Railway Gazette,* 8 July 2008 [Online] Available from: http://www.railwaygazette.com/news [Accessed 16th November 2017].

DEVELOPMENT OF A DECISION SUPPORT SYSTEM FOR ASSESSING ALTERNATIVE AGRICULTURE LAND USES: A CASE STUDY OF THE STELLENBOSCH WINE REGION

J. de Villiers[1*], W.G. Bam[2] & J. van Eeden[3]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
jeanne1566@gmail.com

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
wouterb@sun.ac.za

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
jveeden@sun.ac.za

ABSTRACT

The wine industry is accountable for 1.2 percent of the South African GDP. Financial margins of Stellenbosch wine estates have begun to shrink due to factors such as high production costs and increased competition. To be economically sustainable wine estates need to rethink their current business strategy and consider adopting a diversification strategy. This article identifies a holistic set of considerations that decision-makers in this industry need to evaluate when considering pursuing land use alternatives. It also considers how these factors can be used to develop a decision support system (DSS) to guide farmers through the decision-making process.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

According to Wines of South Africa [2] the South African wine industry dates back to 1655, when the first vine was planted. Marais [3] states that this makes it one of the oldest wine industries in the world apart from Europe and the Mediterranean. In 2016, there existed over 3200 farmers that cultivate 98 597 hectares of vineyard in South Africa [4]. The term wine farm (often referred to as a winery outside of South Africa) refers to a place where grapes are grown, fermented, blended and the wine that is produced from the grapes is bottled [5]. The South African wine industry contributes to the country's GDP and provides job opportunities. The wine industry is one of the biggest agriculture exporters and was responsible for 1.2% of the national GDP [1]. The wine industry, including wine tourism, supported 300 000 jobs (direct and indirect employment) and contributed R36.1 billion to the economy in 2013 [1]. Of the total contribution of R36.1 billion to the GDP, almost R20 billion (53%) was created in the Western Cape. By volume South Africa is the eighth-largest national wine producer in the world [1].

Operating a successful wine business encompasses the anticipation of trends, possible opportunities and apprehensions within the industry, as well as taking into account the views of peers [6]. Constant improvements, and thus changes to a current business strategy are of the utmost importance to keep up with the latest trends and to ensure economic sustainability and revenue growth of wine estates.

Most of South Africa's water sources are under strain and South Africa is accordingly categorised as a dry country [7]. From 2016, the Western Cape had been gripped by a prolonged drought, resulting in the implementation of water restrictions as of 1 November 2016. The Western Cape, facing its worst water shortage in 113 years, was consequently declared a drought disaster zone in May 2017 [8]. The growing pressure on profitability margins of the South African wine industry [1] together with the drought requires wine estates to re-evaluate their business strategies. Many wine estates have recently been investigating diversification opportunities. However, many farmers have limited knowledge and experience outside of the wine industry, making the consideration of alternatives more complex. Thus, there exists an opportunity to develop a decision support system (DSS) that will regard a set of considerations to provide farmers with support when they are assessing possible land use alternatives. Consequently, this study develops a DSS to support farmers who are seeking to adopt a diversification business strategy and are therefore looking for a set of considerations that they need to evaluate when considering an alternative land use option.

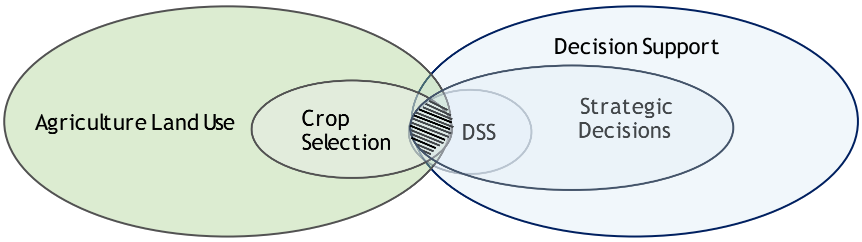


Figure 1: Research Domain

This research aims to identify and validate considerations to be used in the development of a DSS capable of assessing available land use alternatives in the Stellenbosch region to ensure financial success and economic sustainability.

2. METHODOLOGY

To solve the specified problem at hand, a literature study was done which defines strategic decision-making and determines whether decision support tools are applicable in the agriculture field. Specifically, literature was reviewed to determine the different DSSs that can be applied in the agriculture sector. Key considerations and DSS design requirements were considered before developing a novel DSS. The proposed DSS, which allows an end-user to provide tailored inputs for each of the identified considerations, evaluates and compares different selected land use alternatives with each other. After which an illustrative case study was utilised in order to evaluate the proposed DSS. The functionality and the considerations of the DSS were subsequently validated.

Five key stages of the methodology can be identified. These are shown and grouped according to different colours in Figure 2. According to those key stages, the study context is first outlined, after which an extensive literature review is conducted in order to identify and establish areas of importance within the relevant study fields. The information that is obtained from the literature review is subsequently used to inductively define considerations and design requirements by integrating the considerations and design requirements reflecting in existing decision support systems (DSSs). This enable the design of the best practice DSS. The fourth stage includes conducting research to provide context to the illustrative case study as well as using the illustrative case study to apply the developed DSS. Validating the proposed DSS, initially through doing an internal validation, and secondly by having interviews with experts and getting additional inputs from these experts by means of interview questionnaires concluded the fourth stage. The validation process included the research that gave context to the case study and the illustrative case study itself. These two parts as well as the validation of the DSS and the set of considerations were grouped together into one stage.

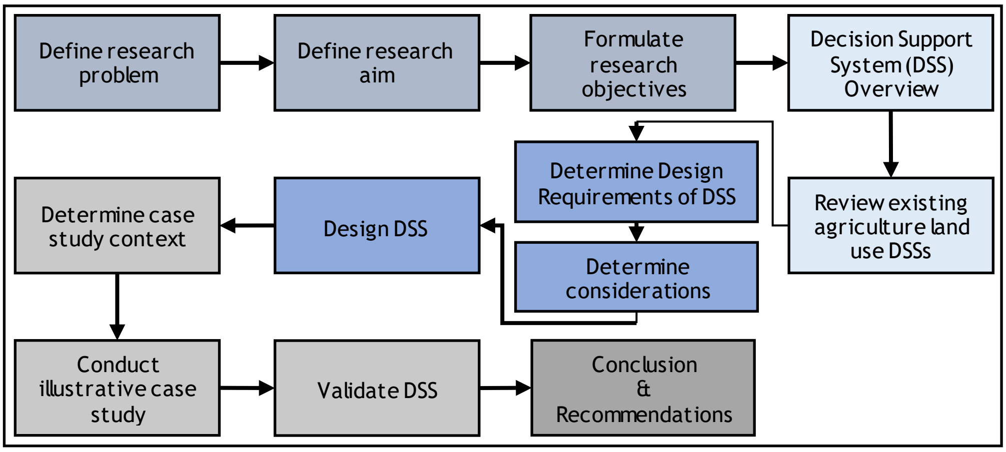


Figure 2: Research Process Flow Diagram

3. DECISION-MAKING SYSTEMS IN GENERAL AND IN AGRICULTURE

This section provides an overview of the literature pertaining to DSSs and the use of DSSs in the agricultural industry.

3.1 Decision Support Systems

Research conducted over the past two centuries has shown that humans assess information in a way that is far from rational [10]. Thus, a set of processes and analytical tools that support systematic and structural thinking especially when it involves difficult choices are involved or required. This set of processes and analytical tools, which are referred to as decision analysis, provides a method where a decision problem can be formed by separating the uncertainties, determining the subjective beliefs of the participants regarding those uncertainties, and then finally building a quantitative decision model [10].

Turban & Aronson [11] defines a DSSs as an 'interactive, computer-based systems, which help decision makers use data and models to solve unstructured problems'. DSSs support the analysis of current statuses or they can give future predictions, or both [12]. They can furthermore assist discussions, store data and models, stimulate learning, and advance internal capacity building [12].

More effective decisions can be made by using DSSs, by leading the user through specific decision stages and portraying the different possible outcomes from various alternatives. Farmers and farm managers can use these tools to efficiently facilitate farm management, by assessing different alternatives based on evidence [13]. Different crop management options and crops must be selected by farmers, which then allocate them to a specific field. These selections are critical, because they influence the productivity and short- and long-term profitability of the farm [14].

Jakeman, Letcher, & Norton [16] suggested an iterative process that consists of ten steps to develop and evaluate a DSS. They proposed certain general steps to ensure credible results and knowledge acquisition for the model, as well as for the community. The authors argue that some of their ten steps might involve the end user as well as the modeller.

Trust in the outcomes of the tool is another important aspect of a DSS that should be included according to [15]. McNie [17] refers to credibility as information that is accurate, valid and of high quality, as a consideration for useful information. Trust in the outcome of the DSS as well as the accuracy, validity, and quality of the information is incorporated into the DSS by consulting decision makers in the agriculture sector, thus validating the model.

3.2 DSSs in Agriculture

A need has been established for a DSS to provide help to farmers considering an agricultural diversification strategy. Especially some who want to evaluate possible land use alternatives that they can employ to stay viable. To support this current literature was reviewed to determine which agricultural focussed DSSs already exist. Table 1 provides an overview of only the leading agriculturally focussed DSSs found. This illustrates DSSs developed for crop selection, and the theory/work a particular DSS is based on.

Hartati & Sitanggang [18] argue that it is not only preferable to apply a DSS for efficient land suitability evaluation and crop selection problems, but also important. Moreover, the DSS assists the decision makers in comprehending the decision problem as well as the effect that their choices have on the enterprise, by allowing them to continuously exchange information between the system and themselves [19 - 21]. The complexity of these systems varies greatly.

Table 1: Decision Support Systems (DSSs) for crop selection

Author/Source	Based on	Focus area	Consideration
Radelescu & Radelescu [30]	Portfolio Theory	Financial risk	↓ Climate risk ↓ Market risk
Collender [31]	Mean variance analysis	Risk estimation	Mean variance characteristics
Salleh [32]	Fuzzy Modelling	Crop selection	Uncertainties during the development of the agriculture DSSs
Hartati & Sitanggang [18]	Fuzzy Modelling	Evaluate land suitability	Land characteristics
Balezentiene, Streimikiene, & Balezentis [33]	Fuzzy MULTIMOORA method	Sustainable energy	Climatic suitability, ↓ Environmental pressure
Nevo & Amir [34]	Rule-based expert system	Crop suitability	Severe uncertainties
Rossing, Jansma, De Ruijter, & Schans [23] van Ittersum, Rabbinge, & van Latesteijn [26] Makowski, Hendrix, van Ittersum, & Rossing [35] Ten Berge et al. [36] Dogliotti, Van Ittersum, & Rossing [37]	Multi goal linear programming	Soil	↓ Erosion, ↑ Organic matter ↑ Rate of change
Annetts & Audsley [23]; Dogliotti, Van Ittersum, & Rossing [38]; Bartolini, Bazzani, Gallerani, Raggi, & Viaggi [39]; Sarker & Ray [40]; Louhichi et al. [41]	Multi-objective optimization problems Process-based simulation model Empirical data	Profit	↑ Gross margin ↑ Annual profit ↑ Income ↑ Net benefit
Dogliotti et al. [23]; Bartolini et al. [39]	Process-based simulation model Empirical data	Labour	↓ Total labour ↓ Casual labour ↓ Cost
Annetts & Audsley [38]; Dogliotti et al. [23]	Multiple objective linear programming Process-based simulation model	Pesticides	↓ Herbicide use ↓ Losses ↓ Pesticide exposure

Model based land use studies should be used to inform debate on development pathways and get an understanding regarding future agriculture development opportunities [22] to help both the formulation of strategy policy objectives [23], as well as strategic planning by farmers, by using trade-offs between economic and environmental objectives [24 - 29].

As illustrated in Table 1 there exist many different DSSs that focus on different aspects in the agriculture field. The aim of this study, however, is to develop a set of considerations as part of a DSS that can provide assistance to decision makers when they are considering adopting a land use alternative. The DSSs identified in this section focused on a few aspects only, thus not taking the whole farming operation into consideration. Thus, a holistic set of considerations that will evaluate land use alternatives needs to be developed and incorporated into the model.

4. DSS DESIGN

Design requirements and a set of considerations have to be developed and need to be included in the design of the proposed DSS. The purpose of the DSS that was developed in this study can be defined as: to help the decision maker to choose suitable crops in a flexible and user-friendly manner, by allowing the user to provide specified input values to fully explore the relationship between the considerations and the land use alternatives. The design requirements discussed are therefore required to be of such a nature that they will make certain that the aim of the DSS is met as well as ensuring that an accurate reflection of the outputs of the proposed model is provided. According to Rose et al. [13] the following factors are important to ensure successful user acceptance of a DSS: 1) ease of use, 2) performance (the usefulness of the tool and whether it works well), 3) the cost of the DSS, 4) trust (whether or not the tool is evidence based), 5) IT education (whether the tool requires good IT skills to use), and 6) habit (whether the tool relates closely with current farming practices.

The design requirements for the proposed DSS in this specific study, taking the above-mentioned aspects (factors?) into consideration, can thus be stated as follows:

1. The proposed DSS is required to inform the end user which possible land alternatives are viable, given specified input values.
2. The proposed DSS needs to tell the end user which of the alternatives he or she could possibly invest in.
3. Land availability: The proposed DSS should be able to evaluate whether there is enough land available for a particular land use alternative to be viable.
4. Practicality: The proposed DSS should be user-friendly and inexpensive. Furthermore, it should be accessible to a range of different farmers.
5. The proposed DSS should include a combination of viable factors (economic, environmental, labour related, pests/diseases) to evaluate the suitability of a land use alternative for a specified region.
6. The prospective DSS should be flexible, thus addressing the limitation regarding the use of DSSs. It should also be efficient and effective.
7. Trust: the planned DSS should make use of accurate trustworthy data.
8. The risk associated with the different alternatives should be assessed.

Identifying a set of considerations that will give farmers assistance when they are considering any land use alternative type marked the first step in the developing process of the proposed DSS. The set of developed considerations is not just confined to one aspect of the farming operation, such as land suitability or climate suitability, but it rather takes the farming business as a whole into account.

The considerations were identified and established by conducting research about each of the selected land use alternatives. Inputs from subject matter experts have also been used in some cases. Each of the identified considerations, with their accompanying category, are shown in

Figure 3.

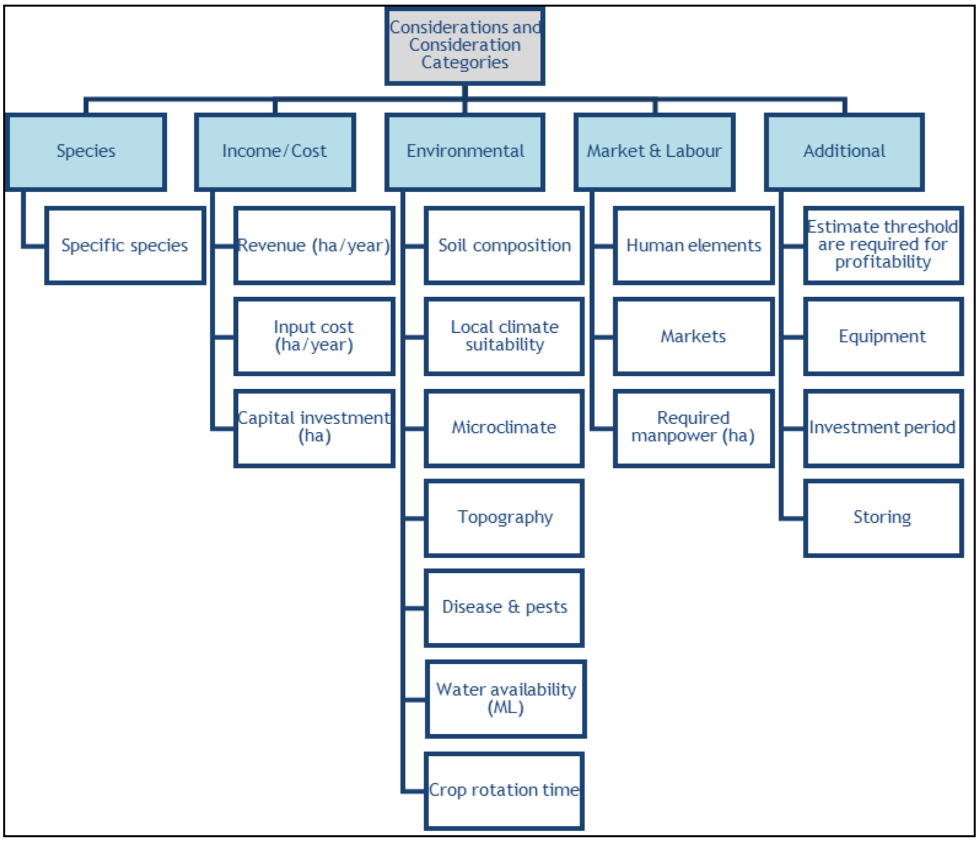


Figure 3: Identified Considerations and main Consideration Categories

The identified considerations play an important part in the feasibility of each of the different land use alternatives and are thus important to examine before implementing any alternative in an intended area. The set of developed considerations that was incorporated into the DSS, are crucial to review before adopting any new land use alternative. Therefore, farm owners should use the DSS as a guide to understand which considerations are important to regard, and subsequently which crops are best suited for their particular region, when they consider adopting a new agriculture diversification strategy.

5. DSS LOGIC

The logic of the DSS provides the reader with an understanding of how the developed DSS works. Process flow diagrams are used to describe the logic of the DSS. Figure 4 provides an overall picture of how the DSS works. It illustrates which information is used and what the DSS accomplishes.

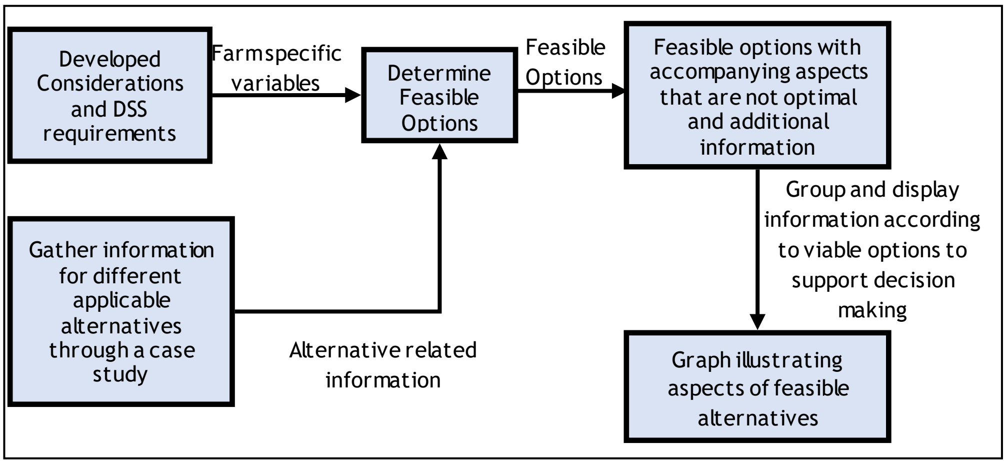


Figure 4: Overall process flow of DSS

Figure 5 shows the process flow of the extended model. The extended model uses the viable options obtained from the DSS as illustrated in Figure 4 together with specified user input. The user can provide the amount of land he or she wants for each of the viable options. Corresponding values of each of the other relative input values are then determined, based on the required provided hectares. A decision maker can thus determine the implications, e.g. cost, for specified hectares, as well as the total of a selected combination of options that he/she wants.

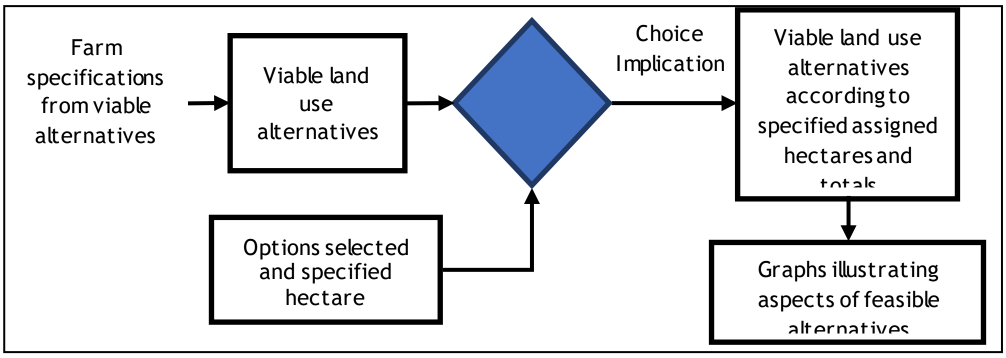


Figure 5: DSS extension process flow

The main purpose of the first developed model is to provide a decision maker viable land-use alternatives that could be adopted according to the user's input data. The information that the researchers gathered to populate the data sheet, together with the set of developed considerations form part of the design of the model. The decision maker is required to provide and fill in data for each of the developed consideration.

The developed DSS and DSS extension makes use of data to provide possible land use alternatives. The purpose of this subsection is to clearly outline the data inputs that the DSS model provides and which data the end user of the model provides.

5.1 Developed DSS data requirements:

The researchers conducted research and held interviews with experts in the agriculture field to populate the data tables that the DSS requires to function. The researchers used a step wise selection process to select a sample of land use alternatives for this study. The research and interviews with experts facilitated populating the data tables for each of the selected land use alternatives. The researchers also provided the annual local sales/exports, annual production, and price data for each of the land use alternatives that were selected for this study. The researchers further developed and provided the set of developed considerations, keys with accompanying meaning, and additional informative information which was included in the DSS. The end user of the DSS has to provide specific user inputs for each of the considerations that are built into the developed DSS. It is important that the decision maker supply this data so that the developed DSS can be tailored to that decision maker's farm/area.

5.2 DSS Extension Data Requirements

The extended model developed in this study, depends on the developed DSS. The model uses the outputs and by implication the data of the DSS. However, the end user is afforded the opportunity to input and compare different hectares allocation scenarios per land use alternative in order to evaluate the expected outcomes of each scenario. After the end user has assigned hectares to each of the alternatives, the DSS extension is programmed to automatically provide the rest of the outputs and generate graphs that are in accordance to the user assigned hectares. The extended model provides the user a choice to manipulate the assigned hectares to evaluate the implication of doing so. For this reason it is important that it is the end user that assigns the hectares to each of the possible displayed land use alternatives and not the DSS.

6. DSS DEVELOPMENT PROCESS

A stepwise process is followed to apply the developed DSS. The process makes sure that the land use options that are selected and intended to be incorporated into the DSS are viable in that specific area. The process steps are shown below and each of the process steps will be executed sequentially.

- Step 1: Identify land use alternatives for the DSS database
- Step 2: Filter Initial land use alternatives for the DSS database
- Step 3: Develop specific considerations for each selected land use alternative
- Step 4: Apply the developed DSS

Figure 6 shows the user interface of the developed DSS with an example of user input values. The input values illustrated are values that a decision-maker chooses, therefore tailored for a specific region. These specified input values are in turn used to provide viable land use alternatives that are applicable in a specific region. The viable land use alternative types which were obtained for the given user inputs of Figure 6, are illustrated in Figure 7. Each of the illistrated land use alternatives shown in Figure 7 are viable according to the tailored values of Figure 6. Thus, for the example input data shown in Figure 6, mandarins, cling peaches, or cabbage are shown as viable options in Figure 7. According to the developed DSS a decision-maker can choose any one of the abovementioned three alternatives for their specific region.

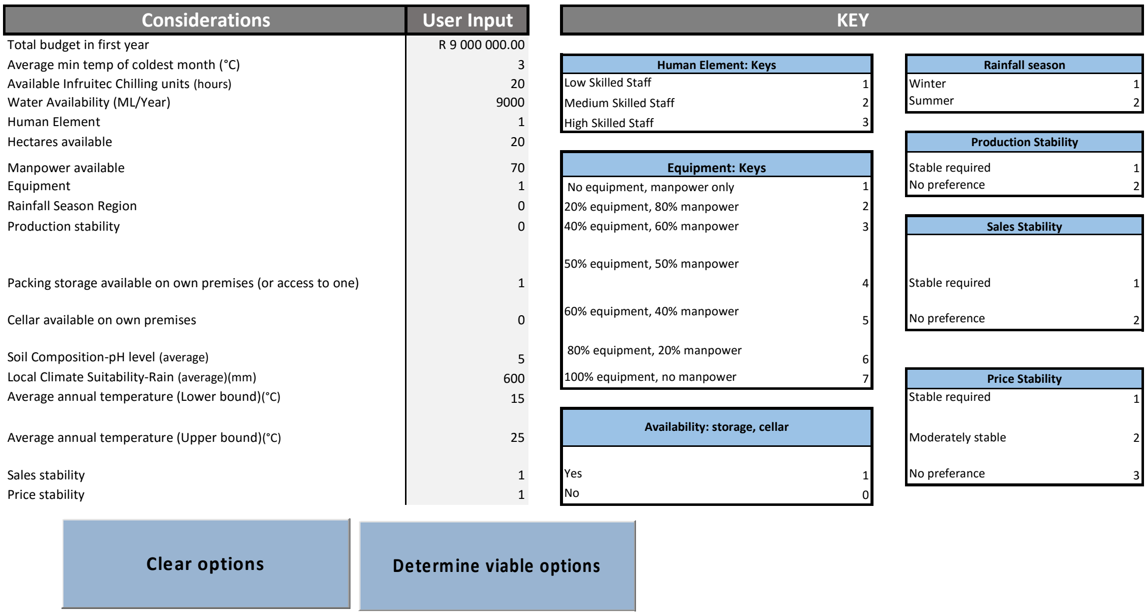


Figure 6: Considerations where user gives an input (left) accompanying user input keys (right)

The viable land use alternatives generated according to the user input of Figure 6 are used to create a graph (Figure 8) which compares the total cost in the first year with the expected gross income per viable option.

VIABLE LAND-USE ALTERNATIVES			
Land-use alternative	**Mandarins**	**Cling Peaches**	**Cabbage**
Annual Gross Income (Based on hectars required)	**R 4 024 725.00**	**R 7 666 666.00**	**R 1 822 485.20**
Capital Investment in first year (Based on hectars required)	R 2 799 750.00	R 2 365 160.00	R 1 300 000.00
Input Cost per year (Based on hectars required)	R 2 923 680.00	R 5 224 320.00	R 451 434.20
Total Cost in first year (Based on hectars required)	**R 5 723 430.00**	**R 7 589 480.00**	**R 1 751 434.20**
Remaining budget	R 3 276 570.00	R 1 410 520.00	R 7 248 565.80
Hectares required	15	20	20
Hectares remaining	5	0	0
Investment Period	2	5	0
Harvest month (start)	4	11	9
Crop rotation time (if required)			3
Risks	Markets regulation changes Change of overseas protocalls and regulations New diseases, specifically Asian Greening Labour related risks	Very sensitive to climate changes, tree develops buttons if not adequate amount of Infruitec chilling Pests/Disease Filed fire if orchard are not kept clean	Pests Plague Weed Hail Market establishment
Known diseases/fungi	Alternaria brown spot Fusarium (secondary fungus) Phytophthora parasitica	Bud mite	Alternaria leaf spot
Known pests	Red scale South African citrus thrips Mediterranean fruit flies Budworm Woolly whitefly	American budworm Snout Beetle	Diamond back moth Cutworm Thrips American bollworm Grey cabbage aphid
Unfavourable conditions	Human Element Equipment Sales stability Price stability Average annual temperature (Lower bound)(°C)	Equipment Sales stability Average annual temperature (Lower bound)(°C) Available Infruitec Chilling units (hours) Rainfall Season Region	Price stability Average annual temperature (Lower bound)(°C) Rainfall Season Region Local Climate Suitability-Rain (average)(mm) Average annual temperature (Upper bound)(°C)
Harvest length (months)	2.5	1	1

Figure 7: Output with provided user input of Figure 6

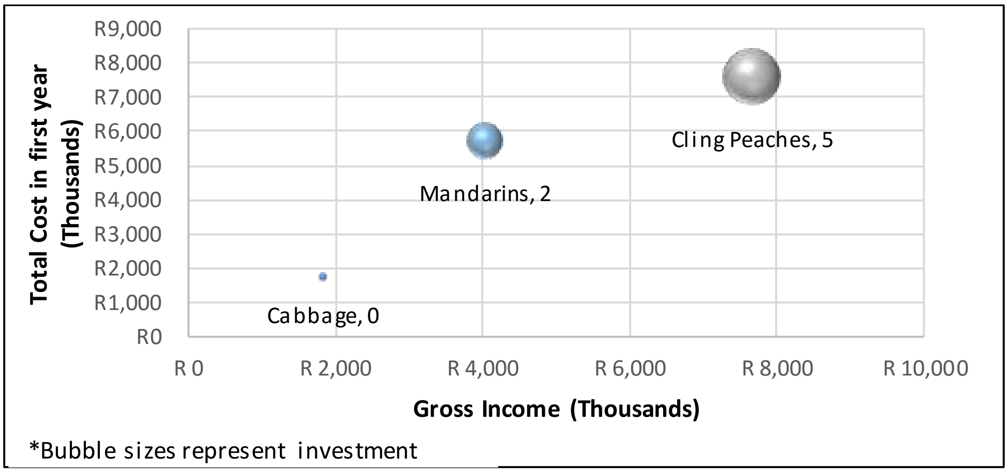


Figure 8: Cost in first year vs annual gross income relation for user input of Figure 6

Furthermore, the sizes of the bubbles in the depicted graph shows the investment period that is required before an alternative is expected to generate income. Figure 9 shows the extended DSS model, which uses the generated viable alternatives of the DSS model together with the user input of Figure 6. The values generated for each of the considerations as depicted in Figure 9 depends on the number of assigned hectares, in this case 20, 15, and 10 hectares that were assigned to mandarins cling peaches, and cabbage respectively. The totals thus reflect the total combined amounts of the different viable options when 20, 15, and 10 hectares are assigned to the respective viable alternatives. Figure 9 show this combination of alternatives with the assigned hectares exceeding the specified budget. The values for different input values and outputs generated can be graphically shown to facilitate visible analysis while using the DSS. Consequently, the farmer can for example phase the combined hectares in over several years, or if (s)he likes, adjust his/her initial amount of input hectares

Considerations	VIABLE LAND USE ALTERNATIVES		
Land-use alternative	Mandarins	Cling Peaches	Cabbage
Hectares Assigned	20	15	10
Hectares Remaining	70	75	80
Annual Gross Income	R 5 366 300.00	R 5 749 999.50	R 911 242.60
Capital Investment in first year	R 3 733 000.00	R 1 773 870.00	R 650 000.00
Input Cost per year	R 3 898 240.00	R 3 918 240.00	R 225 717.10
Total Cost in first year	R 7 631 240.00	R 5 692 110.00	R 875 717.10
Remaining budget	R 1 368 760.00	R 3 307 890.00	R 8 124 282.90
Manpower required	21	19.5	15
Manpower remaining	49	50.5	55
Water required (ML/year)	153	135	0.045
Water remaining (ML/year)	8847	8865	8999.955
	TOTALS		
Hectares Assigned	45 ha		
Hectares Remaining	45 ha		
Annual Gross Income	R 12 027 542.10		
Capital Investment in first year	R 6 156 870.00		
Input Cost per year	R 8 042 197.10		
Total Cost in first year	R 14 199 067.10		
Remaining budget	Insufficient funds. Additional R 5199067.1 required		
Manpower required	56 workers		
Manpower remaining	14 workers		
Water required (ML/year)	288 ML/year		
Water remaining (ML/year)	8712 ML/year		

Figure 9: Populated extended DSS model according to user input of Figure 6

One example of this is show in Figure 10 to quickly compare how much of the available resources each of the viable options use. This is used to evaluate different considerations of the viable options according to the assigned hectares, compares the percentage contribution of the total amount for manpower,water used, input cost, and capital investment in the first year per alternative. If an equal amount of hectares is assigned to each of the viable alternatives, Figure 10 can be used to compare the viable alternatives with each other per indicated considerations.

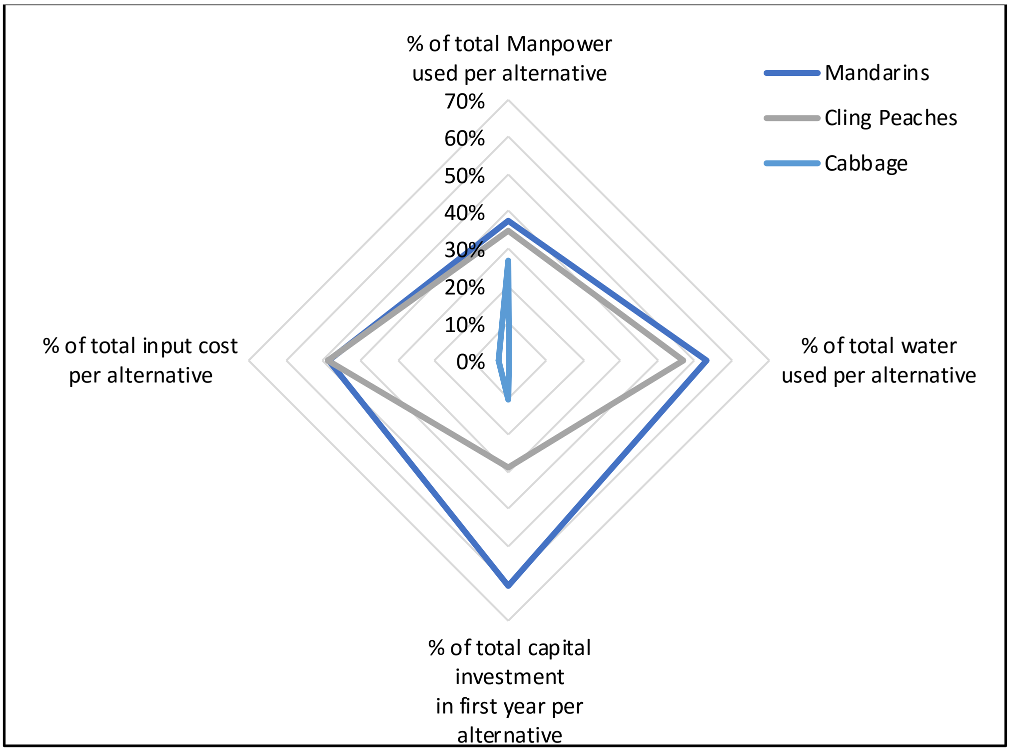


Figure 10: Percentage of available resources used per alternative considered

7. VALIDATION

Junier & Mostert [15] refer to validity as a model's capability to portray reality accurately. Information should therefore be seen as useful. Thus, it should be fit for the purpose, accessible, and user-friendly. The information therefore, is required to be perceived as valid in turn. Additionally, the perceived validity of a model or a DSS is determined by those who work with it, both the developers and the users [14]. Validating something therefore, gives credibility to a claim or statement. Therefore, it is important to validate one's work, to incorporate credibility and quality into it. The purpose of this section is to validate the developed considerations and resulting developed DSS. An internal validation was done after which subject matter experts were consulted to be able to conduct an external validation.

7.1 Internal Validation

To apply this step, an internal validation was conducted by the research team to determine whether the developed DSS provides the expected outputs. For these scenarios the team provided different input values. Logic and extreme conditions tests were done by the researchers to test the DSS model's capabilities to function within the boundaries of resources avaialable and to test the input limitations that could be provided by the end user.

7.2 External Validation

The set of considerations developed was used for validation purposes. The researchers developed the considerations using inputs from experts and research to determine what is important and needs to be taken into account when a decision maker considers undertaking a new land use alternative. Experts were approached to test the DSS model's input parameters and to validate the outputs generated. The experts applied the DSS to their own unique farm environment and found it to be informative and helpful. A number of suggestions and comments on the grpahic user interface was considered and implemented.

8. CONCLUSION AND RECOMMENDATIONS

The literature analysis placed emphases on strategic decisions for landscape alternatives and indicated land use alternative decisions as strategic decisions. The literature suggested that diversification strategies offer a trajectory toward viability, because income is generated from multiple sources which can account for business cycle variations and variation of seasonal income. In this study, diversification was primarily considered as agriculture diversification that is engaged in the cultivation of various crops. This definition excludes farm strategies aiming to relocate and recombine farm resources away from their original farming activities to generate an additional form of non-agricultural income.

Literature further suggested that existing DSSs only focussed on certain aspects regarding the suitability of an agricultural crop, thus not considering the whole farming operation when deciding which crops to select. The aim of the model developed in this study was to develop a holistic set of considerations that will evaluate land use alternatives and which would be incorporated into the model. These considerations are crucial to review before adopting any new land use alternative. Therefore, farm owners should use the DSS as a guide to understand which considerations are important to regard, and subsequently which crops are best suited for their particular region, when they consider adopting a new agriculture diversification strategy.

It is recommended that bankers can greatly benefit in using the developed DSS as a risk management strategy to inform farmers which agriculture strategy would be best to follow when farmers approach a bank for an agriculture loan. The researchers suggests that the best method to populate complete input crop datasets, would be to appoint agricultural consultants and agencies to collect significant amounts of data on their field of expertise. Agricultural consultants and bankers can then use this pool of data with the DSS model, to assess risks and advise farmers. Agricultural consultants and agencies can be used to keep the data pool updated for their crop alternatives.

9. REFERENCES

[1] **SAWIS.** (2015). *Final Report - Macro- economic Impact of the Wine Industry on the South African Economy (also with reference to the Impacts on the Western Cape).* Retrieved from http://www.sawis.co.za/info/download/Macro-economic_impact_study_-_Final_Report_Version_4_30Jan2015.pdf

[2] **Wines of South Africa.** (2015). THREE CENTURIES OF CAPE WINE. Retrieved from http://www.wosa.co.za/The-Industry/History/Three-Centuries-of-Cape-Wine/

[3] **Marais, E.** (2015). What Makes South Africa Unique? Retrieved from http://zar.co.za/world.htm

[4] **WOSA.** (2016). Winegrowing areas. Retrieved 15 April 2016, from http://www.wosa.co.za/The-Industry/Winegrowing-Areas/Winelands-of-South-Africa/

[5] **WINESA.** (2015). Wine Farms. Retrieved from http://www.wine-sa.com/wine-farms/

[6] **PWC.** (2014). The South African wine industry insights survey 2014, (June). Retrieved from http://www.pwc.co.za/en_ZA/za/assets/pdf/wine-insights-survey-2014.pdf

[7] **Pretorius, E.** (2009). Using Gis in the Search for Suitable Sites for the Implementation of fog water collection systems in South Africa, 1–8. Retrieved from

http://icaci.org/files/documents/ICC_proceedings/ICC2009/html/nonref/5_16.pdf
[8] **BBC News.** (2017, May). South Africa's Western Cape declares drought disaster.
[9] **Jabareen, Y.** (2011). Building a conceptual framework. Philopspohy, definition, procedure. *International Journal of Qualitative Methods*, *8*(4), 49–62. https://doi.org/10.1177/160940690900800406
[10] **Hall, W. L.** (2010). Decision Consequence Analysis: Conceptual Model Formulation. In *Sustainable Land Development and Restoration* (1st ed.). Elsevier. https://doi.org/10.1016/B978-1-85617-797-9.00004-6
[11] **Turban, E., & Aronson, J. E.** (2001). *Decision support systems and intelligent systems* (6th ed.). New York: Prentice Hall: Upper Saddle River.
[12] **McIntosh, B. S., Ascough, J. C., Twery, M., Chew, J., Elmahdi, A., Haase, D., … Voinov, A.** (2011). Environmental decision support systems (EDSS) development - Challenges and best practices. *Environmental Modelling and Software*, *26*(12), 1389–1402. https://doi.org/10.1016/j.envsoft.2011.09.009
[13] **Rose, D. C., Sutherland, W. J., Parker, C., Lobley, M., Winter, M., Morris, C., … Dicks, L. V.** (2016). Decision support tools for agriculture: Towards effective design and delivery. *Agricultural Systems*, *149*, 165–174. https://doi.org/10.1016/j.agsy.2016.09.009
[14] **Dury, J., Schaller, N., Garcia, F., Reynaud, A., & Bergez, J. E.** (2012). Models to support cropping plan and crop rotation decisions. A review. *Agronomy for Sustainable Development*, *32*(2), 567–580. https://doi.org/10.1007/s13593-011-0037-x
[15] **Junier, S., & Mostert, E.** (2014). A decision support system for the implementation of the Water Framework Directive in the Netherlands: Process, validity and useful information. *Environmental Science and Policy*, *40*, 49–56. https://doi.org/10.1016/j.envsci.2014.04.004
[16] **Jakeman, A. J., Letcher, R. A., & Norton, J. P.** (2006). Ten iterative steps in development and evaluation of environmental models. *Environmental Modelling and Software*, *21*(5), 602–614. https://doi.org/10.1016/j.envsoft.2006.01.004
[17] **McNie, E. C.** (2007). Reconciling the supply of scientific information with user demands: an analysis of the problem and review of the literature. *Environmental Science and Policy*, *10*(1), 17–38. https://doi.org/10.1016/j.envsci.2006.10.004
[18] **Hartati, S., & Sitanggang, I. S.** (2010). A Fuzzy Based Decision Support System for Evaluating Land Suitability and Selecting Crops. *Computer Science*, *6*(4), 417–424.
[19] **Mallach, E. G.** (2000). *Decision Support and Data Warehouse Systems*. McGraw-Hill.
[20] **Smith, P. N.** (1992). Fuzzy Evaluation of Land-Use and Transportation Options. *Environment and Planning B: Urban Analytics and City Science*, *19*(5), 525–544.
[21] **Wei, C.-C., Liang, G.-S., & Wang, M.-J. J.** (2007). A comprehensive supply chain management project selection framework under fuzzy environment. *International Journal of Project Management*, *25*(6), 627–636. https://doi.org/10.1016/j.ijproman.2007.01.010
[22] **De Wit, C. T., van Keulen, H., Seligman, N. G., & Spharim, I.** (1988). Application of interactive multiple goal programming techniques for analysis and planning of regional agricultural development. *Agricultural Systems*, *26*(3), 211–230. https://doi.org/10.1016/0308-521X(88)90012-1
[23] **Dogliotti, S., Van Ittersum, M. K., & Rossing, W. A. H.** (2005). A method for exploring sustainable development options at farm scale: A case study for vegetable farms in South Uruguay. *Agricultural Systems*, *86*(1), 29–51. https://doi.org/10.1016/j.agsy.2004.08.002
[24] **Castelán-Ortega, O. A., Fawcett, R. H., Arriaga-Jordán, C., & Herrero, M.** (2003). A Decision Support System for smallholder campesino maize-cattle production systems of the Toluca Valley in Central Mexico. Part II - Emulating the farming system. *Agricultural Systems*, *75*(1), 23–46. https://doi.org/10.1016/S0308-521X(01)00110-X
[25] **Pacini, C., Wossink, A., Giesen, G., & Huirne, R.** (2004). Ecological-economic modelling to support multi-objective policy making: A farming systems approach implemented for Tuscany. *Agriculture, Ecosystems and Environment*, *102*(3), 349–364. https://doi.org/10.1016/j.agee.2003.08.010
[26] **Ten Berge, H. F. M., Van Ittersum, M. K., Rossing, W. A. H., Van De Ven, G. W. J., Schans, J., & Van De Sanden, P. A. C. M.** (2000). Farming options for The Netherlands explored by multi-objective modelling. *European Journal of Agronomy*, *13*(2–3), 263–277. https://doi.org/10.1016/S1161-0301(00)00078-2
[27] **Vatn, A., Bakken, L., Botterweg, P., & Romstad, E.** (1999). ECECMOD: An interdisciplinary modelling system for analyzing nutrient and soil losses from agriculture. *Ecological Economics*, *30*(2), 189–205. https://doi.org/10.1016/S0921-8009(98)00116-5
[28] **Wossink, G. A. A., de Koeijer, T. J., & Renkema, J. A.** (1992). Environmental-economic policy assessment: A farm economic approach. *Agricultural Systems*, *39*(4), 421–438. https://doi.org/10.1016/0308-521X(92)90078-3
[29] **Zander, P., & Kächele, H.** (1999). Modelling multiple objectives of land use for sustainable development. *Agricultural Systems*, *59*(3), 311–325. https://doi.org/10.1016/S0308-521X(99)00017-7
[30] **Rădulescu, C. Z., & Rădulescu, M.** (2012). A Decision Support Tool Based on a Portfolio Selection Model for Crop Planning. *Studies in Informatics and Control*, *21*(4), 377–382.
[31] **Collender, R. N.** (1989). Estimation Risk in Farm Planning Under Uncertainty. *American Journal of Agricultural Economics*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=bth&AN=4596241&site=ehost-live
[32] **Salleh, M. N. M.** (2012). A fuzzy modelling of decision support system for crop selection. *ISIEA 2012 - 2012 IEEE Symposium on Industrial Electronics and Applications*, 17–22.

https://doi.org/10.1109/ISIEA.2012.6496622
[33] **Balezentiene, L., Streimikiene, D., & Balezentis, T.** (2013). Fuzzy decision support methodology for sustainable energy crop selection. *Renewable and Sustainable Energy Reviews*, *17*, 83–93. https://doi.org/10.1016/j.rser.2012.09.016
[34] **Nevo, A., & Amir, I.** (1991). CROPLOT-An expert system for determining the suitability of crops to plots. *Agricultural Systems*, *37*(3), 225–241. https://doi.org/10.1016/0308-521X(91)90034-8
[35] **Rossing, W. A. H., Jansma, J. E., De Ruijter, F. J., & Schans, J.** (1997). Operationalizing sustainability: Exploring options for environmentally friendly flower bulb production systems. *European Journal of Plant Pathology*, *103*(3), 217–234. https://doi.org/10.1023/A:1008609624648
[36] **van Ittersum, M. K., Rabbinge, R., & van Latesteijn, H. C.** (1998). Exploratory land use studies and their role in strategic policy making. *Agricultural Systems*, *58*(3), 309–330. https://doi.org/10.1016/S0308-521X(98)00033-X
[37] **Makowski, D., Hendrix, E. M. ., van Ittersum, M. K., & Rossing, W. A. H.** (2000). A framework to study nearly optimal solutions of linear programming models developed for agricultural land use exploration. *Ecological Modelling*, *131*(1), 65–77. https://doi.org/10.1016/S0304-3800(00)00249-0
[38] **Annetts, J. E., & Audsley, E.** (2002). Multiple objective linear programming for environmental farm planning. Journal of the Operational Research Society, 53(9), 933–943. https://doi.org/10.1057/palgrave.jors.2601404
[39] **Bartolini, F., Bazzani, G. M., Gallerani, V., Raggi, M., & Viaggi, D.** (2007). The impact of water and agriculture policy scenarios on irrigated farming systems in Italy: An analysis based on farm level multi-attribute linear programming models. *Agricultural Systems*, *93*(1–3), 90–114. https://doi.org/10.1016/j.agsy.2006.04.006
[40] **Sarker, R., & Ray, T.** (2009). An improved evolutionary algorithm for solving multi-objective crop planning models. *Computers and Electronics in Agriculture*, *68*(2), 191–199. https://doi.org/10.1016/j.compag.2009.06.002
[41] **Louhichi, K., Kanellopoulos, A., Janssen, S., Flichman, G., Blanco, M., Hengsdijk, H., … Ittersum, M. Van.** (2010). FSSIM, a bio-economic farm model for simulating the response of EU farming systems to agricultural and environmental policies. *Agricultural Systems*, *103*(8), 585–597. https://doi.org/10.1016/j.agsy.2010.06.006

ESTIMATING DELAY AT A SIGNALIZED INTERSECTION USING QUEUING MODELS

Z. Mpanza[1*] & H. Ncube[2]

[1]Department of Mechanical and Industrial Engineering
University of Johannesburg, South Africa
zanelem@uj.ac.za

[2]Department of Quality and Operations Management
University of Johannesburg, South Africa
henryn@uj.ac.za

ABSTRACT

Intersections have been identified as the most complex location in a traffic system. This is due to the number of movements that occur within an intersection. Delays have a negative impact on motorists as well as in the economy. Traffic signals are implemented to reduce this burden, but this is determined by the signal timing. The pre-timed traffic signal control assigns the right of way at an intersection according to predetermined schedule; and does not accommodate short-term fluctuations. The purpose of this study was to estimate the delay at an intersection using queuing models. The study uses field data collection.

[1] The author was employed as a Lecturer in the Department of Mechanical and Industrial Engineering, University of Johannesburg
[2] The author was employed as a Lecturer in the Department of Quality and Operations Management, University Johannesburg
*Corresponding author

1. INTRODUCTION

In urban areas, property developments are constantly increasing and this has caused an increase in the number of road users in the urban areas. With an increase in motorists, traffic congestions become inevitable. Traffic congestion contributes to economic, environmental and social issues in urban areas. Due to limited space, it becomes impossible to expand road networks and so traffic engineers have to come up with other mechanisms of making the road network effective. One of these mechanisms is to install traffic control systems.

Signalized intersections are basically points or nodes within a system of surface highways and streets; thus defining appropriate measures of effectiveness to describe the quality of operations within this system is somewhat difficult. Intersections are regarded as the most complex areas of traffic network [1]. This is due to a number of activities that take place within an intersection. A number of measures have been used in capacity analysis and simulation models, all of which quantify some aspect of the experience of traversing a signalized intersection in terms of what the driver comprehends. The most common factors include delay time, queue length and stopping time. Traffic control mechanisms like stop, yield and traffic lights are implemented at intersections to control traffic movement. The effectiveness of a traffic signal determines the overall time spent in a road network. Signal timing, signal synchronization, the arrival and departure rate of vehicles are just some of the influencing factors of the signalized intersection performance. Signal timing optimization is important to reduce traffic delays, relieve congestion, and improve the operation of an intersection. The performance of an intersection is measured by the delay which motorists experience within an intersection [1].

Vehicles approaching a traffic signal have the same characteristics of a queueing system. They have three parts, which are (1) the arrivals or inputs to the system, (2) the queue or waiting line itself, and (3) the service facility [2]. Vehicles arrive at a traffic signal, wait in the queue, receive a service (signal) and depart, as shown in Figure 1. The effectiveness of a traffic signal (service facility) determines the overall time spent in a road network.

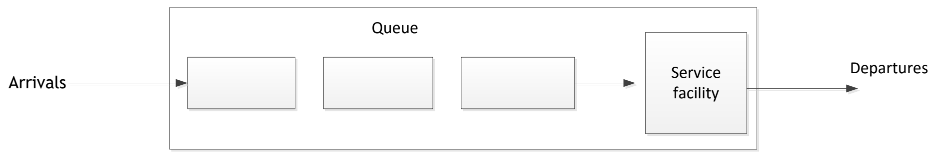


Figure 1: General Queueing system.

At a signalized intersection, the level of service (LOS) or quality of traffic flow can be measured using various parameters [1]. Amongst them is vehicle delay; which is the most important parameter since it indicates the time loss by a vehicle while intersecting an intersection. Travel times are not easy to determine at an intersection and traffic signal control due to stochastic properties of traffic flow and stochastic arrivals and departures. This causes determining delay to be one of the complex task because it is influenced by many variables [1]. Also, delay is one of the activities that are regarded as non-value adding activities.

Traffic delay is no different from any other delay experienced on a daily basis by customers, the delay can be in a processing plant or a service in a bank; they all have a negative impact on customers/users and they also indicate a poor performance of the system as a whole. Not only does it impact negatively on customers or users, it also impacts negatively on the performance of a business as customers end up choosing other alternatives to avoid delay. In traffic, users may choose to use other routes which are not designed for heavy traffic flow, which results in even more congestions. It is therefore important to identify the causes of delay and reduce it as much as possible.

Signal timing optimization is important to reduce traffic delays, relieve congestion, and improve the operation of an intersection. Signal timing is crucial during peak periods because it determines the number of vehicles that can cross a traffic light during green time. If there are fewer vehicles that can intersect, the vehicle queue length increases causing congestion and overflow. In queueing theory, this indicates that the service rate is less than the average arrival rate [2]. This problem is mostly caused by the pretimed traffic signals [1]. Traffic signal are said to be pretimed if it has a fixed signal timing, regardless of the queue length [1]. Pre-timed traffic signals are predetermined even during the peak period which does not solve the problem of queue length. One solution to this is to implement actuated traffic signals. In South Africa, most traffic signals are pretimed which results in traffic overflow on intersections. This study seeks to estimate delay at an intersection using queueing model.

2. LITERATURE REVIEW

In this section, we look at some of the studies and models developed to estimate delay.

There are two basic categories of cost in a queuing situation: those associated with customers waiting for service; in this case motorists waiting to intersect at a traffic signal, and those associated with capacity [2]. The goal of queuing analysis is to balance the cost of providing service capacity with the cost of customers (motorists) waiting for service [2]. The goal of analysis is to identify the level of service (LOS) capacity that will minimize total cost. Much of the success of the analysis will depend on choosing an appropriate model. Model choice is affected by the characteristics of the system under investigation, the main characteristics are arrival and service patterns, queue discipline, number of channels and population source [2]. Waiting lines are a direct result of arrival and service variability. They occur because random, highly variable arrival and service patterns cause systems to be overloaded. Waiting line are most likely to occur when arrivals are bunched or when service times are particularly lengthy. Usually vehicle arrivals at a traffic signal is unlimited. Vehicles arrive randomly since they cannot be predicted exactly, even though during the peak hours a prediction can be made.

Congestion is unavoidable in urban cities due to growth in traffic demand. This has led to the development of many traffic signal control systems which have been tested in a number of countries. In the 1960's, the early stage of systems was implemented in Munich, West London and Toronto and later the urban traffic control system project was completed in the United States [1]. The SCAT system, which has multilevel hierarchical structure, was developed by Australia [1].

A number of studies have been done to estimate delay at a signalised intersection. Wu and Giuliani [3] estimate delay and capacity at a signalised intersection by measuring the cycleflow probability. The authors use the measured cycleflow probability to estimate delays and queueing lengths. Their results show that the cycleflow probability can estimate delays. Vajeeran and De Silva [4] conducted a delay analysis at a signalized T intersection at Katubedda. The authors used Vissim simulation software to model actual data of queue length. They found that signal timing changes and geometric changes can reduce delay. Christofa *et al*. [5] introduced a real-time signal control system that optimizes signal settings based on minimization of person delay on arterials. Tan *et al*. [6] uses a decentralized genetic algorithm (DGA) to minimize delay at signalized traffic network under an oversaturated condition. The results show that the developed DGA is able to reduce network delay by optimizing the traffic signal. Dabiri and Abbas [7] use a Particle Swarm Optimization (PSO) algorithm which is used as an optimizer for generating arterial traffic signal timing parameters. The results reveal that the proposed method is promising and outperforms for various traffic for traffic states. Lu and Yang [8] use a stochastic queue model that considers interdependence relations between adjacent intersections using probability-generating function to analyse delay in signalised networks under different congestion levels. Huang *et al*. [9] calculate intersection delay based on vehicle positioning data and analysed it with DBSCAN and Least Square Fit algorithms, without using the signal or traffic information. Anusha *et al*. [10] use occupancy based method and the queue clearance based method to estimate queue and delay.

Although there has been a tremendous work done in this area, there are very few studies that are published in the South African context. According to White Paper on Roads Policy for South Africa [11], the roads industry in South Africa does not have a formalised system in place to guide the industry at the project implementation level which causes a challenge in implementation of best-practices for road authorities, road designers and builders.
South African traffic engineers conduct these studies but they are seldom published and therefore this limits the availability of data.

3. METHODOLOGY

3.1 Data collection site

Since the data was not readily available, data had to be collected manually for this study. The study was conducted in an intersection situated in Victory Park, which is a suburb in the North of Johannesburg. This site was selected because it is one of the busiest intersections as it connects to an arterial, but yet less attention has been given to improve it. Also, due to the new housing developments in the area, the traffic congestion has increased dramatically as experienced by motorists on a daily basis. The intersection experiences various traffic conditions from saturated to undersaturated with overflow at times. The intersection is a four-leg signalized intersection as seen in Figure 2. Table 1 gives a summary of the information about the intersection. The study was conducted on a Tuesday during peak time of the morning from 7:00 am to 8:00 am. Six cycles were recorded. According to Schroeder *et al*. [12] six cycles is adequate to get conclusive information. The weather condition was clear with no rain or thundershowers. The intersection is a straight paved road. Figure 2 shows the intersection under study. Only the through lane traffic was selected for surveying and therefore the right turn lane was excluded. In Figure 3, the red arrow shows the lane excluded from the survey.

Table 1: Intersection information

Intersection	Cycle length (seconds)	Direction	Green time(seconds)	Approach
Corner Tana Road and 3rd Avenue	70	through	26.06	Eastbound

Figure 2: Intersection selected for study.

Figure 3: Showing the excluded right turning lane group [13].

3.2 Field survey

The data required for this study included queue length, signal timing (green, yellow and red), headways, etc. Before the actual collection of field data, preparation was done days prior to the study which involved observing the traffic flow behaviour, identifying which peak hour (afternoon or morning) was more critical. The morning peak was identified as more critical than the afternoon. The data was manually recorded using a stop watch and the designed recording sheets. Traffic video cameras were the preferred survey method but due to financial constraints these could not be obtained. Nevertheless, the surveyors were conveniently located where they could observe all movements and at a close proximity with the ability to maneuver.

One of the surveyors observed and counted stopped vehicles on an intersection approach at every red signal. This was to get the size of the queue length. A total of eleven sets of queue counts was recorded and these were added to get a total for a 15 minutes interval. The vehicles departing on the green signal were also counted and added up to get a total for a 15 minutes interval. A total of twelve sets of departure counts were recorded. Table 2 shows the total queue length and total departure count for 7:000 am – 8:00 am.

The signal timing was observed and recorded. This intersection has a three-phase signal cycle system with pretimed or fixed cycle time. Table 3 shows the observed signal times, which show green as 26.06 seconds, yellow with 2.94 seconds, red with 41.03 seconds and a cycle length of 70 seconds.

Table 2: Departure count and queue length

Time Period	Total Departure Count (veh)	Queue length (veh)
7:00 – 7:15 am	141	168
7:15 – 7:30 am	107	155
7:30 – 7:45 am	145	172
7:45 – 8:00 am	136	161
TOTAL	529	656

Table 3: Signal times

Signal	Time (seconds)
Green	26.06
Yellow	2.94
Red	41.03
Cycle time	70

When the green signal was initiated, vehicles crossing the curb line were observed and the time between the initiation of the green and the crossing of the first vehicle over the curb line was recorded. These times were recorded for eight successive headways. According to Schroeder *et al*. [12], eight is an ideal number since the headway starts levelling off after the fourth or fifth vehicle. This levelling off is defined as a saturation headway which was recorded as 2 seconds, and we use this to compute saturation flow rate which is the number of vehicles per hour of green time per lane [12]. The start-up lost time and clearance lost time was recorded as 1 second.

The data collected assisted in the calculation of:

- The arrival rate
- Saturation flow rate
- Capacity of lane
- Effective green time
- Effective green time ratio
- Volume to capacity ratio
- Delay time

All the above parameters are calculated to aid in the estimation of delay. The formulas used to calculate these are listed in section 3.3.

3.3 Data Analysis

Webster [14] developed a model for estimating the delay experienced by motorists at signalised intersections using deterministic queuing analysis. The model is presented in Equation 3.1.

$$d = \frac{C(1-\lambda)^2}{2(1-\lambda X)} + \frac{X^2}{2v(1-X)} - 0.65\left(\frac{c}{v^2}\right)^{1/3}[X^{2+5\lambda}] \qquad (3.1)$$

where:

$d = average\ overall\ delay\ per\ vehicle\ (seconds)$,
$\lambda = proportion\ of\ the\ cycle\ that\ is\ effective\ green\ (g/C)$,
$C = cycle\ length\ (seconds)$,
$v = arrival\ rate\ (vehicles/hour)$,

$c = capacity\ for\ lane\ group\ (vehicles\ /hour)$,
$X = volume\ to\ capacity\ ratio\ of\ lane\ group$,
$g = effective\ green\ time\ (seconds)$.

The first term of equation 3.1 represents the average delay to the vehicles assuming uniform arrival. The additional delay, which is due to the randomness of vehicle arrival, is represented by the second term. To correct the delay estimates the third term of the equation was introduced as an adjustment factor.

We calculate delay using the Webster's queueing model [14]. First, we need to compute calculations for saturation flow rate, capacity of lane, effective green time and finally calculate the delay time. We use the following equations:

1) Saturation flow rate:

$$s = \frac{3600}{h} \quad (3.2)$$

where: $s = saturation\ flow\ rate\ (vph$gpl),
$h = saturation\ headway\ (seconds)$,
$3600 = seconds/hour$.

2) Effective green time:

$$g = G + Y - t \quad (3.3)$$

where:
$G = actual\ green\ time$
$Y = sum\ of\ yellow\ plus\ all\ red\ time$
$t_L = total\ lost\ time\ per\ phase$

3) Capacity of each lane group:

$$c = s\frac{g}{C} \quad (3.4)$$

where:
$c = capacity\ of\ lanes\ serving\ movement,(vph\ or\ vhphpl)$,
$s = saturation\ flow\ rate\ for\ movement\ (vphg\ or\ vphgpl)$,
$g = effective\ green\ time\ for\ movement\ (seconds)$,
$C = signal\ cycle\ length,$(seconds).

4) Effective green time ratio:

$$\lambda = \frac{g}{C} \quad (3.5)$$

where:
$\lambda = proportion\ of\ the\ cycle\ that\ is\ effective\ green\ (g/C)$,
$g = effective\ green\ time\ (seconds)$,
$C = cycle\ length\ (seconds)$.

5) Volume to capacity ratio:

$$X = \frac{v}{c} \quad (3.6)$$

where:
$X = volume\ to\ capacity\ ratio$,
$v = arrival\ rate\ (vehicles/hour)$,
$c = capacity\ for\ lane\ (vehicles/hour)$.

4. RESULTS AND DISCUSSION

When estimating delay at an intersection, a number of factors come into play. The basic characteristics used to model any intersection operation are applied the same way when modeling delay, that is, the manner in which vehicles depart, or discharge from the intersection when a GREEN indication is received. The following calculations are necessary to build the model for calculating delay at an intersection.

Figure 4 illustrates a group of vehicles at a signalised intersection, waiting for the GREEN indication.

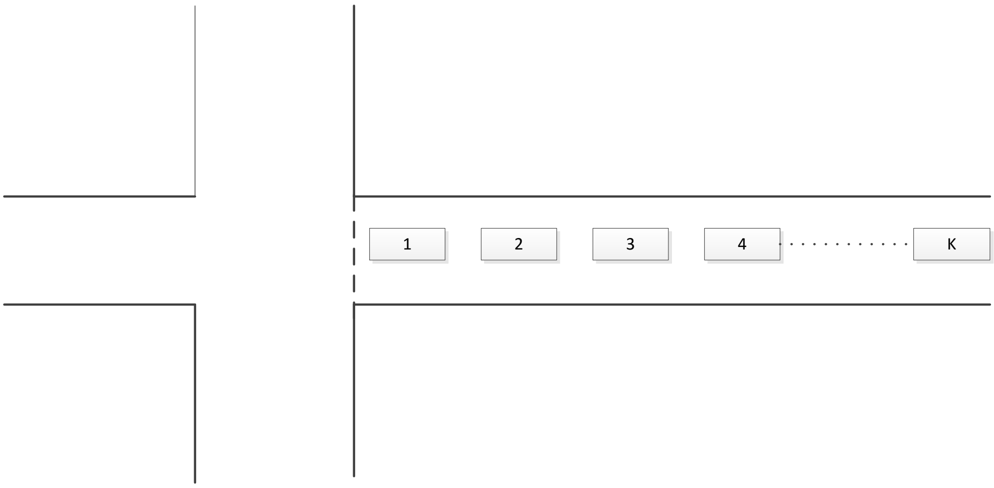


Figure 4: Vehicles waiting in queue for a green signal indication.

When GREEN is initiated, headways between departing vehicles is observed as vehicles cross the curb line. The first headway is the time between the initation of GREEN and the crossing of the first vehicle over the curb line. The second headway is the time between the first and the second vehicles crossing the curb line, etc. As can be seen in the Figure 5, the first headway is longer as it includes the first driver's reaction time, and the time necessary to accelerate. The second headway is shorter because the second driver can overlap her reaction and acceleration time with the first driver's. Each successive headway gets a little smaller and the headways finally tend to level out. The level headway, or saturation headway can be seen from the fifth vehicle in the queue as shown in Figure 5.

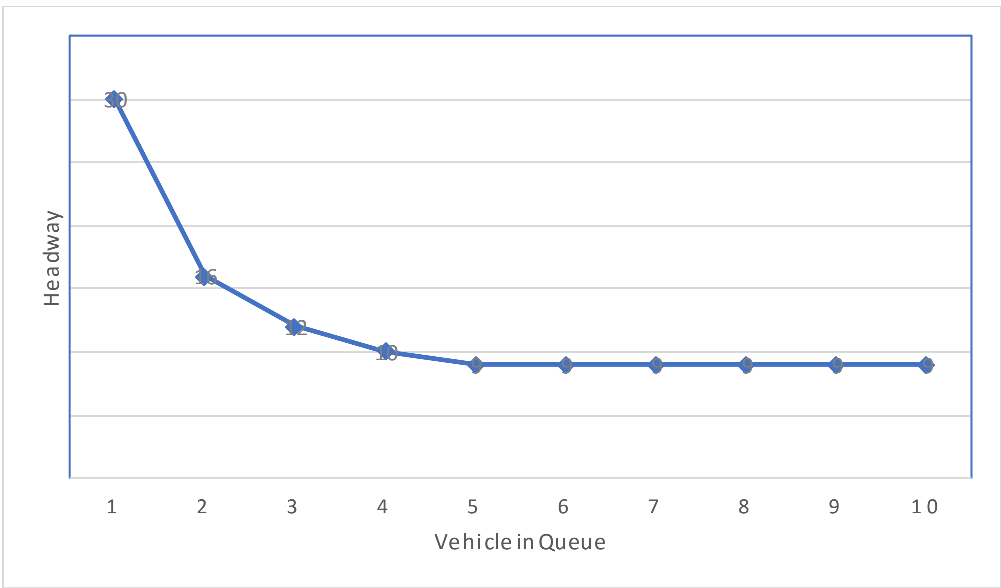


Figure 5: Headways departing signal.

The leveling off is noticed with the fifth headway. From this illustration, we can now calculate the saturation flow rate using equation 3.2 which yields:

$$s = \frac{3600}{h} = \frac{3600}{2} = 1800\, vphgpl$$

This means that 1800 vehicles/hour/lane could enter the intersection if the signal were always green. If there is more than one lane, this number is multiplied by the number of lanes. In reality, the signal is not always green for any movement. Therefore, Miller [15], Akcelik [16] developed mechanism for dealing with the cyclic starting and stopping of a movement at a signal. This entails the calculation of the amount of green time available (effective green time, g) to be used at a rate of one vehicle every 2 seconds (saturation headway). Calculating effective green time g using equation 3.3:

Using data listed in Table 1 of section 3.2 and equation 3.3, we get:

$$g = G + Y - t = 26.02 + 2.94 - (1 + 1) = 27 \; seconds$$

Therefore, the amount of available green time to be used at a rate of one vehicle every 2 seconds is 27 seconds.

We then compute capacity of the lane serving movement, using equation 3.4:

$$c = s\frac{g}{C} = 1800\frac{27}{70} = 694 \; vphpl$$

Thus, 694 vehicles per hour can transverse in each lane, given that the available green time is 27 seconds.

The effective green ratio is computed using equation 3.5:

$$\lambda = \frac{g}{C} = \frac{27}{70} = 0.38 \; seconds$$

The arrival volume is computed by adding the departure volume and the net change in the size of the queue during the counting period as shown in Table 4. For period 1, which is 7:00 – 7:15 am, the arrival volume is calculated by subtracting the total depature from the queue length recorded at the beginning of the study.

Table 4: Arrival Volumes

Time Period	*Queue Length (veh)*	*Total Departure Count (veh)*	*Arrival Volume*
7:00 - 7:15 am	168	141	168-141 = 27
7:15 - 7:30 am	155	107	155+107-141 = 121
7:30 - 7:45 am	172	145	172+145-107 = 210
7:45 - 8:00 am	161	136	161+136-145 = 152
TOTAL		529	510

As can be seen in Table 4, between 7:00 – 7:15 am, 168 vehicles were in the queue at the start of the study period. Out of the 168 vehicles, 141 departed on green signal indication. The arrival volume is then computed by subtracting the departure from the queue length. Between 7:30 – 7:45 am queue length was at maximum with 172 vehicles. This also shows a peak departure flow at 145 veh/15 min with a peak arrival volume of 210 veh/15 min. The results in Table 4 tells us that the traffic flow is not too heavy from 7:00 – 7:15 am. It then starts to increase from 7:15 – 7:30 am, reaching its peak from 7:30 – 7:45 am. The total arrival volume is 510 vehicles per hour. It can be deduced from the results that more vehicles traverse between 7:15 – 8:00 am. Therefore, signal timing can be adjusted to accommodate the traffic conditions during this period to ease congestion.

Using equation 3.6 we calculate volume to capacity ratio:

$$X = \frac{v}{c} = \frac{510}{694} = 0.73$$

Finally, calculate delay using equation 3.1:

$$d = \frac{C(1-\lambda)^2}{2(1-\lambda X)} + \frac{X^2}{2v(1-X)} - 0.65\left(\frac{c}{v^2}\right)^{1/3}[X^{2+5\lambda}]$$

$$= \frac{70(1-0.38)^2}{2(1-0.38*0.73)} + \frac{(0.73)^2}{(2)(510)(1-0.73)} - 0.65(\frac{694}{510^2})^{1/3}\,[(0.73)^{2+5(0.38)}]$$

$$= 18.6 \; sec/veh$$

Thus, each vehicle is expected to experience 18.6 seconds of delay traversing in this intersection.

5. CONCLUSION

In this study, we were able to calculate the saturation flow rate, which is one of the elements that are used to time signals and estimate intersection capacity. Most agencies use standard constant values for saturation flow, but they vary between intersections and may vary significantly in the South African traffic network. More studies need to be conducted at several sites in South Africa to establish saturation rates. The calculations show that the arrival volume reaches peak between 7:15 – 8:00 am.

Delay was calculated to be 18.6 sec/veh, which means that every vehicle joining the intersection delays with 18.6 seconds in the intersection before traversing. This figure is not particularly large but can be eliminated by adjusting signal timing during peak hours. This may entail the use of technology that is able to detect or count the number of cars in the intersection and allowing sufficient green time for vehicles to pass (traffic actuated signal). This can improve the level of service in this intersection thereby reducing travel time for motorists. Reduction in delay may indirectly improve productivity as people will spend less time on the road and more time in the workplace.

More intersection delay studies need to be conducted so that we can compare the results.
Studies similar to this one need to be conducted periodically so that we have enough traffic data for South African road network. Other studies may include delay caused by traffic signal failure in an intersection and the impact of traffic signal syncronization. These type of studies may assist in traffic control as well as transport planning.

REFERENCES

[1] Pande, A. and Wolshon, B. (Ed). 2016. *Traffic Engineering Handbook,* 7th Edition, John Wiley & Sons Inc.

[2] Render, B., Stair R., Hanna, M. and Hale, T. 2015. *Quantitative Analysis for Management*, 12th Edition, Pearson.

[3] Wu, N and Giuliani, S. 2016. Capacity and Delay Estimatin at Signalized Intersections under unsatutaed flow condition based on cycle overflow probability, *International Symposium on Enhancing Highway Performance,* 15(), pp 63-74

[4] Vajeeran, A. and De Silva, G.L.D.I. 2017. Delay analysis at a signaized T intersection. *2017 Moratuwa enginerng Conference (MERCon)*, pp 325-330.

[5] Christofa, E., Ampountolas, K. and Skabardonis A. 2016. Arterial traffic signal optimization: A person-based approach, *Transportation Research Part C* 66(), pp 27-47.

[6] Tan. 1964. *Statistical Inference,* 3rd Edition, McGraw-Hill.

[7] Dabiri, S and Abbas, M. 2016. Arterial traffic signal optimization using Particle Swarm optimization in an integrated VISSIM-MATLAB simulation environment, *201 IEE 19TH International Conference on Intelligent Transportation Systems (ITSC), Windsor Oceanico Hotel, Rio-de Janeiro, Brazil, November 1-4, 2016.*

[8] Lu, Y. and Yang, X. 2014. Estimating dynamic queue distribution in a signalized network through a probability generating model, *IEEE Transactions on Intelligent Transportation Systems,* 15(1), pp 334-344.

[9] Huang, J., Li, G., Wang, Q. and Yu, H. 2013. Real time delay estimation for signalized intersection using transit vehicle positioning data, *13th International Conference on ITS Telacommunications (ITST)*.

[10] Anusha, S., Vanajakshi, S. and Sharma, A. 2015. Performance comparison of two model based schemes for estimation of queue and delay at signalized intersections, *IEEE Intelligenet Vehicles Symposium (VI), June 28 – July 1, 2015, COEX, Seoul, Korea.*

[11] Department of Transport. 2017. *White Paper on National Transport Policy*, Pretoria, South Africa.

[12] Schroeder, B., Cunningham, C., Findley, D., Hummer, J. and Foyle, R. 2010. *Manual of Transportation Engineering Studies*, 2nd Edition, Washington DC: ITE.

[13] Google Maps, available at: https://www.google.com/maps/place, [accessed 31 July 2018].

[14] Webster, F. 1958. Traffic signal settings, *Road Research Technical Paper No. 39, Road Research Laboratory, London, UK.*

[15] Miller, A. 1963. Setting for fixed cycle traffic signals, *Operations Research Quarterly,* 14(4), pp 373-386.

[16] Akcelik, R. 1988. The Highway Capacity Manual delay formula for signalized intersection, *ITE Jounal,* 58(3), pp 23-27.

CONCEPTUAL EVALUATION TOOL FOR ASSESSING THE COMPANY READINESS LEVEL TO IMPLEMENT AN ERP SOLUTION AS A STEPPING STONE FOR THE 4TH INDUSTRIAL REVOLUTION

J.F. Oberholzer[1] & L. Bezuidenhout[2*]

Qmuzik Technologies,
South Africa
[1]francois.oberholzer@qmuzik.com, [2]leri.bezuidenhout@qmuzik.com

ABSTRACT

Enterprise Resource Planning (ERP) solutions aim to increase efficiency and automate certain business processes to help gain a competitive edge in the market. The implementation process of an ERP system is a complex, costly and time-consuming process, with a high failure rate. ERP failure is mostly caused by organisational and social factors, rather than technical factors. With the Fourth Industrial Revolution on the rise, companies need to aim at bringing together digital, biological and physical technologies in new combinations starting with the ERP system. This research aims at identifying the readiness level of a company to implement a system that integrates company wide data and processes management across multiple platforms, departments and locations. The outcome of this research is a suggested framework showing the organisational readiness level for successful Enterprise Resource Planning (ERP) implementation.

*Corresponding Author

1. INTRODUCTION

The need for Enterprise Resource Planning (ERP) systems has grown over the past decade. Companies can choose from several different types of ERP systems that best fit their company need. Integrating an ERP system into any company comes with several risk factors. These risk factors can be present within the ERP system itself or within the organisation.
Before choosing the right ERP system for your company, start by inspecting whether or not your company is ready for an ERP system. We can therefore call it the organisational readiness for an ERP.

In this research, we are discussing the readiness of a SME (Small-Medium enterprise) company to take on an ERP system. The methodology consists of the company requirements for an ERP system and additionally if they are ready to make the required changes. One company is used as a case study in this research.

2. LITERATURE REVIEW

2.1 Enterprise Resource Planning and its role in Industry 4.0

Companies require that one single system be implemented to replace the numerous smaller systems being used by the various functional areas. i.e. the integration of departments and functions into one system [1]. ERP systems are software packages used to run an organisation's processes on a system level [2].

One of the benefits of ERP systems is that all enterprise data are collected during the transaction, stored centrally, and updated in real time [3]. The key functional areas of interest of an ERP system are Financial and Managerial Accounting, HR, Supply Chain Management, Project Management, Customer Relationship Management, and Data Services [4].

Successful ERP implementation depends upon various factors known as Critical Success Factors (CSFs) [2]. CSFs are the critical or non-negotiable features or functionalities that the potential ERP system must be able to fulfil in order to meet the needs of the organisation [5].

This can be categorised by assessing the resulting system against the planned objectives, user expectations, project budget and goals. Shafaei and Dabiri [3] have developed a methodology by establishing an adaption between ERP implementation CSFs and the European Foundation for Quality Management's Excellence Model (EFQM) indicators in assessing organisational readiness for those that have decided to implement an ERP system. An EFQM based model can be used to assess the readiness of an enterprise for effective and successful ERP implementation, using CSFs as main affecting factors of implementation[3]. Similarly, Lien and Chan [6] have developed a hybrid model based on fuzzy logic and an Analytical Hierarchy Process (AHP) methodology, that introduces important factors of a successful ERP selection from both product and a managerial solution perspective of ERP implementation.

Industry 4.0 is the industrial revolutionary move towards digitisation. Industry 4.0 uses Internet of Things and cyber-physical systems having the ability to collect data to further streamline their business processes and minimise waste [7]. Although some concepts and technologies of Industry 4.0 are still under development, there is a risk that they are only being developed for the large manufacturing companies, for instance the automotive industry. This could potentially endanger the SME sector which generally forms the backbone of most economies [8].

2.2 Identification of Organisational Readiness Factors

System requirements must be identified before ERP implementation can begin. These requirements will help in identifying the knowledge, processes and communication requirements of users of a new system. Risks involved when these system requirements are not met include: high system costs that may run over budget, late system delivery that does not meet user expectations, high maintenance costs and possibly an unreliable system [9].

To select an ERP system is a time-consuming task therefore it is necessary to assess the maturity of an enterprise in terms of factors affecting a successful implementation of an ERP system. An ERP implementation impacts the motivation, training and competence of the existing staff [10]. Both the requirements of the technology and the requirements of the organisation must be taken into account simultaneously [10]. Another correlation was found between the acceptance of employees toward technologies and organisational structure factors [11].

Implementing an ERP system also tests an organisations readiness for change. Change readiness can be defined as how an organisation's members recognise the need for change as well as recognising the organisation's own capability to accomplish these changes [12]. Organisational readiness assessment is a method where different dimensions of the organisation are assessed and the readiness of each organisational section for adopting an ERP system is evaluated [13]. Organisational readiness factors include cultural, organisational power, supportive, motivational and information technology infrastructure factors [13]. Other references to organisational readiness factors are included in Table 1.

Table 1: Main Organisational Readiness Factors

Main Organisational Readiness Factors	Literature Reviews
Information technology infrastructure	[14],[3],[15],[16],[13], [10], [11], [17]
Finance/costing/revenue	[16],[13], [17]
Culture/skills and staff/organisation	[14],[3],[18],[13], [10], [12], [11], [17]
Motivation/strategy/environment	[3],[18],[15],[16], [13], [17]
Support/structure and processes/management	[14],[19],[3] [18],[15],[16],[13], [17]

2.3 Organisational Life Cycle

Many different models exist to describe the life cycle of an organisation. Dodge and Robbins [20] describes a four stage model as Formulation, Early growth, Later growth and Stability. Smith et al. [21] proposes their three stage model as being: Start-up, Growth and Maturity. According to these models, it is evident that an organisation moves through similar life cycles. Petch [22] proposes a five stage Organisational life cycle describing it as: Seed and Development, Start-up, Growth and Establishment, Expansion and Maturity and Possible Exit. For the purpose of this study, emphasis will be placed on Stages 3-5, as the need for an ERP system grows once a company is established.

During Growth and Establishment, the main focus of the company is to manage increasing levels of revenue, attending to customers and suppliers and accommodating an expanding workforce [22]. The Expansion phase sees a rapid increase in revenue and cash flow as a result of increased sales and established production processes [22]. Focus during this phase should be to manage the process flow of the company to accommodate this expansion and ensure constant growth. During the Maturity phase, the focus is to maintain the systems that are in place to ensure that the momentum gained in the previous phase does not stagnate to a halt. High-level management and planning systems help ensure that the core of the business can remain constant [23].

3. RESEARCH METHODOLOGY

In order to find an appropriate framework for evaluating the readiness of an enterprise for ERP implementation, it is necessary to assess the maturity of an enterprise in terms of the factors affecting successful implementation of an ERP system.

The research consists of an overview of current literature on the research topic. A framework is developed from the different literature references focussing on readiness levels as well as business need analysis. The different readiness levels are then built up from different author's terminology and might therefore slightly differ to the exact terminology in literature. The framework that is created is used to determine where a company is positioned in terms of business needs and readiness level.

A company is first analysed by reading through the Readiness Level Matrix and positioning themselves within each readiness level. Next a company undergoes a business need analysis based on their current organisational needs.

This research uses a case study to analyse the proposed methodology and suggestions are made on how to match the company's readiness level to the business needs. The ERP solution provides the tool to help the company close the gap between their business needs and their readiness level.

The information gathered is used to indicate how a company can increase their readiness level to match their business needs. The analysis of a company is done in three consecutive steps. Firstly, the company's business needs are determined through a needs and as-is analysis. The second step is to determine the company's readiness level with the help of the readiness matrix in Table 4. The third step is to match the company's readiness level to their business process needs through the help of an ERP solution.

4. SCOPE

This research defines an ERP system as being a software package that can integrate all the complex business processes, update transactions in real time and store data all in one place. Other software packages such as small accounting software packages and less complex MRP systems do not fall into our scope of ERP systems. When looking at readiness leves it is important to note that the term 'readiness' and 'maturity' can be used interchangeably [24],[25],[26]. This research does not look at ERP readiness, but rather organisational readiness for an ERP system.

5. PROPOSED ORGANISATIONAL READINESS FRAMEWORK FOR AN ERP SYSTEM

The proposed organisational readiness framework consists of two main components: the business needs of a company, and the readiness level matrix.

5.1 Business Need Analysis

The business needs of a company are grouped into three stages according to the proposed business life cycle at the time of analysis. During the Growth and Establishment phase, the business processes mainly consist of employee and financial management system to complete day-to-day operations. As a business progresses through its life cycle and expands its operations, it starts to require more complex management systems to keep up with business process needs. Companies that focus on sales and production require systems that can manage their procurement, sales, inventory and manufacturing processes. As a company matures in their business processes, a need arises for higher-level process accountability and planning. Various intelligent reporting tools exist to help with planning and feedback to ensure that the company reaches their goals. Table 2 shows the business needs according to business life cycle model described in 2.3.

Table 2: Business needs according to business life cycle

Business Life Cycle	Business Process Needs
Maturity	Business Intelligence
	Project Management
	MRP
	System Traceability
	Asset Management and Tracking
Expansion	Materials Management
	Sales
	Production
	Procurement
	Inventory Management
	Stores
	Part Master
Growth and Establishment	General Ledger Reports
	Cash Book Transactions
	Accounts Payable
	Accounts Receivable
	Employee Roll

5.2 Readiness Level Matrix

The readiness levels are ranked from Level 1 -6. Level 1 is classified as being ready for a basic ERP. Level 6 is classified as organisational readiness for a full ERP system. At a low readiness level, an organisation can still be ready for some part of an ERP. The readiness level characteristics in this research are formulated from the research done in Table 1. Each of the readiness levels are made up of five characteristics, which are described in Table 3.

Table 3: Readiness Level Characteristics

Readiness Level Characteristic	Description
Systems	The presence of appropriate hardware and communication infrastructure in organisation.
Skills and Staff	Employees' knowledge on ERP systems and the integrated relationship towards the organisation.
Organisations Strategy	Organisation can plan and predict errors. Organisation have ability to train employees that need it. Organisational ability to devote suitable and permanent finance for ERP implementation. The vision of the organisation is in line with the vision and the driving force of employees to grow the business.
Structure and Processes	Support is given throughout the organisation from top management.
Costing and Revenue	Cost models, forecasting models, and revenue expenditure and stakeholder acceptance.

Table 3 is used to measure the readiness of the organisation. This is done in cooperation with the organisation to measure their overall readiness.

The overall readiness is determined by using the lowest readiness level measured. The argument is made that the lowest characteristic is the characteristic that have a negative influence on your overall readiness. By starting to improve on your lowest ranked characteristic level, you are improving the overall readiness level. This will ensure that the organisation systematically builds up its overall readiness instead of having certain outlying characteristics. The Organisational Readiness Level Matrix is described in detail within Table 4.

Table 4: Organisational Readiness Level Matrix

Level	Systems	Skills and staff	Organisation Strategy	Structure and processes	Costing and revenue
6	Systems are top of the range and fully functional to keep the business running. The systems are easy to use and personnel has full understanding of the integrated process between the organisation and the systems.	Strong computer literacy skills such as software understanding and systems are present throughout the organisation. IT management skills are matured.	The ERP system aids in the organisation's goals and objectives. Company is resilient to react to external factors.	Matured business processes that adapt to market change. Top management supports all the business processes. Process flexible to react timeously to change in trends.	The organisation has a self-financed ERP. Revenue sustainably robust through market variations. System cost competitive to drive uptake.
5	The organisation has the technical resources to maintain and develop a fully functioning ERP. All the business process are captured within the ERP system.	Senior management skills can be filtered down to junior staff members. Top management support to employees.	Top management has a clear understanding of when their business goals will be reached. This plan is in line with the organisation's vision and mission.	Top management have the ability to ensure change management throughout the business and ensure employee training. The supply chain is set up for future development.	The business processes are in place for the organisation to become self-financed. The cost model reliably reports the recommended retail price of the product. Product price and value proposition clear and attractive.
4	Previous success in redesigning business processes with upgrade to new systems. Most of the business processes are maintained and most available data can be accessed through the ERP system.	Personnel are disciplined and understand the changes necessary for successful implementation. Communication between employees and the system.	Clear understanding of organisational requirements and functionalities of ERP system. IT as a strategic plan for the future of their business processes.	Business process re-engineering is in place for redesign and rethinking of current business process. The role of IT comes with more responsibility as the business is growing.	Revenue projections based on proven forecasts and accepted commercial data. Product price sustainable to ensure market share increases.
3	The organisation have systems in place that can document more complex processes such as purchasing and sales transactions. Infrastructure can maintain a smooth integration to the upgrade system.	The employees understand the business and the systems enough to be able to adapt to change. The Users' skills and knowledge of the ERP system are increasing.	The need for change is building as the business is growing. Management is aware that training must be in place to help with change management. A risk management strategy is set in place.	Training documentation and support is in place to help personnel understand the systems processes as the business grows. Processes in place to maintain inter-departmental conflicts.	Revenue projections backed by commercial data. Price gaps understood and roadmaps in place to address them. Investors comfortable to secure debt. Cost model is verified to accurately forecast products cost for quotations.
2	The technology infrastructure can support the system and the resources that it needs to maintain. An ERP legacy system is in place and needs to be upgraded.	The organisation has systems in place to train personnel on the use of systems and general business processes.	Personnel are aware that there exists a need for change and improvement of current systems. A clear plan is in place to keep employees motivated during the changes to come.	The organisational hierarchy is starting to build, with more personnel in management roles. The organisational structure can be documented and managed by the system.	The organisation can finance their current systems and think about growing their size to keep up with business change. The cost model is being validated to actual accounting data.
1	The organisation has basic systems in place to do their financing and planning such as MRP systems and Pastel.	Organisation has a small number of personnel that understands the system. Limited availability of key operational skills.	The organisation wants to grow in terms of size and return on investment, but is limited due to the growth of business process complexity.	The organisational structure is becoming more complex. There exist a need for a system that can document processes that the current system cannot. Several high level responsibilities reside in one person.	Key costs based on projections with some actual data to verify. Cost model is being developed to determine the risk management strategy for the feasible region.

5.3 Organisational Readiness Framework for an ERP system

The business needs and the readiness matrix are plotted against each other with the assumption that the relationship between these factors are linear in nature. The hypothesis being described as a company's readiness tends to increase in terms of the characteristics described in Table 3, it progresses through the organisational life cycle. The organisational readiness framework in Figure 1, plots the business needs as a company matures against the company readiness level. The dotted line is a representation of this linear relationship between the two factors.

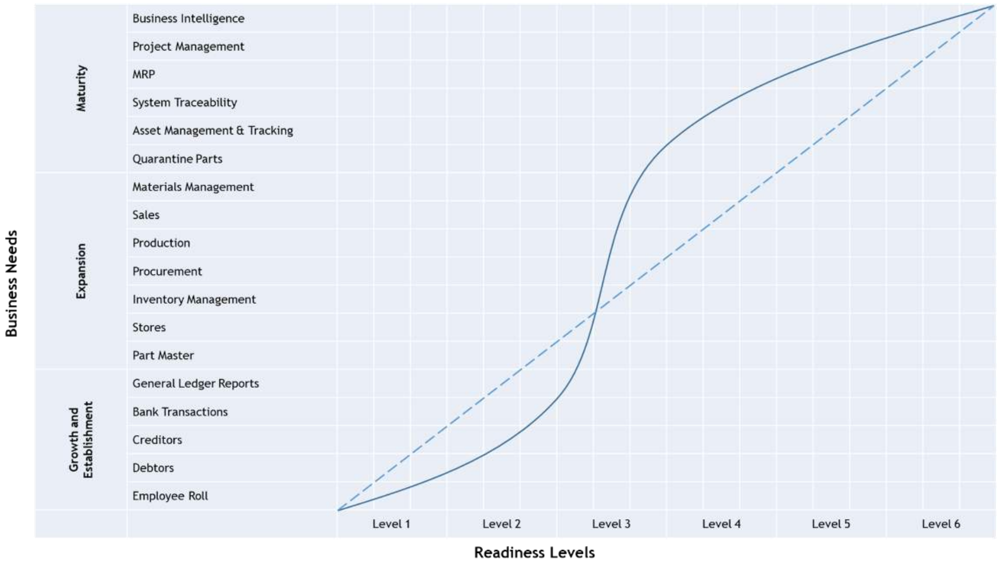


Figure 1: Organisational Readiness Framework

Initially a linear relationship was approximated for the interaction between the two axes in Figure 1. After further investigation, research found that the early stages of a business' life cycle does not yet conform to the necessary system requirements for a full ERP implementation. A more accurate representation of the interaction would be the S-curve, also seen in Figure 1.

6. CASE STUDY RESULTS

6.1 COMPANY A: Micro-precision manufacturer of mechanical Parts

The company studied (refered to as Company A) is active in the micro-precision manufacturing of metal parts industry with their main focus on dental components. Company A is a small enterprise (<50 employees) and primarily manufactures for the leading dental implant company in South Africa. Company A uses various precious and semi-precious metals raw-material stock for their manufactured products. Company A therefore requires inventory tracking at all times. There are different lead times on the various inventory items that require inventory orders to be placed at the correct time. This will ensure that the company does not end up with a shortage of raw materials. Due to the intricasy of the components being manufactured by Company A, each product consists of multiple manufacturing steps to reach the final geometry. The CNC machines are well maintained, but are technilogically outdated. During production, the CNC machines can only be configured for one machining operation at a time and thus poses a scheduling problem.

Not only is it difficult to schedule the machines, but the manpower to operate the machines needs to be skillfully planned as well. There is a 4:1 ratio of machine versus manpower. Specific machine operators are only qualified to complete certain jobs, and must be scheduled along with the machines to achieve the optimal results. The machine scheduling is done by one person on MS Project and printed out each day for the machine operators to follow. When the machine operators see that one machine is causing a delay on the scheduling, he/she needs to inform the person in charge of the scheduling and the entire schedule must be reviewed and possibly altered.

The machine operators track their own work-hours and machine hours by manually loading their hours to the operation assigned to them on a Google Documents sheet. This worksheduling is shared throughout the company for everybody to use and follow. Their biggest customer in South Africa, requires deliveries to be made twice a week, thus the operations must be planned in such a way to accommodate this contractual agreement. The company has recently come under new management with a renewed vision and has set out a three year plan in order to achieve this vision. After speaking and interacting with the employees, the general feeling is that some will be able to adapt to change and follow the leadership of top management. Other employees that have been working there for longer are more resistant to change. Management is aware that ERP training is required for all employees when considering possible improvements of their current systems. The training given to the employees should then be specific training on the chosen ERP system for the specific work of each employee.

Currently their accounting is done by one person on site. Previously they had an accountant that worked off-site and manually picked up and delivered their documents. The company has also only recently set-up a basic cost model reflecting what their actual costs are and how much they should price their manufactured products. There are many competitors in this niche market competing for customers in South Africa. It is important for Company A to be cost effective and ensure a competitive pricing catalogue to remain competitive in their industry. Due to budget and other regulatory costs, they recently lost their ISO accreditation for their products. The loss in ISO accreditation caused international clients not being allowed to buy from them anymore.

6.1.1 Business Needs Analysis

It is clear that Company A is in need of a system to accommodate the various business needs while producing an optimal result. As it is important to keep track of the various materials throughout the process, a materials management system is necessary. The company also needs a financial system that can accommodate sales and procurement.

6.1.2 Readiness Level Matrix

The company readiness level for ERP is measured to the Readiness Matrix in Table 4. The results are shown in Table 5. In Table 5 one can see the level of readiness for each characteristic and the reason for the specific level that was allocated. The company's main area of improvements are within their systems, skills and staff and their structure and processes. All of the characteristics that are seen as a weakness are currently placed at readiness level 1 for an ERP system. Therefore, we can determine that Company A has a Level 1 Organisational Readiness for an ERP system. Company A is placed at a readiness level of 1, as the overall level cannot be higher than the lowest readiness level.

Table 5: Company A Readiness level

Charactetistics	As-is Readiness Level	Reason for achieving readiness level
Systems	Level 1	Accounting on Pastel. Limited inventory management on Excel. Machine scheduling on MS Project. Googe docs for work scheduling. One computer that runs on Windows XP.
Skills and Staff	Level 1	Employees have not been exposed to different work environment and improvement possibilities. Some employees resitant to change. Certain employees only have one responsibility.
Organisational Strategy	Level 3	Company A's vision is mainly the vision of their founder. The employees still admire the founder. 80% of their production goes to one major client. Their main focus is to manage the material flow within processes
Structure and Processes	Level 1	One or two persons having to manage multiple roles in organisation. Not yet resource allocation. Their main business revolves around micro precision manufacturing. Therefore the material cost are very high, a lot of work in progress for a small product. Basic cost model in place.
Costing and Revenue	Level 2	They do their costing based on labour hours plus raw material cost. Accounting is mainly external and processes are being set in place to get internal accounting.

Readiness Level 1 is described as already having some sort of planning system, but also having the need to document more complex processes for their growing business. At this level you are ready for the most basic ERP system there is to offer.

Overall, Company A has many business processes that are still in a development phase, however they have a complex manufacturing process that can benefit from a system that documents and helps with their planning.

6.1.3 Suggested organisational readiness framework for ERP system

Currently, Company A has a higher business need than it is ready for. Figure 2 represents where they are and where they need to be according to their business needs. Currently they are at readiness level 1 and their business needs are classified as being within the maturity phase. The level of readiness they need to obtain in order to correspond to their business needs is at readiness level 4. Company A will have to implement several changes before they will be ready for an ERP system that can delivery the business process complexity they need. They are however ready for an ERP system that can execute the basic administration and finance processes. This solution is in contrast with what they need, but actualy shows what they are ready for at the moment. By increasing their readiness level, they will become ready for the system they actually need at the moment.

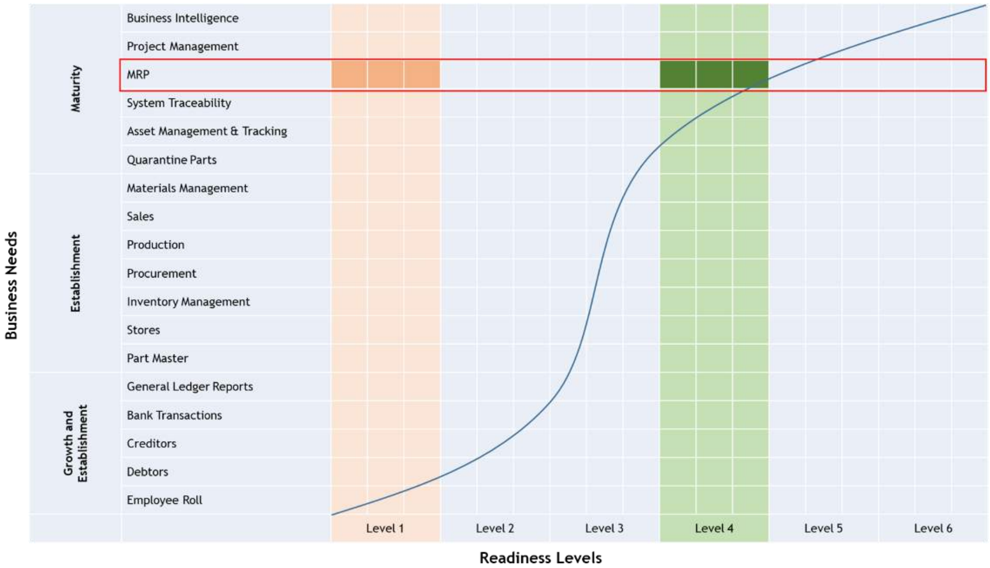


Figure 2: Organisational Readiness Framework for Company A

The suggested improvements for Company A to move to the next readiness level is captured within the Readiness Level Matrix in Table 4. Company A can use Table 4 as a referencing benchmark tool to see the steps necessary to move to the next readiness level. By using this method currently, they will reach the readiness level that fits with their required business needs.

A proposed recommendation would include the implementation of an ERP solution that fits with their current readiness level. The specific ERP solution must allow for the growth of the business as their readiness level escalates. With this approach the company will have a plan in place to grow and also a way of measuring their current readiness level to their future readiness level they want to achieve.

6.1.4 Suggested steps to follow to improve readiness level

The steps to follow in order for Company A to improve their readiness level is summarised in Table 6. These steps are generated from the Organisational Readiness Matrix in Table 4.

Table 6: Suggested steps to achieve To-be readiness level

Charactetistics	To-Be Readiness Level	How to achieve this level
Systems	Level 2	Company A can start by upgrading their current computer software software and also think of purchasing one or two more computers.
Skills and Staff	Level 2	Company A should start by giving their employees basic training on the use of computers and then focussing on gicing their employees training on the spesific ERP system that they choose to implement. The training can be in the form of a class given on every Friday for an hour or a basic "click and do" training doxument on the system.
Organisational Strategy	Level 4	This characteristic is the highest characteristic for company A and therefore do not need as much attention as the other characteristics. If Company A do however choose to improve this characteristic they can start to assign an IT team that can map a strategic plan for the future of their business processes.
Structure and Processes	Level 2	Company A can improve on their business structure by giving more employees more responsibility by not giving them multiple roles. Employees should be rewarded for their wok by moving up in a management position.
Costing and Revenue	Level 3	Company A should put a cost model in place that can accurately forecast product sales as well as costs and quotations. Roadmaps should be set in place and the price gaps should be easily understood by management. Management should also consider to grow in production size to keep up with market share.

The steps mentioned in Table 6 is only a suggestion based on the results from Table 4. Company A has not yet decided on an ERP system to implement. Once they do and the implementation is underway, the suggested Organisation Readiness Matrix tool can be tested.

6.1.5 Suggested ERP solution

A possible ERP system for Company A is the Qmuzik Silver ERP system [27]. This ERP system is a fully functioning ERP package of which functionality/modules can be toggled on and off, depending on the specific company requirements. The Qmuzik Silver ERP system provides companies a full ERP solution with the possibility of expanding integrated modules easily as the specific business grows. The proposed readiness framework integrates a readiness matrix together with a needs analysis of the company to identify the optimal ERP solution that the company requires.

7. CONCLUSION

The solution reflected in this research does not just benefit an organisation in terms of communication, enhanced asset management and efficiency but also provides a constant benchmark for an organisation as it progresses from one ERP readiness level to another. Additionally, the aforementioned research will benefit an organisation in maximising its progress to its preferred readiness level. The ability to implement a flexible ERP system allows a company to gain an advantage through the ability to see the progress of the business processes in real time.

The gathering and feedback of data acts as a stepping stone for the fourth industrial revolution, where the system uses data to allow all the processes to communicate with each other. By implenting a syste m, the human factor becomes less in business processes such as production. Systems allow for more automation within a process. This is but one of the requirements for the fourth industrial revolution.Using this platform through the implementation of an ERP system provides the opportunity for the company to be ready to integrate with the fourth industrial revolution in South Africa.

REFERENCES

[1] M. Shacklett, “Before implementing ERP, understand its many components,” 2015. [Online]. Available: http://searchmanufacturingerp.techtarget.com/feature/Before-implementing-ERP-understand-its-many-components.

[2] K. Cyrus, D. Aloini, and S. Karimzadeh, “How to Disable Mortal Loops of Enterprise Resource Planning (ERP) Implementation: A System Dynamics Analysis,” Multidisciplinary Digital Publishing Institute, Jan. 2018.

[3] R. Shafaei and N. Dabiri, “An EFQM Based Model to Assess an Enterprise Readiness for ERP Implementation,” *J. Ind. Syst. Eng.*, vol. 2, no. 1, pp. 51–74, 2008.

[4] SelectHub, “Key ERP Functional Requirements - Functions of ERP Systems,” 2015. [Online]. Available: https://selecthub.com/enterprise-resource-planning/key-erp-functional-requirements/.

[5] S. Kurkovsky, “Software engineering,” 2017. [Online]. Available: http://www.cs.ccsu.edu/~stan/classes/CS530/Slides/SE-09.pdf.

[6] C.-T. Lien and H.-L. Chan, “A Selection Model for ERP System by Applying Fuzzy AHP Approach,” *Int. J. Comput. Internet Manag.*, vol. 15, no. 3, pp. 58–72, 2007.

[7] M. Moore, “What is Industry 4.0? Everything you need to know,” *World of Tech*, 2018. [Online]. Available: https://www.techradar.com/news/what-is-industry-40-everything-you-need-to-know.

[8] M. Andulkar, D. T. Le, and U. Berger, “A multi-case study on Industry 4.0 for SME’s in Brandenburg, Germany,” in *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018, pp. 4544–4553.

[9] J. Whitten and L. Bentley, “Systems analysis and design methods,” 2007.

[10] G. Stewart, “Organisational Readiness for ERP Implementation,” *AMCIS Proc.*, no. January, pp. 966–971, 2000.

[11] H. Wraikat, A. Bellamy, and H. Tang, “Exploring Organizational Readiness Factors for New Technology Implementation within Non-Profit Organizations,” *Open J. Soc. Sci.*, vol. 05, no. 12, pp. 1–13, 2017.

[12] E. Finch, “Change Readiness,” *Facil. Chang. Manag.*, pp. 17–25, 2012.

[13] P. Hanafizadeh and A. Z. Ravasan, “A McKinsey 7S Model-Based Framework for ERP Readiness Assessment,” *Int. J. Enterp. Inf. Syst.*, vol. 7, no. 4, pp. 23–63, 2011.

[14] A. A. Fadelelmoula, “The Effects Of The Critical Success Factors For ERP Implementation On The Comprehensive Achievement Of The Crucial Roles Of Information Systems In The Higher Education Sector,” *Interdiscip. J. Information, Knowl. adn Manag.*, vol. 13, p. 44, 2018.

[15] F. Authors, “Journal of Enterprise Information Management Article information :,” 2013.

[16] T. Kalantari and F. Khoshalhan, “Readiness assessment of leagility supply chain based on fuzzy cognitive maps & interpretive structural modeling: A case study,” *J. Bus. Ind. Mark.*, pp. 00–00, 2018.

[17] L. Bezuidenhout, “Emerging technologies: commercial readiness index(CRI) for medical additive manufacturing(AM),” no. December, 2017.

[18] W. E. K. Lehman, J. M. Greener, and D. D. Simpson, “Assessing organizational readiness for change,” *J. Subst. Abuse Treat.*, vol. 22, no. 4, pp. 197–209, 2002.

[19] A. J. Shenhar *et al.*, “Toward a NASA-specific project management framework,” *EMJ - Eng. Manag. J.*, vol. 17, no. 4, pp. 8–16, 2005.

[20] R. H. Dodge and J. E. Robbins, “An empirical investigation of the organizational life cycle model for small business development and survival,” *J. Small Bus. Manag.*, vol. 30, no. 1, pp. 27–37, 1992.

[21] K. G. Smith, T. R. Mitchell, and C. E. Summer, “Top Level Management Priorities in Different Stages of the Organizational Life Cycle,” *Acad. Manag. J.*, vol. 28, no. 4, pp. 799–820, Dec. 1985.

[22] N. Petch, “The Five Stages Of Your Business Lifecycle: Which Phase Are You In?,” 2016. [Online]. Available: https://www.entrepreneur.com/article/271290.

[23] C. M. Olszak, “Business Intelligence and Analytics in Organizations,” Springer, Cham, 2015, pp. 89–109.

[24] J. D. Smith, “An Alternative to Technology Readiness Levels for Non-Developmental Item (NDI) Software,” *Proc. 38th Annu. Hawaii Int. Conf. Syst. Sci.*, vol. 00, no. C, p. 315a–315a, 2005.

[25] J. D. Smith II, “ImpACT: An alternative to technology readiness levels for commercial-off-the-shelf (COTS) software,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 2959, pp. 127–136, 2004.

[26] A. Tetlay and P. John, “Determining the Lines of System Maturity, System Readiness and Capability Readiness in the System Development Lifecycle.,” *7th Annu. Conf. Syst. Eng. Res. (CSER 2009)*, vol. 2009, no. April, pp. 1–8, 2009.

[27] Qmuzik, “Qmuzik Technologies (pty) Ltd.,” 2017. [Online]. Available: https://www.qmuzik.com/index.php/enterprise-solutions/.

DEVELOPMENT OF A HEALTH SYSTEM FRAMEWORK TO GUIDE THE ANALYSIS OF INNOVATION ADOPTION IN LOW AND MIDDLE INCOME COUNTRIES

E. Leonard[1*], I.H. De Kock[2] & W.G. Bam[3]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
17026946@sun.ac.za

[2]Department of Industrial Engineering
Stellenbosch University, South Africa
imkedk@sun.ac.za

[3]Department of Industrial Engineering
Stellenbosch University, South Africa
wouterb@sun.ac.za

ABSTRACT

Healthcare systems face numerous challenges that put strain on the system. This is despite the countless resources that are expended on creating innovative healthcare solutions (ranging from innovative healthcare technologies, organisational innovations to pharmaceutical innovations). The literature on innovation and healthcare has shown that the adoption of innovations in practice within the healthcare system is hindered and limited. There is a need to explore and evaluate the role of the innovation system in South Africa insofar as it impacts the adoption of innovations into the national healthcare system. As a starting point, for assessing innovation adoption into health systems, it is necessary to be able to thoroughly describe a health system. In this paper a consolidated health system framework is developed. The purpose of this framework is to be utilised when developing a healthcare innovation adoption framework, i.e. as an input to the healthcare innovation adoption framework. The methodology used to develop the consolidated health systems framework are the eight phases of Jabareen's conceptual framework. Jabareen's framework is a qualitative technique for developing conceptual frameworks. The advantages of using this conceptual framework methodology include its capability to be modified, its flexibility and the focus being placed on understanding, rather than on predictions. The results of this research paper are a consolidated health systems framework which was created by considering existing health system frameworks. The consolidated health systems framework thoroughly describes all aspects of a health system by combining elements from six existing health system frameworks. The elements of the existing frameworks were categorised, integrated and synthesised, as per Jabareen's methodology, to create a complete view of a health system; which includes health processes, building blocks, intermediate objectives and goals. This paper contributes to the field of health systems engineering by providing an extensive list of existing health system frameworks and by providing a framework that combines the major aspects of a health system to thoroughly and completely describe health systems.

[1]The author was enrolled for an MEng(Industrial) degree in the Department Industrial Engineering, Stellenbosch University, South Africa.
*Corresponding author

1. INTRODUCTION

In order to improve a country's health system, a superior quality of research is essential [1]. Health research encourages the improvement of a country's health equity, performance and health systems [1]. Substantial advances have been made in global healthcare during the past few years, and large sums of money have been spent on healthcare research and development. During the 2015/16 fiscal year R11.3 billion was spent on research and development in South Africa, of this 18.1% was used for health research and development [2]. Even with these substantial amounts being spent on developing healthcare innovations, the innovations often do not get implemented into the healthcare system where they are needed [3]. Without good implementation practices, health research is not worth much, as Pressman & Wildavsky [4] state, good ideas have little value if they are not implementable.

The chasm between knowledge and practice means that healthcare stakeholders are not receiving the benefits from health advances - this could be in terms of costs or lifesaving technologies [5]. Clinical and scientific structures have been unable to keep up with the acceleration in new scientific discoveries and technologies. The structures, as they stand, and the available resources and work force, are unable to effectively translate the new discoveries and advancements into practice. This creates missed opportunities where people's lives and health could have been improved [6]. Barriers to innovation adoption can occur on several levels, including at patient level, departmental levels, healthcare organisational level or at policy level [7]. There is a need for a better understanding of the translation process to ensure a higher percentage of health technologies and therapies are successfully implemented [6]. A key part to understanding this translation process is understanding the health system. This paper focuses on developing a framework that effectively describes all aspects of a health system, aspects that could potentially influence the adoption of innovations into practice. This includes determining a health system's functioning and building blocks.

Forms of health systems have existed since societies deliberately attempted to protect their health and themselves from diseases [8]. Health systems as we currently know them have been moulded and refined from the late 19th century health system designs [8]. Health systems are organised differently around the world; however this is not to say that one way of organising a health system is better than another. What is important is that the health system's structure enables good performance of the system's fundamental functions [8]. Health systems are crucial in improving the health of a country [9]. From 1952 until 1992 the World Health Organisation approximated that the implementation of new technologies and knowledge into health systems accounted for half of the improvements in health globally [10]. The healthcare landscape is unstable, the path that healthcare follows is unpredictable; new opportunities, challenges, legislatures and diseases constantly arise. This unstable operating environment needs innovation [11] in order for the healthcare environment to adapt to the ever present changes accordingly.

In this paper a health system framework is constructed using Jabareen's [12] methodology for building conceptual frameworks. The developed health systems framework is the first stepping stone towards understanding how innovation adoption within healthcare works.

2. METHODOLOGY

To develop a consolidated health systems framework, Jabareen's [12] conceptual framework was used; the eight phases of the framework are displayed in Figure 1. Jabareen's [12] framework is a qualitative technique for developing conceptual frameworks. The advantages of using this conceptual framework methodology include its capability to be modified, its flexibility and the focus being placed on understanding, rather than on predictions [12].

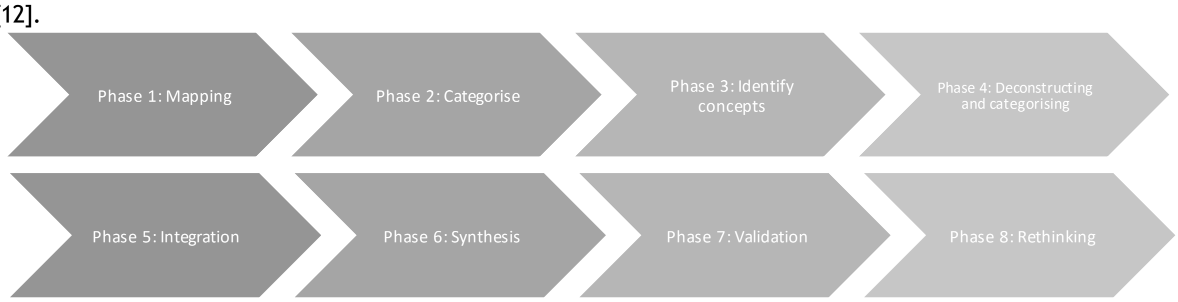


Figure 1 Conceptual framework methodology developed from Jabareen [12]

3. TOWARDS A CONSOLIDATED HEALTH SYSTEM FRAMEWORK

There has been increased international interest in healthcare systems and the frameworks that describe them [13]. How well a healthcare system performs, correlates with the achievement of health and development goals in a country [14]. Globally institutions are realising that even with health improvement initiatives which focus on particular healthcare outcomes, more effective and efficient healthcare systems are needed in order to attain and sustain healthcare goals [13]. The diversity of existing healthcare frameworks emphasise that there is no shared understanding of what health systems are - this can become problematic when different stakeholders interpret health systems differently [15]. This variety is a result of people from different regions, disciplines and timeframes understanding and interpreting health systems differently [16].

The most commonly used definition of a health system [16] is from the World Health Organisation's *World Health Report* where a health system is defined as "all the activities whose primary purpose is to promote, restore or maintain health", [17] this includes the resources, people, institutions and organisations whose principal aim is to improve health [18]. In the World Health Organisation's report *Monitoring the Building Blocks of Health Systems* the definition of a healthcare system continues and includes the supply of promotive, preventative, rehabilitative and curative care, by state and non-state actors [18].

A considerable amount of time and energy has been spent on the development of health system frameworks [15]; leading to numerous health system frameworks being published in healthcare literature [19]. The variety of existing health system frameworks present challenges as each healthcare system framework has been developed with different driving forces, in terms of emphasis, scope, usability, categories, and language [20]. However all of the frameworks aim to offer an enhanced understanding of a healthcare systems (its structure, goals and performance drivers) [15] and provide complementary health system views [20]. Among international health system frameworks there has been a substantial amount of appropriation of preceding frameworks; this suggests that some convergence in parts of the multiple health frameworks has occurred [15].

Through the research that Papanicolas & Smith [15] have conducted on health system frameworks, they deduced that there has been some convergence in the architecture, goals and problem areas of healthcare system frameworks. This convergence suggests that value obtained from developing a completely new framework is low [15]. When analysing a healthcare framework it is necessary to determine what the framework's focus and principles are as well as the understandings the author of each framework had during conceptualisation [15].

3.1. Phase 1: Mapping the selected data sources

In Phase 1 data on health systems frameworks was collected from literature sources; this was done by conducting an extensive review of the health system literature available online. Refer to Table **5** in Appendix A for the comprehensive list of health system frameworks uncovered during the literature review. From the literature search 49 health system frameworks were found. This is not to say that the list contains all of the existing health system frameworks, however this list contains the major/influential frameworks and a large variety of health system framework perspectives. Therefore these frameworks will provide a sufficient overview of the different types of health system framework literature that exists. The numerous existing frameworks serve varied purposes depending on their envisioned use, and on their intended audience; a framework will emphasise certain functions or features of the healthcare system and disregard others [19].

To complete an initial filtration process of the health system frameworks in Table **5**, the abstracts of each framework were read in order to establish whether the frameworks presented new elements or ideas, or whether the frameworks were based too heavily on preceding frameworks. Frameworks whose papers were not freely available were excluded. The frameworks were also screened in order to determine whether they made use of systems thinking, which is deemed necessary, and whether the framework was too specific to be of use (e.g. focusing on one disease). These exclusion criteria output the health system frameworks displayed in Table 6. The frameworks in Table 6 will be considered further to develop a consolidated health systems framework.

3.2. Phase 2: Extensive reading and categorising the selected data

Healthcare frameworks can be categorised as either conceptual or evaluative [15]. A conceptual framework provides an overview of the health system by describing, explaining and providing definitions for the health system [20], i.e. it is a descriptive framework. An evaluative framework is a framework that is based around actions allowing the user of the framework to evaluate and analyse aspects of the health system's performance, functions and factors [20], i.e. it is an interactive framework. A conceptual framework can be used as the foundation of an evaluative framework. Whereas an evaluative framework can not necessarily function as a conceptual framework [15].

Health system frameworks can then be further broken down according to their goals. The framework could be created in order to understand a health system (e.g. the systems' goals, actors, functions) [16], illustrating and providing an overall understanding of the health system, without necessarily showing the manner in which the system operates [20]. The framework could be created to compare health systems (e.g. between countries) [16], by trying to establish which factors influence how efficient the health system's functions are, which allows one to understand why certain systems outperform others [20]. The framework's goal could be to inform change within a health system (e.g. policy changes) [16]; or to evaluate the system [16], by describing and analysing certain features of a health system [20].

Another method of classifying healthcare system frameworks is to determine where the boundaries of the healthcare system lie. There are no clear lines which differentiate between what does and what does not reside within a healthcare system's boundaries [14]. The complexity of health systems makes it difficult to define precisely what components they contain, and what their starting and ending points are [17]. Thus health systems have been described in numerous ways [19]. Dependant on how the health system boundaries have been set, the fundamental responsibility for health improvement would rely on different stakeholders [15]. The advantages and disadvantages of having a framework with wider and narrower boundaries as described by Papanicolas & Smith [15] are presented in Table 1.

Table 1: Implications of health system boundaries on a health system framework, adapted from Papanicolas & Smith [15]

	Narrow boundary	Wide boundary
Advantages	Stakeholders held accountable more easily.	More realistic view of the factors that impact healthcare.
	Areas where stakeholders are capable to make changes can be identified.	Relationships between institutions, people and sectors are identified.
Disadvantages	A large portion of factors that influence healthcare are not represented.	Elements included are often difficult to change in a short timeframe.
	Difficulties identifying the effect the elements have on the environment they are in.	Managerial roles are not clarified.
		Challenging to allocate responsibility, and to hold role-players accountable.

A narrow boundary allows for greater accountability of healthcare system role-players during improvement initiatives; however a narrow boundary can also introduce accountability complications seeing that many healthcare determinants fall outside of narrow boundaries [15]. A wider boundary allows for a more complete understanding of healthcare factors [15]. Health system boundaries can be divided into three categories [16]:

- Sub-frameworks focus on specific parts of a healthcare system;
- The frameworks category encompasses the whole healthcare system;
- Supra-frameworks are frameworks outside the limits of traditional healthcare systems: these frameworks consider how the healthcare system interacts with other societal systems.

The health system frameworks have been categorised according to their goals (understanding, comparing, informing change or evaluating) and where their boundaries have been set (sub-framework, framework, or a supra-framework). For example a framework can be categorised as an understanding supra-framework. Refer to Table 6 for the categorisation of the 26 existing health system frameworks that were output after Phase 1. A visual representation of the possible categories is show in Figure 2.

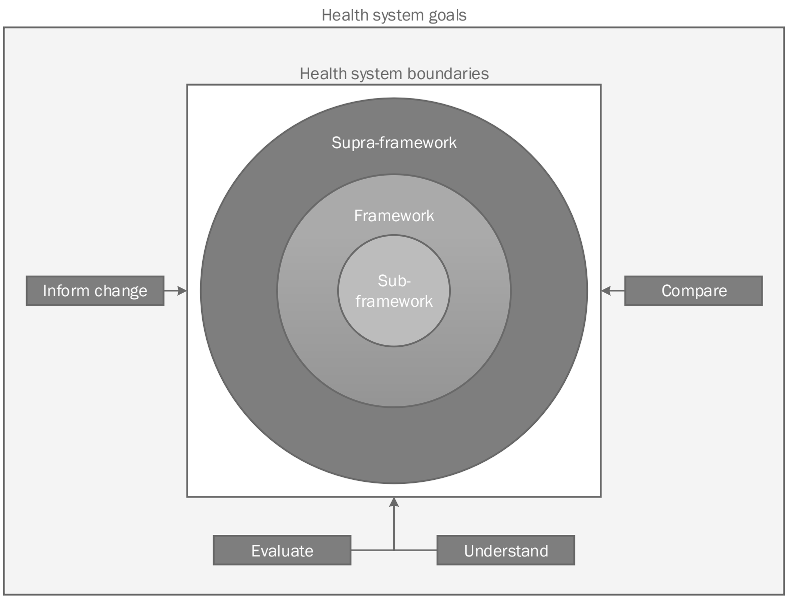


Figure 2: Framework categorisation

Each of the frameworks in Table 6 were considered in more detail in order to determine which frameworks would be used to develop the consolidated health systems framework. The criteria used to assess each framework in Table 6 consists of the following:

- The framework must account for all applicable stakeholders' perspectives [15];
- The framework should describe the objectives of the healthcare system [15];
- All significant elements of the healthcare system should be considered [15], i.e. the framework should not be too broad or vague;
- A systems thinking mind-set should be adopted[1];
  - The framework must incorporate the links between the system and the environment it exists in [9];
  - The framework must provide a holistic view of a health system (it should not be too specific);
  - The framework should enable an understanding of health systems and how the different components of the system relate and interact with one another;
- The framework should have the ability to support the assessment of healthcare systems in low and middle income countries;
- The activities included within the health system framework's boundary must be clearly identifiable [15];
- The framework's goal should correspond with the aim[2] of this investigation to the extent that without this correspondence, considering the framework further would not be beneficial towards developing the consolidated health system framework.

In order for a healthcare system framework to make it through this filtration step, it needs to adhere to two or more of these previously mentioned criteria. The frameworks that remained after applying the stated criteria to the health system frameworks in Table 6 are described in Table 2.

Table 2: Health systems frameworks after second round of filtration

Framework	Category	Description	Source
Health systems building block framework	Understanding framework	This framework's goal is to develop a common understanding of what a health system consists of as well as areas where health strengthening measures can be applied [21]. This framework describes six building blocks (service delivery; workforce; vaccines, products and technologies; information; financing; governance and leadership), which a health system is composed of [21]. The building blocks are founded on the 2000 World Health Report, *Health Systems: Improving Performance*, each building block is necessary to improve health outcomes [21].	[21]
Control knobs framework	Evaluating framework	Roberts et al. (2002) conceptualise health systems in terms of control knobs. These control knobs are a metaphor for factors that influence a health system's performance, they are discrete aspects that significantly impact health system performance [14]. Changing the control knobs' settings (health system factors) determine how the health system functions [14].	[14]
Health systems context framework	Understanding supra-framework	The health systems context framework provides an understanding of the connections between health systems and the environment in which the system exists [9].	[9]
Health systems in transition	Comparing framework	The health systems in transition framework provides countries the ability to generate thorough descriptions of their health systems in a standard set up [22].	[22]
Health systems strengthening framework	Evaluating framework	The health systems strengthening framework is built on a foundation of four health system components. These components are stewardship and governance, monitoring and evaluation, financing system and health services [23]. Each component consists of a combination of health system processes, elements and functions; these components are identified as being the areas where health system strengthening activities can take place [23]. The health systems strengthening framework emphasises that the components are inter-related, and that adjusting one part of a component will have repercussions elsewhere in the system.	[23]
Converging health systems frameworks	Understanding Supra-framework	While Shakarishvili et al. [24] did not propose a framework in their paper *Converging Health Systems Frameworks: Towards A Concepts-to-Actions Roadmap for Health Systems*	[24]

[1] It is important to assume a systems thinking perspective for healthcare systems, healthcare systems display the significant features of a complex dynamic system [9]. Systems thinking considers the context in which a system is operating and the system itself as a complex entity of interdependent and interconnected parts [9]. Systems thinking is the capability to view a system as a whole, that contains multiple interdependent and interconnected parts, and not just the individual components [64].

[2] The overall aim of this research is to develop a framework which analyses the system of innovation and the healthcare system in low and middle income countries insofar as these systems influence the adoption of pharmaceutical and technological innovations into the healthcare system.

		Strengthening in Low and Middle Income Countries, they compiled elements from various health system frameworks that correspond with each other.	

3.3. Phase 3: Identify concepts

Each of the frameworks described in Table 2 offer complementary perspectives of health systems. The health systems building blocks framework provides a succinct way of understanding health systems. The influence of this framework can be seen in numerous other health system frameworks. The control knobs framework takes the approach of identifying which aspects of a health system can be influenced. The health systems context framework emphasises the importance of understanding the environment which the health system is operating in, as the environment will impact how well the health system operates. The health systems in transition framework is very practical in the way that it provides a structured and a reliable method of analysing a health system. The health systems strengthening framework provides an exhaustive list of health system elements and highlight the fact that all of the elements are interlinked. The converging health systems framework provides a detailed overview of health system elements.

After analysing numerous health system frameworks the concepts of a healthcare system framework that have been deemed necessary for a complete health systems view are: context, building blocks (functions), control knobs (processes), intermediate objectives and goals as displayed in Figure 3. Some frameworks would only include one of these concepts, while others would include multiple ones.

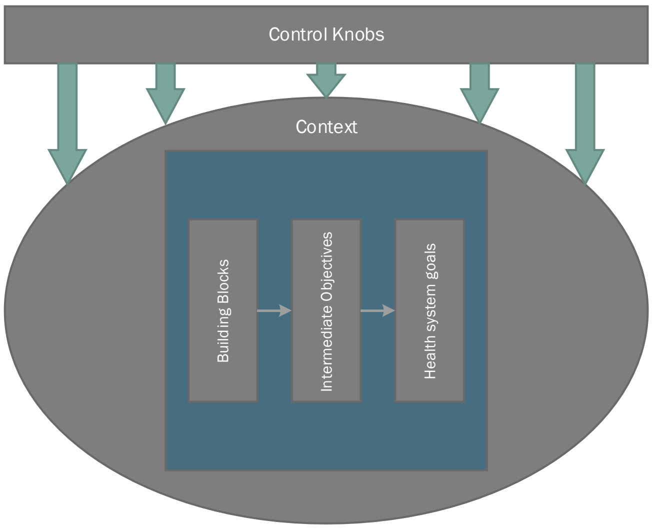


Figure 3: Consolidated health system concepts

3.4. Phase 4: Deconstructing and categorising

In order to determine what the five health system components should consist of (i.e. the health system elements that each health system component is made up from), the six frameworks in Table 2 were analysed. The health system frameworks in Table 2are deconstructed into their basic elements, and these elements are categorised according to the health system components. Table 3 displays the elements from the health system frameworks categorised according to the proposed health system components.

Table 3: Health system framework elements categorised according to health system components

		Health system elements					
		Health systems building block framework [21]	**Control knobs framework** [9]	**Health systems context framework** [14]	**Health systems in transition** [22]	**Health systems strengthening framework** [23]	**Converging health systems frameworks** [24]
Health system component	**Context**		Economic; Legal and regulatory; Political; Demographic; Technological; Epidemiological; Socio-demographic;		Political; Health status; Socio-demographic; Economic context; Geography;		

			Environmental;				
	Control knobs - Processes		Financing; Organisations and regulations; Resource; Provision;	Financing; Payment; Organisation; Regulation; Behaviour;			Resource creation; Resource allocation; Payment; Organisation; Integration; Regulation; Behaviour;
	Building blocks - Functions	Service delivery; Health workforce; Information; Medical products, vaccines, technologies; Financing; Leadership and governance;			Organisation and governance; Financing; Physical and human resources; Provision of services; *(has sub-elements for each building block)*	Health services; Stewardship and governance; Financing system; Monitoring and evaluation; *(has sub-elements for each building block)*	Services; Health workforce; Health information; Technologies and commodities; Demand generation; Financing; Governance;
	Intermediate objectives	Access; Coverage; Quality; Safety;	Equity; Choice; Efficiency; Effectiveness;	Efficiency; Quality; Access;			Equity; Efficiency; Sustainability; Quality; Access; Coverage; Safety; Choice;
	Health system goals	Improved health; Responsiveness; Social and financial risk protection; Improved efficiency;	Health; Financial risk protection; Consumer satisfaction;	Health status; Customer satisfaction; Risk protection;			Better health; Financial protection; Responsiveness; Satisfaction;

3.5. Phase 5: Integrating

All of the health system elements in Table 3 were deliberated in order to create the consolidated health system elements displayed in Table 4. Table 4 shows the consolidated elements of each health system component and provides a brief description of how the elements were chosen.

Table 4: Consolidated health system elements

	Elements	Comments
Context	Political [9] [22] Health status [22] Sociodemographic [9] [22] Economic context [9] [22] Geography [22]	The factors from Thomson et al. [22] framework were used as the base for the context component; their report made it clear as to what these factors entail. However there are a lot of Atun & Memable's [9] factors that overlap.
Control knobs: Processes	Resource creation [9] [24] Resource allocation [24] Payment [14] [24] Financing [9] Organisation [9] [14] [24] Integration [24] Regulation [14] [24] Behaviour [14] [24]	Shakarishvili et al. [24] combined elements from both Atun & Memable [9] and Roberts et al. [14], it was therefore deemed appropriate to use Shakarishvili et al. [24] elements as the basis for the control knobs component with the exception of financing which was added to the list.
Building blocks:	Service delivery [21] [22] [23] [24] Information [21] [24] Physical resources [21] [22] [24] Human resources [21] [24] [22] Financing [21] [22] [23] [24] Leadership and governance [21] [22] [24]	WHO's [21] building block elements were used as the basis for the building blocks component. The health systems building blocks framework has been influential in health systems framework literature, this can be seen through Thomson et al. [22], Shakarishvili et al.

		[23], Shakarishvili et al. [24] building blocks, which all relate to WHO's [21] building block elements.
Intermediate objectives	Equity [9] [24] Efficiency [9] [14] [24] Sustainability [24] Quality [21] [14] [24] Access [21] [14] [24] Coverage [21] [24] Safety [21] [24] Choice [9] [24]	Here Shakarishvili et al. [24] successfully combined WHO [21], Roberts et al. [14] and Atun & Memable [9] intermediate objectives.
Health system goals	Improved health status [21] [9] [14] [24] Responsiveness [21] [24] Social and financial risk protection [21] [9] [14] [24] Improved efficiency [21] Consumer satisfaction [9] [14] [24]	The health system goals elements are based on WHO's [21] elements with the exception of customer satisfaction which was first proposed by Roberts et al. [14]. Again it can be seen that multiple frameworks have over lapping health system goals.

3.6. Phase 6: Synthesise

In this phase the concepts, components and elements are synthesised into a theoretical framework. The consolidated framework was developed by considering the following objectives:

- The framework can support the assessment of healthcare systems in different sub-Saharan African countries;
- The framework can be linked to or integrated with a system of innovation;
- The framework provides a holistic view of the healthcare landscape (framework cannot be too specific);
- The framework enables an understanding of the health system and how the different components of the system relate and interact with one another;
- The framework contributes to the development of a tool that assesses the role of a healthcare system and an innovation system in the adoption healthcare innovations.

Figure 4 shows the Consolidated Health Systems Framework (CHS Framework) and the interactions between the health system components.

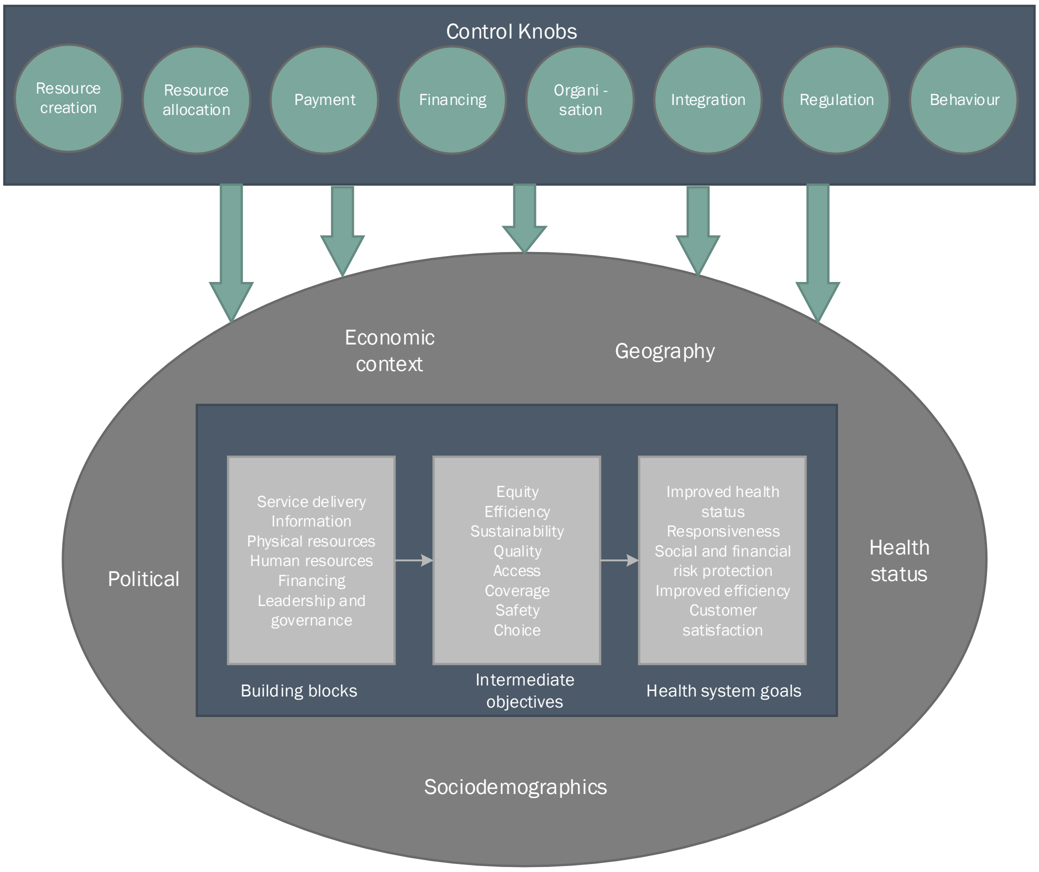


Figure 4: Consolidated health system framework (CHS Framework)

3.7. Phase 7 and 8: Validating and rethinking

Phases 7 and 8 do not come into play yet. Validation and rethinking the CHS Framework will take place when the framework for health innovation adoption is developed. The consolidated health systems framework is an input for the development of the innovation adoption framework. Validation and rethinking of the CHS Framework (along with the other inputs) will have to occur within the context of innovation adoption, when developing the overall innovation adoption framework.

4. CONCLUSIONS

A considerable amount of time and energy has been spent on the development of health system frameworks [15]; leading to numerous health system frameworks being published in healthcare literature [19]. The variety of existing health system frameworks present challenges as each healthcare system framework has been developed with different driving forces, in terms of emphasis, scope, usability, categories, and language [20]. Among international health system frameworks there have been a substantial amount of appropriation of preceding frameworks; this suggests that some convergence in parts of the multiple health frameworks has occurred [15].

The aim of this research paper was to develop a consolidated health systems framework using existing frameworks which were found in literature to create a framework that combines all major aspects of a health system to thoroughly and completely describe health systems. In Section 3.2 the existing health system frameworks were categorised. One of the criteria for the categorisation was where the boundaries of the health system framework have been set. The boundaries could be at sub-framework level (focus on specific parts of a health system), at framework level (encompasses the whole healthcare system), or at supra-framework level (considers how the healthcare system interacts with other societal systems). Sub-frameworks were deemed too specific, supra-frameworks were deemed too vague, and frameworks were deemed to be missing aspects or elements necessary to appropriately describe a health system; an ideal framework would include the details of the framework level while still considering the aspects of a supra-framework i.e. the context of the health system. Rather than choosing an existing health system framework, combining numerous frameworks to create a comprehensive framework was deemed best, as none of the analysed frameworks contained all the elements that were found across the various evaluated frameworks. This meant being able to get the best aspects of the supra-framework and of the framework categories in the CHS framework.

As per the name of the developed framework, the *Consolidated Health System Framework*, this framework consolidates all aspects (elements and components) of previous frameworks to provide a comprehensive, consolidated view of the health system. Jabareen's [12] methodology for building conceptual frameworks was followed. The 49 frameworks found through the literature search were filtered down to six health system frameworks. The details of these six frameworks were used, deconstructed, categorised and integrated to form a single consolidated health systems framework. The consolidated health systems framework is the first stepping stone towards understanding how innovation adoption within healthcare works. The CHS Framework differs from the existing health system frameworks in the way that it thoroughly and comprehensively describes a health system. This is due to the CHS Framework being developed by combining elements from preceding health system frameworks. The consolidated health systems framework is an improvement on existing frameworks as it displays an overview of a health system, using concepts and elements from the existing frameworks in such a way that a new thorough representation of a health system, that can be used for a variety of purposes, has been created. None of the analysed frameworks in Table 2 contain all of these elements and concepts in a single framework. The developed framework can be used by future researchers when an overview of a healthcare system is required. Recommendations for future work would be to use this framework in the context of an actual healthcare system to determine whether there are still aspects missing. Further the consolidated health systems framework needs to be rethought and validated by health industry experts in future work.

5. REFERENCES

[1] F. Paruk, J. M. Blackburn, I. B. Friedman, and B. M. Mayosi, "Health and finance: National expenditure on health research in South Africa: What is the benchmark?," *South African Med. J.*, vol. 104, no. 7, pp. 468–474, 2014.

[2] PATH, "Health Research and Development Budget Allocations and Expenditures in South Africa: A Baseline Report," 2016.

[3] J.-L. Denis, Y. Hébert, A. Langley, D. Lozeau, and L.-H. Trottier, "Explaining Diffusion Patterns for Complex Health Care Innovations," *Health Care Manage. Rev.*, vol. 27, no. 3, pp. 60–73, 2002.

[4] J. Pressman and A. Wildavsky, *Implementation: How Great Expectations in Washington Are Dashed in Oakland*, 3rd ed. California: University of California press, 1973.

[5] A. Bergström *et al.*, "Health system context and implementation of evidence-based practices-development and validation of the Context Assessment for Community Health (COACH) tool for low- and middle-income settings," *Implement. Sci.*, vol. 10, no. 1, 2015.

[6] S. A. Waldman and A. Terzic, "Clinical Translational Science 2020: Disruptive Innovation Redefines the Discovery-Application Enterprise," *Clin. Transl. Sci.*, vol. 4, no. 1, pp. 69–71, Feb. 2011.

[7] L. J. Damschroder, D. C. Aron, R. E. Keith, S. R. Kirsh, J. A. Alexander, and J. C. Lowery, "Fostering implementation of health services research findings into practice: a consolidated framework for

advancing implementation science," *Implement. Sci.*, 2009.
[8] P. Musgrove *et al.*, "The World Health Report 2000 Health Systems: Improving Performance," 2000.
[9] R. Atun and N. Memable, "Health Systems and Systems Thinking," *Heal. Syst. Chall. Commun. Dis. Exp. from Eur. Lat. Am.*, pp. 121–140, 2008.
[10] World Health Organization, "The World Health Report Making a Difference," 1999.
[11] T. O. Akenroye, "Factors influencing innovation in healthcare: A conceptual synthesis," *Innov. J.*, vol. 17, no. 2, pp. 1–21, 2012.
[12] Y. Jabareen, "Building a Conceptual Framework: Philosophy, Definitions, and Procedure," *Int. J. Qual. Methods*, vol. 8, no. 4, pp. 49–62, 2009.
[13] World Health Organization, "The Global Fund Strategic Approach to Health Systems Strengthening," *Rep. from WHO to Glob. Fund Secr.*, p. 46, 2007.
[14] M. Roberts, W. Hsiao, P. Berman, and M. Reich, "Getting Health Reform Right: A Guide to Improving Performance and Equity," *Get. Heal. Reform Right A Guid. to Improv. Perform. Equity*, no. October, pp. 1–344, 2002.
[15] I. Papanicolas and P. C. Smith, *Health System Performance Comparison: An Agenda for Policy, Information and Research*. 2013.
[16] S. Hoffman *et al.*, "A Review of Conceptual Barriers and Opportunities facing Health Systems Research to inform a Strategy from the World Health Organization Facing Health Systems Research to Inform a Strategy from the World Health Organization," *Syst. Res.*, 2012.
[17] P. Musgrove, A. Creese, A. Preker, C. Baeza, A. Anell, and T. Prentice, "Health Systems: Improving Perfomance," *World Health Organization*, vol. 78, no. 1. pp. 1–215, 2000.
[18] World Health Organization, "MONITORING THE BUILDING BLOCKS OF HEALTH SYSTEMS: A HANDBOOK OF INDICATORS AND THEIR MEASUREMENT STRATEGIES," *World Heal. Organ.*, pp. 1–92, 2010.
[19] A. Ergo, R. Eichler, M. Koblinsky, and N. Shah, "Strengthening Health Systems to Improve Maternal, Neonatal and Child Health Outcomes : A Framework," no. May, 2011.
[20] G. Shakarishvili, R. Atun, P. Berman, W. Hsiao, and C. Burgess, "Building on Health Systems Frameworks for Developing a Common Approach to Health Systems Strengthening," *World bank, Glob. Fund, GAVI Alliance*, 2009.
[21] WHO, "Everybody's business: strengthening health systems to improve health outcomes: WHO's framework for action.," *Production*, pp. 1–56, 2007.
[22] S. Thomson, B. Rechel, and E. Van Ginneken, "Health Systems in Transition," *World Health*, 2010.
[23] G. Shakarishvili *et al.*, "Health systems strengthening: A common classification and framework for investment analysis," *Health Policy Plan.*, vol. 26, no. 4, pp. 316–326, 2011.
[24] G. Shakarishvili, R. Atun, P. Berman, W. Hsiao, C. Burgess, and M. A. Lansang, "Converging Health Systems Frameworks: Towards A Concepts-to-Actions Roadmap for Health Systems Strengthening in Low and Middle Income Countries," *Glob. Heal. Gov.*, vol. 3, no. 2, 2010.
[25] R. Evans, "Incomplete Vertical Integration in the Health Care Industry: Pseudomarkets and Pseudopolicies," *Annals of the American Academy of Political and Social Science*, vol. 468, no. July. pp. 60–87, 1983.
[26] B. Marchal *et al.*, *Analysing Health Systems To Make Them Stronger*. 2010.
[27] S. Siddiqi *et al.*, "Framework for assessing governance of the health system in developing countries: Gateway to good governance," *Elsevier Heal. Policy*, vol. 13, no. 2, 2008.
[28] L. A. Aday, C. E. Begley, D. R. Lairson, C. H. Slater, A. J. Richard, and I. D. Montoya, "A Framework for Assessing the Effectiveness, Efficiency, and Equity of Behavioral Healthcare," *Am. J. Manag. Care*, vol. 5, pp. 25–44, 1998.
[29] A. Anell and M. Willis, "International comparison of health care systems using resource profiles," *Bull. World Health Organ.*, vol. 78, pp. 770–778, 2000.
[30] M. Roemer, "National health systems throughout the world," *Annu. Rev. Public Health*, vol. 14, pp. 335–353, 1993.
[31] British Columbia: Ministry of Health Services, "A Framework for Core Functions In Public Health Resource Document Population Health and Wellness Ministry of Health Services Province of British Columbia," no. March, 2005.
[32] J. Kutzin, "A descriptive framework for country-level analysis of health care financing arrangements," *Elsevier*, vol. 56, pp. 171–204, 2001.
[33] J. Frenk, "Dimensions of health system reform," *Health Policy (New. York).*, vol. 27, no. 1, pp. 19–34, 1994.
[34] M. Feldstein, B. Friedman, and H. Luft, "DISTRIBUTIONAL ASPECTS OF NATIONAL HEALTH INSURANCE BENEFITS AND FINANCE," *Natl. Tax J.*, vol. 25, no. 4, pp. 497–510, 1972.
[35] D. E. Yett, L. Drabek, M. D. Intriligator, and L. J. Kimbell, "Health manpower planning: an econometric approach.," *Health Serv. Res.*, vol. 7, no. 2, pp. 134–47, 1972.
[36] M. Feldstein and B. Friedman, "The Effect of National Health Insurance on the Price and Quantity of Medical Care," *Natl. Bur. Econ. Res.*, vol. The Role o, pp. 505–541, 1976.
[37] P. Khaleghian and M. Das Gupta, "Public Management and the Essential Public Health Functions," *Elsevier*, vol. 33, no. 7, pp. 1083–1099, 2005.
[38] C. Ramagem and J. Ruales, "Essential public health functions as a strategy for improving overall health systems performance: Trends and challenges since the public health in the Americas Initiative, 2000-

2007," *World Heal. Organ.*, no. April, p. 21, 2008.
[39] The Commonwealth Fund, "Framework for a high performance health system for the United States." Commission on a High Performance Health System, New York, 2006.
[40] World Health Organisation and The World Bank, "A common framework for monitoring performance and evaluation of the scale-up for better health," *Briefing note for H8*. 2008.
[41] D. W. Bettcher, D. Yach, and G. E. Guindon, "Global trade and health: key linkages and future challenges," *Bull. World Health Organ.*, vol. 78, pp. 521–534, 2000.
[42] G. Marchildon, "Evaluating Health Policy and System Performance: Are We Moving to a Network Model?" Research Gate, Canada, 2014.
[43] G. Seidman and R. Atun, "Aligning values and outcomes in priority-setting for health." Journal of Global Helalth, 2016.
[44] I. Mikkelsen-Lopez, K. Wyss, and D. de Savigny, "An approach to addressing governance from a health system framework perspective." BMC International Health and Human Rights, 2011.
[45] A. Cassels, "Health Sector Reform: Key Issues in Less Developed Countries," *J. Int. Dev.*, vol. 7, no. 3, pp. 329–347, 1995.
[46] J. Hurst and M. Jee-Hughes, "Performance Measurement and Performance Management in OECD Health Systems." OECD Publishing, Paris, 2001.
[47] I. A. Agyepong *et al.*, "Spanning maternal, newborn and child health (MNCH) and health systems research boundaries: conducive and limiting health systems factors to improving MNCH outcomes in West Africa." BioMed Central, 2017.
[48] W. C. Hsiao and P. S. Heller, "What Macroeconomists Should Know about Health Care Policy." International Monetary Fund, Washington DC, 2007.
[49] J. Nixon and P. Ulmann, "The relationship between health care expenditure and health outcomes," *Eur. J. Heal. Econ.*, vol. 7, no. 1, pp. 7–18, Mar. 2006.
[50] C. Sicotte, J. Barnsley, and F. Beland, "A Conceptual Framework for Analysis of Health Care Organizations' Performance," *Artic. Heal. Serv. Manag. Res.*, 1998.
[51] E. Docteur and H. Oxley, "Health-care systems: lessons from the reform experience," *OECD Health Working Papers*. OECD, Paris, 2003.
[52] S. Anand and T. Bärnighausen, "Human resources and health outcomes: cross-country econometric study," *Lancet*, vol. 364, pp. 1603–1609, 2004.
[53] J. Veillard, A. Moses McKeag, B. Tipper, O. Krylova, and B. Reason, "Methods to stimulate national and sub-national benchmarking through international health system performance comparisons: a Canadian approach.," *Health Policy*, vol. 112, no. 1–2, pp. 141–7, Sep. 2013.
[54] T. Boerma, C. Abou-Zahr, E. Bos, P. Hansen, E. Addai, and D. Low-Beer, "Monitoring and evaluation of health systems strengthening: An operational framework." World Health Organization, Geneva, 2009.
[55] O. A. Arah, G. P. Westert, J. Hurst, and N. S. Klazinga, "A conceptual framework for the OECD Health Care Quality Indicators Project," *Int. J. Qual. Heal. Care*, vol. 18, no. SUPPL. 1, pp. 5–13, 2006.
[56] J. van Olmen *et al.*, "Primary Health Care in the 21st century: primary care providers and people's empowerment," *Trop. Med. Int. Heal.*, vol. 15, no. 4, pp. 386–390, Feb. 2010.
[57] T. Savel *et al.*, "A Public Health Grid (PHGrid): Architecture and value proposition for 21st century public health," *Int. J. Med. Inform.*, vol. 79, no. 7, pp. 523–529, 2010.
[58] J. H. M. Veillard, A. D. Brown, E. Bariş, G. Permanand, and N. S. Klazinga, "Health system stewardship of National Health Ministries in the WHO European region: Concepts, functions and assessment framework," *Health Policy (New. York).*, vol. 103, no. 2–3, pp. 191–199, 2011.
[59] A. Mills, F. Rasheed, and S. Tollman, "Strengthening health systems," *Dis. Control priorities …*, no. chapter 8, pp. 87–102, 2006.
[60] Londono and Frenk, "Structured pluralism: towards an innovative model for health system reform in Latin America," *Health Policy (New. York).*, vol. 41, no. 1, pp. 1–36, 1997.
[61] T. R. Frieden, "A framework for public health action: the health impact pyramid.," *Am. J. Public Health*, vol. 100, no. 4, pp. 590–5, Apr. 2010.
[62] The World Bank, "What is a health system?," *Heal. Dev. World Bank Strateg. HNP Results*, pp. 1–4, 2007.
[63] WHO, "World health report 2008 Primary Health Care," 2008.
[64] J. D. Sterman, "System Dynamics Modeling: TOOLS FOR LEARNING IN A COMPLEX WORLD," *Calif. Manag. Rev. Repr. Ser.*, vol. 43, no. 4, 2001.

6. APPENDIX A: Health system frameworks

Table 5 List of existing health system frameworks

	Framework	Source
1	Actors framework	[25]
2	Analysing health systems to make them stronger	[26]

3	Assessing governance in developing countries' health systems	[27]
4	Behavioural healthcare framework	[28]
5	Comparing healthcare systems with resource profiles	[29]
6	Component elements of health systems	[30]
7	Control knobs framework	[14]
8	Converging health systems frameworks	[24]
9	Core Functions framework	[31]
10	Country level analysis of healthcare financing	[32]
11	Dimensions of health system reform	[33]
12	Distributional aspects of national health insurance	[34]
13	Econometric model of the healthcare system	[35]
14	Effect of National Health Insurance on Medical Care	[36]
15	Essential functions of public health	[37]
16	Essential Public Health Functions	[38]
17	Framework for high performance health system in the United States	[39]
18	Framework for monitoring and evaluating performance	[40]
19	Global trade and health	[41]
20	Health policy and system performance	[42]
21	Health priority setting	[43]
22	Health system framework to improve maternal, neonatal and child health (MNCH)	[19]
23	Health system functions and goals	[17]
24	Health system governance	[44]
25	Health system key institutional components	[45]
26	Health system performance measurement and management	[46]
27	Health system shelter	[47]
28	Health systems and their context	[9]
29	Health systems in transition	[22]
30	Health systems strengthening framework	[23]
31	Healthcare and the macro-economy	[48]
32	Healthcare expenditure and health outcomes	[49]
33	Healthcare organisation performance framework	[50]
34	Healthcare system reform	[51]
35	Human resources and health outcomes	[52]
36	International health system performance comparison	[53]

37	Monitoring and evaluating framework of health systems strengthening	[54]
38	OECD Health Care Quality Indicators Framework	[55]
39	Primary healthcare	[56]
40	Public health grid	[57]
41	Stewardship health system framework	[58]
42	Strengthening health systems	[59]
43	Structured pluralism model of healthcare systems reform	[60]
44	The Global Fund health systems strengthening	[13]
45	The health impact pyramid	[61]
46	The World Bank: healthy development	[62]
47	WHO health performance framework	[8]
48	WHO health system building blocks	[21]
49	WHO primary healthcare framework	[63]

Table 6 Health systems frameworks after first round of filtration

	Framework	**Type of framework**	**Source**
1	Actors framework	Understanding framework	[25]
2	Analysing health systems to make them stronger	Informing change framework	[26]
3	Behavioural healthcare framework	Evaluating framework	[28]
4	Component elements of health systems	Understanding framework	[30]
5	Control knobs framework	Evaluating framework	[14]
6	Converging health systems frameworks	Understanding supra-framework	[24]
7	Core Functions framework	Informing change framework	[31]
8	Dimensions of health system reform	Informing change sub-framework	[33]
9	Econometric model of the healthcare system	Understanding sub-framework	[35]
10	Essential Public Health Functions	Evaluating framework	[38]
11	Framework for monitoring and evaluating performance	Evaluating framework	[40]
12	Health system framework to improve maternal, neonatal and child health (MNCH)	Evaluating sub-framework	[19]
13	Health system functions and goals	Understanding framework	[17]
14	Health system governance	Understanding framework	[44]
15	Health system key institutional components	Informing change supra-framework	[45]
16	Health systems and their context	Understanding supra-framework	[9]
17	Health systems in transition	Comparing framework	[22]

18	Health systems strengthening framework	Evaluating framework	[23]
19	OECD Health Care Quality Indicators Framework	Evaluating supra-framework	[55]
20	Public health grid	Informing change sub-framework	[57]
21	Stewardship health system framework	Understanding supra-framework	[58]
22	Structured pluralism model of healthcare systems reform	Informing change framework	[60]
23	The Global Fund health systems strengthening	Evaluating supra-framework	[13]
24	WHO health performance framework	Evaluating framework	[8]
25	WHO health system building blocks	Understanding framework	[21]
26	WHO primary healthcare framework	Informing change sub-framework	[63]

UTILIZING DECISION FOREST REGRESSION MACHINE LEARNING ALGORITHM TO PREDICT FILLING LINE UTILIZATION FOR OPTIMAL MAINTENANCE SCHEDULING

L. N. Smith[1] & C.M. Zastron[2] & P.J. Vlok[3*]

[1]Department of Industrial and Systems Engineering
University of Tennessee, USA
lsmit194@vols.utk.edu

[2]Pragma South Africa (Pty) Ltd
mauritz.zastron@pragmaworld.net

[3]Department of Industrial Engineering
University of Stellenbosch
pjvlok@sun.ac.za

ABSTRACT

Small margins within the packaging industry mean financial success in this field relies on high equipment availability. To achieve this high equipment availability, maintenance schedules should be carefully planned to minimize downtime. A key component of maintenance schedule planning is predicting equipment utilization. This can prove very difficult as there are many variables such as market demand, seasonality of products, capability and diversity of equipment, and inherent reliability, to name a few. Even some of the leading players in the packaging industry treat the complexities and chaos involved with predicting equipment utilization as a topic best avoided. Current approaches to this problem range from no prediction at all to only a simple linear extrapolation.

This paper investigates the merits of using machine learning algorithms to predict equipment utilization in the packaging industry with the aim of optimizing maintenance schedules. Machine learning entails pattern recognition of past data and inclusion of pertinent variables in the present to forecast behaviour. This paper begins with a brief literature review of the field before using data, obtained from a multinational packaging company, to test some of the most promising methods of machine learning in a case study.

[1] The author was enrolled for an B Eng (Industrial) degree in the Department of Industrial and Systems Engineering, University of Tennessee
[2] Business Intelligence Engineer
[3] Extraordinary Professor, Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION

1.1 1.1 Problem Statement

Successful packaging companies depend on high availability of their equipment for financial success. Breakdowns are detrimental to availability and maintenance must be performed regularly, in a minimally disruptive manner, to achieve high equipment availability. To ensure equipment availability, maintenance plans are created for machines in an attempt to replace parts before they break. These maintenance plans normally contain three types of tasks, namely: condition based, time/calendar based, and usage based. Condition based tasks usually have certain limits in which the condition of the part must be. If the part's condition falls outside the specified limits, the part is replaced. Common features which are measured include vibration, heat, wear and tear, as well as noise. Time or calendar based tasks are used when parts need to be cleaned or replaced periodically, regardless of their use or the time on the asset. Finally, usage based tasks are used to replace parts which should be replaced after it has been used for a specified number of cycles or running hours. Usage based scheduling will be the only maintenance type investigated in this paper. The time of work order (collection of tasks) execution needs to be predicted before the tasks are due so that the planning department can ensure all resources are available at the time of the service. Currently, the date at which a usage based service will be done is forecasted by looking at the asset's average usage over a number of past readings and then using that average usage per day when making a future prediction. The number of readings used to calculate the average usage can either be a predefined number or all of those that fall within a fixed time period. The current method employed by the packaging company uses a predefined number of meter readings to calculate the average daily usage. Machine learning, with its unique ability to adapt to different problems and harvest information from large data sets, has the potential to create more accurate usage predictions than the current method. While other solutions do exist, this paper will focus solely on the merits of machine learning when applied to machine usage prediction

With the creation of such software as Microsoft's Azure, IBM's Watson, and TensorFlow, machine learning integration has never been easier or more accessible. As computational resources are constantly becoming more readily available, numerous applications are still being discovered. Machine learning has been utilized by power plants to predict electricity usage and detect malicious energy usage [1]. It has been used to diagnose and classify cancers, with higher accuracy and precision than doctors [2]. Machine learning can also be used to optimize manufacturing processes while increasing product quality and decreasing process testing costs [3]. It has been used to create better marketing and pricing strategies within the steel industry by predicting raw steel prices [4]. Machine learning has even been used to predict stream flows, providing important information for hydrological studies [5]. The success machine learning has had in these and other applications warrants an investigation into how it can improve machine usage predictions.

This paper begins investigating the merits of predicting equipment usage with machine learning with the aim of optimizing maintenance schedules by conducting a literature review. Since the review revealed no sources that considered this use of machine learning, similar applications were researched. From this research, a generic method was constructed to use machine learning to predict equipment usage. This method is then tested in a real world case study to test the performance of machine learning and investigate if and by how much it can improve on the current prediction method.

1.2 Machine Learning: An Overview

Machine learning can be divided into two categories, namely: unsupervised and supervised. When given unlabeled data points, machine learning can find an underlying pattern. This is known as unsupervised machine learning. When given labeled data points, machine learning can detect anomalies, make classifications, or predict numerical values. These are types of supervised machine learning and are called anomaly detection, classification, and regression respectively. To create usage predictions, numerical values need to be predicted, so supervised regression machine learning will be used in this study.

Supervised regression begins with the collection and preparation of data. During the preparation phase, missing and incorrect values in the dataset are identified and corrected. This data is then split into a training set and a testing set. The training data is used to train the chosen algorithm. Every algorithm creates predictions differently and uses the training data to create, optimize, and validate these prediction decisions. Features can be selected before the algorithm is trained, known as the fitler method, or while training the algorithm, known as the wrapper method. Once the algorithm has been trained with the chosen features, the algorithm then makes prediction on the testing data set. The accuracy of predictions is calculated by a prediction error, or the difference between predicted and actual outcome. The lower the error, the more accurate the algorithm is in making predictions. The steps outlined above are summarized in the Figure 1, seen below.

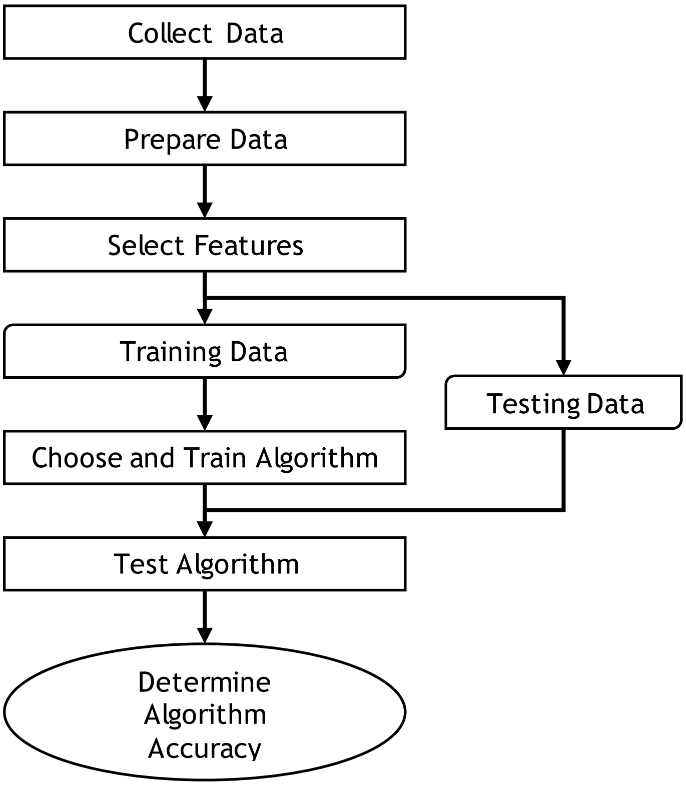


Figure 1: Overview of Machine Learning

1.3 Literature Review

There is currently very little research on using meter readings for predicting the future use of packaging machines using machine learning. However, the application of machine learning to improve prediction accuracies is prevalent in the literature. The most similar application found was the use of machine learning to predict electricity usage. These sources provided the framework for the literature review with other sources supplementing the specific considerations of machine learning.

After acquiring electricity usage data, one source identified incorrect values as those above the average of a continuous period of time [1]. The incorrect values were then replaced with that average [1]. Missing values were filled in with interpolation by one study [1] and linear extrapolation in another [6]. After preparing data, one study considered the distribution of the data with Q-Q plots and the Shapiro Wilks test [1]. This study considered using the Kolmogorov-Smirnov test as well, but did not because this test can only be used if the cumulative density function is known [7].

To divide the data into a training and testing set, the past 500 hours of data was used as the training data to predict the next 48 hours of testing data [1]. In another instance, the training set was composed of four years' worth of data and the testing set was composed of the most recent year of data [6]. Data sets can also be split using random splits [8] and K-folds [9]. If the data set is extremely large, learning curves can be used to limit the training data. As data set size increases, algorithm accuracy increases greatly at first. Eventually, algorithm accuracy sees marginal improvements as data set size continues to increase [10].

Feature selection for predicting electricity usage was done using the relieff algorithm [1] and the statistical properties of the features [6]. Both of these are examples of filter algorithms which are better suited for large datasets [11]. A wrapper method of feature selection is more time intensive, but provides information about the interaction of features [12].

The two main types of wrapper selection algorithms are forward selection and backward elimination [13]. Although these algorithms work quickly, they do not always return the most helpful subset of features [13]. The accuracy of these algorithms can be improved by using a combination of the two algorithms [12].

Boosted decision tree [6], support vector machine [1], and neural network [1] algorithms have all been used to predict electricity usage. The efficiency at which the decision forest regression method operates also makes a popular choice of algorithm for large data sets [14].

Although the literature review did not reveal any publications in which machine learning was used for the same objective as this paper, enough elements of the larger machine learning process were found to construct a methodology in which the problem under review can be addressed. In the next section a proposed solution is constructed with which the paper objectives can be achieved.

2. PROPOSED SOLUTION

The findings from the literature review were used to create a proposed solution for more accurate machine usage predictions with machine learning. Data will be prepared by filling missing values with linear extrapolation [6]. Incorrect values rarely occur as meters are maintained regularly and data is validated upon capturing to draw attention to possible mistakes. The most recent year of data will be withheld for testing and the rest will be used for training [6]. Learning curves based on data set size will be used to further reduce the testing set size and shorten training time [10] if necessary. Features will be selected with a combined forward selection and backward elimination wrapper algorithm [13]. The boosted decision forest regression algorithm will be used to create predictions because of its short training time, high accuracy, and insensitivity to parameters [14]. Success will be measured by reduction of total prediction error. This error can be further classified into over maintenance and under maintenance. When a maintenance task is executed before it is required, it is called over maintenance. This inaccuracy results in increased replacement part costs and increased labor costs. A maintenance task occuring after the scheduled meter reading is known as under maintenance. This results in a higher probability of a breakdown occurring. Breakdowns can be very costly as they result in lost production time, unavailability of artisans, costly expedition of replacement parts, and in some cases loss of the product in the machine at the time of the breakdown. The proposed solution is outlined in the following figure (figure 2).

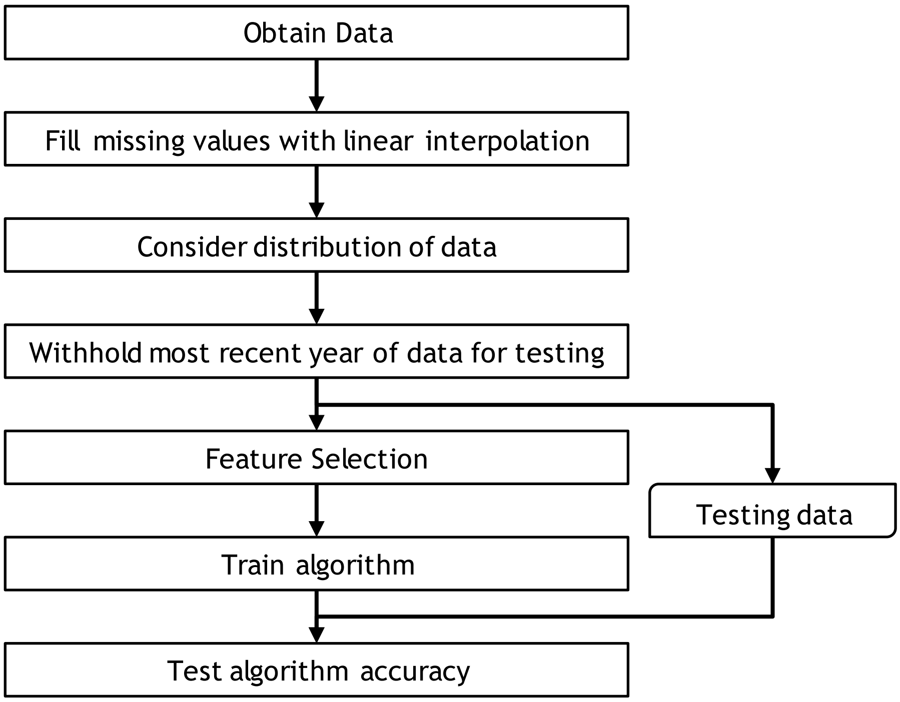


Figure 2: Flowchart of Proposed Solution

3. CASE STUDY

An algorithm can be trained to predict meter readings in a variety of ways. This paper will consider three different methods. First, a reading will be predicted for a certain future date. Secondly, the amount produced between the current date and future date will be predicted, then added to the current reading to construct the future reading. Lastly, daily rates will be predicted for every day between the current and future date, summed, and then added to the current reading to predict the future reading. These rates are referred to as Average Daily Rates (ADRs) within the industry.

3.1 Case Study Data

Data for this case study comes from a multinational packaging company with over 50,000 assets. These assets are fitted with meters which record the total hours that a machine has been operational. When a reading is taken, the date, unique meter code, and total hours of operation are recorded. Additional information such as meter location and brief description of product is obtained from the asset attached to the meter. The time span between readings varies by meter location and date. More recent readings have a shorter time span than older readings. If an asset is refurbished, the meter is often adjusted (reset) to reflect the age of the asset after the refurbishment. For the case study, 500 meters without meter resets were selected. These meters were evenly divided among the five cluster locations to ensure an adequate distribution.

Also used in this study was the most recent year of scheduled work order data. Work order data includes a work order code, meter code, date of work order creation, and date the work order was scheduled for (due date). The work orders are scheduled for certain hour readings and the due date of the work order is calculated by extrapolating the last x number of meter readings to predict when the meter will reach the reading of the work order.

3.2 Reading Predictions

To create a training data set for reading predictions, the span between between work order creation and the work order due date was calculated for each work order. These values were then averaged for each meter. The training data was constructed to reflect this average time span. For every reading and date, a previous reading and date were selected such that the difference between date and previous date closely matched the average span found from the work order data. This created two additional features of previous reading and time between. When a value was missing the preceding reading and date were used.

All features of the reading data were included as well as the days between and previous reading features. Combined forward selection and backward elimination was used to select the most helpful subset of features. They were determined to be days between and previous reading. This created a relatively small training set, so all data except for the most recent year was used for training.

The decision forest regression algorithm was then trained with this data. Parameters were selected using a partial grid sweep. This algorithm proved to be much less accurate than the current prediction method. One algorithm was then trained for each cluster to reflect locational differences. These cluster specific algorithms created more accurate results than the single algorithm, but were still not more accurate than the current prediction method.

3.3 Amount Predictions

The same data preparation technique used to construct the predict readings data was again used to construct the predict amounts training data. Instead of of using the previous reading feature, an amount feature was used. This amount feature was constructed by subtracting the previous reading from the reading. Using forward selection and backward elimination, days between and cluster were selected as the best subset of features. With only two features, the data set size was quite small, so all data after the most recent year was used for training.

After training one decision forest regression algorithm, no improvement in prediction accuracy was seen against the current method. Therefore, 5 algorithms were trained, one per cluster. Parameters were selected with a random grid sweep. Using these algorithms, the total absolute error of one out of the five clusters was reduced by 26 610 units. All other clusters saw an increase in total absolute error. As the total absolute error was not decreased for these algorithms, maintenance costs would not be reduced either.

3.4 ADR Predictions

The final prediction method in this case study was predicting average daily usage. Data was prepared by finding the ADR between each reading. As readings are not taken every day, missing values were filled with linear interpolation.

In addition to the eight features included with a reading, five additional features were mined from the data. To account for seasonality, the month of the reading was added. Additionally, ADRs were averaged based on day of the year, for each of the four location features. Using combined forward selection and backward elimination; site code, meter, reading on, customer plant description, description, and month were selected as the most helpful features.

The training data set size for this prediction type was quite large, so a learning curve was constructed for data set size. Using data from 100 meters, percentages of the meters were used to predict the most recent year of all meters. Ten meters, representing 10% of the data, were randomly selected and used to train the ML alogrithm. The resulting trained algorithm was then used to predict ADR for all 100 meters and the mean error was recorded. This process of random selection was then repeated five times for each 10% increment. The five mean error values for each increment were then averaged and plotted in Figure 3. From the graph, it was determined that a subset of 50% of the meters would provide sufficient accuracy.

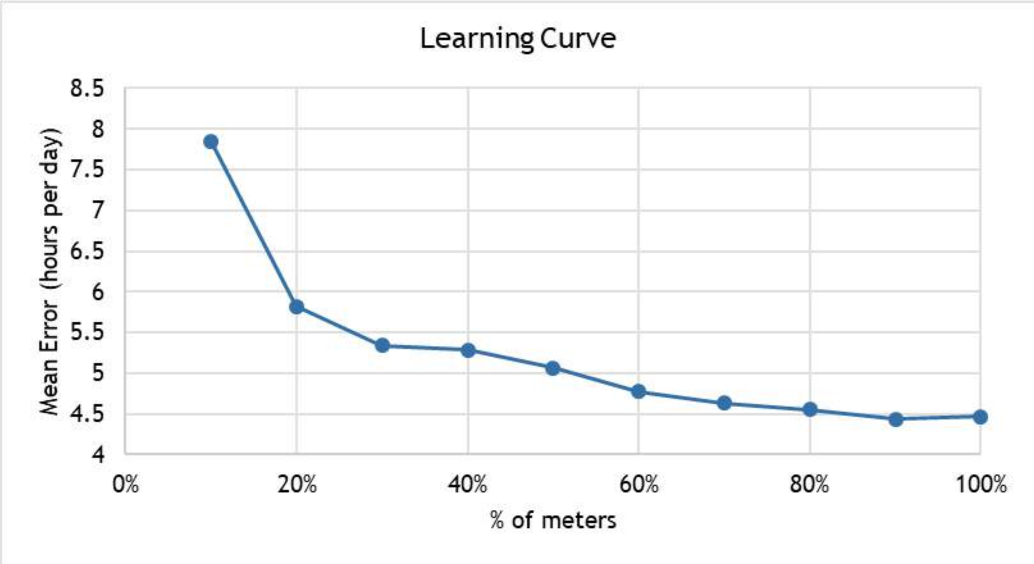


Figure 3: Mean error per day as a function of the percentage of available meters used to train the machine learning algorithm.

To increase accuracy, one algorithm was trained for each location cluster. Parameters were chosen for each algorithm with a partial grid sweep. The accuracy of the algorithms was then calculated with the most recent year of work order data. This was done by summing the ADRs between the most recent reading date to the work order creation date and work order due date, then adding this value to the most recent reading to the work order creation date.

The current method of meter reading prediction saw a total absolute error of 360 032 hours. If the machine learning method had been used over the past year, it would have had a total absolute error of 296 100 hours. Thus, machine learning was able to reduce the total absolute error by 63 931 hours (17.8%). These hours can be further divided into over maintenance and under maintenance hours. The current method had 272 811 hours of under maintenance, while the machine learning method had 113 087 hours of under maintenance, a reduction of 159 724 hours. Over maintenance hours for the current method were 87 221. The machine learning method had 183 031 hours, an increase of 95 810 hours. These errors are further summarized in the box and whisker plot in Figure 4.

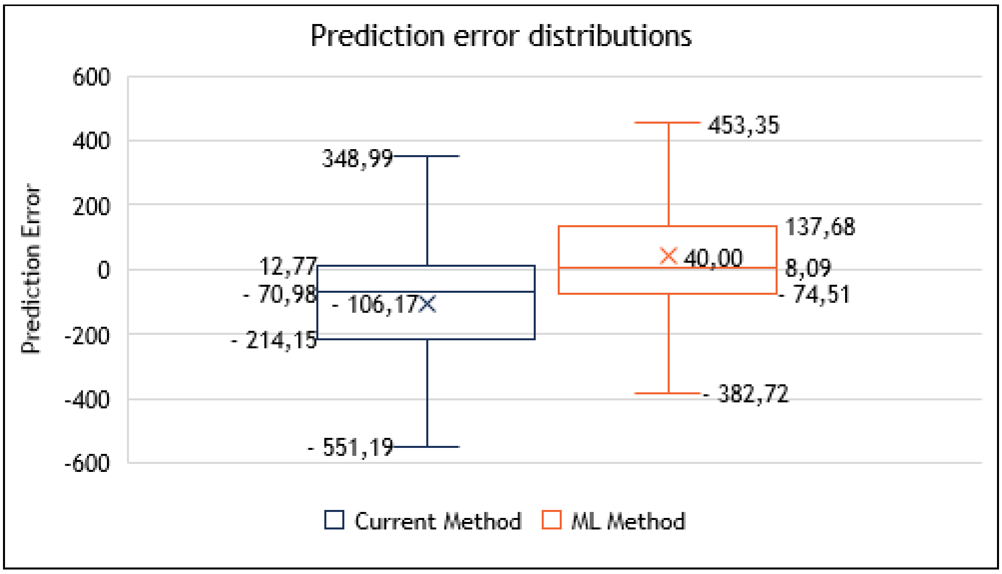


Figure 4: Box and whisker plot showing the distributions of the mean error of the current prediction method and the machine learning algorithm.

An ideal prediction method would have a total absolute error of zero. As no prediction method will be ideal, a reasonable accurate prediction method will have prediction errors centered around zero error, specifying accurate predictions, and a small spread from the zero error, translating to a precise error. The centering of the errors can be considered by the median and mean of the errors. The machine learning method had a mean error of 40,00 and a median error of 8,09. The current prediction method had a mean error of -106,17 and a median error of -70,98. As the absolute value of the mean and median error for the machine learning method is less than the absolute value of the mean and median error for the current method, machine learning creates more accurate predictions than the current prediction method.

The spread of the errors can be considered by adding the absolute values of the upper and lower whiskers on the box plot. The machine learning method had a spread of 836,07. The current prediction method had a spread of 900,18. As the machine learning method had a smaller spread, it is the more precise prediction method. Because the machine learning method was more closely centered on the zero error line and had a smaller spread, it is both more accurate and precise than the current prediction method.

The monetary savings of using machine learning instead of the current method are difficult to quantify, particularly for under maintenance. The costs associated with under maintenance vary greatly based on the part or parts that break, availability of those parts, and lost production, to name a few factors. However, a breakdown does not occur every time a machine is under maintained. Therefore, the true cost of under maintenance is the probability of a breakdown occurring for each hour under maintained multiplied by all associated costs of a breakdown. As breakdowns are not always recorded, and their costs are so variable, data was not available to construct an average cost for each hour of under maintenance.

Calculating over maintenance is much simpler as it only involves the cost of labor and parts. These values are known so the average cost of over maintenance can be calculated. However, without knowing the cost of under maintenance, this value does not provide an accurate picture of cost saved. Therefore, reduction of total error, error spread, and median error value should be used as the determiner for success.

4. CONCLUSION

Like most industries, financial success within the packaging industry relies on high equipment availability. In turn, this high availability relies on maintenance schedules to minimize downtimes. One way of creating these maintenance schedules is through machine usage predictions. The current method of predicting machine usage considers the last certain number of meter readings to create a future prediction. With a total prediction error of 360 032 hours for 500 meters over the past year with this method, there is great room for improvement.

Machine learning was considered as a viable option for creating improved usage predictions in this paper. A literature review was conducted and a proposed solution was created from that review. This proposed solution was then applied to three prediction types. First, readings were predicted. This method produced less accurate results than the current prediction method. Secondly, production amounts between two dates were predicted. This method created slightly more promising results. It did create improved predictions for one out of the five clusters, but for every other cluster it created less accurate predictions. The final method tested was ADR predictions based on the month. When scored on the most recent year of work order data, is reduced the total error by 63 931 hours. Extrapolating that across all 50 000 meters, that translates to a total amount of hour reduction of 6 393 100. The only drawback of the machine learning method was that it is more prone to under maintenance than the current method. Unfortunately, the costs associated with these hours was not able to be calculated due to unavailability of data.

While the total costs saved could not be calculated, machine learning still shows improvements from the current method of machine usage predictions. It reduced the total absolute error of predictions on the 500 meters selected for testing. Furthermore, it shows every indication of extending that prediction accuracy to all 50 000 meters. In conclusion, machine learning was able to create more accurate machine usage predictions than the current method, enabling the creation of more accurate maintenance schedules.

5. REFERENCES

[1] **Yu, Wei & An, Dou & Griffith, David & Yang, Qingyu & Xu, Goubin**. 2015. Towards Statistical Modeling and Machine Learning Based Energy Usage Forecasting in Smart Grid, Applied Computing Review, 15(1). pp 6-16.

[2] **Wong, Derek & Yip, Stephen**. 2018. Machine Learning Classifies Cancer, International Journal of Science, Nature, 555, pp. 446-447.

[3] **Lu, Y. & Rajora, M. & Zou, P. & Liang, S.Y.** 2017. Physics-Embedded Machine Learning: Case Study with Electrochemical Micro-Machining, Machines, 5, pp. 4.

[4] **Ou, Cheng, Chen, & Perng**. 2016. Dynamic cost forecasting model based on extreme learning machine - A case study in steel plant, Computers & Industrial Engineering, 101, pp. 544-553.

[5] **Yaseen, Zaher Mundher & Jaafar, Othman & Deo, Ravinesh C. & Kisi, Ozgur & Adamowski, Jan & Quilty, John & El-Shafie, Ahmed**. 2016. Stream-flow forecasting using extreme learning machines: A case study in a semi-arid region in Iraq, Journal of Hydrology, 542, pp 603-614.

[6] **Bansal, Anshul & Rompikuntla, Susheel Kaushik & Gopinadhan, Jaganadh & Kaur, Amanpreet & Kazi, Zahoor Ahamed**. 2015. Energy Consumption Forecasting for Smart Meters. BAI Conference

[7] **Hazewinkel and Michiel**. 2001. Kolmogorov-Simirnov test. 1st edition. Springer

[8] **Bloice, M.D. & Holzinger, A**. 2016. Machine Learning for Health Informatics, 1st edition, Springer.

[9] **Fushiki, Tadayoshi**. 2011. Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, Springer. 21. pp. 137-146

[10] **Senior, Andrew & Heigold, Georg & Ranzato, Marc' Aurelio**. 2013. An Empirical Study of Learning Rates In Deep Neural Networks For Speech Recognition, Acoustics, Speech and Signal Processing (ICASSP) IEEE International Conference, pp. 6724 - 6728.

[11] **Hall, Mark A**. 2000. Correlation-based Feature Selection for Discrete and Numeric Class Machine Learning, Working Paper Series, ISSN 1170-487X, Working Paper 00/8

[12] **Bolón-Canedo, Verónica & Sánchez-Maroño, Noelia & Alonso-Betanzos, Amparo**. 2015. Feature Selection for High-Dimensional Data, 1st Edition, Springer.

[13] **Aboudi, N. El & Benhlima, L**. 2016. Review on Wrapper Feature Selection Approaches, International Conference on Engineering & MIS (ICEMIS), pp. 1-5.

[14] **Criminisi, Antonio & Shotton, Jamie & Konukoglu, Ender**. 2011. Decision Forests: A Unified Framework for Classification, Regression, Density Estimation, Manifold Learning and Semi-Supervised Learning, Foundations and Trends in Computer Graphics and Vision., 7(2-3), pp. 81–227

AN INFORMATION SYSTEMS APPROACH TO THE PROACTIVE MANAGEMENT OF SUBSCRIBER IDENTIFICATION MODULES IN INDUSTRY

J. Prinsloo[1*], M. J. Mathews[2], J. N. du Plessis[3], J. C. Vosloo[4]

[1-4]North-West University, CRCED-Pretoria, Pretoria, South Africa

[1]jprinsloo@researchtoolbox.com, [2]mjmathews@researchtoolbox.com, [3]jduplessis@researchtoolbox.com, [4]jcvosloo@researchtoolbox.com

ABSTRACT

The Industry 4.0 paradigm focuses heavily on decentralisation, modularity and interoperability. These principles form the foundation for an effective Industry 4.0 ecosystem. The South African heavy industry still relies heavily on the traditional approach to centralised monitoring and control. Various Industry 4.0 initiatives have however been proposed and implemented in the heavy industry sector. For these initiatives to operate effectively, a stable network layer is required to facilitate reliable data transfer. To achieve data transfer over large geographical distances, most industrial information systems make use of mobile cellular networks, which implies the use of subscriber identification modules (SIMs) and supporting hardware such as network routers. The mismanagement of SIM cards can however result in excessive communication costs, which can in turn hamper the effectiveness of wireless data transmission. This paper presents a comprehensive information and asset management system to assist with overall SIM card management. The system allows for the accurate analysis of communication costs and ensures transparency both in the financial domain and in the physical domain. Results in this paper illustrate the effectiveness of this system in isolating communication anomalies that result in excessive data costs. The results show that the system can improve the quality and reliability of wide-area wireless telemetry and maintain indirect expenses that are associated with industrial information system communications. The proposed system assisted in reducing overall wide-area wireless communication cost with approximately 62%.

This work was sponsored by ETA-Operations (Pty) Ltd.

[1] Mr. Jaco Prinsloo holds a B.Eng. degree in Computer Engineering. He is currently enrolled for his Masters studies at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr. Marc Mathews holds a PhD in Mechanical Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr. Johan du Plessis is a registered professional engineer and holds a PhD in Electronic and Computer Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[4] Dr. Jan Vosloo is a registered professional engineer and holds a PhD in Electrical Engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

*Corresponding author

1. INTRODUCTION

The term Industry 4.0 is a relatively new concept that is still considered to be in its infancy. Industry 4.0, initiated by the German government and presented at the Hannover Trade Fair in 2011, aims to address growth challenges in the European manufacturing sector [1]. The primary objective of the Industry 4.0 paradigm is to promote digitisation and integrate technological collaboration between various industrial systems [2]. For Industry 4.0 initiatives to be successful, various principles such as decentralisation, real-time production, modularity and interoperability need to be integrated into traditional methods of production [3].

The South African heavy industry, such as the mining and steel industry, still makes use of traditional centralised approaches to monitoring and control of production and assets [4]. This makes a full shift to decentralised and autonomous monitoring and control infeasible since the infrastructure to support Industry 4.0 initiatives must still be incorporated into the current industrial architecture. Various Industry 4.0-related initiatives and methodologies have however been proposed and implemented in the South African heavy industry. These initiatives mostly focus on indirect support and service delivery and includes condition-based maintenance systems, energy management information systems, automated bill analysis systems and remote monitoring and maintenance systems [5–8].

A fundamental component required to facilitate Industry 4.0 is the “Internet of Things” (IoT) to achieve its main goals [9]. The interconnection of individual devices allows for the transmission and reception of data. Data gathered by individual devices can then be used for intelligent decision-making purposes. This layer of intelligence, realised by IoT, allows devices to operate collectively in order to achieve predefined goals [10].

The fundamental component to the systems mentioned above is the communication network used to relay raw data from industrial sites to decentralised storage and computing infrastructure. To achieve data transmission, these systems make use of wide-area wireless telemetry networks, such as mobile cellular networks [11]. To utilise mobile cellular networks for data transmission and reception, subscriber identification modules (SIMs), as well as related hardware systems such as network routers, must be implemented at each remote industrial endpoint [12]. Mismanagement of these modules can result in additional expenditures and network instability, which in turn can effectively hamper the quality and reliability of Industrial IoT (IIoT) information systems discussed above.

Based on the context and challenges discussed above, the objective of this paper is to present and emphasise the value of an information and asset management system for SIM cards used by existing IIoT applications. To achieve this objective, the following sections will:

- present an architectural overview of industrial information systems in the context of the Industry 4.0 domain;
- discuss key challenges that can hamper the network layer quality and stability of industrial information systems;
- introduce a comprehensive information and asset management system to address SIM management; and
- analyse results of the information and asset management system to emphasise its value in industry.

2. AN OVERVIEW OF INDUSTRIAL INFORMATION SYSTEMS

This section aims to introduce an architectural overview of industrial information systems. A generic description of industrial information systems is first presented, after which two information system examples are provided to highlight critical points in the overall operation of these systems.

Industrial information systems typically follow an input-output service delivery model. For these systems to provide valuable feedback for decision-making purposes, data must first be provided from various sources. Raw data is then stored and processed in a decentralised storage and computing environment, after which information is relayed to visual dashboards for user notification purposes. Figure 1 gives an overview of the operation of an industrial information system.

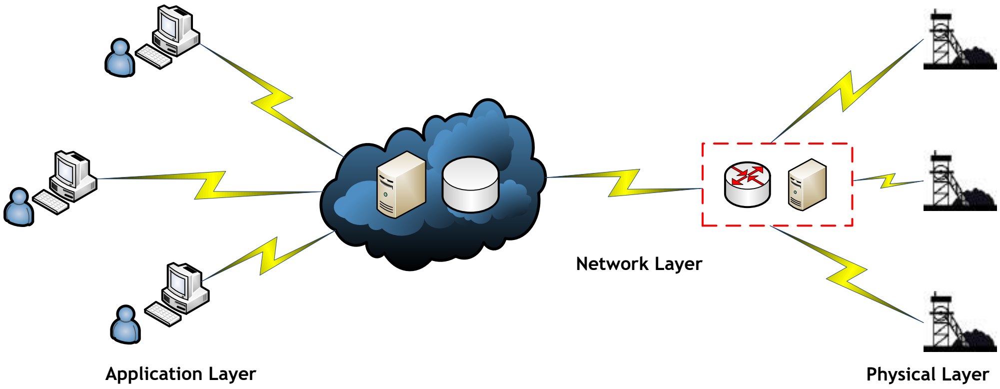


Figure 1: An overview of an industrial information system.

An industrial information system can be divided into three main parts namely, the physical layer, the network layer and the application layer. The collection of raw data takes place in the physical layer. Various sensors and actuators can be used to perform this task, but it is also possible to make use of an existing industrial supervisory control and data acquisition (SCADA) system to collect data.

The data obtained from industrial endpoints must then be transmitted to a decentralised storage scheme for data processing and storage. Cloud-based systems are used to allow for decentralised accessibility. Wide-area wireless communications, such as the Global System for Mobile communication (GSM), is used to transmit data to the cloud-based facility. This requires industrial GSM routers to be installed at industrial endpoints. Each router must also be fitted with a SIM card for connection to a mobile service provider network. Various works in literature emphasise the importance of wide-area wireless communications in IoT applications [11, 13–16]. Without a stable network layer, industrial information systems will be rendered inoperable.

Processed data can be relayed to users at the application layer. From here, system statuses can be viewed and informed decisions can be made. The decentralised nature of the storage and processing subsystem allows users to gain access without being restricted to certain geographical locations. All users are provided with system credentials for security purposes. Figure 1 illustrates a basic overview of a typical industrial information system. Two examples of existing systems in industry are provided below. An architectural overview of each system is given. Limitations and benefits of each system are also presented.

2.1 Example A: An Energy Management Information System (EMIS)

Goosen et al. [7] proposed a comprehensive energy management information system with the objective of providing industry with the ability to comply with various energy management standards, legislation and incentive structures. The system facilitates data collection and storage processes, as well as data analysis methods to provide users with accurate energy reports. Input data is provided to the system from external metering parties, the state-owned electricity utility Eskom, as well as direct input from industrial SCADA systems. Accurate energy reports are then provided as output. The energy reports can be viewed in document format, or the data can be viewed on real-time electronic dashboards. Figure 2 below gives an operational overview of the EMIS.

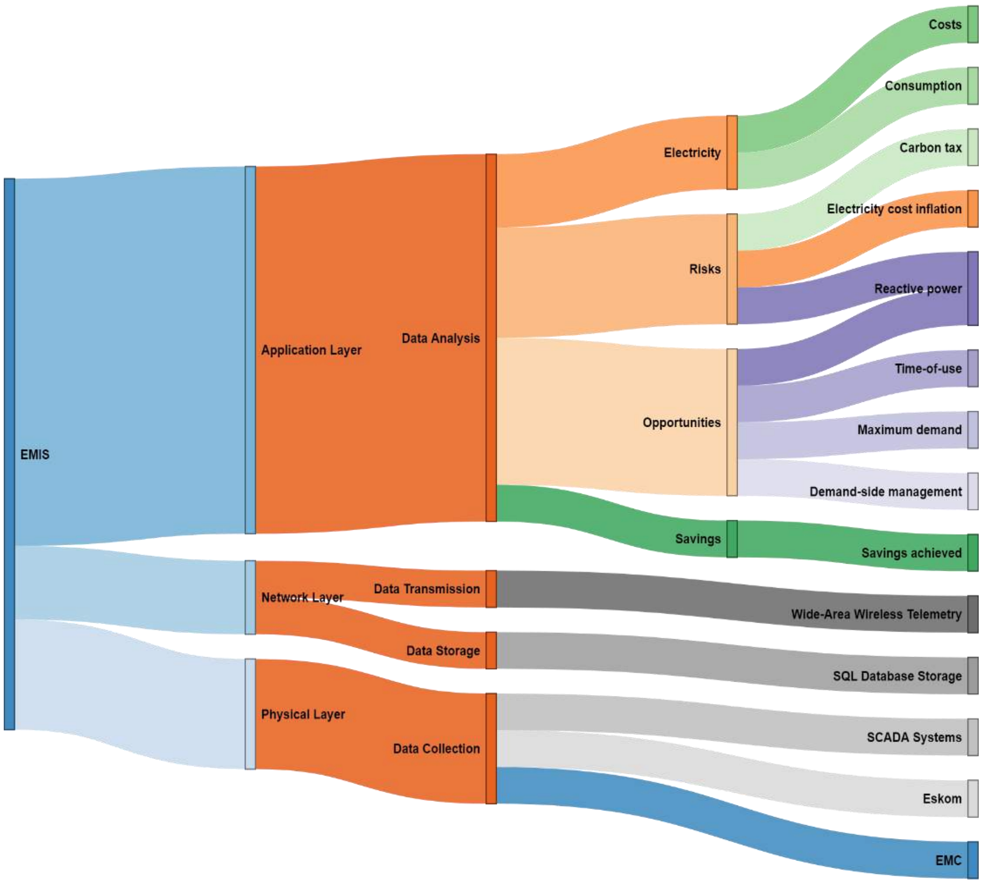


Figure 2: An operational overview of an EMIS (adapted from [7]).

The system proposed by Goosen et al. [7] provides an efficient and time effective way of allowing energy consumers to track their energy consumption and compare it to historic profiles. This system promotes decentralisation and is easily expandable. Goosen further expanded this system by incorporating an automated process where electricity bills are analysed [6]. This further improved the robustness of the EMIS and provided energy consumers with valuable information. The EMIS relies heavily on the GSM network for data transfer. Should intermittent connectivity issues present themselves, the EMIS will effectively be offline. The quality and stability of wide-area wireless communications is therefore considered to be a critical component to the effective operation of this system.

2.2 Example B: A condition-based maintenance system

Van Jaarsveld et al. [8] proposed a condition-based maintenance system with the objective of providing industry with the ability to accurately determine in real-time fashion the status of vital assets, such as pumps, compressors and fridge plant units. This system follows a similar approach to that of the EMIS proposed by Goosen et al. Only one source of input is used, which is the industrial SCADA system. Various status and efficiency parameters are monitored, such as vibration characteristics, temperature levels and trip counters. Data is transmitted in real-time fashion to a decentralised processing and storage unit, after which the processed data is relayed to visual dashboards. The transmitted data is processed for exception reporting, as well as “Safe, Caution, Risk, Failure” (SCRF) assessments. Figure 3 below provides an operational overview of the condition-based maintenance system.

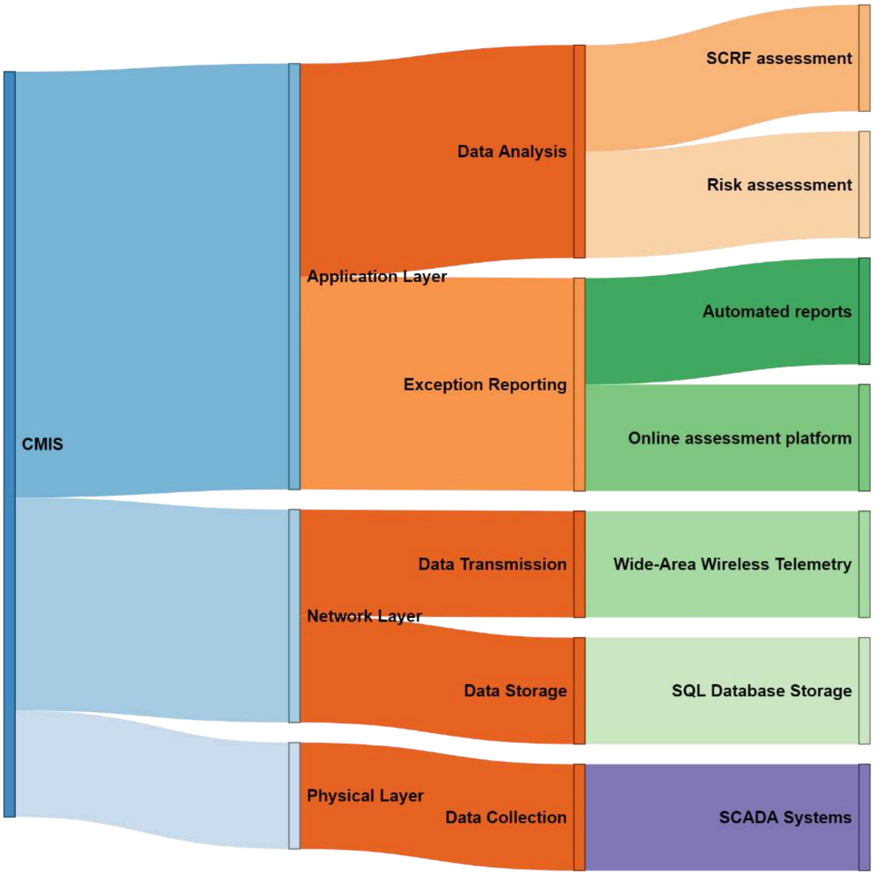


Figure 3: An operational overview of a condition-based maintenance system (adapted from [8]).

The system proposed by van Jaarsveld et al. [8] provides a unique and effective way of allowing industry players to know the operational status of critical assets. This improves maintenance processes and prevents unnecessary capital expenditures. As with the EMIS discussed above, this condition-based maintenance system also relies heavily on the GSM network for data transfer. Intermittent connectivity will hamper the effectiveness and reliability of this system. Emphasis must therefore be placed on the quality and reliability of the network layer.

2.3 Summary

This section aimed to provide an overview of industrial information systems. An industrial information system consists of three layers: the application layer, the network layer and the physical layer. The use of the GSM network and Internet Protocol technologies for Machine-to-Machine (M2M) communication and industrial wireless telemetry is a broadly accepted method for achieving connectivity and can therefore be considered as a key enabler for Industry 4.0 applications [10, 17–19]. Two examples of industrial information systems were presented. Both systems utilise the GSM network for wide-area wireless telemetry. The network layer is the backbone of an industrial information system. The quality and reliability of the network layer is therefore crucial for industrial information system stability. Based on the information discussed above, there is a need for a comprehensive information and asset management system that ensures transparency in the management of SIM units and other related network layer equipment. This system must ensure the stability in network layer communication and keep communication costs at a minimum.

3. CHALLENGES THAT DISRUPT DATA TRANSMISSION

Several challenges exist that inhibits the ability of industrial information systems to relay critical data in a real-time fashion. Mismanagement of SIM cards, the mismanagement of GSM routers on site and general hardware failures are among the few challenges that can result in intermittent connectivity. These challenges are discussed below and examples are provided.

3.1 Hardware failure

The most common challenge present in the network layer of industrial information systems is hardware failures. Industrial GSM routers are required to facilitate data transmission from site to cloud infrastructure. No transmission is possible if hardware failure occurs. Examples of router hardware failure include on board electronic failures, as well as power supply failures. In both cases, the router must be retrieved from the industrial site and a new router must be requisitioned to replace the defective unit. Replacing a router results in communication downtime, which is unacceptable in the context of mission-critical information systems. Hardware failure occurrences is however inevitable and proactive measures must be implemented to prevent interruptions in data transmission. A solution would be to incorporate network layer redundancy where more than one router is utilised for data transmission purposes. This however has cost implications, both on the hardware side as well as on the bandwidth side.

3.2 SIM card mismanagement

The mismanagement of SIM cards is also a key concern that can hamper network layer stability. For industrial GSM routers to relay data, they must be connected to the backbone network of a mobile service provider. As discussed in Section 2, SIM cards are used to facilitate this uplink. The mismanagement of these modules can result in additional expenses, either in the form of new mobile broadband contracts that need to be acquired, or purchasing of additional data bundles for existing mobile broadband contracts. Mismanagement can also lead to potential theft of both routers and SIM cards. It is therefore required to accurately monitor monthly data contract bills on a consistent basis. Monthly expenses must be analysed to determine the current state of individual modules. The bills must then be compared to the operational status of router units to determine whether these modules are crucial to information system operations or an unnecessary expense.

3.3 Router mismanagement

Router mismanagement often occurs with the repositioning of equipment. One of the prominent parameters that can directly be affected by router repositioning is network signal strength. A low signal strength results in intermittent connectivity. Remote industrial locations often have weak spots in the wireless network. It is therefore crucial to first perform an investigation to determine the optimal positions where GSM routers will have maximum signal strength. Interference on site can also hamper the signal strength of a GSM network router. This should be considered when determining the best location for network equipment. To ensure that no defective router is provided to an industrial site, accurate asset management procedures must be implemented that provide details of the status of all equipment. Router theft is also a key concern that can disable the network layer of an industrial information system. GSM routers must always be installed in locations with restricted access and these locations should be recorded for future reference.

3.4 Summary

This section presented several challenges that can disrupt the transmission of data in an industrial information system. Table 1 below summarises the main causes and provides examples that are related to the listed challenges.

Table 1: A summary of the challenges that can disrupt data transmission in an industrial information system.

Challenges that disrupt data transmission	
Challenges	***Examples from industry***
Hardware failure	• Power supply failure • On board electronic failure • Antenna failure
SIM card mismanagement	• Theft • Failure to analyse costs and budgets
Router mismanagement	• Repositioning of routers • Issuing of defective routers • Theft

4. A COMPREHENSIVE INFORMATION AND ASSET MANAGEMENT SYSTEM

Based on the sections above, a comprehensive information and asset management system is required that facilitates the effective management of SIM cards in industry. The information and asset management system can be divided into two main functional blocks. The general management block serves to link various SIM cards to industrial sites or personnel. The module management block identifies individual SIM units and serves to provide cost and administrator details of individual SIM cards. This allows for a cost and budget analysis to be conducted. The following sections describe each functional block in detail.

4.1 Module management

The module management subsystem focuses primarily on providing accurate cost analysis functionalities based on the provided input data. Figure 4 gives an overview of the module subsystem.

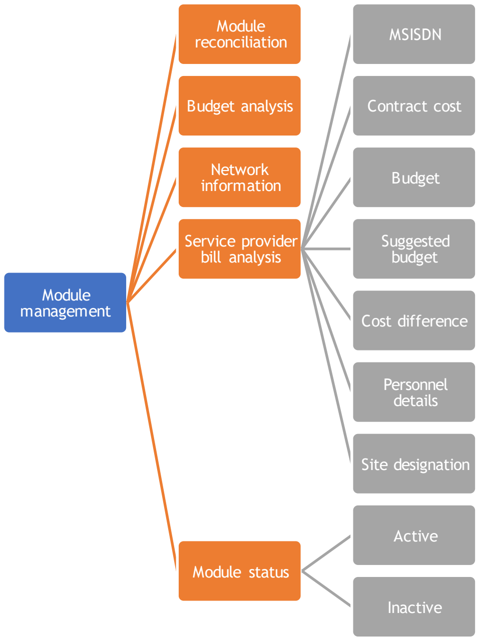


Figure 4: An overview of the module management subsystem.

For the subsystem to provide accurate cost analysis results, data contract invoices, as well as mobile service provider summary sheets must be provided as input on a monthly basis. In addition to cost analysis functionalities, the module management subsystem includes a module reconciliation section that reconciliates individual SIM cards with the corresponding module IP address. The system uses the mobile service provider summary sheet to update all MSISDN numbers with the corresponding IP address. The module reconciliation section will relay all mismatches to the user. The module management subsystem also displays network information for all SIM cards. This information includes the mobile service provider name, the contract type, and the network IP address that is linked to the module.

The cost analysis functionality relies on a separate service provider bill analysis section. The service provider bill analysis section uses the data contract invoices as input and extracts the module MSISDN number and the monthly expense for that module. This data is then compared to a suggested budget. From this comparison, a cost difference can be calculated to determine the usage condition of the module. If the module cost exceeds the suggested budget, the module is flagged as a negative expense and further investigation must be done to determine the nature and possible reason for exceeding the budget. In most cases, the sending of short message system (SMS) messages is the reason for exceeding stipulated budget values. Industrial information systems usually issue alarms via SMS to relevant personnel. Since data contracts do not cover SMS costs, an additional value is levied on the monthly contract amount.

4.2 General management

The main objective of the general management subsystem is to link a SIM card to a user. Figure 5 gives an overview of the module subsystem.

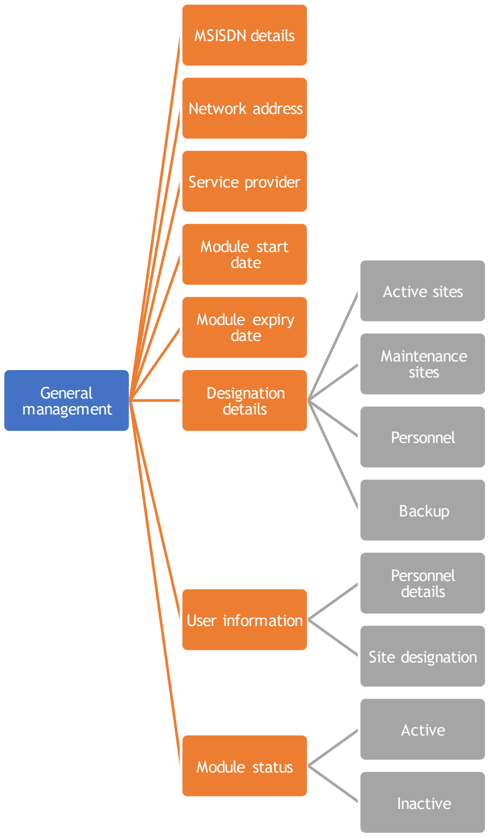


Figure 5: An overview of the general management subsystem.

The user can either be personnel or an industrial site. Personnel often require SIM cards for remote monitoring and maintenance of site equipment and systems. A SIM card with a designated IP address allows a project engineer to securely access on-site servers from a remote location. Modules that are allocated to industrial sites are utilised for data transmission purposes and forms an integral part of the network layer of industrial information systems. The general management subsystem lists module details such as the MSISDN number, the network address, the mobile service provider and the module activation and expiry date.

In addition to the information listed above, the general management subsystem also links group information to SIM cards. The purpose of the group information is to accurately indicate where the module is in use. The options that are available include active sites (sites where industrial information systems are currently active), maintenance sites, personnel and backup. Maintenance sites usually require modules to assist with remote monitoring and maintenance purposes. The general management subsystem also contains user information. If a module is labelled as a personnel module, the user information section will contain the personal details of the user, usually the name and surname, of the user. If the module is labelled as a site module, either an active site or a maintenance site, the site designation is provided. The site designation includes the industrial client group name, as well as the area of operations. Finally, the module is labelled as active or inactive. All modules issued either to personnel or to industrial sites must be in an active state.

The information and asset management system presented in this section allows for the effective management of SIM cards. The primary benefit of this system is the automated cost analysis functionality. Results of this functionality is presented below.

5. RESULTS

The information and asset management system presented in Section 4 above has been implemented on an existing telemetry network that sustains the network layer of multiple industrial information systems. These information systems are actively used by industrial client groups in the Republic of South Africa. The results presented below in Figures 6 to 9 are cost analysis results based on monthly data contract costs. An Energy Services Company (ESCo) is responsible for maintaining the communication uplink to the different industrial client groups.

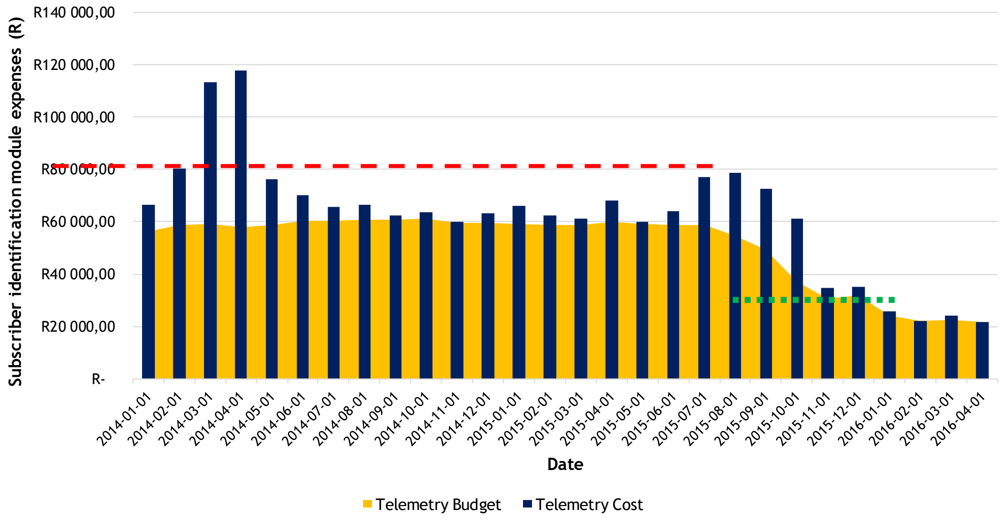


Figure 6: A cost analysis of existing SIM cards in circulation for information system operations.

Figure 6 represents a global cost analysis in South African Rand of sustaining telemetry uplinks to multiple industrial client groups. The actual cost per month is shown by the bar chart and the budgeted value is shown by the area chart. Multiple crossover points can be seen, where the total cost for the month exceeded the budgeted amount. This is because multiple SIM cards exceeded their monthly data limit. This resulted in out-of-bundle data rates coming into effect, which further increased the communication costs for the applicable months. The information and asset management system also assisted in the identification of SIM units on industrial sites that do not require any monitoring and maintenance. These SIM units were marked for decommissioning, which led to the decrease in the overall telemetry cost illustrated in Figure 6. The dashed line represents the average communication costs which equates to an amount of R71,698.93. After decommissioning unused and unnecessary cards, the average communication cost dropped to an average amount of R27,287.48, which is illustrated by the dotted line. This signifies a decrease in overall communication cost of approximately 62%.

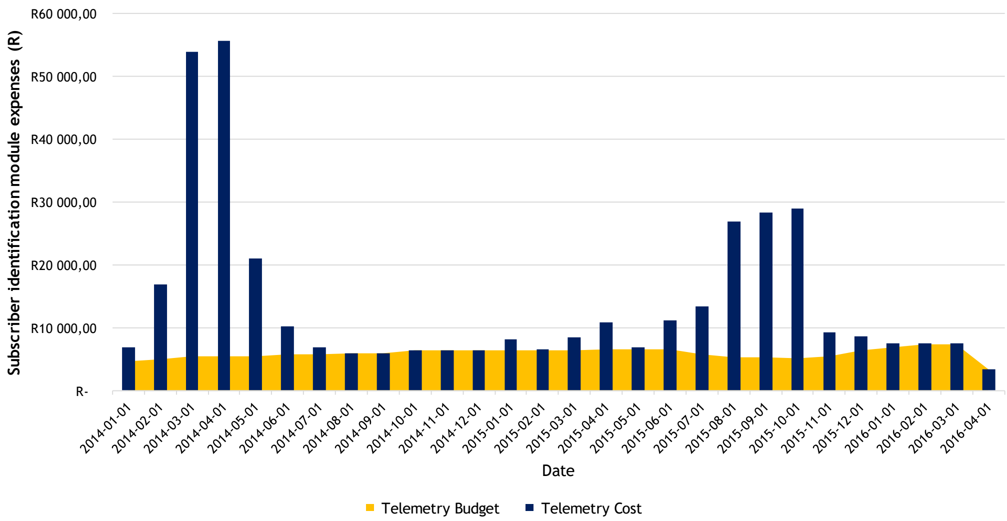


Figure 7: A cost analysis of SIM cards in circulation for a specific pumping scheme.

Figure 7 shows a specific cost analysis that was applied to a single industrial client group that was responsible for the operation of a water pumping scheme. The SIM cards utilised for sustaining the telemetry of the pumping scheme project exceeded the budgeted amount. The information and asset management system effectively allowed the ESCo to isolate the cause for excessive data bundle expenses seen in Figure 6. An individual cost analysis of the water pumping scheme telemetry network isolated the water pumping scheme network as the main cause for going over-budget. An investigation was launched by the ESCo to determine the exact cause for excessive data usage in both instances and the issues were rectified.

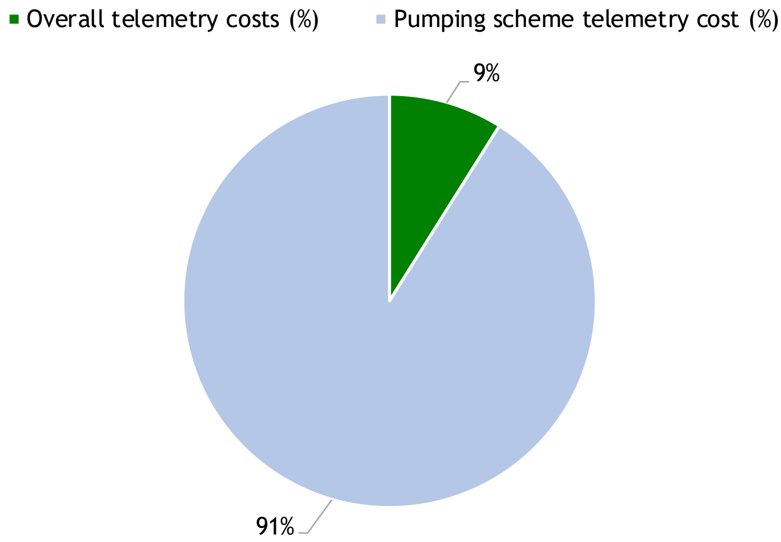


Figure 8: A cost comparison between overall telemetry costs and the telemetry cost of a pumping scheme network with exceptional data usage.

The proposed information and asset management system also allows for weight analyses to be done on different industrial client groups. Figure 8 illustrates the total weight of the water pumping scheme during the excessive data usage periods. The cost of maintaining the water pumping scheme telemetry network clearly outweighs the combined cost of other industrial client groups. This analysis can be used as a risk metric to determine the status of the telemetry network of a specific industrial client group. By doing a weighted analysis, excessive data usage can be highlighted and pinpointed to an industrial client. Investigations can then be initiated to determine the cause of excessive usage.

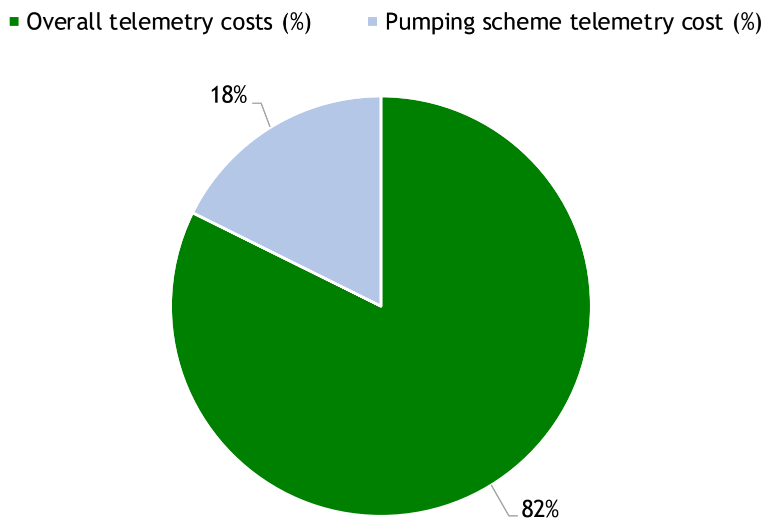


Figure 9: A cost comparison between overall telemetry costs and the telemetry cost of a pumping scheme network with normal data usage.

Figure 9 illustrates the same detailed cost analysis represented by Figure 8 however, this cost analysis was executed in a period where normal data usage was present. The total cost to maintain the telemetry network for the water pumping scheme is a mere 18% compared to the combined telemetry costs of all industrial client groups. This confirms that the telemetry uplink for a specific client is operating within acceptable budget parameters.

6. DISCUSSION OF RESULTS

A review of existing systems from literature and industry shows that there is a need for a comprehensive information and asset management system to address the management of SIM cards in industry. This system forms an integral part of the network layer of any Industry 4.0 information system that utilises the GSM network for real-time wireless data transfer. The information and asset management system proposed in this paper allowed for accurate cost analyses to be executed on SIM contracts. The system automatically generates these profiles based on billing invoices provided as input. The results presented in the previous section clearly shows the value of the system. The automated cost analyses can provide valuable insight into the health of the network layer and it can identify communication anomalies that has a direct effect on the cost of maintaining a wide-area wireless telemetry uplink.

The system proposed in this paper will ensure that industrial information systems, such as remote monitoring and maintenance systems proposed by Yang et al. [20] and condition-based maintenance systems as proposed by Van Jaarsveld et al. [8] will always be able to maintain real-time communication, whilst keeping communication costs at a minimum. In addition to the pumping scheme case study presented in the previous section, the system also identified SIM cards that were subject to theft. Module theft is usually flagged by excessive data costs. One investigation showed that a module linked to a router that was not operational still showed excessive data costs. An on-site investigation revealed that an unauthorised SIM swap was done with the replacement SIM being defective. The system proposed in this paper approaches network layer maintenance and stability from a cost and budget point of view. Not only does this improve the quality and reliability of wide-area wireless telemetry, but it also maintains indirect expenses that are associated with industrial information system communications.

The system is currently implemented as a standalone desktop system. This implies that the system can only be used locally. To increase operational and management efficiency, this system must be incorporated into existing information systems. An example would be to incorporate this system into the EMIS proposed by Goosen et al. [7]. This would not only allow industrial client groups to view telemetry cost analyses online, but it will also allow for the incorporation of these cost analyses into generated reports. This integration will expand the reach of the system and will allow clients to take responsibility of the communication costs.

7. CONCLUSION

The network layer is the backbone of any industrial information system. In the context of Industry 4.0, most information systems rely on wide-area wireless telemetry to achieve data transmission and decentralised storage and processing of data. The GSM network is currently an acceptable method to achieve this. Challenges are however present that can hamper the effectiveness of industrial information systems. These challenges include the mismanagement of hardware devices, mismanagement of SIM cards and hardware failures.

This paper presented an information systems approach to the management of SIM cards utilised in industry. The system consists of two main foundational blocks which allows for automated cost analyses to be calculated. Cost analyses contain actual expenses as well as budgeted values, which allows for budget comparisons to be made. Weighted analyses are also possible which assists in determining exceptions where communication anomalies are present.

ACKNOWLEDGEMENTS

This work was sponsored by ETA-Operations. The authors also gratefully acknowledge and thank the reviewers for the provided suggestions, which helped to improve the overall quality of this paper.

REFERENCES

[1] **Hofmann, E.** and **Rüsch, M.** 2017. Industry 4.0 and the current status as well as future prospects on logistics, *Computers in Industry*, vol. 89, pp. 23–34.

[2] **Sackey, S. M.**, **Bester, A.**, and **Adams, D.** 2017. Industry 4.0 Learning Factory Didactic Design Parameters for Industrial Engineering Education in South Africa, *South African Journal of Industrial Engineering*, vol. 28, no. 1, pp. 114–124.

[3] **Barenji, A. V**, **Barenji, R. V**, and **Hashemipour, M.** 2014. A Framework for Structural Modelling of an RFID-Enabled Intelligent Distributed Manufacturing Control System, *South African Journal of Industrial Engineering*, vol. 25, no. 2, pp. 48–66.

[4] **du Plessis, J. N.**, **Prinsloo, I. M.**, and **Groenewald, H. J.** 2013. Results from implementing a remote diagnostic and maintenance solution on energy management systems, in *2013 International Conference on Industrial and Commercial Use of Energy (ICUE)*, pp. 35–40.

[5] **du Plessis, J. N.**, **Pelzer, R.**, and **Kleingeld, M.** 2012. An automated diagnostic system to streamline DSM project maintenance, in *2012 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, pp. 7–10.

[6] **Goosen, P.**, **Mathews, M. J.**, and **Vosloo, J. C.** 2017. Automated Electricity Bill Analysis in South Africa, *South African Journal of Industrial Engineering*, vol. 28, no. 3, pp. 66–77.

[7] **Goosen, P.**, **Swanepoel, J. A.**, and **du Plessis, J. N.** 2016. The Need for a Comprehensive Energy Management Information System for Industries, *South African Journal of Industrial Engineering*, vol. 27, no. 3, pp. 1–11.

[8] **van Jaarsveld, S.**, **van Heerden, S. W.**, and **van Rensburg, J. F.** 2017. Development of a condition monitoring information system for deep level mines, in *SAIIE28 Proceedings*, pp. 3422-1-3422–13.

[9] **Vaidya, S.**, **Ambad, P.**, and **Bhosle, S.** 2018. Industry 4.0 - A Glimpse, in *2nd International Conference on Materials Manufacturing and Design Engineering*, vol. 20, pp. 233–238.

[10] **Miorandi, D.**, **Sicari, S.**, **De Pellegrini, F.**, and **Chlamtac, I.** 2012. Internet of things: Vision, applications and research challenges, *Ad Hoc Networks*, vol. 10, no. 7, pp. 1497–1516.

[11] **Dhillon, H. S.**, **Huang, H.**, and **Viswanathan, H.** 2017. Wide-area Wireless Communication Challenges for the Internet of Things, *IEEE Communications Magazine*, no. February, pp. 168–174.

[12] **du Plessis, J. N.**, **Pelzer, R.**, and **Vosloo, J. C.** 2015. Sustaining the performance of diverse energy management systems through reactive maintenance, in *2015 International Conference on the Industrial and Commercial Use of Energy (ICUE)*, pp. 44–49.

[13] **Barnaghi, P.** and **Sheth, A.** 2016. On Searching the Internet of Things : Requirements and Challenges, *IEEE Intelligent Systems*, pp. 71–75.

[14] **Wenwei, L.**, **Dafang, Z.**, **Jinmin, Y.**, and **Gaogang, X.** 2007. On evaluating the differences of TCP and ICMP in network measurement, *Computer Communications*, vol. 30, pp. 428–439.

[15] **Suciu, G.**, **Butca, C.**, and **Suciu, V.** 2015. M2M sensors for Future Internet of Things monitoring, *13th International Conference on Engineering of Modern Electric Systems, EMES 2015*.

[16] **Rawat, P.**, **Singh, K. D.**, and **Bonnin, J. M.** 2016. Cognitive radio for M2M and Internet of Things: A survey, *Computer Communications*, vol. 94, pp. 1–29.

[17] **Gazis, V.** 2017. A Survey of Standards for Machine-to-Machine and the Internet of Things, *IEEE Communications Surveys & Tutorials*, vol. 19, no. 1, pp. 482–511.

[18] **Shrouf, F.**, **Ordieres, J.**, and **Miragliotta, G.** 2014. Smart factories in Industry 4.0: A review of the concept and of energy management approached in production based on the Internet of Things paradigm, in *IEEE International Conference on Industrial Engineering and Engineering Management*, pp. 697–701.

[19] **Xu, L. Da**, **Member, S.**, **He, W.**, and **Li, S.** 2014. Internet of Things in Industries: A Survey, *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2233–2243.

[20] **Yang, S. H.**, **Dai, C.**, and **Knott, R. P.** 2007. Remote maintenance of control system performance over the Internet, *Control Engineering Practice*, vol. 15, no. 5, pp. 533–544.

THE SUITABILITY OF SAP ERP TO GENERATE A MAINTENANCE ZERO BASED BUDGET

C.A. Saffy[1] AND Dr J. Bührmann[2]
[1]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
colin.saffy@gmail.com

[2]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
joke.buhrmann@wits.ac.za

ABSTRACT

To determine the ability of SAP ERP to generate a maintenance zero-based budget, a case study was conducted at Company X, using the FY17 maintenance budget (approximately R138M). From the information obtained it was determined that the maintenance module (within SAP ERP) could only accurately allocate 35.3% of the maintenance costs for FY17. Further findings found that by changing the cost centre structure of the maintenance module in SAP ERP one could increase the cost visibility to from 35.3% to 88%. It was therefore concluded that SAP ERP is suitable to generate a maintenance zero-based budget, if setup correctly.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Mechanical, Industrial and aeronautical engineering, University of Witwatersrand, South Africa
[2] The author is a senior lecturer at the School of Mechanical, Industrial and Aeronautical Engineering, University of Witwatersrand, South Africa
*Corresponding author: joke.buhrmann@wits.ac.za

1. INTRODUCTION

Why is it necessary to budget? Simply put, every organisation only has a limited number of resources to achieve their goals and objectives [1]. Resources include funding in terms of income from operations, capital investment from owners or loans. If the cash on hand is less than expenditures, it becomes difficult to pay the bills of the company, as well as source more funding. Budget planning indicates what management has prioritised by highlighting the allocation of planned expenses in the foreseeable future – as the saying goes 'failing to plan is planning to fail'. This also accentuates the importance of planning to ensure that the company always has cash on hand [1]. However, in the maintenance field, there have been many findings of inadequate budgetary control. Maintenance budgets are often created simply by using an incremental budgeting method (the previous year's costs plus inflation), calling into question the detail of, and control over these budgets [2], [3].

Lack of planning can be quite daunting given that maintenance is supposed to provide critical support for heavy and capital intensive industries by keeping the productivity performance of plants and machinery in a reliable and safe operating condition [2], [4]. There are also costly consequences of catastrophic failure if equipment is not maintained to standard [5], [4], [6]. Due to an increase in global competition, technological complexity and ever evolving Health Safety and Environmental (HSE) regulations, maintenance costs are ballooning. For example, BHP, a world leading mining company spends US$3.5 billion annually to maintain more than 3 000 mobile machines (many larger than the average house) across its operations [7]. One can imagine the damage caused (both from a safety and production point of view) when equipment that large and powerful fails. As another example, some airline companies claim that 40-45% of their operating expenditure is maintenance related [1]. It therefore seems logical that maintenance costs should be examined and monitored to ensure sustainability, competitiveness and even safety of the company [2]. Budget planning is one method where costs can be monitored and controlled, because it provides a benchmark against actual expenditure.

The shift of maintenance from being seen as a 'necessary evil', to a value adding department that sustain long term profitability has gain momentum, especially in large industrial companies [4], [8]. In fact, there is already a maintenance standard that has been setup to align companies with best maintenance practices – ISO 55000. It is expected to become a business requisite, which obliges companies to be certified before being eligible for insurance and may soon become as important to companies as the ISO 9001 and ISO 14001 standards [9]. The backbone of ISO 55000 is the establishment of leadership roles to ensure sustainable practices are implemented. To get this right there is a focus in evidence based decision making, which provides guidelines for data identification, collection, record keeping and control [10], [9]. For large companies it is therefore vital to have a system of some sort for all this data, as well as a solid understanding of the system [8]. Because of ISO 55000 and the other complexities associated with maintenance, Computerised Maintenance Management (CMM) systems have become standard practice and necessary. CMM systems are able to store and collate all the various maintenance information. However, without analysing this data to improve the process, these systems become nothing other than an expensive and glorified spreadsheet tool [4], [11], [12].

Today, maintenance software, such as CMM systems, are commercially available, either as standalone packages, or as modules in an integrated system. This includes SAP ERP, a well-known Enterprise Resource Planning (ERP) tool [13]. Ideally CMM systems enables data to be utilised from multiple maintenance functions and manage it all through one system. By processing this large volume of data, managers are able to make more analytical decisions and implement the most effective corrective actions, thereby increasing availability and ensuring a safer work environment [5]. Again, looking at BHP - they have stated that by FY2023 (within the next 5 years) they expect to accumulate a total saving of US$1.2 billion in maintenance costs while increasing production by 20%. They plan to do this by using their CMM system and other big data platforms to optimise maintenance [7]. This shows the confidence placed in these systems when managed and implemented correctly.

With the successful implementation of these systems, management should be able to justify the value gained from maintenance, justify equipment and resource investments, investigate and identify maintenance and HSE (Health, Safety and Environment) issues or trends, focus on the essentials and adapt to the changing organisational strategies [4], [8]. However, despite investments by large companies to implement CMM systems, there are still gaps and problems with collecting and understanding the data, especially when it comes to the maintenance function [14]. One of these problems is accurately accounting of the costs within the maintenance department. This lack of cost visibility often reinforces the negative perceptions associated with the maintenance function as a 'necessary evil'.

1.1 Problem Statement

For some operations, costs can be managed manually. However, this case study was done for a company with a maintenance budget close to R 140 million and over 2000 maintainable pieces of equipment. This company will hereafter be referred to as Company X. It is clear that for a company of this size, it will become difficult to keep track of the various maintenance activities and costs without a CMM system. Company X uses a CMM system called Work Management, the maintenance module used in the SAP ERP system. Although it is required that all maintenance activities and costs are processed through Work Management, Company X was still unable to account for its complete maintenance expenditure through this system.

Due to the lack of maintenance expenses data, Company X was unable to get a holistic view of all its maintenance activities. It was also not possible to generate a quantifiable means to allocate resources, reduce breakdowns or manage effective expenditure of maintenance costs. The benefits of each strategic decision could therefore not be determined. If a CMM system fails to track maintenance expenditure accurately, then maintenance improvements are based on subjective decision-making. These kinds of decisions prevent continuous improvement and foster a culture of firefighting, as maintenance teams can often not keep up with demand due to the lack of proper planning.

1.2 Objectives

Given all the functionality that already exists in CMM systems, it appears that the necessary data is available to be used to generate a zero-based budget, as described in Section 2.1. Thus, the objective of this case study was to determine the potential of SAP ERP Work Management, to help plan and prioritise future work, as well as its suitability to be used as a tool to produce a maintenance zero-based budget.

1.3 Scope

The rest of this research includes background information on zero-based budgeting and the SAP ERP Work Management. This is followed by a literature survey to evaluate maintenance practices involving the use of CMM systems for budgeting, as well as their challenges and downfalls. Company X's maintenance data, software and culture was used as a case study to investigate these aspects.

The data from Company X, used in this case study, was reduced due to the size of the maintenance function. Only the execution maintenance costs (corrective, planned and breakdown) for the FY17 figures were analysed. Project costs, maintenance support costs and engineering costs were excluded from the study so that the focus could remain on implementing a zero-based budgeting system for execution maintenance costs.

2. BACKGROUND

2.1 Zero-based budgeting

Zero-based budgeting, strictly speaking means to budget from zero. By wiping the slate clean, all expenditure can be analysed line by line. This is a tedious, costly and time-consuming process which requires specialised skills and often additional external assistance [15], [16]. The benefit of using zero-based budgeting is that it forces one to consider the changes of the operating environment and look for cost effective alternatives, like technological improvements, alternative service providers and Activity Based Costing (ABC) analysis [17]. One company realised an 11% saving in its operating budget within the first four months of a new zero-based budgeting programme [16]. This was due to increased visibility into labour costs, as well as establishing contracts for certain services.

It should be seen as a meticulous process that allows absolute visibility into all costs. This can then be used to identify unnecessary waste within organisations, while forcing a process that prioritises resources, money and effort to achieve set targets [16]. When successful, zero-based budgeting frees up unproductive costs, which can then be redirected to where most important. The process therefore challenges the cultural mind set of cutting costs, from the "do more with less" mentality to the "do the right thing with the right amount" mentality [15].

Despite the interest in zero-based budgeting and the potential benefits, it has not been recognised as a cost-effective option for many organisations. This is because zero-based budgeting is more than a cost cutting exercise. In some cases, even if done correctly, the results will show that the budget may need to be increased. This could be due to a variety of reasons including critical maintenance needs [15], [16]. In fact, zero-based budgeting is mostly used (and successful) in large private companies which have experienced rapid growth while costs have not been controlled [15]. A full zero-based budgeting exercise is rarely required and improvements can be made without having to analyse every line item of the entire company's costs [15]. Most companies and agencies only use zero-based budgeting for selected departments or components in their periodic review [15].

2.2 Maintenance background

When using SAP ERP Work Management to analyse maintenance work, it is important to ensure that all the necessary information is captured in the system. This will ensure the availability of data to sufficiently analyse the maintenance expenditure. If a job is completed and the data captured through Work Management, one can determine: the cost of the job, which resources were used, on which equipment maintenance was performed, the components used, the services rendered, the time to completion etc. This provides a cost visibility desired for aspects of zero based budgeting. Once this information has been stored one can easily go back and manipulate the data for the purposes required.

The SAP ERP Work Management system can be set up to capture the relevant maintenance data by giving each piece of maintainable equipment a unique number, known as a function location (see Figure 1.) and the equipment is given to a plant manager and a supervisor to maintain. This is done by linking the plant manager's cost centre and the supervisor's work centre (not shown) to the equipment.

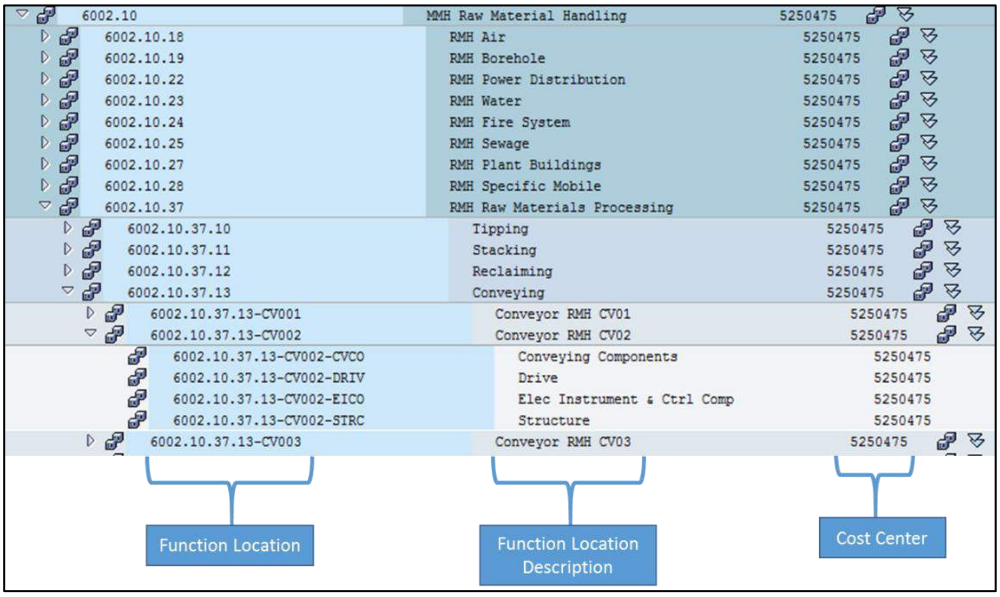


Figure 1: Typical breakdown of the function location structure showing the linkage to a maintenance cost centre.

If any maintenance work is required a notification is created by the work identifier on the required function location. This notification will alert the responsible supervisor to review, plan and execute the requested work. All costs will be allocated to the manager's cost centre. However, only additional costs over a certain threshold (i.e. external labour, services, spare parts etc.) will flow to the responsible plant manager for approval before work can commence.

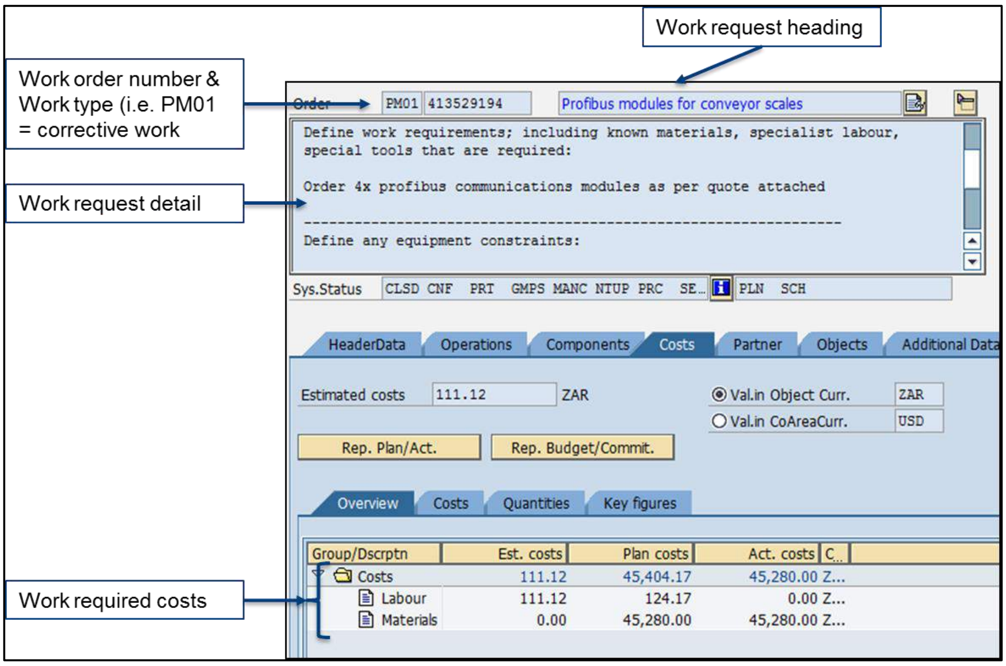


Figure 2: Screenshot of completed maintenance job on a conveyor captured in the SAP ERP Work Management system.

Employees are forced to use the system if no work can be allocated without an automatic number generated by the SAP ERP system to issue a permit, referred to as the work order number (see Figure 2.). This process was implemented to comply with the Occupational Health and Safety Act of 1993 which states that no work is to be conducted without the permission of one's supervisor and/or a risk assessment [18] . After the work has been completed, the supervisor will finalise the work details on the system and adjust any costs as necessary. This ensures that the relevant people receive the correct payment for the work they have done. Figure 2 below shows a snap shot of a work order created to purchase and install a communication module for scales. The labour and the material costs are shown, with the work order number required to conduct the work.

All this information is stored in SAP ERP for further analysis. However, if the work is not analysed, maintenance becomes isolated from the rest of the business functions and the benefits of the ERP system are never experienced. Therefore, it is important to have a maintenance support service team that can analyse these costs. In this way maintenance becomes more automated and efficient by integrating all the other business functions, therefore ensuring continuous improvement. Figure 3 was composed to illustrate how analysis of completed maintenance work, with the integration of the other business functions (e.g. Supply, Production and Finance), can lead to the continuous improvement of the maintenance function.

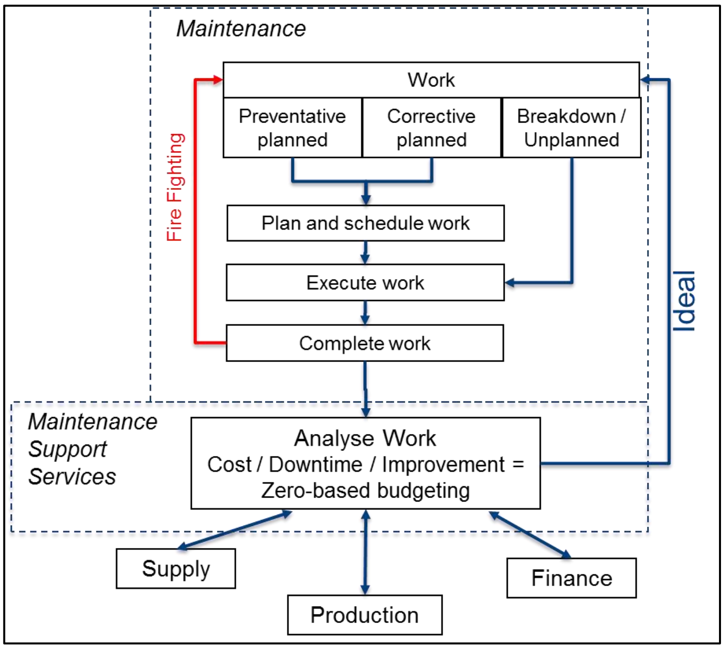


Figure 3: A simplified work process for the various different maintenance work types in a CMM system.

3. LITERATURE REVIEW

3.1 Purpose of CMM Systems

Due to size and complexity of many organisations or companies, it is necessary to ensure that all maintenance departments run through a CMM system. CMM systems no longer need to be justified, but are deemed essentials to achieving world-class maintenance, by offering a platform for objective and quantifiable decision making [11], [3]. If implemented correctly CMM systems are able to monitor any and all transactions, enabling it to identify the source and purpose of all costs and provide a platform on which this data can be analysed for strategic focus and future improvements [3]. This objective data will help management implement and sustain new maintenance techniques and strategies, as it will help them focus on the long-term benefits of the plant instead of giving way to the short-term needs, which often result in a maintenance culture of firefighting [2].

3.2 Maintenance Prioritisation Methods

It has been stated that the best practical mechanism for controlling the overall maintenance effort would be a well-designed maintenance costing system [3]. If the costing system is set up correctly, historical data can be used to build up a budget from scratch (i.e. zero-based budgeting). As an example, a case study conducted on an alumina refinery divided the plant into small sections or process units [3]. All work performed on any of these identified units, as well as downtime hours, were captured and aggregated, using manual job cards. A process flow diagram of the entire plant was plotted and the high cost as well as high downtime areas were easily identified. Resources were dedicated to those areas to ensure that they were receiving the correct maintenance, being operated according to procedure, double checking the maintenance parameters and re-measuring the costs. After going through this process and identifying root causes, the necessary actions could be taken to reduce costs and/or increase availability [3]. This process was then repeated until all high cost or low availability areas had been studied - ensuring continuous improvement. This case study was conducted using manual job cards before the availability of CMM systems. Therefore, if a cost prioritisation process implemented manually brought so much improvement, the benefits of using an electronic capture system like a CMM system, is clear.

However, it is not feasible to control maintenance from a cost perspective only. If managements' Key Performance Indicators (KPIs) are incentivised by short term initiatives only, it can fail to look at the long-term factors associated with maintenance [4]. In order to get the maintenance costing system right, one would need high level and long term oriented targets to monitor and analyse the maintenance effort, with lower level Performance Indicators (PIs) needed for shorter time intervals. This would bridge the suggested gap between maintenance strategy and business strategy, while allowing for quicker response time to imminent problems [2], [19], [3]. However, despite significant research done in maintenance, implementing PIs in both a financial and non-financial terms have been met with limited success in a practical environment, in all industries [8].

Another study tried to determine if any useful data could be extracted from multinational mining companies' CMM systems using the job card text [10]. The conclusion was that, through a rule based cleansing exercise, some companies could produce useful data, while some could not. However, the effort required to manipulate the data into something useful was identified as a major roadblock. It was concluded that a system should first be set in place that can measure data quality, to ensure that only useful and correct data is being captured.

3.3 Maintenance Budgeting Techniques

Various types of budgeting techniques have also been used to try and justify maintenance expenditures and identify areas of focus. The literature shows the following common maintenance budgeting methodologies: asset replacement value budgeting, insured value budgeting, capital cost plus inflation and incremental budgeting [6], [19]. However, because maintenance typically consists of many small and relatively inexpensive purchases or transactions, in comparison to few but large and expensive orders from other company functions, the maintenance budget is normally calculated with incremental increases to the current budget [2], [3]. This method of budgeting is often used because other methods are too time consuming. This results in crude assumptions at best [3]. It gives the impression that management is more worried about the overall maintenance spend than the details of the budget, because the maintenance function is often not linked to any particular business strategy owned by upper management i.e. increase availability, safety or extended equipment life etc. [3].

3.4 Problems Surrounding CMM systems

CMM systems have a tendency to be more financial, accounting and IT orientated, as opposed to ergonomically designed, making them less user friendly to the maintenance personnel [3], [11], [14]. Therefore, many maintenance modules are either ineffective or never fully implemented - even though they contain all the maintenance functionality necessary [3]. These systems are often seen as 'black holes' from a maintenance perspective [11]. The users insert all the necessary data, but receive nothing (or next to nothing) in return in terms of decision support, therefore wasting their time and often giving confusing and inconsistent results [3], [5], [11].

There are many other reasons for the failures of CMM systems, or the lack of utilisation of the maintenance modules in CMM systems. The main reason is the underestimation of the time, effort and money required to train the company staff to use it [3], [12]. This suggests that, although CMM systems may be capable of performing all the necessary functions, essential input data may be omitted depending on the staff's competency, training and motivation, often causing one to question the quality of the data [3], [4], [5], [10].

A quick look at SAP's website (the biggest leading CMM system worldwide for ERP suppliers) shows that they are confident they can increase asset return by 10%, reduce maintenance inventory by 11% and reduce unplanned downtime by 16% in the mining sector alone - although no academic literature could be found to back this claim [20]. BHP also stated that they are aiming to reduce maintenance costs, by using their CMM system [7].

An article written by a SAP ERP consultant, however, explains that the asset returns are possible but only 10% of companies that implement the SAP ERP system ever reach this stage (see Figure 4) [21]. It is explained that each phase requires additional work. There is also a steep learning curve for system users, due to the extensive functionality of the system to move to the next phase. Due to these challenges, 70% of all organisations do not generate any benefit from the SAP ERP system, as they do not enter Phase 3 and 4, where the benefits of the system lie [21].

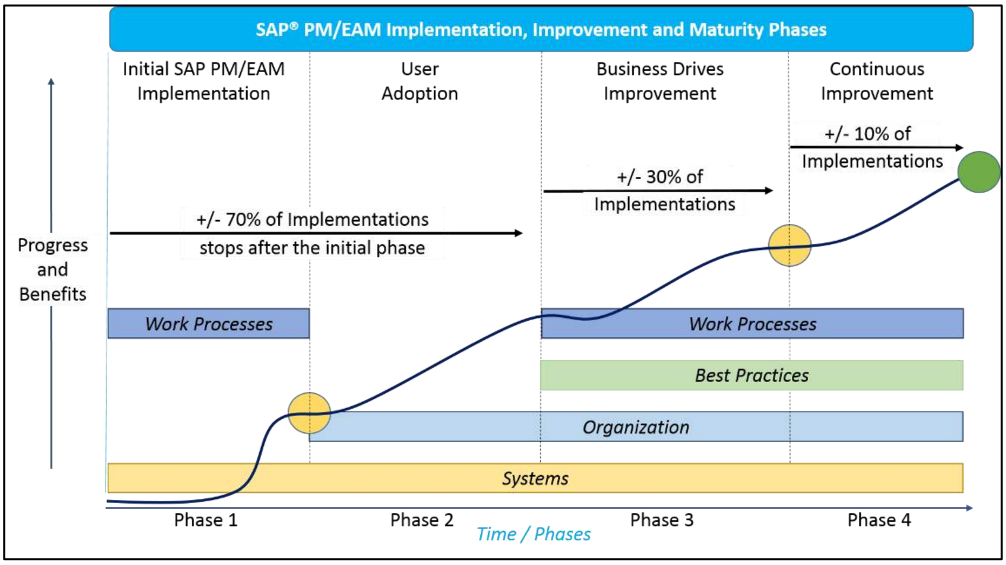


Figure 4: The SAP ERP Implementation, Improvement and Maturity phases (modified for ease of reading from [21]).

Therefore, if the system users are not able to turn the data into knowledge or actionable information to be analysed, then the purpose of the system fails and the CMM system turns into a glorified spreadsheet used to store equipment information [4], [11], [12].

3.5 Literature review summary

As can be seen, there is a lot of literature on maintenance practices, trying to find ways to improve maintenance while reducing costs [2], [3], [6], [11], [19]. Costs cannot be controlled if they cannot be monitored and improvements cannot be motivated if there is no reference or benchmark. For these reasons, and others, CMM systems have become a necessity for many companies. The reviewed literature also accentuated why CMM systems fail or fall short of their stated potential. It seems there is a gap in expectations between the software developers and users. This gap puts the two against each other as they both end up blaming the other for the failure of the system.

The majority of the research reviewed looked at the failure of CMM systems due to a lack of skills and knowledge by the user, with little or no proof of the practical success of the system [3], [4], [5], [10], [12], [21]. It would therefore be interesting to see if the software is in fact able to live up to its claimed potential. The reason for this case study was therefore to understand the implemented CMM system and investigate if an improvement in the usage of the CMM system could indeed lead to better cost savings.

Claims from companies like BHP that they are aiming to reduce maintenance costs by using their CMM system, does however show that these systems, if implemented and maintained correctly, may actually be able to provide better benefits than most companies are seeing [7], [21]. Additionally, there are many new companies coming into the market with add-on solutions, which will supposedly improve the system and allow for more accurate data collection. However, are all these new additions to the system necessary? Perhaps we just need to learn how the system works and use it accordingly.

4. METHOD

Formal and informal interviews were conducted with employees of Company X, as well as a case study approach to explore the potential for data collection from the SAP ERP system. The aim was to understand SAP ERP's maintenance system (Work Management) and how it links up with SAP ERPs financial system, while using the interviews as a means to guide and assist in fully understanding SAP and exploring its full potential.

The formal interviews were limited to twelve participants - five planners and seven supervisors. The five planners and seven supervisors were chosen because they are considered the functional experts, as they use Work Management to plan and execute the maintenance work on site and because this is all the planners and supervisors that Company X had on site. Informal interviews were held with one Principle Maintenance Specialist and two financial experts.

The formal interview questions were set out to determine how each of the participants found using the current Work Management system. Questions included: what their daily activities involved, their user limitations and if the system was changed, how would that impact their current use of the system. However, the main reason for these interviews was to determine what sort of impact any changes to the system might have on the direct stakeholders.

The informal interviews were used to ensure a good understanding of how the maintenance and financial function worked and integrated in SAP ERP. From this understanding one could ensure the correct data was extracted from the system. All the maintenance data was extracted for FY17 using the Business Improvement (BI) functionality in SAP, including 18 396 line items generated inside Work Management and 5 463 line items generated outside of Work Management.

The data that was generated through Work Management was used for analysis. This had two purposes – one, to determine any data quality issues and two, to determine if the data can be used as a means to account for costs. This information could help prioritise work based on a variety of factors (highest spend machinery, highest spend per breakdown machinery, highest spend per vendor, etc.) and can be used for benchmarking purposes. The data can then be used, not only to motivate future work and maintenance focus points, but also to reinforce the importance of having absolute cost visibility within the maintenance department.

After analysing the data, discussions were held with the maintenance and financial experts to determine the reasons for non-compliance when using the system and ways to improve the data accuracy of the system. It was then determined that SAP ERP is a suitable tool for generating a maintenance zero-based budget i.e. is it possible to obtain 100% visibility of all maintenance spend by utilising the maintenance module in SAP ERP properly.

5. RESULTS AND DISCUSSION

The total maintenance spend for FY17 was extracted from SAP ERP, totalling R 138.1 million, using the inbuilt BI functionality. The data was imported into MS Excel where it could easily be organised into a variety of ways to help prioritise work and identify data quality, behavioural and/or system issues. This was used to highlight areas that would prevent SAP ERP from being able to generate a zero-based budget. The following finding were observed:

5.1 Observation 1 – Work Management usage

The first observation from the data showed that only 47% of all maintenance expenditures were captured in Work Management (see Figure 5.). This meant that more than half of the FY17 maintenance expenses could not be properly accounted for. The rest of the expenses could only be identified through the amount captured in the General Ledger (GL) account, the vendor and the cost centre amounts were allocated to. This indicated that it was not possible to create a maintenance zero-based budget for the FY17's maintenance expenditure as the expenses could not be linked to specific equipment. One small finding of the direct maintenance costs was that roughly 2% of these costs were spent on sundries and tools. Therefore, one can only take into account 51% of the direct maintenance costs.

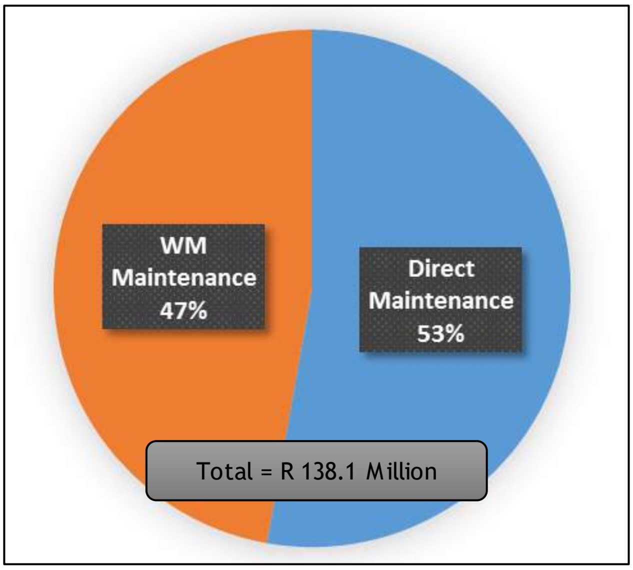


Figure 5: Total maintenance spend for FY17 divided between costs captured in the CMM system (Work Management) and costs not captured.

5.2 Observation 2 – Reasons for bypassing Work Management

Further investigation was conducted as to why 53% (R73.2 Million) of the total maintenance budget was not being spent through Work Management. It was found that the CMM system (Work Management) worked exactly like it should. However, the maintenance personnel had found ways to bypass Work Management by spending money directly and not through the processes in place

The reasons for the maintenance team bypassing Work Management was determined through interviews. There were two main reasons ascertained. The first was due to employees trying to bypass specific maintenance KPI's. The second was due to monthly maintenance contracts being set up outside work management. Therefore, the lack of cost visibility (53% of maintenance spend being spent outside of Work Management) was due to behavioural issues and a lack of knowledge of how the system is supposed to work.

The investigation also showed that only some employees could bypass Work Management while others could not. From the interviews it was determined that 27% of all the maintenance cost centres were able to bypass Work Management while the rest could not. The reason for such a small percentage of the cost centres accounting for such a large portion of the budget was because those cost centres serviced the biggest plants within the company.

The reason for the inconsistency was due to the cost centre structure. Cost centres that were able to bypass Work Management were categorised differently within SAP ERPs Finance system than those who could not. From the interviews it was determined that this difference came about when the company went through multiple changes and there were not enough people who understood the implications of these changes. The finance department therefore set up the cost structures as it suited them, but did not understand the implications it caused the maintenance department. Given that no one understood the maintenance implications these changes went unchallenged.

By changing the structure of those few maintenance cost centres that could bypass Work Management, one could force all maintenance work through the system - ensuring a greater cost visibility. From the twelve maintenance experts interviewed it was determined that ten preferred to work through Work Management and four of them had no idea how to generate orders outside of Work Management.

5.3 Observation 3 – Lack of granularity

To determine the potential of the SAP ERP Work Management system, the 47% of the maintenance budget that was captured in Work Management was analysed to determine the cost visibility. This cost could be analysed in many different ways to provide a benchmark for any given maintenance strategy. One could look at the costs of breakdown jobs, or focus only on the highest maintenance costs and ensure that the best strategy is in place for the identified equipment, or one could consider the vendors most utilised to determine if a contract should be drawn up to ensure the most competitive prices and best service. There are many different ways to use the data that comes out of Work Management, as it is comprehensive, and what is important will be different for each company depending on their business requirements.

A Pareto cost analysis was used to breakdown the costs for each piece of equipment. This is a vertical bar chart where the costs per resource type are plotted in decreasing order from left to right [22]. Figure 6 illustrates that if Work Management was used correctly the data can be sorted easily to determine which pieces of equipment are the most expensive, therefore helping management prioritise where to focus their time.

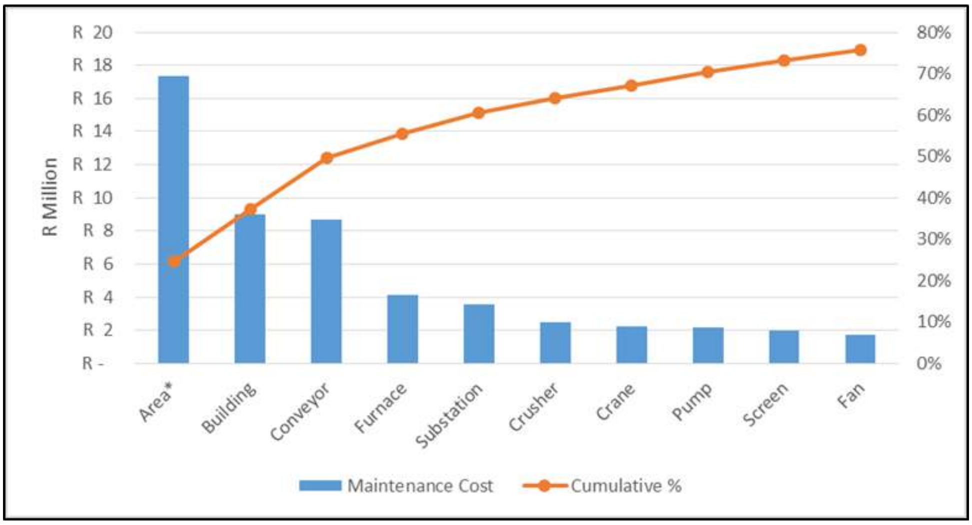


Figure 6: A Pareto cost analysis of total maintenance cost by equipment type.

Figure 7 gives a breakdown of the costs shown in Figure 6. Although Figure 7 only shows the costs of the overall labour, services and materials, the costs can easily be further analysed by resource type (mechanical, electrical or instrumentation), vendor and maintenance criticality (breakdown, planned corrective and planned scheduled) etc. Once management decided where to focus, that specific equipment (or vendor, trade, etc.) could further be analysed to understand the intricacies of that focus area.

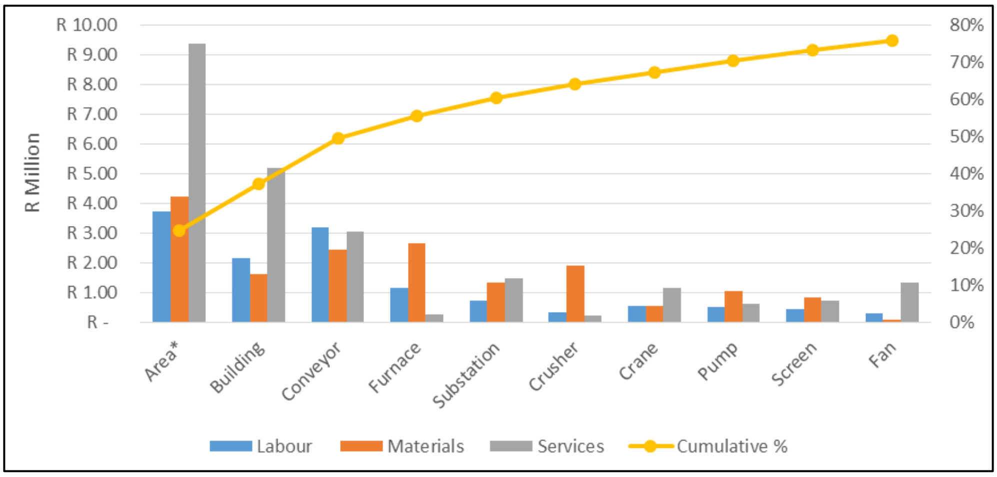


Figure 7: A breakdown of the maintenance cost analysis by equipment type into labour, materials and services costs.

The Pareto analysis was conducted for the entire expenditure captured in Work Management to determine if there were any costs that could not be linked to a specific equipment Figure 6 and Figure 7. indicated that, unfortunately, the data from Work Management also had further limited cost visibility. In particular, the Area* resource type indicated that 25% of the captured costs have not been allocated to a specific equipment. This reduced the total visible maintenance budget down, from 47%, to 35% (R 48.7 million).

Again, the reasons for job information in Work Management being allocated to an area as opposed to specific equipment are many, and were determined by the interviews. The primary reason was that no one really understood why allocating jobs to the specific equipment was important. The allocation to a specific equipment is done in the work request heading field. This is a free text field (as was seen in Figure 2.). Unless this heading was non-descriptive or unclear, the supervisors did not bother to change it.

Unlike changing the cost centre structures, which force employees to use the Work Management system, it is recommended that continuous monitoring of cost allocation be implemented and supplemented with ongoing training for all supervisors and planners to ensure that the maximum cost visibility can be obtained from the system. One way of doing this would be to set up a KPI which would monitor the cost of jobs that are not allocated to a specific equipment. This value could for example be targeted to be below 5 - 10% on a monthly basis. By doing this, one can then at least be confident that the captured data will be 90 - 95% reliable.

6. CONCLUSION AND RECOMMENDATIONS

The findings of this study reinforced what has been reported in the literature. The primary conclusions are:

- There is a lack of understanding of the CMM system, which resulted in data being incorrectly inserted or captured [4], [11], [12].
- The maintenance cost centre system was not set up correctly, which has resulted in the system being misused [11], [14].
- KPIs have been set up without the supervisors' agreement, this led to employees finding ways to bypass the system [8].
- There is a lack of supervision and leadership driving the correct use of the system, therefore it has deteriorated from its intended purpose. Even if employees capture all the required data correctly there was no reporting in place. This gave the impression that management was not interested in their efforts and not concerned about continuous improvement [3], [4], [8], [9], [11], [12].
- No maintenance efforts could be justified through the use of the SAP ERP Work Management system as progress cannot be tracked or benchmarked from previous trends [4], [8].

The purpose of this case study was to determine the suitability of the SAP ERP Work Management system to generate a zero-based budget. It was postulated that if all costs could be traced back to a specific type of equipment, then one could claim to be able to generate a zero-based budget.

The results showed that in the FY17, Company X spent R 138.1 million on maintenance, with 53% not captured in the system. The reasons for bypassing Work Management varied, but the primary motive was that certain KPIs could not be tracked and therefore was not reported on. It was also discovered that certain cost centres could bypass the SAP ERP system due to a lack of knowledge and skills when changes in restructuring were implemented.

Of the 47% that was captured within the Work Management system, only 25% was correctly linked to a specific equipment instead of an area. This lack of granularity effectively meant that visibility into the maintenance expenditure was only accountable for 35.25% (R 48.7 million) of total expenditures.

It therefore seemed that the SAP ERP Work Management system was not suitable to generate a zero-based budget as only 35.25% of the maintenance expenditure was accountable to a specific equipment. However, by implementing the recommendations mentioned in the discussion one could increase the maintenance cost visibility from 35.25% to 88%. This can be done by the following recommendations:

- Implementing cost KPI's and providing training and ongoing supervision to all personnel using the CMM system. This will ensure the majority of the maintenance costs will be visible through the BI functionality in SAP ERP. Should a compliance of 90% be achieved then the cost visibility could increase from 35.25% to 42.3% (90% of the 47% visible costs).
- By restructuring the maintenance cost centres that can bypass the Work Management system, one could force all maintenance work to be completed through the system. This would increase the maintenance cost visibly by retrieving the 45.9% (90% of the 51% spent outside of Work Management) of costs that have bypassed the Work Management system. This change would increase the maintenance cost visibility from 42.3% to 88.2%.

The benefits of implementing these recommendations are threefold

1) It can be used to provide quantitative data to determine where to focus resources, so as to reduce maintenance costs and production downtime. This can be filtered by work type, equipment type, work trade or vendors;
2) It can then be used as a benchmarking tool to track progress on any strategy that has been implemented; and
3) It will automatically put the company on a path to becoming compliant with ISO 55000 standard, as a greater percentage of the maintenance costs will be able to be monitored and therefore be used to assist in quantitative decision making.

It can therefore be concluded that SAP ERP is a suitable tool that can be used to generate a zero-based budget, provided that it is correctly implemented and maintained.

A final recommendation should be made to repeat the analysis two years from now to determine if the implemented recommendations provided the desired output by improving the system's ability to account for the majority of the maintenance spend.

7. REFERENCES

[1] Battles, B. "Budgeting and Maintenance," AVIATIONPROS, 01 July 2002. [Online]. Available: http://www.aviationpros.com/article/10387550/budgeting-and-maintenance. [Accessed 05 July 2017].

[2] Cooke, F.L. "Plant maintenance strategy: evidence from four British manufacturing firms," *Journal of Quality in Maintenance Engineering,* vol. 9, no. 3, pp. 239-249, 2003.

[3] Kelly, A. Maintenance Systems and Documenation, Burlington: Elsevier Ltd., 2006.

[4] Parida, A. Kumar, U. Galar, D. and Stenstrom, C. "Performance measurement and management for maintenance: a literature review," *Journal of Quality in Maintenance Engineering,* vol. 21, no. 1, pp. 2-33, 2015.

[5] Tretten, P. and Karim, R. "Enhancing the usability of maintenance data management systems," *Journal of Quality in Maintenance Engineering,* vol. 20, no. 3, pp. 290-303, 2014.

[6] Gupta, P. Gupta, S. and Gandhi, O.P. "Annual maintenance budget estimation for a plant system (Digraph model and matrix approach)," *Journal of Quality in Maintenance Engineering,* vol. 20, no. 2, pp. 193 - 210, 2014.

[7] BHP, "Driving value through excellence in maintenance," 20 September 2017. [Online]. Available: https://www.bhp.com/community/case-studies/driving-value-through-excellence-in-maintenance. [Accessed December 2017].

[8] Kumar, U. Galar, D. Parida, A. and Stenstrom, C. "Maintenance performance metrics: a state-of-the-art review," *Journal of Quality in Maintenance Engineering,* vol. 19, no. 3, pp. 233-277, 2013.

[9] O'Hanlon, T. and Abad, M. "Asset Management PAS 55/ISO 55000," maintenanceonline.co.uk, 26 March 2015. [Online]. Available: http://www.maintenanceonline.org/article.asp?id=9813&typ=all. [Accessed 27 May 2017].

[10] Hodkiewicz, M. and Tien-Wei Ho, M. "Cleaning historical maintenance work order data for reliability analysis," *Journal of Quality in Maintenance Engineering,* vol. 22, no. 2, pp. 146-163, 2016.

[11] Labib, A.W. "A decision analysis model for maintenance policy selection using a CMMS," *Journal of Quality in Maintenance Engineering,* vol. 10, no. 3, pp. 191-202, 2004.

[12] Braglia, M. Carmignani, G. and Frosolini, M. "AHP-based evaluation of CMMS Software," *Journal of Manufacturing Technology Management,* vol. 17, no. 5, pp. 585-602, 2006.

[13] Garg, A. and Deshmukh, S. G., "Maintenance management: literature review and direction," *Journal of Quality in Maintenance Engineering,* vol. 12, no. 3, pp. 205-238, 2006.

[14] Nikolopoulos, K. Metaxiotis, K. Lekatis, N. and Assimakopoulos, V. "Integrating Industrial Maintenance Strategy in ERP," *Industrial Management & Data Systems,* vol. 103, no. 3, pp. 184-191, 2003.

[15] Deloitte, "Zero-Based Budgeting: Zero or Hero?," 2015. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/.../us-cons-zero-based-budgeting.pdf. [Accessed July 2017].

[16] McKinsey&Company, "Five myths (and realities) about zero-based budgeting," October 2014. [Online]. Available: http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/five-myths-and-realities-about-zero-based-budgeting. [Accessed 26 July 2017].

[17] Issac, S. "Airline Budgeting," 05 February 2011. [Online]. Available: https://www.slideshare.net/samissac/airline-budgeting. [Accessed 05 July 2017].

[18] The South African Labour Guide, "Health and Safety," 2017. [Online]. Available: http://www.labourguide.co.za/component/tags/tag/2-health-and-safety. [Accessed July 2017].

[19] Galar, D. Parida, A. Kumar, U. Baglee, D. and Morant, A. "The measurement of maintenance function efficiency through financial KPI," *Journal of Physics: Conference Series,* vol. 364, no. 1, 2012.

[20] SAP, "Mining," SAP, [Online]. Available: https://www.sap.com/africa/industries/mining.html. [Accessed 02 July 2017].

[21] Van Wyk, D. "Making Sense of Plant Maintenance," 13 March 2013. [Online]. Available: http://www.maintenancetechnology.com/2013/03/making-sense-of-plant-maintenance/. [Accessed 08 August 2017].

[22] Rouse, M. "Pareto chart (Pareto distribution diagram)," TechTarget, March 2011. [Online]. Available: http://whatis.techtarget.com/definition/Pareto-chart-Pareto-distribution-diagram. [Accessed 03 March 2018].

HOSPITAL INVENTORY MANAGEMENT: A SYSTEMATIC REVIEW OF THE LITERATURE

M.R. Higgo[1*] & S.S. Grobbelaar[2] & J. van Eeden[3]

[1*]Department of Industrial Engineering
University of Stellenbosch, South Africa
matthewhiggo@gmail.com

[2]Department of Industrial Engineering AND DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP)
University of Stellenbosch, South Africa
ssgrobbelaar@sun.ac.za

[2]Department of Industrial Engineering
University of Stellenbosch, South Africa
jveeden@sun.ac.za

ABSTRACT

Hospitals and clinics provide an essential service to the populace, assisting people to overcome a range of ailments. In order to do so physicians are reliant on the tools and inventory they have at their disposal. When stock levels appear frighteningly low or even become depleted, physicians begin to order more than the ideal amount causing overstocking and consequently, wastage. This paper performs a systematic literature review in order to identify the causes for the unsatisfactory inventory management currently experienced in South African healthcare facilities.

[1] The author is enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, University of Stellenbosch
[2] The author is an Associate Professor in the Department of Industrial Engineering, University of Stellenbosch
[3] The author is a Senior lecturer in the Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION

Health services provide the essential medication, vaccines and life changing operations which citizens rely on during times of need. In South Africa there are 5,083 dispensaries[4] which provide roughly 119,155 hospital beds to patients [1]. Each of these dispensaries require a great deal of medical supplies in order to treat patients on a daily basis. Stock orders are placed which exceed the necessary amount to ensure stock availability. According to research conducted by Western Cape News Online [2] approximately R14.2 million in medication was discarded between April 2011 and April 2012 from dispenaries overstocking. Despite this, Sowsetan newspaper [3] reported that six of South Africa's nine provinces experienced antiretroviral[5] shortages during 2012. It was estimated that during 2016 South Africa experienced 110,000 deaths as a result of AIDS, had 270,000 new HIV infections and 7.1 million citizens which were still living with HIV [4].

2. PROBLEM STATEMENT

In order to accommodate the sizeable demand resulting from the HIV/AIDS epidemic, as well as other ailments, many hospitals are resorting to overstocking as an immediate solution. Overstocked inventory leads to expired products which get discarded, wasting both money and valuable resources. Dispensaries need to take another look at their inventory management policies, re-evaluate their order quantities, consider implementing modern decision support systems, and review their organisational structures.

3. OBJECTIVES

In order to learn more about the causes of poor inventory management being experienced in dispensaries this paper will perform a detailed review on the available literature surrounding the problem.

4. METHODOLOGY: AN INTRODUCTION TO THE SYSTEMATIC LITERATURE REVIEW

The *Systematic Literature Review* (SLR) is a powerful tool for acquiring relevant research papers befitting the topic with reproducible results [5], [6]. This is to say that should another researcher pursue the same topic at the same time, they should ultimately arrive at the same conclusions. The SLR is performed sequentially by a five step process [8].

1. **Scoping**: An introduction to the topic. The research objectives and target audience should be clearly defined. Do an initial search for anyone who might have already conducted a SLR on the defined topic.
2. **Planning**: Identify the topic interests and create a list of primary keywords which will be used as search terms during the *Searching* step.
3. **Searching**: Select at least one appropriate search engine from which literature will be acquired. Using the search terms defined in the *Planning* step, perform several filtered searches using the selected search engine database(s). Do not perform searches which merely filter through literature titles, authour keywords or abstracts, but rather explores all fields.
4. **Screening**: Export details of the found literature for further review. Some important information would include the record title, abstract, year, citation count and authour name(s). The final selection process involves reading the title and abstract of each record to identify topic related literature.
5. **Analysis**: Obtain the actual documentation of the remaining (chosen) literature. Carefully read through the entirety of each chosen paper and draw relevant information for the study.

5. SCOPING

The topic has been described in Sections 1, 2 and 3. A search through *Google*, *Google Scholar*, *Scopus* and *Web of Science* produced no indication of any similar SLR having been conducted. The SLR will review the following parameters:

1. The research dates of the literature.
2. The types of records found to be discussing this topic.
3. The industries to which these records are focussed.
4. The full topic list under discussion in the found literature.
5. The geographic results of where the literature was conducted, tested or observed.

6. PLANNING

Filters were used to find the most appropriate, topic-related literature. The keywords used in the Searching step of the SLR can be grouped into three individual searches, as shown in Table 1. Larger clinics employ a

[4] Hospitals and clinics.
[5] Medication used to treat HIV/AIDS.

number of similar operations to hospitals. For this reason, all three searches had to include either the keyword “hospital” or “clinic” and helps to broaden the scope of the search to other relevant literature.

7. SEARCHING

Two search engines, namely *Scopus* and *Web of Science*, were selected to each perform the three searches. Both search engines are use reliable academic databases and provide all record information (author, title, abstract, citations, year, type, etc.). This information will be required during the Screening step.

Table 1: Search terms and findings, 25 May 2018.

#	*Search terms*	*Scopus*	*Web of Science*	*# Duplicate records*	*Final count*
1	("hospital" OR "clinic") AND "inventory management" AND "decision support system"	10	3	2	11
2	("hospital" OR "clinic") AND "inventory" AND "pharmacy" AND ("policy" OR "lead time" OR "order quantity" OR "lot size")	68	8	7	69
3	("hospital" OR "clinic") AND "ward" AND "organizational structure"	108	6	6	108
	TOTAL				**188**

Table 1 identifies the number of records found from each search using the two search engines. Only articles, technical reports, journals, book chapters and conference proceedings were considered. No restrictions were set on dates. Records which were found in both search engines (duplicates) were identified, reducing the findings to 188 records. No duplicate records existed between the different searches, indicating that each search defined a unique concept.

8. SCREENING

The 188 record abstracts were collected from either Scopus or Web of Science before thoroughly being read through. This step was important to ensure that only relevant literature would be selected for the full review. It was found that 41 of the 188 records were relevant to the topic and chosen for acquisition. This selection process is described in Figure 1.

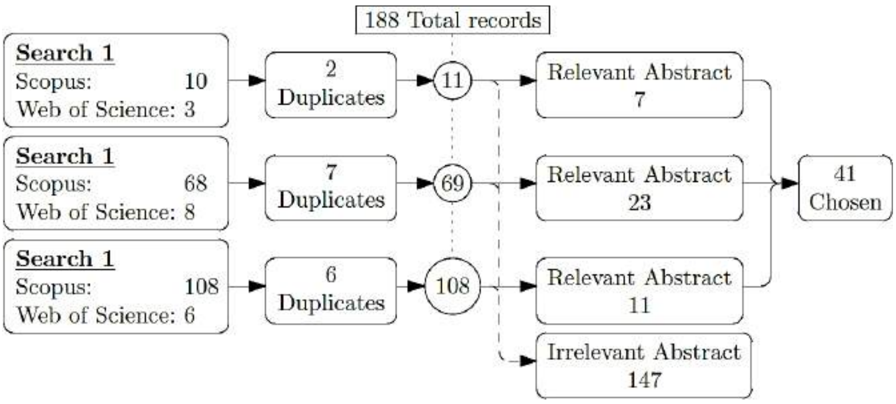


Figure 1: Selection of chosen records.

Some topics which were encountered and rejected include:

- Determining practical solutions to paediatric pain control.
- Manufacturing policies for drug production.
- Negative patient feedback.
- Nurse experience surveys.
- Organisational requirements for mass burn incidents.
- Patient experience surveys.
- Patient safety questionnaires.
- Training of advanced nurse practitioners.
- Trauma centre characteristics effect on patient outcomes.

Figure 2 shows the acquisition process. There were 19 openly accessible English records, of which one was an abstract list. Two records were only available in a foreign language and twenty records had to be specifically requested. Requests were sent both to the Stellenbosch University Library and directly to the authors via ResearchGate. Eight of these records were made available bringing the number of acquired records to 26.

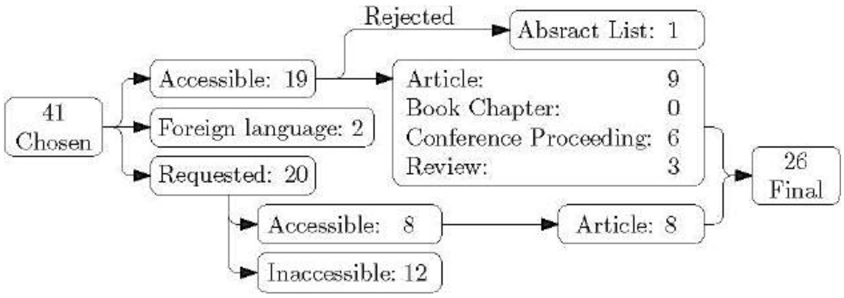


Figure 2: Acquisition of records.

Figure 3 takes a comparative look at the publication dates for the 188 records and chosen 41 records. The earliest found record dates back to 1972, and the earliest chosen record was 1974. More publications on the topic arose in 1981, but had settled down by 1990s which only contributed two records towards the chosen list. The largest activity can be seen from the start of the twenty-first century and become densely discussed between 2008 and 2018.

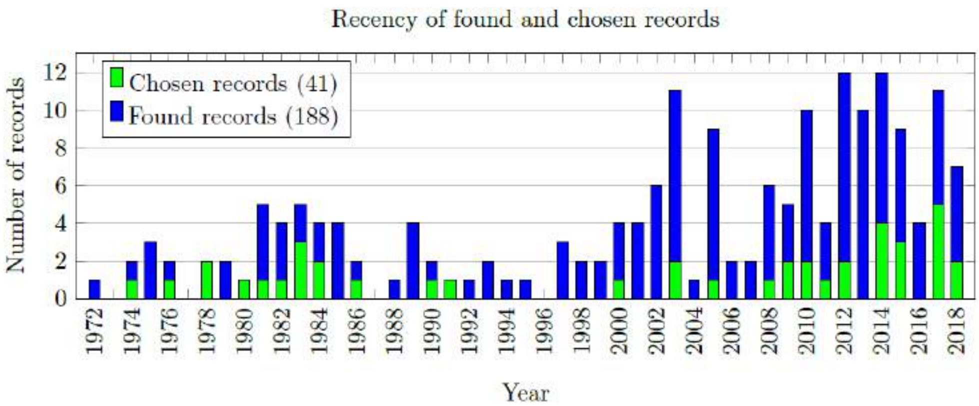


Figure 3: Number of found and chosen records, sorted by date of publication.

Figure 3 helped to understand the time-spread of the chosen records. The chosen literature may be useful to assist both modern and old-fashioned systems. An additional chart reviewing the dates of the acquired 26 records is presented in Figure 4. The acquired records appear to remain well distributed across the timeline, with exception to the gap from 1985 and 2002. Six of the earliest records were still obtainable and will be able to provide insight into past studies. The majority of acquired records were published during the twenty-first century.

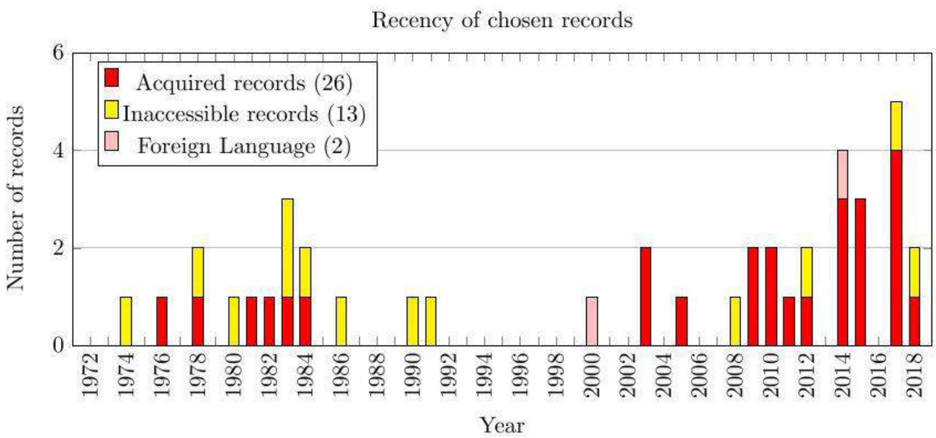
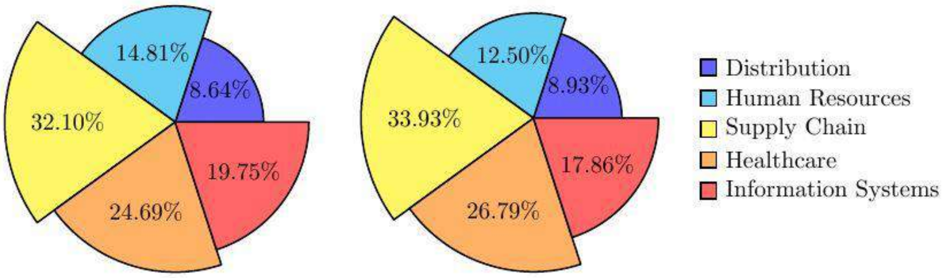


Figure 4: Number of acquired and inaccessible records, sorted by date of publication.

Knowing the source which published the record can be helpful for researchers that may want to find more papers in that field of study. 73% of the found literature were articles. Articles accounted for 65% of the acquired literature. Conference proceedings contribute 17% of the found literature and 23% of the acquired literature.

The final four records consist of three reviews and only a single book source. Only one source came up more than once, the "American Journal of Hospital Pharmacy", and provided five of the chosen records and four of the acquired records.

Five industry types were identified in the abstracts and author-defined keywords: Distribution; Human resources; Supply chain; Healthcare; Information systems. Records may exist in more than one industry type. Figure 5 (a) shows the industry distribution the 41 chosen records in the form of a polar pie chart. Similarly, Figure 5 (b) shows this for the acquired 26 records. The polar pie charts appear very similar, which indicate that the acquired literature still embodies the same proportion of industries as the chosen literature. The two most prominent industries are Supply Chain and Healthcare.



(a) *Chosen records* (b) *Acquired records*

Figure 5: Classification of records by industry type.

Authors often use keywords to help researchers to find their work. A count of the author keywords is listed in Figure 6 for the chosen 41 records, sorted in descending order of frequency. Similar terms were grouped together, such as "Medical care" being grouped with "Healthcare". The number of keywords assigned to records vary. This is why there appears to be such a small difference between the number of keywords in the acquired and inaccessible records shown in Figure 6. The bar chart will therefore only provide a rough idea of the discussions covered in the literature.

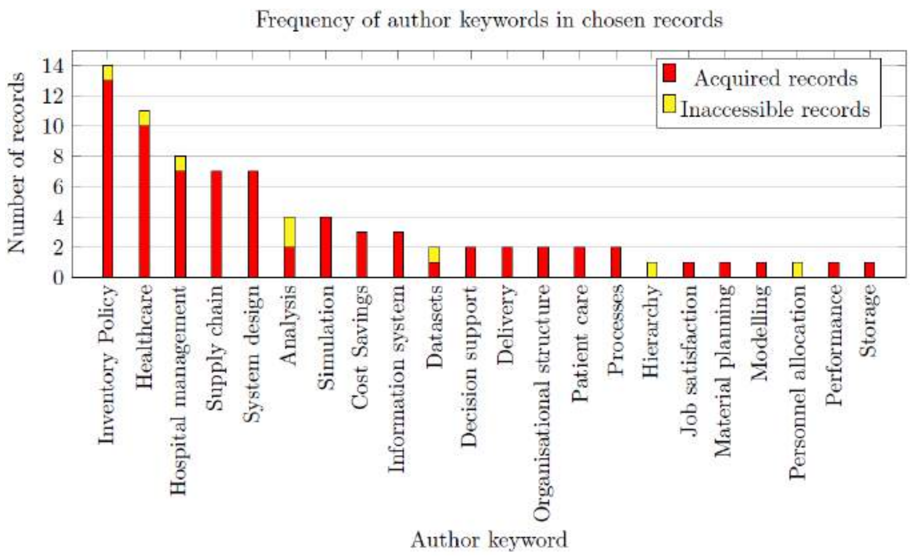


Figure 6: Number of author-defined keyword occurrences.

Countries differ with regards to infrastructure, law, religion(s) and population densities. Studies performed in one region might not be applicable to another region. For example, a wealthy institution in a developed country may publish research on expensive robotics used for manufacture. This information is not helpful to a poor institution in an underdeveloped country which can only afford cheap labour.

Figure 7 provides a visual representation of these regions on the world map (shaded black). None of the acquired records conducted their research within African, much less South Africa. 50% of the acquired records were conducted in the United States of America (USA). Figure 7 shows three distinctive clusters which can be grouped by continent. This continental view of the geographic research shows that Europe also contributed a considerable amount of the research: 50% USA; 38% Europe; 12% Asia.

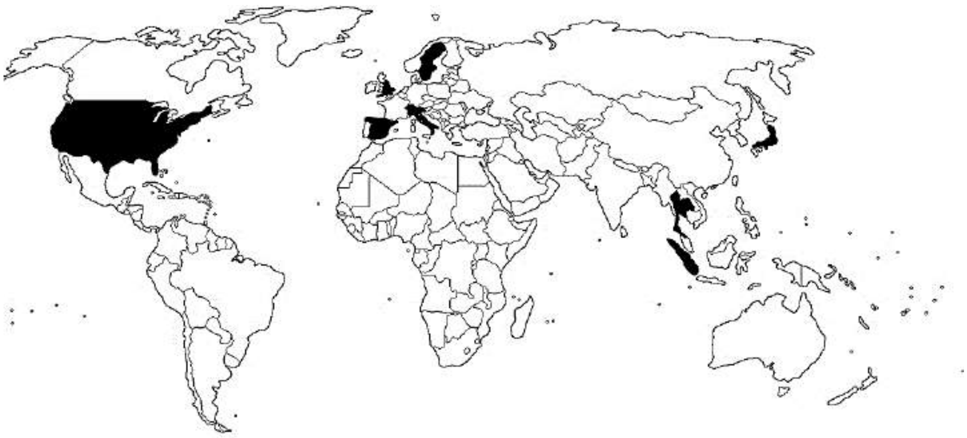


Figure 7: Geographic locations of acquired records, defined by shaded regions.

9. ANALYSIS

After reading the entirety of each record, the literature was divided into exact. The most evident topic discussed in the literature were “Inventory Policies” which featured in fourteen of the twenty-six acquired records (54%). This corresponds with the bar chart developed in Figure 6. Due to the magnitude of the literature, only “Inventory Policies” will be discussed further for this paper.

9.1 ABC Inventory Control

The first inventory management concept to be discussed from the literature is the ABC classification. This control method allows managers to focus on the minority of products which are responsible for the majority of inventory investment. As the name “ABC” suggests, there are three categories which products can be divided into [9]. Figure 8 graphically represents the ABC concept. *Category A* refers to the smallest number of products (10-15%) that make up the majority of the inventory costs (70-80%). These are the products which managers should direct most of their attention to. *Category B* populates 20-25% of the inventory and constitute 15-20% of the inventory costs and *Category C* are the majority of products (60-70%) which comprise the smallest inventory costs (10-15%) [10], [11].

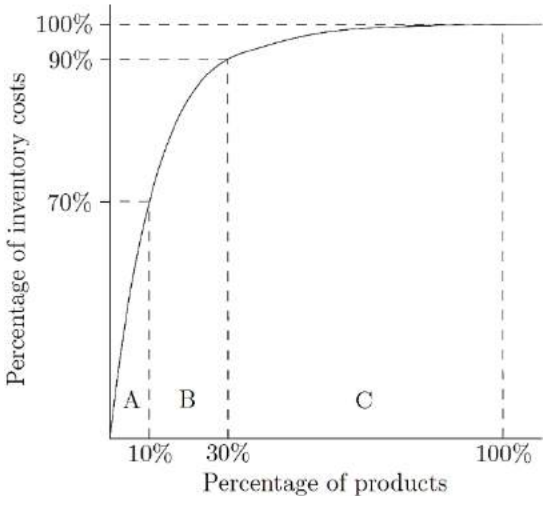


Figure 8: ABC classification.

9.2 Economic Order Quantity

The Economic Order Quantity (EOQ) was discussed in six of the fourteen inventory policy records (43%). It refers to the stock quantity ordered that exists at the point where the inventory carry costs are equivalent to the ordering costs, yielding the lowest total annual costs for that product [12], [13]. This is shown in Figure 9.

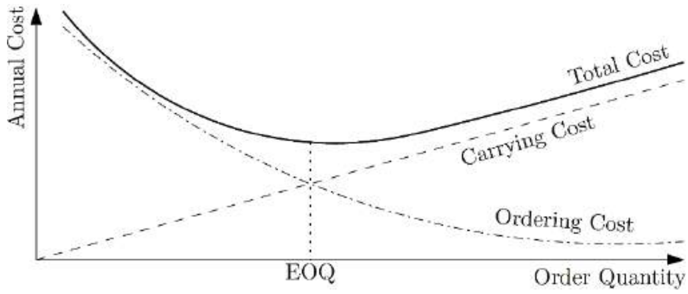


Figure 9: Economic Order Quantity curve, adapted from [9].

The EOQ is calculated using Equation 1. Additionally, the total annual cost can be calculated using Equation 2 [10], [13], [14]. The equation variables are: annual demand (D); the purchase cost of each unit (P); the annual inventory carrying cost expressed as a percentage (C); and the order cost per order (A); actual size of order placed (Q). Q is often the *EOQ* value rounded up to the nearest specified batch size.

$$EOQ = \sqrt{\frac{2DA}{CP}} \tag{1}$$

$$\text{Total cost} = \text{PD} + \frac{DA}{Q} + \frac{QCP}{2} \tag{2}$$

9.3 Types of Inventory Policies

According to Wilson *et* al. [15] inventory policies can be categorised into two primary control models namely "continuous review" and "periodic review". Continuous review refers to a system which undergo unceasing inventory level checks. This means that the moment an inventory level is reduced to the reorder point, s, a new order will be placed. Periodic reviews only perform inventory checks intermittently. This interval (period) between reviews can be as short as one day or as long as one month (30 days) and is commonly referred to as the review period, R.

Noel [9] states that the periodic inventory method offers users simplicity and may generate lower costs, but creates a lack of control and causes unnoticed shortages to occur. Wilson *et* al. [15] further subdivides these two review methods into "fixed order quantity" and "order-up-to" models. The fixed order quantity model orders pre-determined lot sizes, Q, when the inventory is smaller than or equal to the reorder point. The order-up-to model changes the lot size of each new order based on a pre-determined maximum par level, S.

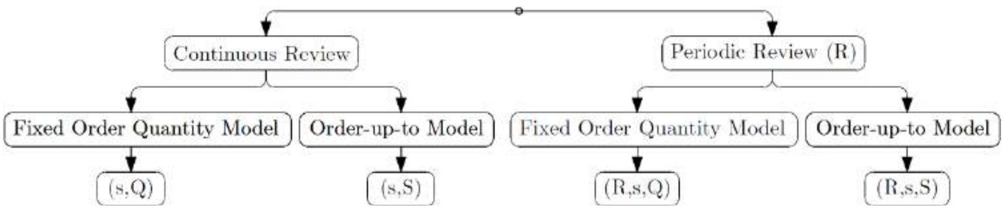


Figure 10: Types of Inventory Policies.

Figure 10 is a flow diagram describing this classification of the four possible inventory policies. The literature described several variations of determining the s, S and Q and the consequent inventory policies thereof. The follow variables will be used generically for the policies listed in Table 3, Table 3 and Table 4.

$\alpha = service\ level$ [%]
$\alpha_c = cycle\ service\ level$ [%]
$\mu = average\ daily\ demand$ $[units/day]$
$\sigma = standard\ deviation\ of\ daily\ demand$
$\sigma_L = standard\ deviation\ of\ last\ L\ days$
$\sigma_{L+1} = standard\ deviation\ of\ last\ L + 1\ days$
$C_{BI} = cost\ of\ inventory\ levels\ at\ beginning\ of\ year$ [\$]
$C_{EI} = cost\ of\ inventory\ levels\ at\ end\ of\ year$ [\$]
$C_P = cost\ of\ inventory\ purchases\ during\ the\ year$ [\$]
$COGS = cost\ of\ goods\ sold$ [\$]
$D = annual\ demand$ $[units/year]$
$DL = expected\ demand\ during\ lead\ time$ $[units/L]$
$DR = expected\ demand\ during\ review\ period$ $[units/R]$
$DOH = days\ on\ hand$ $[days]$
$E_L = expected\ number\ of\ expired\ items\ during\ lead\ time$ $[units/L]$

$EOQ = economic\ order\ quantity\ (see\ page\ 6)\ [units]$
$I = inventory\ level\ (stock\ on\ hand\ (SOH))\ [units]$
$I_L = expected\ inventory\ level\ one\ lead\ time\ away\ [units]$
$L = lead\ time\ [days]$
$OL = operational\ levelling\ factor$
$R = review\ period\ [days]$
$s = reorder\ point\ [units]$
$S = par\ level\ [units]$
$SS = safety\ stock\ [units]$
$TO = inventory\ turnover\ ratio$
$U = number\ of\ units\ below\ the\ reorder\ point\ (undershoot)\ [units]$
$z = z\ score\ (normal\ distribution)$

Table 2: Inventory policies [#1-2] found in the acquired literature.

#	Description	Type	Model
1	The Par inventory model is a popular order-up-to model which operates using periodic reviews [16]. During each period review the material handler must perform an inventory count of the stock on hand. Each product has a predetermined par-level, S, which is determined using historic demand. Once the material handler has finished counting the stock on hand, orders are placed for the exact number of items which will bring the inventory levels back up to par. Orders are made whenever stock levels fall beneath the par value [17]. This model can thus be mathematically described by [15].	(R,s,S)	$s = S - 1$ $Q = S - I$
2	The Kanban Two-Bin model is designed to make items quick and easy to access to physicians [17]. This method uses two identical bins to hold stock (one lot size, Q) of each product. The containers shelved one behind the other. Physicians take stock out of the front container. Once the last item has been picked the physician must place the empty container to one side (allocated for empty bins) and pull the second container forward [16]. Refilled bins are placed behind the current front container [18]. This model allows busy physicians to take stock without the need to check with the MH every time [15], [19].	(R,s,Q)	$DL = \mu \times L$ $s = DL + SS$ $SS = 2.3\sqrt{DL}$ $Q = s + EOQ - I$

Table 3: Inventory policies [#3-7] found in the acquired literature.

#	Description	Type	Model
3	Another periodic review fixed-order-quantity policy was described by vanDerLinde [11]. This tested A-category products, according to the ABC inventory control (see page 6), at St. Luke's Memorial Hospital in 1980.	(R,sQ)	$DL = \mu \times L$ $s = DL + SS$ $SS = 2.3\sqrt{DL}$ $Q = s + EOQ - I$
4	Wilson, Hodge and Bivens [15] developed a model which provides the decision maker with two choices for the reorder point. These choices will define the probability that the demand will exceed the inventory level. Using $\mu + 3\sigma$ will yield the lowest chance of stock-out, but may be more costly than using $\mu + 2\sigma$.	(s,Q)	$s = \begin{cases} \mu + 2\sigma, & 2.270\%\ prob(demand > I) \\ \mu + 3\sigma, & 0.135\%\ prob(demand > I) \end{cases}$ $S = 2s$ $DR = \mu \times R$ $U = \left(\frac{(DR)^2}{2(DR)}\right) - \left(\frac{1}{2}\right)$ $Q = S - s + U$
5	In order to get a comparative model for the (s,Q) model described in Policy 4, Jensen and Bard's [20] mathematical framework will be considered.	(s,Q)	$DL = \mu \times L$ $s = DL + (z \times \sigma)$ $SS = s - DL$ $Q = EOQ$
6	Kelle, Woosley and Schneider [21] designed a model to find an optimal schedule in order to minimise the overall costs to the hospital. To compare this model fairly with the rest of the literature, only the equations concerned with ordering will be considered to prioritise meeting demand rather than optimising costs.	(R,s,S)	$s = \mu(L+1) + (p(y) \times \sigma_{L+1}) - MAX\left(\frac{\sigma^2}{\mu} - 1;\ 0\right) \times \left(\frac{A + By}{C + Dy}\right)$ $y = \left(\frac{Q \times (1-\alpha)}{\sigma_{L+1}}\right)$ $p(y) = \left(\frac{a_0 + a_1 w + a_2 w^2 + a_3 w^3}{b_0 + b_1 w + b_2 w^2 + b_3 w^3 + b_4 w^4}\right)$

	A, B, C, D, a_i and b_i are constants calculated using Schneider's Approximation formulas [22].		$w = \sqrt{\ln\left(\frac{E}{y^2}\right)}$ $S = s + Q - U$ $Q = EOQ$ $U = \left(\frac{\mu^2 + \sigma^2}{2 \times \mu}\right)$
7	Wilson, Hodge and Bivens [15] modelled a periodic order-up-to policy for use in a cancer centre's ambulatory care clinic. The model is intended to operate mechanically in the same manner as the Kanban Two-Bin model (see policy #2). The *DOH* equations were found at *Accounting-Explained* [23] and *Finance Train* [24].	(R,s,S)	$COGS = C_{BI} + C_P - C_{EI}$ $TO = \left(\frac{COGS}{Average\ Annual\ Inventory}\right)$ $DOH = \left(\frac{365}{TO}\right)$ $OL = \left(\frac{DOH}{2}\right)$ $s = (\mu + 2\sigma)OL$ $S = 2s$ $U = \left(\frac{(DR)^2}{2(DR)}\right) - \left(\frac{1}{2}\right)$ $Q = S - s + U$

Table 4: Inventory policies [#8-11] found in the acquired literature.

#	Description	Type	Model
8	There are two variations of this inventory policy described by Gebicki *et* al. [14]. The first policy (this one) makes use of the ABC inventory control to select an appropriate z-value.	(R,s,Q)	$DL = \mu \times L$ $s = DL + (\sigma \times z \times \sqrt{L})$ $Q = s + EOQ - I$ $S = s + Q$ $z = \begin{cases} 1.96, & \alpha = 97.5\%\ (Category\ C) \\ 2.33, & \alpha = 99.0\%\ (Category\ B) \\ 3.09, & \alpha = 99.9\%\ (Category\ A) \end{cases}$
9	This policy is the second variation of policy #8 described by Gebicki *et* al. It uses an equation to calculate an appropriate z-value.	(R,s,Q)	Same as Policy #8 $z = \left(\frac{s_{t-1} - DL}{\sigma L}\right)$
10	The final policy tested by Gebicki *et* al. achieved the lowest average cost and stockout values making it the most promising policy discussed in the paper. For this model Gebicki *et* al. fixed the z-value for all products to the highest service level.	(R,s,Q)	$DL = \mu \times L$ $s = DL + (\sigma \times z \times \sqrt{L})$ $I_L = I - DL - (\sigma \times z \times \sqrt{L}) - E_L$ $Q = s - I_L$ $S = s + Q$ $z = 3.09\ (\alpha = 99.9\%)$
11	An additional inventory model not found in the SLR, but worth testing is presented by Chopra and Meindl [25]. In this inventory model a cycle service level, α_c, is calculated based on the expected demand during the lead time and used instead of the predefined service level, α. This cycle service level, α_c, determines the probability that the current reorder point, s, will be able to support the expected demand during lead time, DL.	(R,s,Q)	$\sigma_L = \sigma\sqrt{L}$ $DL = \mu \times L$ $\alpha_c = Prob(DL \leq s)$ $= F(s, DL, \sigma_L)$ $[Excel] = NORM.DIST(s, DL, \sigma_L, 1)$ $SS = F^{-1}(\alpha_c, DL, \sigma_L) - DL$ $[Excel] = NORM.INV(\alpha_c, DL, \sigma_L) - DL; \quad \alpha_c \neq 1.0$ $s = SS + DL$ $Q = EOQ$

10. TESTING THE INVENTORY POLICIES

This section will test and evaluate the inventory policies discovered during the SLR. South Africa's department of health released a "Master Procurement Catalogue" on the 13th July 2018 [26] which lists 1174 individual contracts commissioned between 2013 and 2018. Each contract defines a unique order agreement (product, package size, etc.) and discusses roughly 1062 medicines and vaccines provided by approximately 78 suppliers. The promised lead times are shown in Figure 11. There are only 25 items expected to be delivered in less that 7 days. The majority of products are delivered after roughly two weeks (14 days). There are 10 items which take three weeks for delivery, three items which take four weeks for delivery and two items which take three months for delivery. These lead times are much larger than described in the literature, which regularly assumed lead times of less than a week.

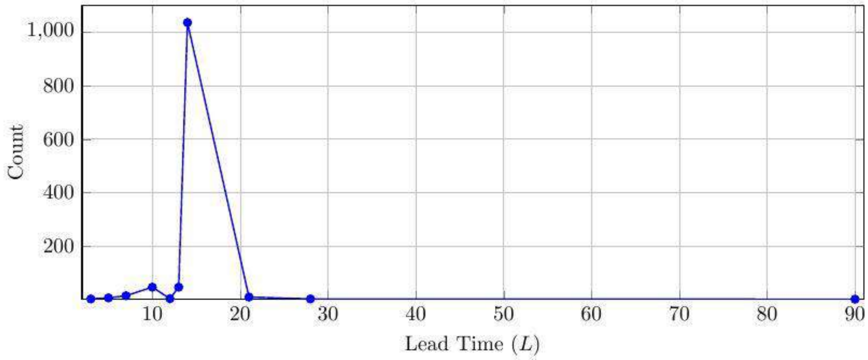


Figure 11: Lead time frequency for medical products.

10.1 Test parameters

Before proceeding to test the inventory policies found in the SLR, four custom policy models have been created by integrating different aspects of the equations found in the literature. These four custom inventory policies are listed as policies 12–15.

Policy 12 - (s,Q)

$SS = z \times \sigma \times \sqrt{L}$
$DL = \mu \times L$
$s = DL + SS$
$Q = EOQ$

Policy 13 - (s,Q)

$SS = z \times \sigma \times \sqrt{L}$
$DL = \mu \times L$
$s = DL + SS$
$S = 2 \times s$
$Q = S - I$

Policy 14 - (R,s,S)

$E_L = I - \mu$
$DL = \mu \times L$
$SS = z \times \sigma \times \sqrt{L}$
$I_L = I - DL - SS - E_L$
$s = DL + SS$
$Q = s - I_L$
$S = s + Q$

Policy 15 - (R,s,S)

$DL = \mu \times L$
$SS = z \times \sigma \times \sqrt{L}$
$s = DL + SS$
$S = 2 \times s$
$Q = S - I_L$

All fifteen policies were tested with six demand sets containing 1095 days (3 years) data. Figure 12 shows the first 365 days (1 year) data for these six sets. The fifteen policies were tested using all combinations of lead times (L) and review periods (R) ranging from 1 to 28 days. This means that each policy was tested $28 \times 28 = 784$ times in each demand set.

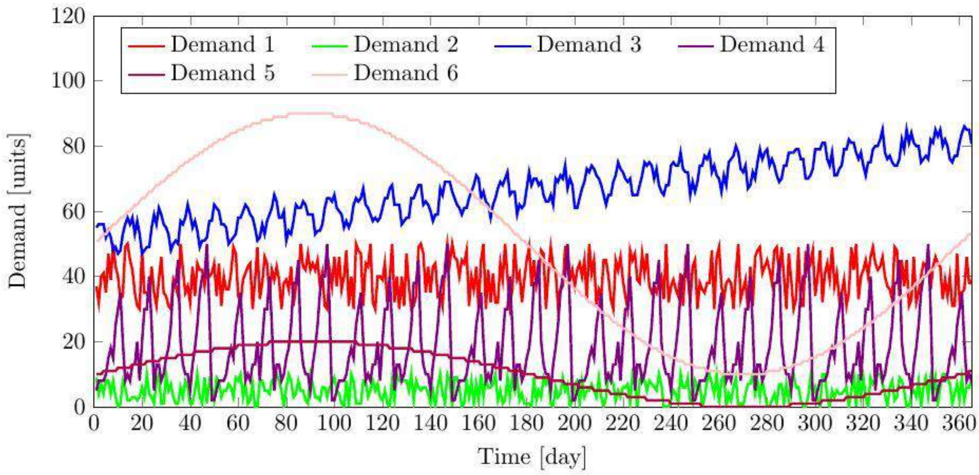


Figure 12: First year demand sets used to test Policies 1−15.

These tests will not consider minimising costs, but rather focus on determining how well the policies can meet patient demand. For this reason the test results were compared by looking at the percentage of days for which the policy was unable to meet the full demand (E). For example, a policy that failed to meet the full demand for 100 days of the three year period scored $E = \frac{100}{1095} \times 100\% = 9.1\%$. Additionally, the maximum and minimum inventory levels experienced over the three-years was recorded. Assumptions had to be made for certain models:

- In all cases, the starting inventory level was 350 units.

- Unmet demand must still be accommodated. Patients unable to get medication at that time are likely to ask when new stock will be available and return for it.
- The average daily demand, μ, was calculated using a moving average of the last 25 days' demand.
- The service level, α, was fixed at 99%, unless stated otherwise.
- For the EOQ: The purchase cost per unit was set to $P = \$100.00$, the annual inventory carrying cost was set to $C = 10\%$, and the order cost per unit was set to $A =$ R20.00.
- The days on hand, *DOH*, equation used in Policy 7 requires company-specific values which cannot merely be made-up. For this reason the days on hand was changed to twice the longest contributing period, $DOH = 2 \times MAX(L; R)$.

10.2 Findings

Firstly, policies using the *EOQ* appeared to perform fine for smaller lead times (L ≤ 3 days) and review periods (R ≤ 3 days), but failed completely for any longer periods. The reason for this is that the *EOQ* equation does consider the lead time, nor review period. The order quantity will only scale with the current annual demand value, calculated using the last 365 days. Any recent developments in demand would barely have an effect on the order quantity. The *EOQ* value thus remained fairly unchanged throughout all tests. This means policies 2, 5, 6, 11 and 12 (which all use the *EOQ*) were already experiencing *E* values exceeding 90% by L<7 and R<14 days, and as early as L=2 and R=2 days.

Policy 11 experienced a problem with increasing demand sets. As the demand continued to increase the cycle service level, α_c, would gradually decrease. As α_c decreased the safety stock, SS, decreased. Additionally, as the demand increased, DL increased and *SS* further decreases. This relationship between the SS and DL values caused the reorder point, $s = SS + DL$, to remain constant despite the change in demand. The increasing difference between the *DL* and *s* values gradually reduced α_c to zero causing an error in the "NORM.INV" function. The rest of the code would eventually fail as a result.

Policies 2, 5, 6, 11 and 12 at this point have all failed to perform acceptably. The greatest factor influencing the results was the review period. Even for the lowest lead time (L = 1 day) *E*-values were often above 20% for R ≥ 3 days. The best results were most often achieved from demand sets 2 and 4. This was likely due to the lower demand values experienced in the sets. Demand set 5 was also showing low demand values, but due to the consistent change in seasonality the policies battled to manage the demand.

The greatest problem was a lack of recovery. None of the inventory policies incorporated unmet demand into the next order quantities. This meant that policies were assuming unmet demand is gone forever. At a dispensary, patients will be likely to ask when new stock will be available and return for it. New stock should first be used to accommodate the waiting demand and then still be sufficient to satisfy any new demand. Figure 13 identifies two examples of where inventory levels were unable to replenish because the order quantities had not considered the unmet demand.

The best results were obtained by policies 7, 10, 13 and 15. Each of these policies appeared to perform well within certain parameters. Policy 7 performed best overall and obtained near-perfect results in demand sets 2 and 4. This was because of the order quantity's relation to both the review period (with the undershoot value) and the lead time (with the reorder point that is related to the operational levelling factor). However, for the remaining demand sets policy 7 only performed very well for R >> L. This may have been a result of changing the DOH formula. Regardless, Policy 7 shows promise for determining lot sizes with relation to lead time and review period values.

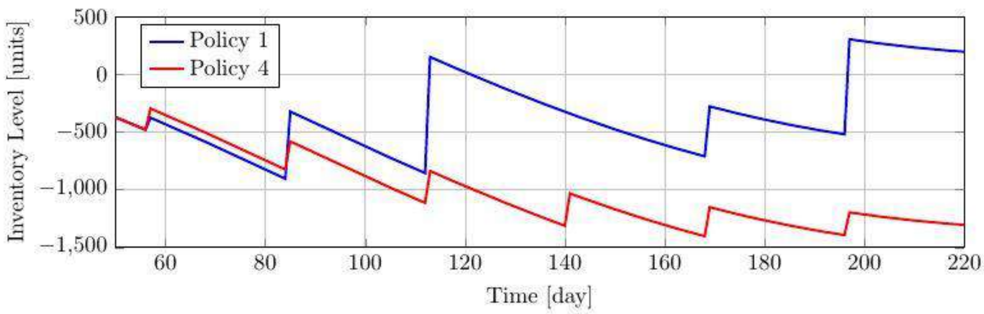


Figure 13: Inventory levels over the first 220 days for Policies 1 and 4 in demand set 5 with L=28 and R=28 days.

Policies 10, 13 and 15 achieved their best results where L >> R. The reason for this was how each of these policies' order quantities scaled with the lead time. Additionally, shorter review periods produce more frequent orders being placed, allowing the policies to respond better to unexpected demands.

11. CONCLUDING REMARKS

Hospitals and clinics should not be treated as a regular “supply and demand” industry. Patients are reliant on despensaries to acquire the necessary medication and treatment they cannot get elsewhere. Inventory policies should thus include unmet demand in its reorder decision making process.

The inventory policies found during the systematic literature review came primarily from first world countries with short lead times and using electronic systems to provide accurate, regular reviews. South Africa, however, is still in the process of improving the healthcare supply chain and has to be aware of the effects of longer lead times and review periods. The current inventory policies used in South Africa are causing a large amount of wastage from discarding expensive products. New policies should be explored to prioratise meeting demand rather than minimising cost. By reducing stockouts and the degree of overordering, despensaries can save money from the lowering of wastage.

12. REFERENCES

[1] T. Makombo, “Fast Facts: Public health sector in need of an antidote,” *Fast Facts*, vol. 6, no. 298, pp. 1–6, 2016.

[2] M. Kachwee and D. Hartmann, “Hospital Supply Chain Management and Optimisation,” *SAIIE25 Proc.*, vol. 511, no. 1, pp. 1–12, 2013.

[3] SAPA, “Aids drug shortage in 6 provinces,” 2012. [Online]. Available: https://www.sowetanlive.co.za/news/2012-05-30-aids-drug-shortage-in-6-provinces/. [Accessed: 14-Jun-2018].

[4] Avert, “HIV and AIDS in South Africa,” 2018. [Online]. Available: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/south-africa. [Accessed: 16-Jun-2018].

[5] D. Budgen and P. Brereton, “Performing systematic literature reviews in software engineering,” *Int. Conf. Soft. Engin.*, pp. 1051–1052, 2006.

[6] K. Khan, R. Kunz, J. Kleijnen, and G. Antes, *Systematic Reviews to Support Evidence-based Medicine: how to review and apply findings of healthcare research*, 1st ed. London: Royal Society of Medicine Press, 2003.

[7] B. Kitchenham, “Procedures for performing systematic reviews,” *Keele, UK, Keele Univ.*, vol. 33, no. TR/SE-0401, pp. 1–28, 2004.

[8] A. Siddaway, “What is a systematic literature review and how do I do one?,” Stirling, UK, 2014.

[9] M. Noel, “Inventory control: a review for hospital pharmacists,” *Hosp. Pharm.*, vol. 13, pp. 617–631, 1978.

[10] P. W. Abramowitz, “Controlling financial variables - purchasing, inventory control, and waste reduction,” *Am. J. Hosp. Parmacy*, vol. 41, no. 17, pp. 309–317, 1984.

[11] L. P. VanDerLinde, “System to Maximise Inventory Performance in a Small Hospital,” *Am. J. Hosp. Pharm.*, vol. 40, no. 3, pp. 70–73, 1983.

[12] D. O. Francher and D. B. Beasley, “Cost cutting through the EOQ.” pp. 36–37, 1981.

[13] R. Ballentine, R. L. Ravin, and J. R. Gilbert, “ABC inventory analysis and economic order quantity concept in hospital pharmacy purchasing,” *Am. J. Hosp. Parmacy*, vol. 33, pp. 552–555, 1976.

[14] M. Gebicki, E. Mooney, S. J. (Gary) Chen, and L. M. Mazur, “Evaluation of hospital medication inventory policies,” *Health Care Manag. Sci.*, vol. 17, no. 3, pp. 215–229, 2014.

[15] K. J. Wilson *et al.*, “Reducing Stockouts in a Cancer Center’s Ambulatory Care Clinics,” *Eng. Manag. J.*, vol. 27, no. 3, pp. 99–108, 2015.

[16] N. Esmaili, B. A. Norman, and J. Rajgopal, “A Heuristic for Selecting Multi-item Inventory Review Policies,” pp. 1050–1060, 2015.

[17] Southwest Solutions Group, “Par vs Kanban Hospital Inventory | Two Bin Supply System for Nursing Supplies,” 2018. [Online]. Available: http://www.southwestsolutions.com/healthcare/par-vs-kanban-hospital-inventory-two-bin-supply-system-for-nursing-supplies. [Accessed: 26-Apr-2018].

[18] T. Lyon, “Two Bin Supply System at St Clair Hospital - YouTube,” 2012. [Online]. Available: https://www.youtube.com/watch?v=yjSwwPF5BUU. [Accessed: 26-Apr-2018].

[19] C. R. Rosales, M. Magazine, and U. Rao, “The 2Bin system for controlling medical supplies at point-of-use,” *Eur. J. Oper. Res.*, vol. 243, no. 1, pp. 271–280, 2015.

[20] P. A. Jensen and J. F. Bard, “Inventory Theory - The (s, Q) Inventory Policy,” in *Operations Research Models and Methods*, 1st ed., Australia: John Wiley and Sons, 2003, pp. 1–4.

[21] P. Kelle, J. Woosley, and H. Schneider, “Pharmaceutical supply chain specifics and inventory solutions for a hospital case,” *Oper. Res. Heal. Care*, vol. 1, no. 2–3, pp. 54–63, 2012.

[22] H. Schneider, “Methods for Determining the Re-order Point of an (s , S) Ordering Policy when a Service Level is Specified,” *J. Oper. Res. Soc.*, vol. 29, no. 12, pp. 1181–1193, 1978.

[23] AccountingExplained, “Days’ Inventory on Hand Ratio,” 2013. [Online]. Available: https://accountingexplained.com/financial/ratios/days-inventory-on-hand. [Accessed: 06-Jul-2018].

[24] Finance Train, “Inventory Turnover and Days of Inventory on Hand (DOH),” 2018. [Online]. Available: https://financetrain.com/inventory-turnover-and-days-of-inventory-on-hand-doh/. [Accessed: 06-Jul-

2018].
[25] S. Chopra and P. Meindl, "Managing Uncertainty in a Supply Chain: Safety Inventory," in *Supply Chain Management: Strategy, Planning, and Operation*, 6th ed., Boston Pearson, 2015, pp. 328–347.
[26] South African Department of Health, "Medicine: Medicine List," 2018. [Online]. Available: http://www.health.gov.za/index.php/medicine. [Accessed: 19-Jul-2018].
[27] G. Gordon and R. Tibshirani, "Gradient descent revisited," *(Unpublished) Presentation*. pp. 1–31, 2012.
[28] M. Dotoli, M. P. Fanti, G. Iacobellis, L. Martino, A. M. Moretti, and W. Ukovich, "Modeling and Management of a Hospital Department via Petri Nets," in *2010 IEEE Workshop on Health Care Management (WHCM)*, 2010, pp. 1–6.
[29] Jonrinaldi, M. Y. Primadi, and R. A. Hadiguna, "A joint inventory policy under permissible delay in payment and stochastic demand (Case study: Pharmacy Department of Pariaman Hospital)," in *American Institute of Physics (AIP) Conference Proceedings*, 2017, pp. 1–10.
[30] J. L. Lu He, J. L. T. Cuevas, M. Z. Kokash, E. A. Banimustafa, and M. T. Khasawneh, "Optimization in Pharmaceutical Supply Chain Inventory management for Disaster Planning," in *67th Annual Conference and Expo of the Institute of Industrial Engineers 2017*, 2017, pp. 1583–1588.
[31] J. M. Maestre, M. I. Fernández, and I. Jurado, "An application of economic model predictive control to inventory management in hospitals," *Control Eng. Pract.*, vol. 71, pp. 120–128, 2017.
[32] R. Iannone, A. Lambiase, S. Miranda, S. Riemma, and D. Sarno, "Cost savings in hospital materials management: look-back versus look-ahead inventory policies," *Int. J. Serv. Oper. Manag.*, vol. 22, no. 1, p. 60, 2015.
[33] J. M. Woosley, S. Wiley-Patton, P. Kelle, and H. Schneider, "Decision support in healthcare supply chain management and pharmaceutical inventory control," in *15th Americas Conference on Information Systems 2009, AMCIS 2009*, 2009, vol. 6, pp. 1–10.
[34] T. Awaya *et al.*, "Automation in Drug Inventory Management Saves Personnel Time and Budget," *Yakugaku Zasshi*, vol. 125, no. 5, pp. 427–432, 2005.
[35] N. Saelim, P. Porouhan, and W. Premchaiswadi, "Improving organizational process of a hospital through Petri-net based repair models," in *International Conference on ICT and Knowledge Engineering*, 2016, pp. 109–115.
[36] E. Xu, M. Wermus, and D. B. Bauman, "Development of an integrated medical supply information system," *Enterp. Inf. Syst.*, vol. 5, no. 3, pp. 385–399, 2011.
[37] A. Asllani, E. Culler, and L. Ettkin, "A simulation-based apheresis platelet inventory management model," *Transfusion*, vol. 54, no. 1, pp. 2730–2735, 2014.
[38] M. W. Cooke, J. Higgins, and P. Kidd, "Use of emergency observation and assessment wards: a systematic literature review," *Emerg. Med. J.*, vol. 20, no. 2, pp. 138–142, 2003.
[39] C. Cordenier, "To what extent is the Panopticon a suitable analogy for power?," 2011. [Online]. Available: http://www.e-ir.info/2011/03/02/to-what-extent-is-the-panopticon-a-suitable-analogy-for-power/. [Accessed: 15-Mar-2018].
[40] R. L. Kingston, C. E. Pfeifle, R. J. Cipolle, D. E. Zaske, C. B. Eells, and H. L. Johnson, "Comprehensive Pharmaceutical Services for a State Correctional Facility," *Am. J. Hosp. Pharm.*, vol. 39, pp. 86–89, 1982.
[41] F. Llopis Roca *et al.*, "Register of short-stay units in Spain, the REGICE project, study 1: Location, organizational features, and staffing," *Emergencias*, vol. 26, no. 1, pp. 57–60, 2014.
[42] T. Mettler and M. Eurich, "A 'design-pattern'-based approach for analyzing e-health business models," *Heal. Policy Technol.*, vol. 1, no. 2, pp. 77–85, 2012.
[43] D. Myny, D. Van Goubergen, V. Limère, M. Gobert, S. Verhaeghe, and T. Defloor, "Determination of standard times of nursing activities based on a nursing minimum dataset," *J. Adv. Nurs.*, vol. 66, no. 1, pp. 92–102, 2009.
[44] T. O. Paulussen, N. R. Jennings, K. S. Decker, and A. Heinzl, "Distributed patient scheduling in hospitals," in *IJCAI International Joint Conference on Artificial Intelligence*, 2003, pp. 1224–1229.
[45] A. M. Pegram, M. Grainger, K. Jones, and A. E. While, "An exploration of the working life and role of the ward manager within an acute care hospital organisation," *J. Res. Nurs.*, vol. 20, no. 4, pp. 312–328, 2015.

A SYSTEMS APPROACH TO SUSTAIN RSA SECTION 12L TAX INCENTIVE PROJECTS

L. Coetzee[1], W. Booysen[2] & J.F. van Rensburg[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
lcoetzee@rems2.com
[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbooysen@rems2.com
[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
jvanrensburg@rems2.com

ABSTRACT

Section 12L of the South African Income Tax Act (12L) incentivises companies towards improved energy efficiency (EE). Research shows that two key issues generally prevent eligible EE projects from achieving maximum 12L benefits: 1) poor data integrity which affects application compliance and 2) savings are not sustained throughout the assessment period. A need therefore exists to make use of intelligent data driven analyses to effectively identify these issues in order to maximise 12L potential.

This paper presents the development of a performance tracking and reporting system. This system monitors and analyses project data to identify potential anomalies. Once an anomaly has been detected, it triggers a report to notify and inform the end-user of the specific issue. This detection increases the opportunities to sustain targeted savings and improve data integrity.

The functionality of the system is illustrated by using three case studies. These case studies test whether the system can correctly identify and report anomalies. The value of the system is further validated by assessing the potential increase in savings due to timely corrective action. The paper concludes with a discussion of the benefits, including a potential increase of R 63.9 million in 12L related value.

This work was sponsored by ETA Operations

[1] Mr. L. Coetzee holds a Master's of engineering from the North-West University.

[2] Dr. W. Booysen is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr. J.F. van Rensburg is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

Section 12L of the Income Tax Act (12L) incentivises energy efficiency by offering tax certificates valued at 95c per kWh for quantified energy savings [1]. However, this attractive offer is governed by strict Regulations and Standards [2][3]. Furthermore, the overall application and review process is diligently managed by SANEDI to ensure that the final claim is a compliant and conservative reflection of the true energy savings [4].

The value offering of 12L has made it an attractive research topic, driving further development and innovation in several areas including measurement & verification (M&V), data analytics, decision making and support systems [5 - 10]. One of these studies conducted by Campbell investigated the 12L feasibility of 47 industrial case studies [7]. An overview of the results is presented in the figure below.

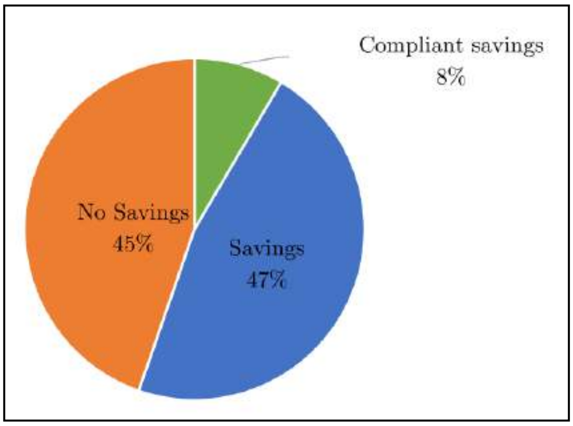


Figure 1-1: 12L feasibility evaluation results [7]

The research identified two critical challenges: 1) 47% of the case studies showed viable savings potential, but could not be claimed due to poor data compliance and quality, 2) 45% of the case studies failed to achieve and sustain target savings for the full year of assessment [7][10][11]. Ultimately only 8% of the assessed case studies managed to receive a 12L benefit. It is therefore clear that a need exists to understand and address the issues limiting viable claims.

The topics of data quality and the sustaining of EE savings formed the core discussion and motivation of the authors' research and subsequent methodology development. This paper presents a condensed version of the work performed [10]. The following sections will briefly discuss supporting literature before moving on to an overview of the systems approach developed to address the challenges. Finally results from several case studies are presented and discussed.

2. LITERATURE OVERVIEW

2.1 Introduction

The amount of data being generated by industries is set to double every two years [12]. The abundance of information potentially allows the impact of energy saving actions (ESAs) to be quantified at increasingly high levels of certainty [5]. However, the requirements stipulated for data used in 12L assessments is very strict. The availability of high quality, 12L compliant data is therefore critical for all potential claims. Section 2.2 will further discuss the role of data quality in M&V.

12L requires an application to show sustained savings for a full year of assessment [Regulations]. Unfortunately, research shows that the sustainability of energy efficiency projects over extended periods remains a challenge [11][13]. Failing to sustain the maximum project performance for at least a full year will affect claimant's ability to receive the full 12L benefit. Section 2.3 will further discuss research relating to sustainability, as well as techniques for corrective action. The following sections present condensed versions of a detailed literature study conducted by the author [10].

2.2 The role of data quality in M&V

All 12L applications are assessed based on the M&V process prescribed by the SANS50010. Essentially, this process requires a point of reference (baseline) that can be adjusted to objectively compare operational performance pre- and post-energy saving action (ESA) [3].

The figure below illustrates the concept of how the energy savings can be quantified using a baseline (based on pre-ESA data) adjusted for post-ESA conditions (adjusted baseline). The difference between the adjusted baseline -how the system "would have" operated- and the actual energy consumption, depicts the energy saving attributed to the relevant ESA.

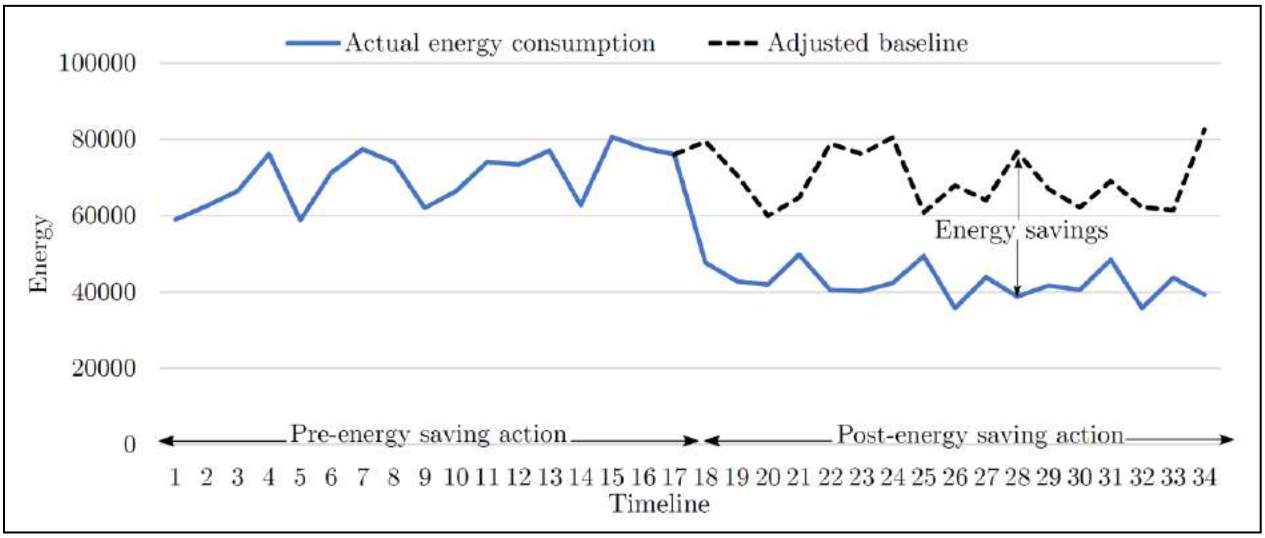


Figure 2-1: ESA general assessment approach [10]

There are four basic boundary options that can be selected when developing a representative baseline [3][14]. These options are a) key parameter, b) all parameter, c) whole facility and d) calibrated simulation. The usability of the different boundaries depends on the nature of the ESA and available points of measurement. Ideally option B will be selected to fully measure and monitor all parameters encapsulating a specific ESA. The Figure below illustrates the application of the different boundaries applied to an industrial water cooling system [10].

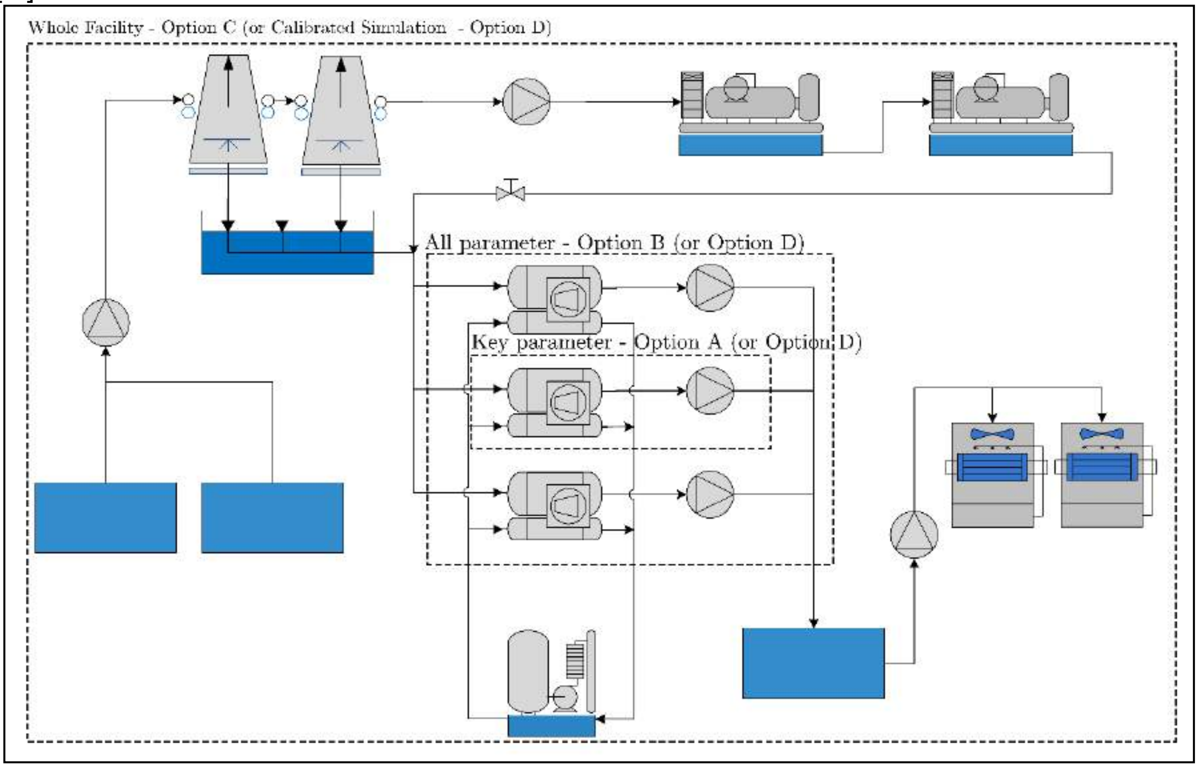


Figure 2-2: Layout of industrial water cooling system illustrating different boundaries [10]

The availability of multiple points of measurement combined with the different boundary selection options creates the possibility for several different forms of assessments [5 - 10]. The challenge for the M&V practitioner and other stakeholders is therefore to select an option that provides sufficient encapsulation and data for high-quality analyses (i.e. as many points as possible) and keep long-term compliance and sustainability in mind (i.e. using minimal points simplifies maintenance requirements).

The nature of data compliance and quality is not static. Unfortunately, a good data source can become a useless liability due to several factors. Data quality assessment is therefore critically important to ensure accurate results [5][9][11]. The figure below illustrates some common abnormalities that can occur due to faulty measurements or communication failures.

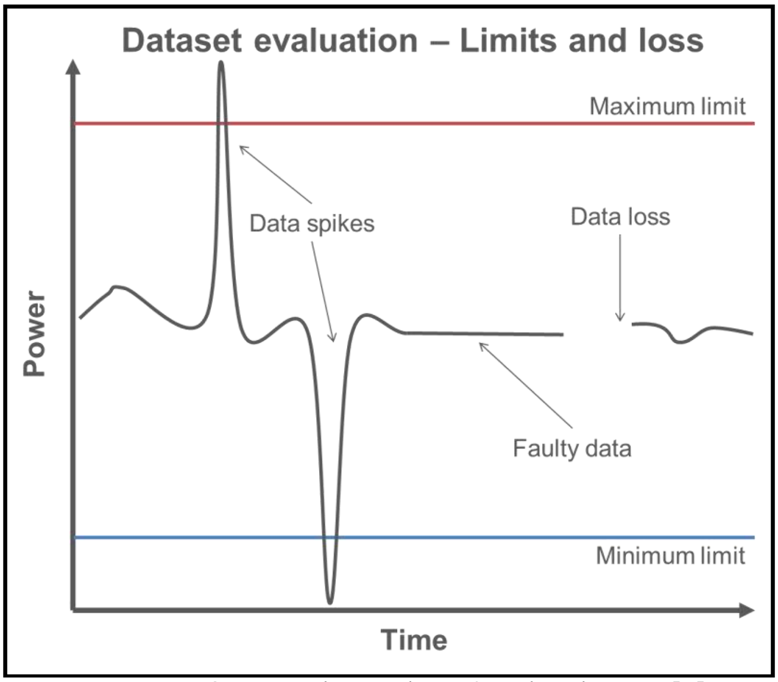


Figure 2-3: Common abnormalities found in datasets [9]

These abnormalities exhibit known characteristics which can be used to quickly identify them in a dataset. Unfortunately, meters can also experience a drift in accuracy over time which will not necessarily be clearly discernible and therefore require regular checks and calibrations. The figure below illustrates the use of a cumulative sum (CUSUM) graph which can be used to track to effect of variance on a measurement or the consistency in a project's performance [9]. A natural variance will result in a constant profile being maintained. However, a significant change will result in a visible departure from the normal trend.

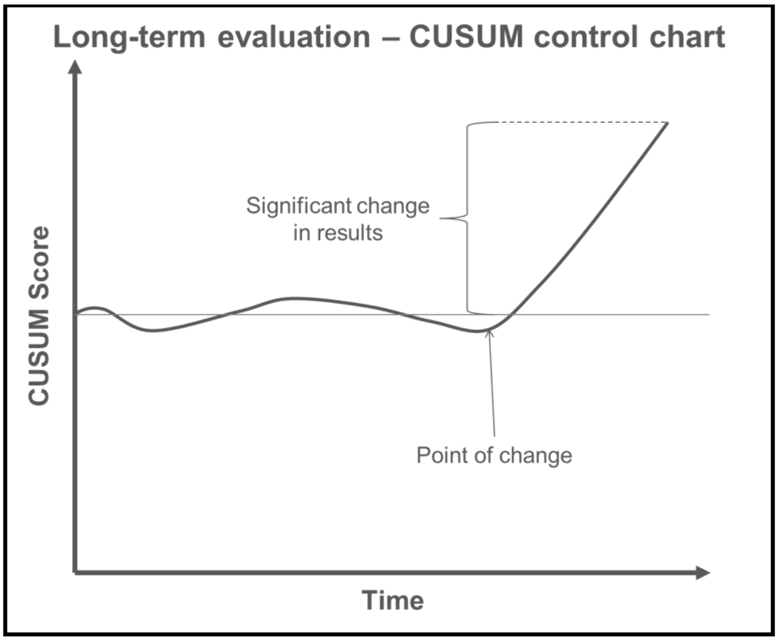


Figure 2-4: CUSUM graph illustrating a change in results [9]

The information briefly conveyed in this section supports the argument that data quality is fundamentally important, not only for 12L purposes, but also for general M&V work. Over time, the multiple combinations of boundaries and measurement points, together with varying compliance and quality lead to a significant amount of permutations that need to be assessed and monitored. These and other assessment techniques will be used as part of the methodology presented in the next section.

2.3 Sustainability of EE projects

Grobbelaar investigated the long-term sustainability of energy cost saving projects implemented by an energy services company (ESCo) [11][13]. The study found that projects generally performed well during the initial performance assessment (PA) period. The PA was typically conducted over three months under the strict supervision of the ESCo and client engineers. Unfortunately, the projects failed to sustain their initial performance assessment (PA). The failure was attributed to the lack of consistent supervision and poor data availability.

The figure below illustrates a case study's monthly performance [11][13]. The green line shows the average PA performance and can be used as an indicator of the projects "best case" performance. The red line shows the contractual target and can be used as an indicator of a "reasonable" performance. The actual monthly performance is shown as blue bars and clearly illustrates the project's failure to sustain either the PA or contractual target.

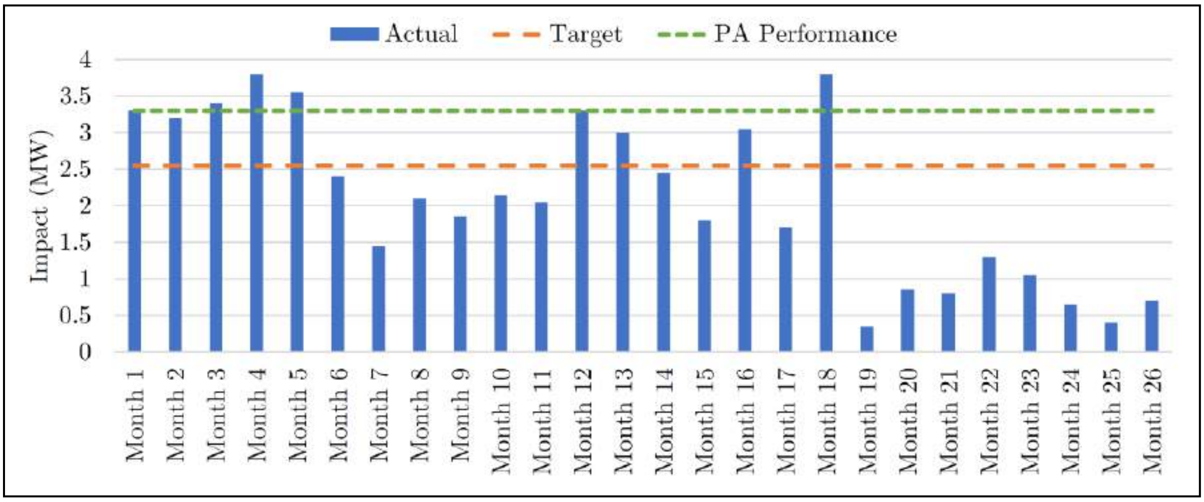


Figure 2-5: Monthly performance assessment results [10, 11]

Groenewald also evaluated the sustainability of five mining projects (similar to the case study presented above) [11]. The study quantified an overall monetary loss (linked to electricity costs) approaching R12.7 million over a 12-month period [13]. The significant financial losses, attributed to the lack of sustainable savings, prompted Groenewald to develop a performance centred maintenance strategy [11]. This strategy was based on the ISO 50001 plan-do-check-act (PDCA) cycle which is illustrated in the figure below.

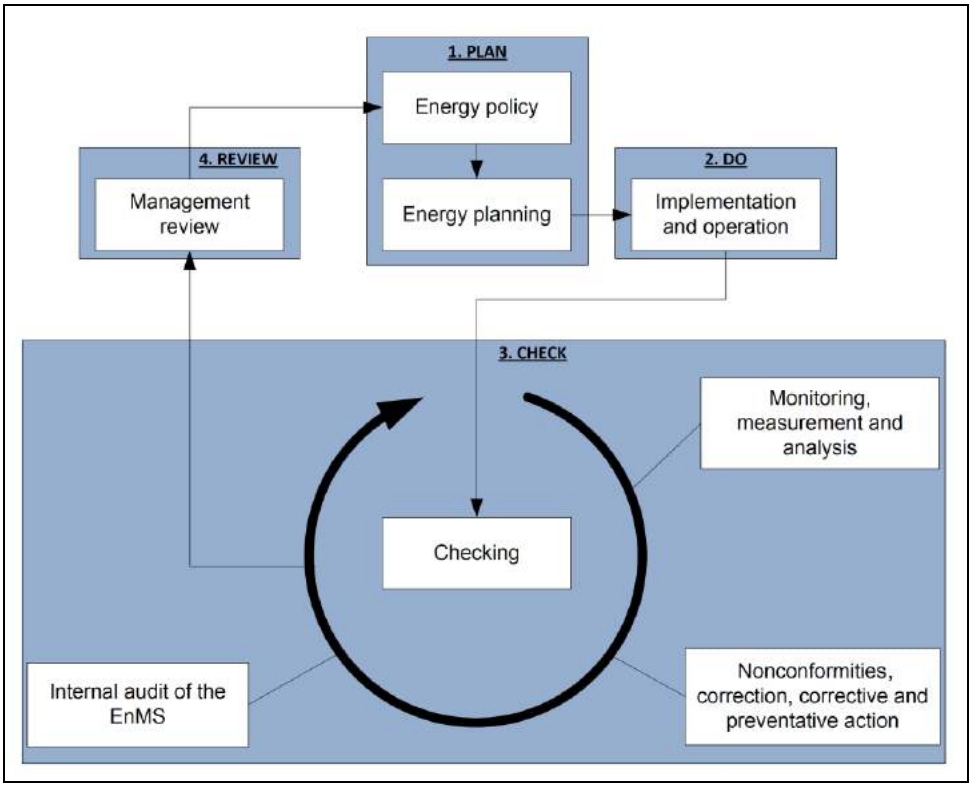


Figure 2-6: The applied ISO 50001 approach [10]

The application of the ISO 50001 approach was deemed to be beneficial and helped to sustain savings. However, the overall process was still constrained in terms of time and cost by the various "human factors" present throughout the process. The next challenge is therefore to balance the cost, time and resulting quality presented by the abovementioned solution. The figure below illustrates the different trade-offs between these three value metrics [5].

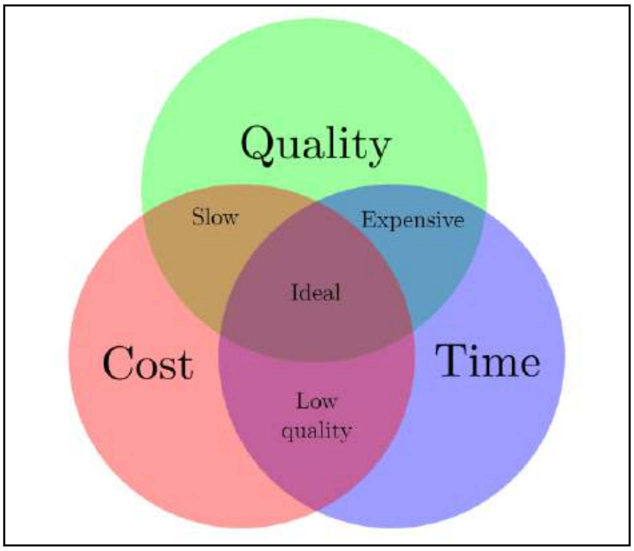


Figure 2-7: Balancing the three value metrics [10, 5]

From the figure it is clear that the ideal balance lies in the middle between cost, time and resulting quality. Fortunately, the availability of new software technology allows industry to significantly decrease costs, processing and review time without negatively affecting quality.

2.4 Summary

The previous sections briefly touched on examples and illustrations obtained from available literature. The findings show that data quality plays a critical role in overall M&V work. A new 12L linked challenge is to ensure that data compliance and quality is sustained over the two-year period (baseline and assessment) typically required for a 12L application. Furthermore, an additional problem beyond data quality is the sustainability of achieved savings. Fortunately, the literature also shows that several separate solutions have already been developed for these issues [10]. However, the challenge is to ensure that these solutions are implemented with the optimal balance between quality, cost and time in mind. The following section will discuss the development of a system that integrates these existing solutions in an efficient manner.

3. GENERAL METHODOLOGY

The previous sections support the need for the development of an intelligent, data driven system to effectively monitor and support the sustainability of data quality and project performance. This section will give an overview of the methodology approach followed to address the need. The methodology is based on ISO 50001 PDCA principles and focuses on three key areas, namely; 1) Measure, 2) Monitor and 3) Analyse.

The methodology is designed on the premise that it will be incorporated into a software system and repeated on a continued basis (e.g. daily). Each key area (measure, monitor and analyse) will request specific user action only when needed (i.e. inputs or corrective action) before flowing to the next step. This approach will ensure that the system can autonomously operate as long as the specified criteria are met. Each focus area will now be discussed in more detail.

3.1 Measure

The first key area of the developed system is the collection of relevant measurement data. The architecture of the system is designed so that all raw data is centralised before being processed. The basic flow of data from point of measurement to the central server is illustrated in the figure below.

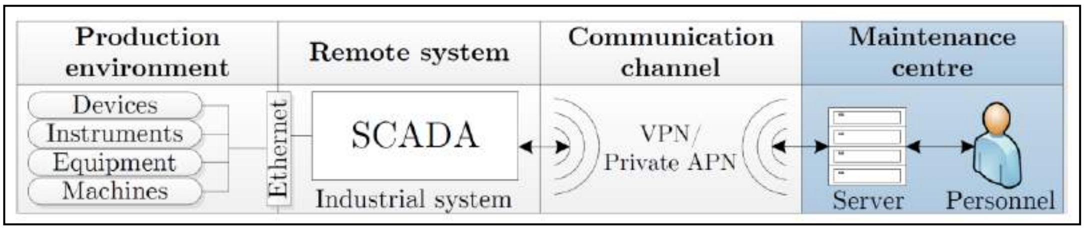


Figure 3-1: Communication of data from site to server [10]

In addition to the transfer of data from site to server, the data also needs to be contextualised before it can be effectively processed. The contextualising process (e.g. data type, unit of measure, description etc.) is done for each point of measurement and is further enhanced by completing an additional checklist. The checklist serves as a tool to gather extra supporting information such calibration dates, expiry dates, certificates etc. Finally, all the relevant components can be combined using a processing script, producing a data file for further analysis. The process is illustrated in the figure below.

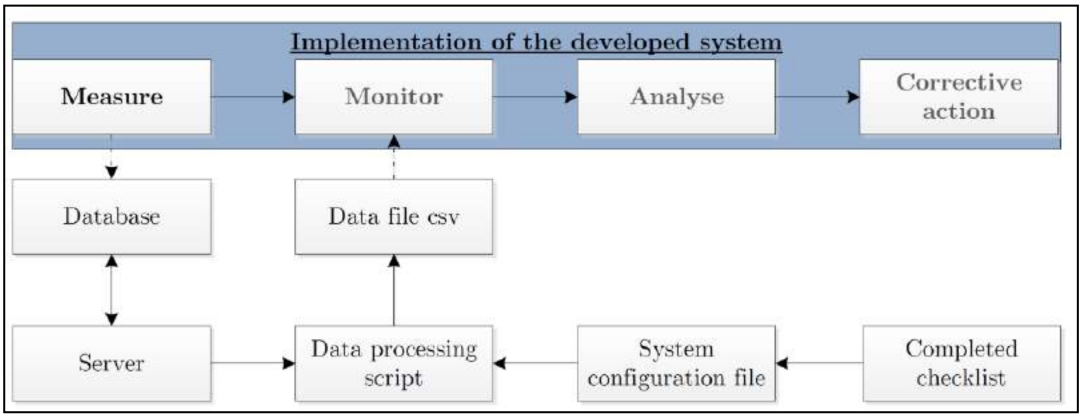


Figure 3-2: Methodology - Flow of data from Measure to Monitor phases [10]

Figure 3-2 shows how the measured data obtained from the production environment and then centralised on a server (as illustrated in **Figure 3-1**) is combined with the user generated configuration (contextualised) data and finally sent to the Monitor phase as an integrated data file for further processing. The system configuration file is created when the system us setup for the first time, thereafter it only requires updates when system components / characteristics change.

3.2 Monitor

The next phase is to monitor the collected data. Here the tools discussed in Section 2.2 were incorporated into the digital system to quickly assess all measurements based on selected sets of criteria. If the monitored data passes the evaluation process it will move on to the Analyse phase. If not, an exception will be generated for corrective action. The figure below illustrates three basic test criteria (min/max, hanging and missing) that can be applied to all incoming data streams. The criteria can also be further refined to assess and identify more specific data anomalies.

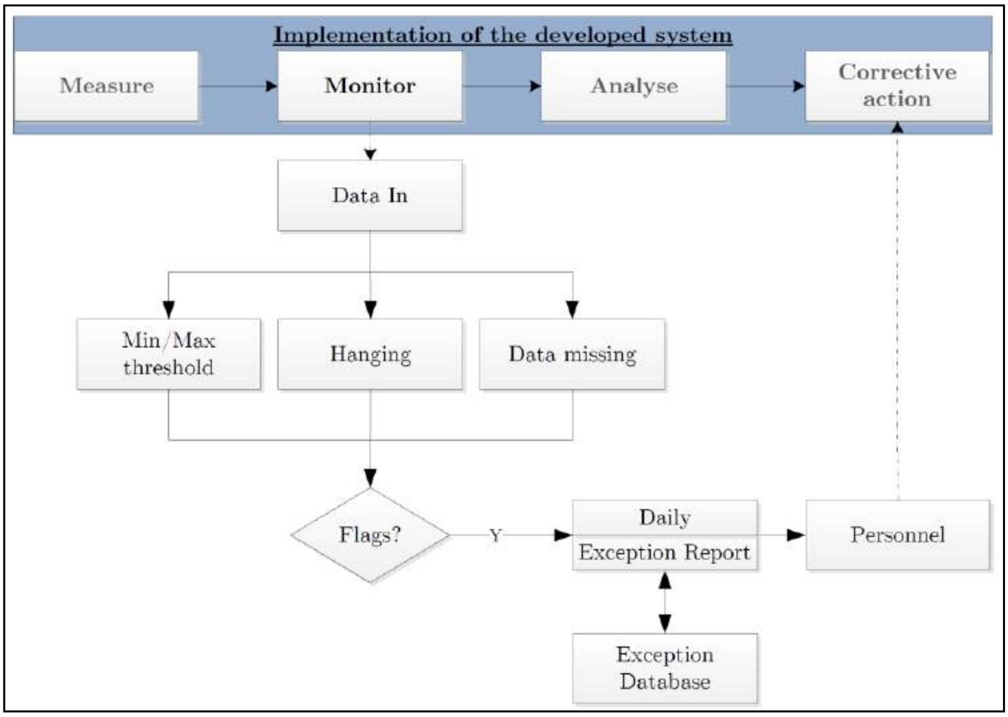


Figure 3-3: Methodology – Monitor phase [10]

A flag is created when the selected test criteria identifies a potential abnormality. The generation of a flag will also result in the generation of a daily exception report (summarising all the relevant flags for the specific day). The exceptions will furthermore be added to an "Exception Database" for later follow-up notifications.

The goal of the exception driven reporting is to selectively request human interaction only in the event of an abnormality occurring. This greatly reduces the amount of man hours previously required to monitor and assess all measured streams and allows operators more time to focus on the problematic areas.

This step ensures that only high-quality datasets are used for subsequent analyses (final step). This is achieved by using the test criteria as a filter, passing quality data and flagging potentially faulty data. The corrective action step allows for users to correct the faulty data, or at least assess the source of the fault so that future data will be useable.

3.3 Analyse

The final step in the methodology is to analyse the performance of the specific focus area / intervention. This step produces several outputs for corrective action and review in the form of performance tracking - and exception-reports. Ideally the analysis delivers a savings report that can be distributed to interested parties for review. Corrective action can then be implemented at user discretion. However, to ensure that any underperformance / abnormality is quickly identified two additional checks are implemented.

The example process illustrated in the figure below combines relevant measurements using specific, predetermined calculations (e.g. methods discussed in section 2.2 and 2.3) to identify abnormalities. The complexity of this calculation depends on the complexity of the underlying system and selected analyses. Regardless, the use of an automatic performance assessment produces outputs that can prompt users to perform corrective action when needed.

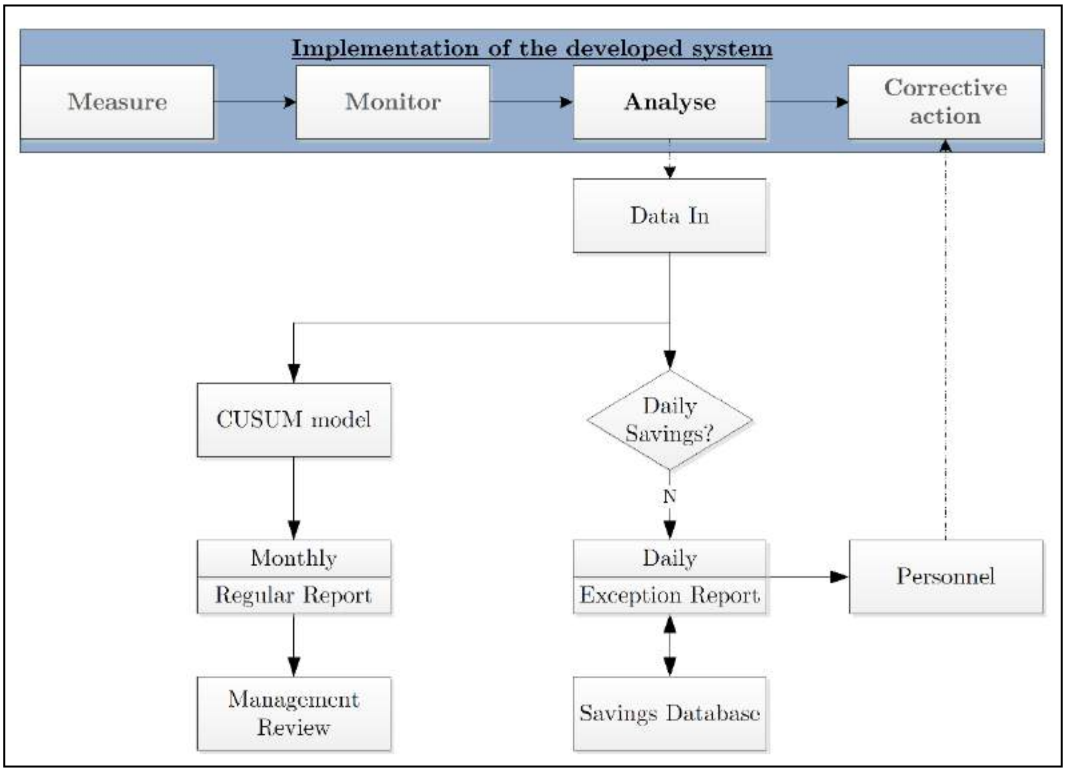


Figure 3-4: Methodology – Analyse phase [10]

The results from this assessment are reviewed using two separate approaches. The first merely tests whether the project has achieved its pre-set daily target. An exception report is generated when it fails to achieve the target. This report (similar to previous flags) will be sent to the relevant personnel for corrective action.
The combined results are also included in a CUSUM assessment model. The goal of this assessment model is to review general project performance over an extended period. Various factors can impact daily performance, potentially creating erratic noise and thereby clouding the ability to do a holistic assessment. The CUSUM model dampens these erratic events, producing a general trend illustrating overall project performance. A significant change in the slope of the trend can be used to indicate recurring changes in project performance. The assessment and interpretation of the CUSUM tool requires a management level review. It is therefore included in a regular monthly feedback report.

3.4 Summary

The general methodology was developed to address two key concerns, data quality and sustained savings. The approach focused on centralising and integrating site data for automatic, predetermined calculations. Human intervention is only requested when exceptional events occur. This automation significantly reduces the amount of human time required for general assessments, instead allowing the time to be spent focussed on understanding and solving the flagged problems. The next section will illustrate the working of the methodology on industrial case studies.

4. DISCUSSION OF RESULTS

4.1 Overview

The developed methodology was tested by applying it in the mining and metal processing industries. All identifying details have been removed to honour confidentiality agreements signed with the various parties. The results presented in the next section will only showcase selected examples from the case studies. This is done with the focus on brevity and with the goal to illustrate basic functionality. More detailed descriptions and discussions of each case study is available in the author's full work [10]

4.2 Case studies

The first step in the methodology is to collect and contextualise the relevant measurements. The table below illustrates an example of the general information captured as part of the initial checklist / setup process.

Table 4-1: Case study result - Illustration of an "Input Checklist" [10]

Input Checklist	
System	Case study C site
Components	Facility
Boundary Selection	Whole facility
M&V Model	Intensity
Energy Driver	Production [T]
Driver Resolution	Daily
Energy Carrier	Facility electricity usage [kWh] Reductants used [kWh]
Carrier Resolution	Daily
Maximum Capacity	Production: 400 T Electricity: 100 MW Reductants: 1 000 MW
Minimum Capacity	Production: 1650 T Electricity: 5 880 MW Reductants: 9 700 MW
Compliance	Eskom Invoices
Target savings	130 GWh
Implementation start date	N/A
Baseline start date	1 January 2014
Assessment start date	1 January 2015
Corrective action personal	N/A
Project manager	N/A

The next step is to monitor and asses the quality of the various streams. This was done using the following basic test criteria: missing, hanging, below minimum, or above maximum. The figure below illustrates the aggregate "Error count" for the various events each day. A quick assessment of the results clearly shows that there was a significant increase in errors in August 2014.

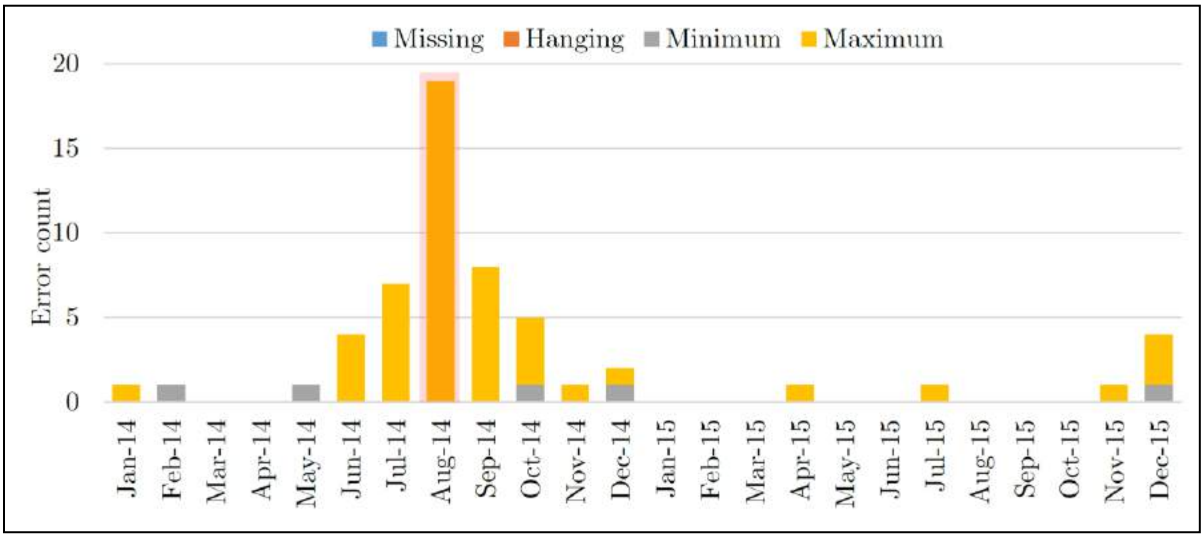


Figure 4-1: Case study result - Monthly aggregated error count [10]

The figure above focused on the overall error counts, but the same process can also be applied to individual points of measure over the total aggregate of available data. The figure below illustrates the total error count for several meters. It is clear that the "Air Flow" meter has significant issues with missing data. Presenting results in this format enables the relevant persons to focus their attention on the primary, and most significant causes of flags.

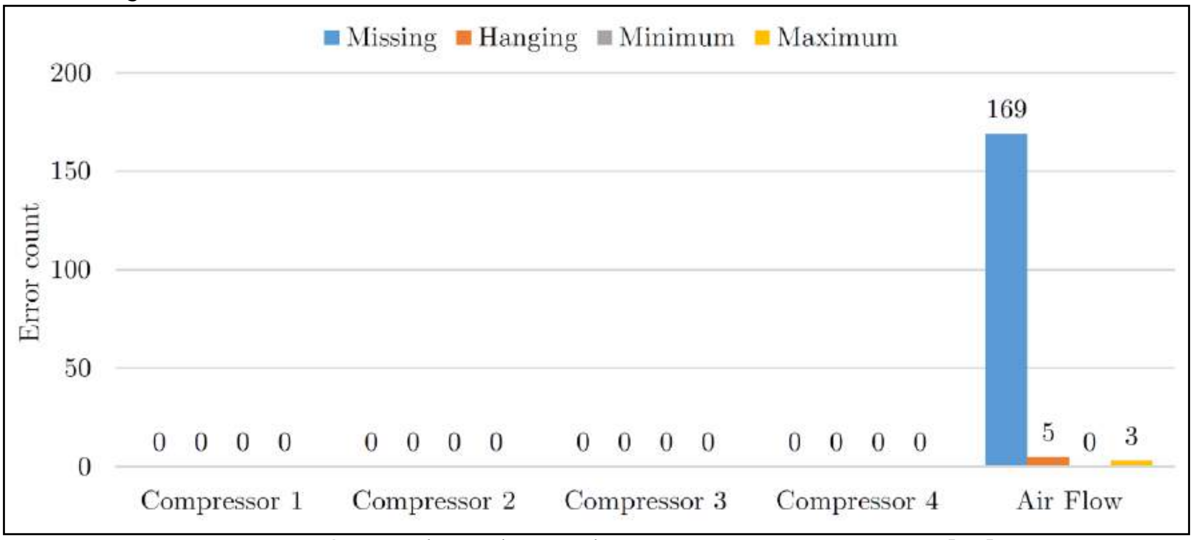


Figure 4-2: Case study result – Total error count per measurement [10]

The analyse step in the methodology focuses on evaluating the performance of the project with the intent of sustaining savings. The figure below illustrates the savings assessment for a specific project.

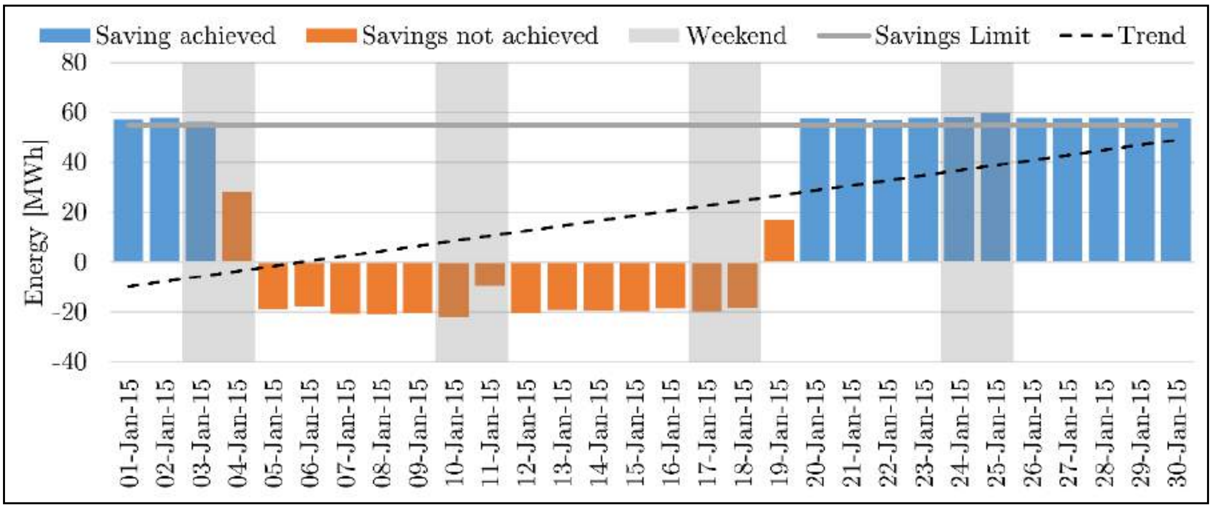


Figure 4-3: Case study result – Daily savings [10]

Assessing the daily results shows that the project initially achieved its target, but started to underperform on the 4th of January. At this point the relevant personnel will have been alerted to the underperformance. However, it is clear that the project continued to underperform for a full two weeks before the issue could be corrected.

The final step in the process is to assess the overall performance of the project. The figure below shows the baseline period (blue), followed by the assessment period (orange). The example was selected because it clearly illustrates the change in trend in two areas between the baseline and assessment (i.e. when the project started working) and post August 2015 when data loss occurred (red block).

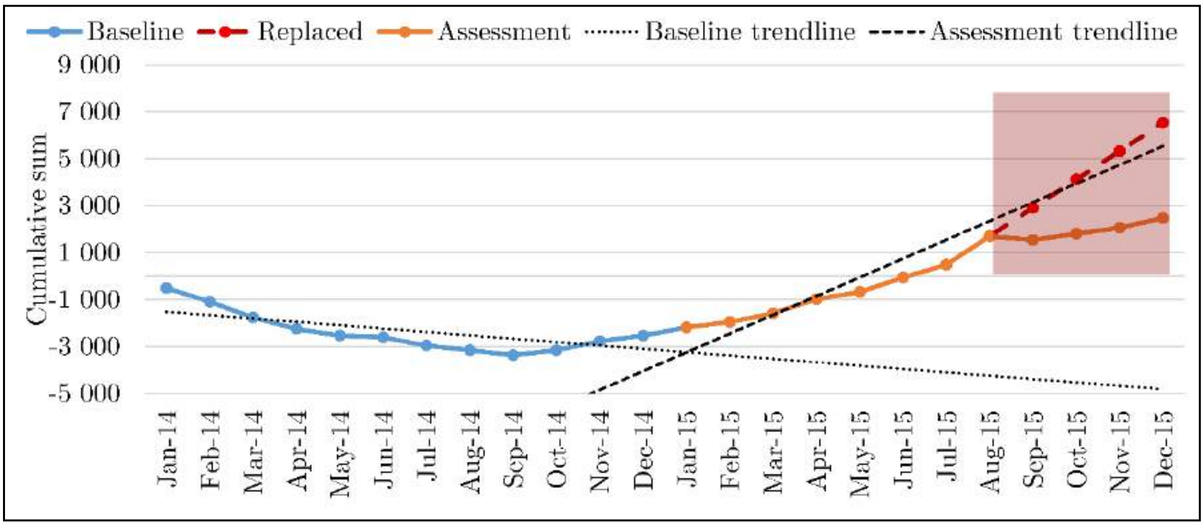


Figure 4-4: Case study result – CUSUM trend [10]

The figure (above) also shows the effect of correcting the data loss (orange vs. red line). The general trend of the graph (illustrated by black dotted lines) gives a point of reference on project performance. If the graph follows the general trend; it indicates that performance is sustained. If the graph angles above the line, it indicates an increase in performance and vice versa where dipping below the line indicates a decrease.

4.3 Potential value

This paper started with an excerpt from Campbell's feasibility study on 47 potential 12L claims [7]. The study highlighted two main challenges / reasons for failure: the sustainability of savings and poor data quality. The table below gives a breakdown of the assessed case studies.

Table 4-2: Case study result - Illustration of an "Input Checklist" [10]

Industry	Total number of cases	Reason for failure	
		Lack of savings	Poor data integrity
Cement industry	9	8	1
Platinum industry	6	1	5
Gold industry	23	8	15
Steel industry	5	2	1
Chrome industry	3	2	0
Petrochemical industry	1	0	0
Total	**47**	**21**	**22**

The case studies presented in section 4.2 highlight some of the outputs from the author's primary research [10]. Based on these case studies, the estimated avoidable energy efficiency savings loss (per project) was between one and two GWh per annum. If these results are extrapolated back to the case studies presented in the table above the estimated energy efficiency savings lost exceeds 60 GWh, or R60-million in 12L tax certificate equivalent value (roughly R20-million in actual tax savings). Furthermore, the additional energy cost impact was estimated at R43-million per annum, based on Eskom TOU tariffs [10]. The quantification figures are a rough estimate but still clearly illustrate the potential benefits.

5. CONCLUSION

The Section 12L energy efficiency tax incentive presents significant value to industry. Previous research highlighted two major challenges, namely; data quality and the sustainability of savings. This paper gave a basic overview of literature-based techniques and result in order to convey the relevant background and context.

The presented methodology strives to follow a systems approach to integrate the various solutions obtained from literature. The final solution adds additional value by limiting the need for human interaction. This is achieved by flagging selected events thereby focusing attention where it is required.

The methodology was applied to three industrial case studies. Selected results illustrate examples of graphical outputs to showcase the functionality of the system. The potential value of the solution is estimated by extrapolating the avoidable losses (energy savings) quantified by the case studies back to the 47 case studies originally evaluated by Campbell. The resulting potential exceeds R60-million in 12L certificate value plus an additional R43-million in electricity costs. Ultimately the solution shows that there are definite benefits to using new technology together with exiting research to improve on current challenges.

6. BIBLIOGRAPHY

[1] **Minister of Finance**. 2015. Draft Carbon Tax Bill, *National Treasury*, Cape Town, pp. 1-43.

[2] **National Treasury**. 2013, Regulations in terms of section 12L of the South African Income Tax Act, 1962, on the allowance for energy efficiency savings, *Government Gazette*, vol. 37136, no. R.971, pp. 3-10.

[3] **SABS Standards Division**. 2011. SANS 50010: 2011 Measurement and Verification of energy savings, version 1.

[4] **South African National Energy Development Institute**. 2008. National Energy Act, Act No. 34 of 2008, Republic of South Africa

[5] **W. Hamer**. 2014. A practical approach to quantify RSA Section 12L EE tax incentives for large industry, *PhD dissertation*, North-West University.

[6] **L. Botes. 2017.** Objective evaluation of industrial energy efficiency models for the RSA Section 12L tax incentive, *Masters thesis*, North-West University.

[7] **K. Campbell**. 2016. Evaluating the feasibility of the 12l tax incentive for energy intensive industries, *Masters thesis*, North-West University.

[8] **M. J. van Rensburg**. 2015. Structuring mining data for RSA section 12L EE tax incentives, *Masters thesis*, North-West University.

[9] **W. Booysen**. 2014. Measurement and verification of industrial DSM projects, *PhD dissertation*, North-West University.

[10] **L. Coetzee**. 2017. A systems approach to sustain RSA Section 12L tax incentive projects, *Masters thesis*, North-West University.

[11] **H. Groenewald**. 2015. A performance-centred maintenance strategy for industrial DSM projects, *PhD dissertation*, North-West University.

[12] **V. Turner, J. F. Gantz, D. Reinsel,** and **S. Minton**. 2014. The digital universe of opportunities: Rich data and the increasing value of the internet of things, *IDC Analyze the Future*, EMC Framingham, United States of America.

[13] **H. Grobbelaar**. 2014. Maintenance procedures on DSM pumping projects to improve sustainability, *Masters thesis*, North-West University.

[14] **Efficiency Valuation Organisation**. 2012. International performance measurement and verification protocol. concepts and options for determining energy and water savings, *Efficiency Valuation Organisation*, Toronto.

[15] **South African Bureau of Standards**. 2011. South African National Standard: ISO 50001, Energy management systems requirements with guidance for use, South Africa, version 1, Jun. 2011, pp. 1-22.

A REAL-TIME SCHEDULING SYSTEM FOR A SENSORISED JOB SHOP USING CLOUD-BASED SIMULATION WITH MOBILE DEVICE ACCESS

S. Snyman[1*], J. Bekker[2] & J.P. Botha[3]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
snymans@sun.ac.za

[2]Department of Industrial Engineering
Stellenbosch University, South Africa
jb2@sun.ac.za

[3]Department of Electrical and Electronic Engineering
Stellenbosch University, South Africa
jp17619084@gmail.com

ABSTRACT

Scheduling is a challenge that continues to trouble management of the operational phase of the manufacturing life cycle and can be attributed to the complex, dynamic and stochastic nature of a manufacturing system. Computer simulation is often used to assist with scheduling, as it can sufficiently mimic complex, discrete, dynamic, stochastic processes. We propose and develop a prototype real-time simulation scheduling system for a sensorised factory, which is to serve as a decision support tool for real-time rescheduling of machine steps in a job shop.

---

[1]The author was enrolled for an MEng(Industrial) degree in the Department Industrial Engineering, Stellenbosch University, South Africa.
[3]The author was enrolled for an MEng(Electrical and Electronic) degree in the Department Electrical and Electronic Engineering, Stellenbosch University, South Africa.
*Corresponding author

1. INTRODUCTION

Manufacturing is defined by Groover [1] as the transformation of materials into items of greater value by one or more processing and/or assembly operations. The transformation is accomplished by combining machinery, tools, power, and manual labour. From this description of manufacturing, the process can be considered a complex engineering endeavour, where the coordination of people, material, equipment, and information to accomplish a manufacturing goal demands considerable time and effort.

The manufacturing system life cycle can be divided into several phases which include the design, planning, implementation, operation and termination phases [2]. The coordination challenges faced in the design and implementation phases can be overcome by careful planning; however, such challenges continue to persist over the operational phase. For many manufacturing systems, scheduling is such a challenge, which can be attributed to the complex, dynamic, and stochastic environments exhibited by these systems. Many technological advances, including cloud-based computing and the improved capabilities of sensor networks, have offered the opportunity of designing a real-time scheduling system that can overcome the scheduling challenges, as well as the opportunity for the creation of software architectures to support these real-time scheduling systems.

This paper forms part of a research project with the objective to develop a prototype real-time simulation scheduler of a sensorised job shop, which is to serve as a decision support tool so that unexpected disturbances in the shop can be overcome by the generation of new schedules with real-time data from the shop. This paper was preceded by [3], which provided the relevant literature and initial architecture for the proposed solution. This paper will focus on the development and implementation of the proposed solution, as well as the testing and validation of the solution.

We subsequently provide a brief summary of the literature, as documented in [3], in Section 2, which is followed by a revised architecture of the proposed scheduler in Section 3. The development and implementation of the proposed solution will be described in Section 4, while the validation and operational testing will be discussed in Sections 5 and 6. Finally, the conclusion of the research project will be discussed in Section 7.

2. LITERATURE

The literature for this project was comprehensively discussed in [3], therefore only the most important literature will be reiterated. When considering scheduling in a job shop environment, there are a number of factors that need to be taken into account. These factors are described by [4] as:

- If more work is accepted per day than the organisation can complete per day, then the overall work in progress inventories will increase, causing shop congestion, an erosion of the firm's output rate, and a lengthening of job completion times.
- If completion times or dates are promised to customers, then estimates of lead times for each job must be determined, and jobs must be started early enough to complete the job by the promised date.
- Facilities can finish more jobs per period and satisfy more customers if they work on the shortest jobs first. However, longer jobs will ultimately be completed late or behind schedule.
- More highly valued customers may require earlier completion dates which will then be given processing priority in the shop. This will in turn make it more difficult to estimate accurate completion dates for other jobs.
- Interruptions such as machine breakdowns, employee absence, poor raw-material quality, and processing errors, can cause unforeseen delays in processing.

To address these factors, machine-level controlling measures, such as sequencing, dispatching rules and performance indicators, must be in place. The dispatch rules guide the production sequence of the jobs within the shop, and ensure that operators know the sequence in which the jobs must be processed. Dispatching rules, as mentioned by [4] and [5], include:

1. shortest process time,
2. first-come-first-served,
3. most-important-job-first,
4. earliest due date,
5. critical ratio, and
6. minimum slack time per operation.

The decision of which dispatching rule should be implemented is dictated by the job shop performance indicators that need to be optimised. The most common performance indicators where identified by [4] and [5] as:

- Average flow time, where the flow time begins when a job arrives at the shop and ends when it leaves. This flow time is averaged over a number of jobs.

- Average queue time, where the queue time is the total flow time minus the process time of the job. The queue time is then averaged over a number of jobs.
- Average job lateness, where lateness is the difference between the completion date and the due date. The lateness is averaged over a number of jobs.
- Average job tardiness, where tardiness is the amount of time a job finishes beyond the due date. Tardiness is averaged over a number of jobs.
- Makespan, where makespan is the total elapsed time to complete a number of jobs.

These dispatching rules and performance indicators are then used to define the job shop scheduling problem. The problem was previously mathematically defined and illustrated and therefore the reader is referred to [3]. This concludes the summary of the literature, and the following section will present the revised system architecture.

3. ARCHITECTURAL DESIGN OF THE PROPOSED SIMULATION SCHEDULER

We explain the concept using an informal schematic diagram, followed by a formal architectural description.

3.1 System overview

Figure 1 shows the system overview of all the components and how they interact with each other. The overview illustrates that there is a shop floor where different machines are installed. At each machine, a sensor is used for logging operation status changes. The sensors can log one of three possible states (*i.e. Waiting*, *Processing* or *Completed*). There is also a RFID card linked to a specific job, which can be swiped by the operator of the machine when a status change needs to occur. The state is then transmitted to a gateway which in turn transmits the state to the cloud-based information system. The scheduler (*i.e. Tecnomatix Plant Simulation*), also located in the cloud, consumes data from the information system and generates new schedules. The web interface can also use and log data on the information system. Finally, there is a mobile device that can access both the scheduler and the web interface.

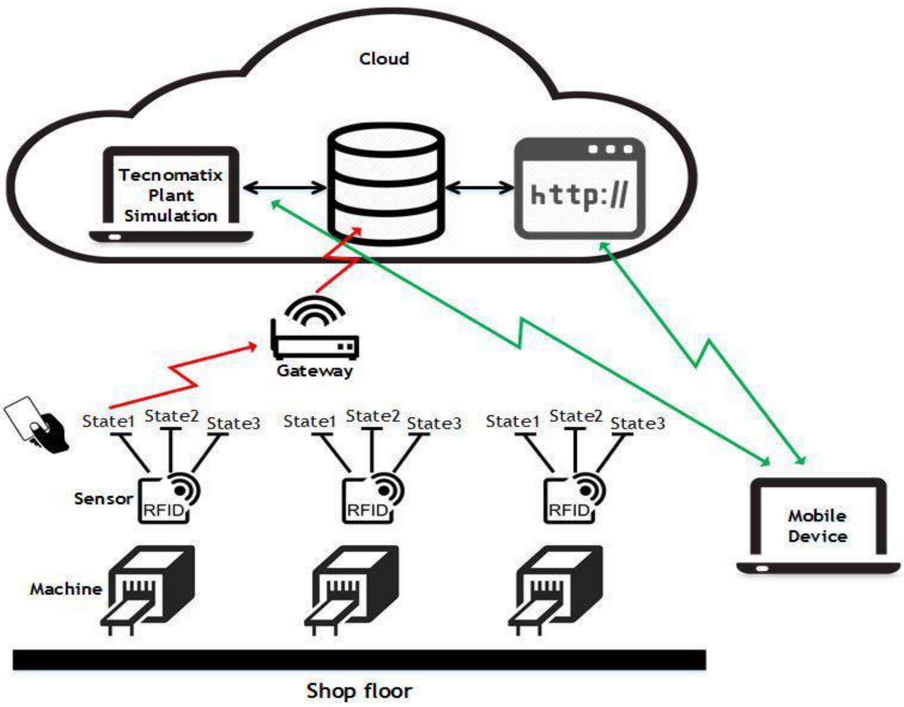


Figure 1: System overview

The initial architecture for the proposed scheduler was developed by [3], however this architecture was revised as illustrated in Figure 2. The architecture is described using the Object-Process Methodology (OPM) [6].

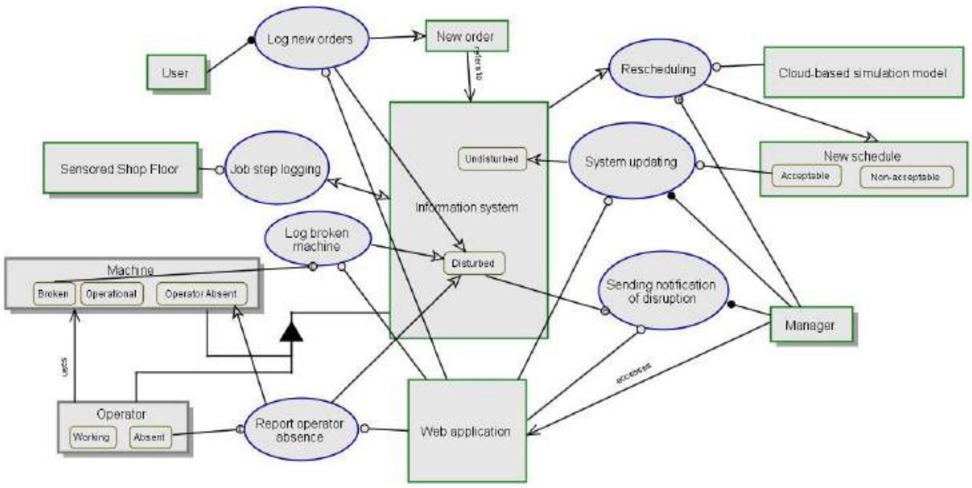


Figure 2: Top-level architecture of proposed scheduler (revised from [3]).

The semantical representation of the revised architecture is as follows:

User is physical.
User handles Log new orders.
New order refers to Information system.
Information system can be Undisturbed by default or Disturbed.
Information system consists of many Machines and many Operators.
  Machine is physical.
  Machine can be Operational by default, Broken, or Operator absent.
  Operator is physical.
  Operator can be Working by default or Absent.
  Operator uses Machine.
Information system triggers Sending notification of disruption when it enters Disturbed.
Sensored shop floor is physical.
Manager is physical.
Manager accesses Web application.
Manager handles System updating and Sending notification of disruption.
Manager triggers Rescheduling.
New schedule can be Acceptable or Non-acceptable.
Log new orders requires Web application.
Log new orders yields Disturbed Information system and New order.
Log broken machine occurs if Machine is Broken.
Log broken machine requires Web application.
Log broken machine yields Disturbed Information system.
Report operator absence occurs if Operator is Absent.
Report operator absence requires Web application.
Report operator absence yields Disturbed Information system and Operator absent Machine.
Sending notification of disruption requires Web application and Disturbed Information system.
Rescheduling requires Cloud-based simulation model and Manager.
Rescheduling consumes Information system.
Rescheduling yields New schedule.
System updating requires Acceptable New schedule and Web application.
System updating yields Undisturbed Information system.
Job step logging requires Sensored shop floor.
Job step logging affects Information system

4. DEVELOPMENT AND IMPLEMENTATION

This section will describe the development and implementation of the different components of the proposed scheduling system. Reference will be made to the integration of the different hardware components into a sensor prototype, as well as the software components which will include the information flow through the sensor and gateway, the information system, the web interface and the scheduler. Finally, the section will contain a system overview that will graphically illustrate the functionality of the proposed scheduler.

4.1 Hardware development and implementation

In this section the hardware components used to develop the sensor will be discussed with a complete explanation of the sensor integration.

4.1.1 Components

The main components that form part of the sensors and the additional components used to complete the hardware setup are discussed.

Raspberry Pi

The computer used to capture the sensor data and send it to the cloud is a Raspberry Pi, also referred to as the *system gateway*. This computer is used because it is well known and thus has sufficient software support. The Raspberry Pi runs on a *Debian* platform that forms part of the *Linux* operating system (OS) [7]. The *Linux* OS environment integrates seamlessly with the hardware of the Raspberry Pi through Python. The *serial peripheral interface* (SPI) bus of the Raspberry Pi is used to communicate to the radio frequency transceiver, RFM98. The RFM98 is discussed later in this section.

Arduino Pro Mini

The Arduino Pro Mini consists mainly of the Atmel Atmega328P micro-controller [8]. The operating voltage is 3.3V, which is ideal for this low-power sensor due to the use of only 3.3V components. The extremely low cost of this micro-controller with its functionality that supports the sensor outcomes made it the ideal choice [9]. The Arduino serves as the controller of the sensor, which captures the data from the RFID module and buttons, and sends this data to the RFM98 module. The RFM98 and RFID modules communicate via the SPI bus lines to the Arduino. The buttons trigger *general purpose input output* (GPIO) pins on the Arduino for user interface.

LoRa (RFM98)

The RFM98 module is a breakout board that holds the Semtech LoRa *integrated circuit* (IC). This module is used due to its compact design and design support. LoRa, short for Long Range, is a long-range, low-power radio frequency communication platform which is at the forefront of technology in network communication [10]. The confirmed line of sight communication distance is between 15 and 20 km [11]. It makes use of the low-frequency bandwidth that is part of the unlicensed radio spectrum, thus it is free to use, which holds great advantages. One disadvantage with this technology is that due to its low frequency the communication speed is limited [12]. However, this application makes use of low data rates which makes the RFM98 module the ideal choice. The module is used with the Raspberry Pi and Arduino Pro Mini to establish the communication network.

RFID (RC522)

The RFID module that is used in this application is the RC522 module. It identifies and communicates card or tag ID data to the Arduino via the SPI bus. In this application passive cards and tags are used as they are the cheapest and do not require a battery. The cards and tags utilise the radio energy that is transmitted by the reader to energise its circuit and to provide its ID [13].

Programmer

In this application an external programmer is used as it reduces the cost of each sensor [8]. The programmer consists of the FTDI FT232 UST to serial IC. The programmer is used to program the Arduino with its application specific functionality. The programmer makes use of the serial connection on the Arduino to write the developed code onto the Atmel micro-controller.

Power Supply

The sensors were developed to be seamlessly integrated into the proposed system by means of battery power. To make this possible and user-friendly a charging module and 18650 Lithium-Ion battery are added to the sensors. The battery has a capacity of 2600 mAh and is widely available [14]. A typical power bank charging module is used with the battery. It enables the user to easily charge the sensor through USB.

4.1.2 Gateway design

The gateway consists of the Raspberry Pi and a RFM98 module. A small *printed circuit board* (PCB) breakout board for the RFM98 was designed to connect directly onto the Raspberry Pi. The software used to design this PCB was *Altium Designer*. This was done to eliminate prototype wires and to ensure the stability of the gateway with good component connections. Figure 3 illustrates the PCB with numbers and descriptions as follows:

1. The connection pins that connect to the Raspberry Pi to the LoRa module to receive the data from the module through the SPI bus.
2. The LoRa module that receives the data from the sensors.

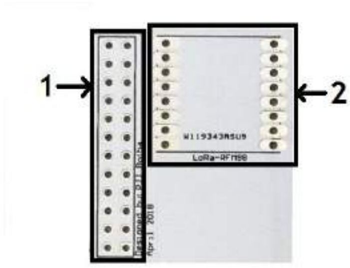


Figure 3: Designed PCB for gateway

4.1.3 Sensor design

The sensor consists of the Arduino Pro Mini, LoRa, RFID and the power supply modules. A PCB was designed to connect all of these components. The PCB serves as the sensor component integration that also has the advantage of good connections, which provide sensor stability. Figure 4 illustrates the front and back of the developed PCB with numbers and descriptions. The components are matched to the numbers as follows:

1. The RFID that reads the ID on the card or tag and sends this data to the Arduino through the SPI bus.
2. The three buttons that enable the user to define the status (*i.e. Waiting*, *Processing*, *Completed*) of an operation.
3. The charging module of the power supply that supplies power to the sensor and charges the Lithium-Ion battery.
4. The LoRa module that handles all the communication through low-power and low-frequency communication.
5. The Arduino Pro Mini that serves as the master of all the connected components.

Figure 4: Front and back view of the developed PCB for sensor

A casing was also designed in *Autodesk Inventor*, as illustrated in Figure 5, to enclose the sensor and all its components. The casing holds the battery safely underneath the sensor. Push buttons are designed to enable the easy use of the input buttons on the sensor within the case. The casing is 3D printed.

Figure 5: CAD design of sensor casing

4.2 Software development and implementation

The development of the software components of the proposed scheduler will now be discussed.

4.2.1 *Information flow in sensor and gateway*

The information flow through the sensor is illustrated by the flow diagram in Figure 6. The information flow starts off when the operator presses the button which indicates the new status and processes the ID of the button to the Arduino. The operator can then swipe the RFID card linked to the job that is then read by the RFID reader which processes the ID of the card to the Arduino. Only when both the button ID and card ID are received by the Arduino can a data package be created. The data package is then communicated to the LoRa module, which will transmit the package to the gateway. The data package will consist of a string of numbers that includes the sensor ID, card ID and button ID, where the sensor ID corresponds to the machine ID it is located at.

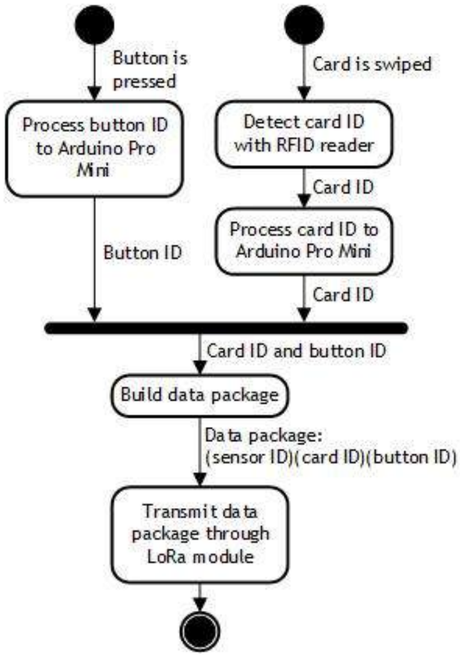


Figure 6: Information flow through sensor

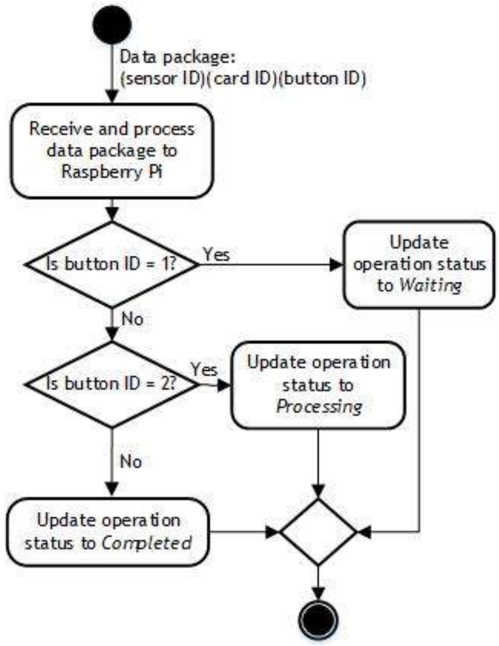


Figure 7: Information flow through gateway

The information flow through the gateway is also illustrated through the flow diagram in Figure 7 and starts off when the LoRa module receives the data package from the sensor. This package is processed by the LoRa module to the Raspberry Pi which then logs the data change in the cloud-based information system. The Raspberry Pi performs this data change at the operation that is associated with the machine ID and card ID provided in the data package. If the button ID = “1”, the status of the operation will be updated to *Waiting;* while if the button ID = “2”, the status of the operation will be updated to *Processing*. If neither of these conditions is met, the button ID must be “3” and the status of the operation will be updated to *Completed*.

4.2.2 *Information system and web interface design*

The information system is the database where all data from the job shop is recorded and stored. It is necessary to track the status of the shop floor, as it is required by the digital twin (the simulation model) when scheduling the existing set of jobs in progress. Due to the system requirement which stipulates that the information system must be cloud-based, a cloud server needed to be created. *Google Cloud Platform* (GCP) was chosen as a cloud-computing service provider which provided sufficient functionality required by the authors, and due to GCP making use of *MySQL* as the information system development tool, a MySQL server was created. MySQL is an open-source, relational database system which enables the delivery of reliable, high-performance and scalable web-based applications [15].

The information system was created for the documentation of information regarding machines, operators, jobs and operations in the shop which can either be used by the scheduler or displayed on the web interface. The web interface is what the user of the system will use to log data changes, view scheduler results and choose new schedules. The web interface was created through *Microsoft Visual Studio* and the functionality of the web pages that were incorporated into the interface include:

- logging of a broken machine,
- logging of a new job entering the shop,
- changing of an operator status,
- displaying of scheduler results,

- manual capturing of sensor data,
- addition or withdrawal of a machine, and
- selection of a new schedule.

4.2.3 Simulation scheduler

The simulation scheduler is the component in the proposed solution that generates new schedules according to the different dispatching rules. New schedules are created each time the shop floor is disturbed by events such as a machine failure, a new order arriving, an operator being absent, *etc*. First of all, the scheduler is required to be in the cloud, therefore a cloud server was created to run the model. Following the creation of the server, the model was created with the use of *Tecnomatix Plant Simulation* which is a simulation software package. The scheduler incorporated different elements, which include:

1. MySQL data import, which is used to import data from the MySQL server created for the information system. This will ensure that the scheduler uses the current state of the shop.
2. Dispatching rules, which are used by the scheduler to generate machine schedules according to the different dispatching rules.
3. Machine schedules, which are the data tables the machines refer to, to determine the next operation to start processing.
4. Machine behaviour, which includes the various machines and ensures that each machine adheres to the sequences in the generated schedules.
5. Experiment inputs, which are controlled by a variable in the scheduler. The variable informs the scheduler as to which dispatching rule must be used.
6. Experiment outputs, which are the results of the performance indicator calculations for each dispatching rule. These values will be compared to the corresponding values of the other dispatching rules, and the best performing dispatching rule can then be identified.

5. DEVELOPMENT VALIDATION

This section will serve to describe the validation and testing of the developed system. Reference will be made to the validation of the scheduler as well as the validation of the sensor and gateway.

5.1 Scheduler validation

The validation of the scheduler starts off by defining test scenarios that are used to determine whether the dispatching rule generated the correct schedules. Thereafter the results of the generated schedules can be used to determine the best-performing dispatching rule for each performance indicator. The test scenarios that were defined include:

- **Scenario one:** five jobs are entered into the system, where four jobs have similar expected processing times. There is, however, one job which is a clear outlier because the processing times of its operations are considerably longer than those of the other jobs. (A small number of jobs is selected in order to be able to analyse the results manually.)
- **Scenario two:** five jobs are entered into the system, each having similar expected processing times.
- **Scenario three:** five jobs are entered into the system, each with exactly the same expected processing time.
- **Scenario four:** 50 jobs are entered into the system, having similar expected processing times. This scenario is used to test the robustness of the scheduler, i.e. whether it can handle a relatively large number of jobs.

It was decided that there will be no machine failures or absent operators during these tests. Each of these scenarios was then run by the scheduler and a schedule was generated for each dispatching rule. For explanation purposes and due to space limitations, only the schedule generated for the shortest processing time dispatching rule of test scenario one will be compared to a manually drawn schedule, to determine whether it generated the correct sequence. First the manually drawn schedule was created using the data from the information system, where meticulous care was taken to follow the dispatching rule. Figure 8 illustrates the schedule that was manually drawn. The schedule that was generated for test scenario one using the shortest processing time dispatching rule is illustrated in Figure 9, where the different machines are denoted by M1–M8. Figure 9 also illustrates the shift start and end time (*i.e.* 08:00 and 17:00), between which the machines will not accept any new operations. If there is an operation that already started on the machine before 17:00, the machine will continue processing until the operation is finished, but it will not accept a new operation. The generated schedule was then analysed to determine whether the sequence in which jobs were processed was correct, by comparing it to a manually drawn schedule for the same test scenario and dispatching rule. Both figures illustrate the same schedule which means that the scheduler did adhere to the dispatching rule and generated the correct schedule. The only difference between the schedules is that the manually drawn schedule does not incorporate the stoppage time between shifts, while the schedule generated by the scheduler does incorporate the stoppage time. This same evaluation process was done for each schedule that was generated, however, due to space limitations it will not be shown.

The same process was repeated for each dispatching rule and test scenarios, which resulted in a total of six schedules that were generated per test scenario. The schedules from the different dispatching rules could then be compared when referring to each performance indicator, and the best-performing dispatching rule could then be identified. The results from each test scenario are shown in Table 1, Table 2, Table 3 and Table 4. The results include the performance indicator value, in "days:hours:minutes:seconds", for each dispatching rule. Depending on the performance indicator, the user can select the best performing dispatching rule by finding the shortest duration for that performance indicator. This concludes the validation of the scheduler; following is the validation of the sensors and gateway.

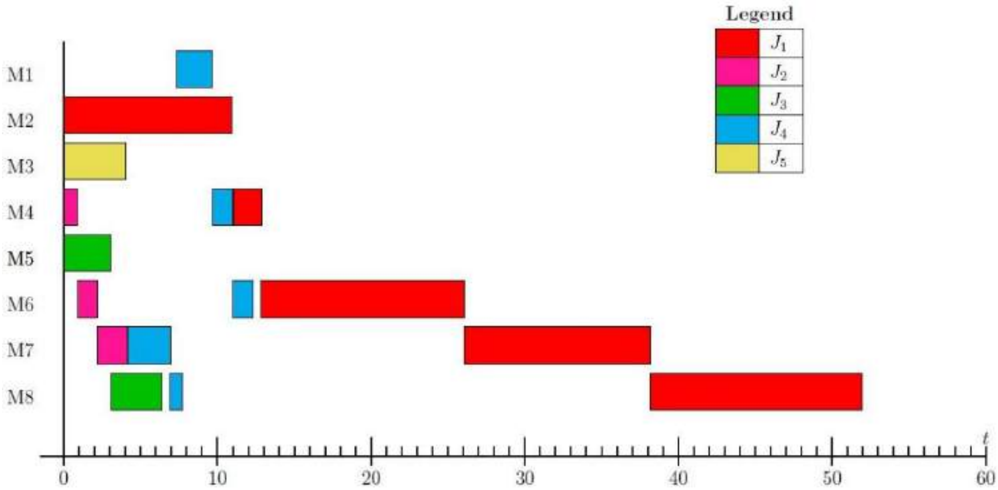


Figure 8: Schedule drawn manually through shortest processing time dispatching rule for Test scenario 1

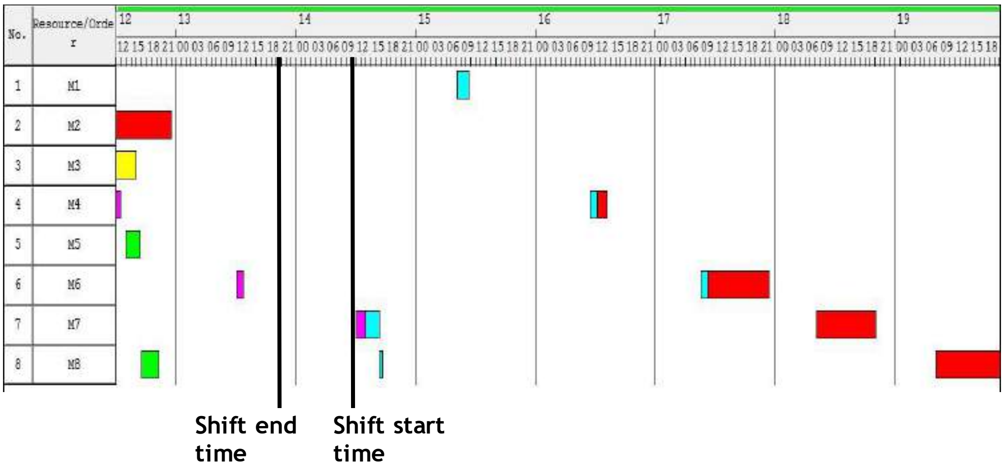


Figure 9: Schedule generated through shortest processing time dispatching rule for test scenario one

Table 1: Test scenario 1 results

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	9:21:57:13	9:07:02:19	05:36	-2:12:29:38	14:08:08:57
First-come-first-serve	10:05:56:27	9:15:01:33	42:17	-3:00:36:47	11:20:25:03
Most-important-job-first	10:00:20:30	9:09:26:18	00:30	-2:13:01:40	14:07:46:39
Earliest due date	9:05:11:34	8:14:17:36	27:53	-3:02:07:11	11:20:20:43
Critical ratio	10:05:50:39	9:14:55:46	41:59	-3:00:36:49	11:20:19:23
Minimum slack time	9:23:00:27	9:08:05:33	55:15	-2:12:48:52	14:07:48:57

Table 2: Test scenario 2 results

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	1:07:02:14	16:53	0	-3:02:32:16	2:01:51:34
First-come-first-serve	1:04:07:26	22:05	0	-2:22:58:57	1:22:51:11
Most-important-job-first	1:04:49:25	03:27	0	-3:01:21:32	2:00:57:56
Earliest due date	1:01:19:29	34:00	0	-3:09:36:49	1:09:32:53
Critical ratio	1:05:49:46	04:24	0	-3:04:34:41	1:17:45:17
Minimum slack time	1:04:31:22	46:01	0	-3:01:25:26	2:00:36:51

Table 3: Test scenario 3 results

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	1:05:33:34	47:28	0	-2:22:17:24	2:00:06:06
First-come-first-serve	1:04:51:38	05:32	0	-2:22:53:06	1:22:32:52
Most-important-job-first	1:04:16:50	30:45	0	-3:01:27:08	2:00:11:56
Earliest due date	1:03:48:45	02:39	0	-3:04:45:58	1:17:02:09
Critical ratio	1:07:20:16	34:11	0	-3:02:50:03	1:23:46:05
Minimum slack time	1:04:11:05	24:59	0	-3:01:27:31	2:00:04:44

Table 4: Test scenario 4 results

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	7:14:05:41	7:06:19:44	21:28	-13:12	14:22:40:41
First-come-first-serve	7:15:35:53	7:07:49:56	30:00	34:21	13:22:19:15
Most-important-job-first	7:23:10:02	7:15:24:03	22:28	09:52	14:20:53:59
Earliest due date	8:11:02:44	8:03:16:51	0	-3:34:30	15:06:17:49
Critical ratio	7:19:52:56	7:12:06:59	0	-4:03:59	14:06:09:51
Minimum slack time	6:12:47:58	6:05:02:01	11:09	-3:34:29	15:03:51:36

5.2 Sensor and gateway validation

The validation of the sensors and gateway entails testing whether the status of an operation changes in the IS when the sensor is evoked. To accomplish the validation, a job was entered into the system which was then allocated several operations and linked to the ID of a RFID card. All the operations of the job will have a status of *Pending* at the start of the test. Having the job arrive at the machine of its first operation, it is expected that the status of that operation should change to *Waiting*. To accomplish the status change, the button corresponding to the status *Waiting* was selected and the RFID card linked to the job was swiped. The information system was consulted after the card was swiped, and it could be concluded that the sensor succeeded in changing the operation status. This concludes the development validation section. The following section will discuss the operational testing of the scheduler.

6. OPERATIONAL TESTING

For the operational testing of the system, the authors logged disturbances in the information system to determine what effect they will have on the schedules. The disturbances that were tested, include:

- logging a machine as broken,
- logging an operator as absent, and
- adding a new job with many operations.

For the purpose of the operational testing, a test scenario was created with ten jobs that have exactly the same expected processing times. The results of this scenario, where there are no disruptions, are illustrated in Table 5. These results can be used as reference values for the results of the tests where disruptions occurred. The results include the performance indicator value, in "days:hours:minutes:seconds", for each dispatching rule.

Table 5: Performance indicator results with no disruptions

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	2:09:49:58	2:02:03:52	9:21:37:52	11:29	4:02:15:18
First-come-first-serve	1:22:59:25	1:15:13:20	9:13:25:35	2:14:32:35	3:03:15:51
Most-important-job-first	1:20:09:22	1:12:23:16	9:08:26:43	26:38	3:21:54:20
Earliest due date	1:12:56:36	1:05:10:30	9:03:53:09	21:57	2:19:39:05
Critical ratio	1:23:20:17	1:15:34:11	9:11:54:31	34:16	3:19:39:05
Minimum slack time	1:18:22:08	1:10:36:03	9:08:55:00	19:33	3:00:09:30

When the authors logged a machine as *Broken*, it is expected that the scheduler will start with the broken machine as unavailable and not utilise the machine before it is operational again. The jobs will therefore be allocated to the identical machine, if applicable. However, when the machine is operational again, the scheduler should divide the remaining jobs allocated to the identical machine between the two machines. Table 6 provides the results of the different performance indicators when one of the milling machines (*i.e. M3*) was broken for four days. From the table it is evident that the scheduler changed the schedule to accommodate the disruption, because the values are longer than those where no disruption occurred.

Table 6: Performance indicator results when M3 is broken for four days

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	3:04:57:13	2:21:11:07	10:17:44:59	10:17:44:59	5:00:55:18
First-come-first-serve	2:05:19:51	1:21:33:45	9:19:28:48	13:34	3:17:09:10
Most-important-job-first	2:08:01:38	2:00:15:33	9:20:53:08	9:20:53:08	4:17:04:22
Earliest due date	2:06:15:33	1:22:29:27	9:20:22:52	24:20	3:17:59:05
Critical ratio	2:02:32:48	1:18:46:42	9:16:41:49	02:25	3:17:59:05
Minimum slack time	2:03:12:48	1:19:26:43	9:19:03:00	37:14	3:01:24:15

When the authors logged an operator as *Absent*, it is expected that the scheduler will start with the machine where that operator is located as unavailable and not utilise the machine before the operator is available again. The jobs will therefore be allocated to the identical machine, if applicable. However, when the machine is operational again, the scheduler should divide the remaining jobs allocated to the identical machine between the two machines. Table 7 provides the results of the different performance indicators when an operator at a grinding machine (*i.e. M6*) was absent for three days. From the table it is evident that the scheduler changed the schedule to accommodate the disruption, because the values are longer than those where no disruption occurred.

Table 7: Performance indicator results when an operator is absent for three days

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	2:10:21:38	2:02:35:32	10:00:00:50	27:02	4:00:35:18
First-come-first-serve	2:22:42:53	2:14:56:47	10:11:50:54	10:11:50:54	4:22:06:04
Most-important-job-first	2:00:35:46	1:16:49:41	9:14:22:59	24:58	3:23:48:46
Earliest due date	2:02:22:54	1:18:36:48	9:16:36:45	13:34	3:19:39:08
Critical ratio	2:02:09:33	1:18:23:27	9:16:26:22	11:26	3:19:26:20
Minimum slack time	2:04:24:39	1:20:38:33	9:18:47:48	21:06	3:18:39:40

Finally, when the authors added a new job with several operations to the system, it is expected that the schedule generated by the scheduler will incorporate the new job into the new schedule. The job that was added has a total processing time that is similar to those of the other jobs. Table 8 provides the results of the different performance indicators when a new job was added to the system, and from the results is is evident that the scheduler changed the schedule to accommodate the disruption, because the values are longer than those where no disruption occurred.

Table 8: Performance indicator results when a new job was added

Dispatching Rule	Average Flow Time	Average Queue Time	Average Job Tardiness	Average Job Lateness	Makespan
Shortest processing time	2:15:22:12	2:07:10:18	10:16:38:43	2:09:50:55	4:21:30:00
First-come-first-serve	2:01:51:21	1:17:39:26	10:05:31:54	3:19:34:50	3:19:00:00
Most-important-job-first	2:19:47:19	2:11:35:25	10:20:37:28	10:20:37:28	5:16:29:10
Earliest due date	2:02:50:59	1:18:39:04	10:04:48:49	09:32	3:17:36:41
Critical ratio	2:05:04:38	1:20:52:43	10:06:24:40	57:54	4:00:51:08
Minimum slack time	2:02:27:55	1:18:16:01	10:03:51:11	42:36	4:00:34:20

Each of these disturbances that were tested, proved that the scheduler adapted and generated a new schedule that incorporated each disturbance. The scheduler can therefore be assumed to be working correctly. This concludes the operational testing of the scheduler and the following section will provide the conclusion of the research project.

7. CONCLUSION

This paper described the development and testing of a prototype of a real-time simulation scheduling system for a sensorised job shop. A revised architecture of the proposed system was provided, which describes the working of the system. The different hardware and software components required for the development of the system, were also identified and discussed. Finally, the validation and operational testing of the system was discussed. It was our purpose to build a real-time simulation scheduler of a sensorised job shop so that unexpected disturbances in the shop can be overcome by the generation of new schedules with real-time data from the shop. The results from the testing of the system proved that the authors were successful in creating the desired system, and ultimately succeeded in fulfilling the purpose of the research project.

REFERENCES

[1] **Groover , M. P.** 2013. *Principles of modern manufacturing*, 5th Edition, US: John Wiley and Sons.
[2] **Mitsuishi, M., Ueda, K. & Kimura, F.** 2008. *Manufacturing systems and technologies for the new frontier: The 41st CIRP Conference on Manufacturing Systems May 26-28, 2008, Tokyo, Japan*. London: Springer.
[3] **Snyman, S., Bekker, J.** 2017. Real-time scheduling in a sensorised factory using cloud-based simulation with mobile device access, *South African Journal of Industrial Engineering*, 28(4), pp 161-169.
[4] **Wisner, J. D.** 2016. *Operations Management: A supply chain process approach*. California: SAGE

Publications.
[5] **Dominic, P. D. D., Kaliyamoorthy, S. & Kumar, M. S.** 2004. Efficient dispatching rules for dynamic job shop scheduling, *International Journal of Advanced Manufacturing Technology*, 24(1), pp. 70–75.
[6] **Dori, D.** 2002. *Object-Process Methodology: A Holistic Systems Paradigm*. New York: Springer.
[7] **Raspberry Pi.** 2014. Teach, learn, and make with Raspberry Pi, [accessed 20 November 2017], available at: https://www.raspberrypi.org/help/faqs/
[8] **Arduino.** 2017. Getting started with the Arduino Pro Mini, [accessed 1 November 2017], available at: https://www.arduino.cc/en/Guide/ArduinoProMini
[9] **Scott Tattersall.** 2018. How to build custom IoT hardware with Arduino, [accessed 10 February 2018], available at: https://opensource.com/article/17/12/how-build-custom-iot-hardware-arduino
[10] **Semtech.** 2018. Lora transceivers, [accessed 10 February 2018], available at: https://www.semtech.com/products/wireless-rf/lora-transceivers
[11] **Murata.** 2018. Lora module FAQ, [accessed accessed 10 February 2018], available at: https://www.murata.com/support/faqs/products/lpwa/lora/hardware/0008
[12] **RF Wireless World.** 2017. Advantages and Disadvantages of Lora or LoRaWAN, [accessed 5 November 2017], available at: http://www.rfwireless-world.com/Terminology/Advantages-and-Disadvantages-of-Lora-or-LoRaWAN.html
[13] **NXP.** 2016. MFRC522 data sheet, [accessed 1 December 2017], available at: https://www.nxp.com/docs/en/data-sheet/MFRC522.pdf
[14] **AA Portable Power Corp.** 2004. Specification Cylindrical Li-ion 18650 2600mAh Rechargeable cell, [accessed 1 March 2018], available at: http://www.batteryspace.com/prod-specs/18650_2600.pdf
[15] **Oracle.** 2013. MySQL, [accessed 10 October 2017], available at: http://www.oracle.com/technetwork/database/mysql/index.html

DEVELOPMENT OF A DATA ANALYSTICS-DRIVEN SYSTEM FOR INSTANT, TEMPORARY PERSONALISED DISCOUNT OFFERS

Z. Els[1]* & J. Bekker[2]

[1]Department of Industrial Engineering
University of Stellenbosch, South Africa
zandalineels@gmail.com

[2]Department of Industrial Engineering
University of Stellenbosch, South Africa
jb2@sun.ac.za

ABSTRACT

The innovation of targeting customers with personalised discount offers has been incorporated into business strategies in order to ensure a competitive advantage amongst peers along with ensuring customer experience. In this article, a demonstrator model was developed which provides a holistic view of an individual customer's behaviour in retail outlets. The demonstrator creates instant, temporary personalised discount offers based on the purchasing tendencies of that customer in retail outlets. The model illustrates the utilisation of customer behavioural data and data analytics to identify unique cross-selling and upselling opportunities to ultimately improve customer experience. This article also includes the architecture of the proposed model along with the results from the demonstrator model.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION

In the time-intensive world we live in today, one would be lost without any communication and technology. This is caused by the dependency on technology that the world has fallen into. An attitude change towards new technology has become necessary in order to accommodate the ongoing rush to achieve more.

The International Data Corporation (IDC) identified the so-called Third Platform in 2007. This platform is built on four technology pillars namely; mobile computing, cloud services, big data analytics and social networking [1]. This Platform has a series of innovation accelerators that depend on the Platform, where the Internet of Things (IoT) is one of the most promising ones. The IoT can be explained as all devices that connect and communicate with one another via the internet. This new communication channel has created new means for the creation and transfer of Big Data and with that, a new world of innovation exists.

The challenge to influence potential users and initiate a change in user attitude towards these ne innovations are ideally what industrial engineers are good at. The prospect of these new innovations creates new opportunities for industrial engineers.

In the United States of America, Walgreens partnered with Aisle 411 and Google Project Tango to create a 3D augmented reality to Walgreens. Aisle411 helps the customer map and search products to where they are located on the shelf and Project Tango can determine the user's location within the store. This proposes a game-like shopping experience and is limited to a Tango device and Walgreens store [2]. This innovation does not include personalised offers based on customer historical data and usage.

As the world became more advanced with technology and new innovations, the cost of living also increased over time. Almost all retail stores now propose discount offers and loyalty programmes to attract customers. A reason for this is the competitive attitude that started to exist among retail stores. In the past, discount offers were sufficient to still have a profitable business, but this is not so anymore. To maintain a competitive advantage, innovations must be considered to ensure customer experience is superior.

A case study provided by TM Forum [3] was the starting point of the new initiative to improve customer experience by using data analytics. In the case study, customers receive personalised discount offers (PDO) based on their buying behaviour and acceptance history. The case study was redefined according to the following scenario:

As a customer walks into a participating store, the customer will receive PDOs via a mobile device on certain items in that store. These discount offers are only valid for this specific individual at this point in time in the store mentioned. This can be made possible by analysing the purchasing behaviour of customers and determining which items they would be likely to buy or need at a specific point in time. Using historic information, customer profiles are created and personalised special offers can be determined. Along with the customer profiles, the efficiency of marketing can also be analysed and improved. To better understand the scenario described here, consider the following specific example:

Jo decides to subscribe to this personalised offer platform (which is free) and allows the tracking of his buying behaviour. Jo buys a certain brand of shampoo every time his is finished. After a certain period of time, a buying trend for Jo is identified. The trend indicates that he buys this specific shampoo every three weeks on average. One day Jo walks into one of many participating retail outlets and passes the shelf displaying all the shampoos. Jo had not gone to the store to buy shampoo since he still had a little bit left. As Jo passes this shelf, he receives a notification presenting him with a one-time PDO on his specific brand of shampoo. Jo finds this offer appealing and argues that he might as well buy shampoo now for less rather than coming back to the store in two days' time. Jo accepts this PDO and buys his shampoo for less along with his other groceries. Jo experiences a satisfied feeling towards his trip to the store and the money he saved by using his personalised offer.

Apart from being offered discount on a specific product, the client could also receive discount offers for alternative products, which are referred to as *cross-selling*. Also, discounts on similar but more expensive products could be offered, thus earning more revenue from the client if he accepts the offer. This is called *upselling* and is the ultimate goal of PDOs.

These offers must be proposed at a point in time when individuals are most susceptible to them. A large quantity of data must be processed and analysed in order to create customer profiles, track buying behaviour and specify personalised offers. The problem at hand is thus to develop a demonstrator that creates these customer profiles and determines the most suitable PDOs for specific individuals in real time. Along with this demonstrator, a simulator is necessary to generate historical, stratified data to be analysed by the demonstrator during development and testing. The authors work in the Unit for Simulation Modelling and Analysis (USMA). This research group forms part of the Department of Industrial Engineering at Stellenbosch University.

This article consists of a short literature review in Section 2. This section elaborates on customer profiling, cross-selling and upselling and offers identification, which is necessary to understand the fundamentals of this study. The proposed model for the demonstrator is described in Section 3. Section 4 presents a toy problem created by the authors. Section 5 describes the development of the simulator and demonstrator followed by the results in Section 6. A business case is provided in Section 7 and the last section concludes the research work discussed in the article.

2. SHORT LITERATURE REVIEW

The short literature review includes information regarding customer profiles, cross-selling and upselling and offers identification.

2.1 Customer profiling and segmentation

An overview of customer profiling and segmentation is presented in this section. Approaches for developing customer profiles follows thereafter.

2.1.1 Overview of customer profiling and segmentation

Customer profiling is used to describe the customer based on some predefined attributes. Each customer has a unique profile based on their factual and behavioural data. Profiling is used to personalise individuals to better understand their needs. Customer service and satisfaction are improved by understanding the customers' desires and requirements [4], [5]. Marketers use profiling for targeted marketing strategies in which customers are presented an offer they would be interested in. The behaviour of customers is predicted by the method of discovering the similar patterns in the behavioural data. Profiling thus attempts to discover knowledge within customer data that was not known before [5]—[7].

Customer segmentation is done when customers are divided into homogeneous groups based on shared attributes and habits [5]. As in customer profiling, segmentation is used to identify certain properties for a certain group of customers. With the amount of data that must be analysed in the technological world of today, it becomes more and more tedious to segment customers based on similar traits. In order to identify the needs of each individual and choose marketing strategies appropriately, profiling individual customers seems to be the solution [8].

Bounsaythip *et al*. [5] state that customer profiling is performed after customer segmentation. The authors do not agree that this is necessarily the case. The full advantage of segmentation is done by profiling the different segments and designing marketing strategies accordingly. In the case of this study, the authors plan to go beyond customer segmentation and profile the buying behaviour of the customer.

2.1.2 Approaches to develop customer profiles and segmentations

According to Jansen [9], customer segmentation can commence without the knowledge of the data being analysed. This, unfortunately, does not apply in the case of customer profiling. Customer profiling can only commence once a complete set of individual customer data is available. The availability of data dictates which features are used for customer profiling. The factual profile is derived from the demographical data of the customer such as name, age, address, *etc*. It can also contain information derived from transactional data such as preferences. The behavioural data profile is derived from the transactional data of the customer which contains the records of the purchasing history of the customer for a certain time period [4].

Data mining is the phase where machine learning tools and techniques are used to develop customer profiles and segmentations. This is made possible by the advances in computer technology. The machine running the data mining software automatically searches large datasets to identify hidden correlations in data [6]. Machine learning techniques used for profiling differ from those used for segmentation. Customer segmentation is achieved by using unsupervised clustering models whereas customer profiling uses supervised classification models [10].

The next section describes where and how customer profiling and segmentation can be used by reviewing cross-selling and upselling.

2.2 Cross-selling and Upselling

In the domain of Customer Relationship Management (CRM), one of the core activities is customer retention. A well-known method to achieve this is by using cross-selling and upselling techniques.
Cross-selling can be defined as offering alternative products or services to customers during their current buying process. This is a strategy used to ensure an enterprise captures a large share of the consumer market. The company effectively increases the number of products the customer purchases from the company and guarantees a competitive advantage amongst its peers [11].

According to Schiffman [12], ***Upselling*** is a technique used when asking a customer who has already purchased something to purchase more of it or more of something else. This technique motivates customers to acquire more expensive versions of what is already owned. A promotion is one of the most popular upselling methods. Another method is alerting customers about alternative products by including information with the original purchase [13], [14].

Effective upselling and cross-selling occur when the needs of the customer are fully understood. This creates an opportunity for the application of customer profiling and segmentation to identify the specific needs of customers. Customer data is analysed by the practice of data mining which is an effective approach to identify cross-selling and upselling opportunities. It delivers knowledge about which products to promote and the most appropriate time to do so [13], [15].

There are three objectives of identifying cross-selling and upselling opportunities. The first is to understand the acquisition pattern of the customer. The second objective is to identify the factors which influence the repurchase decision of the customer and thirdly, forecasting the time of the possible repurchases [12]. It is important to understand these objectives as cross-selling and upselling not only ensure customer retention within the CRM domain, they also ensure profit growth for the enterprise.

In order to identify cross-selling and upselling opportunities one needs holistic information regarding customers. This was described in Section 2.1. The following section sheds light on how holistic customer segmentations are created and the process used for identifying cross-selling and upselling opportunities.

2.3 Offer identification

This subsection is an overview of RFM and Market Basket Analysis which can be used in identifying discount offers.

2.3.1 Recency, Frequency and Monetary

RFM (recency, frequency and monetary) is a common approach used for analysing customer purchase behaviour and making predictions based on the findings. The customer behaviour is usually contained within databases of enterprises and this is where the behavioural profile mentioned in Section 2.1 is developed from [10], [16].

The database must first be sorted by each attribute of RFM. After this the customers are divided into five equal segments. The different segments have different values for each RFM attribute depending on the customer's behaviour. ***Recency*** refers to the number of periods since the last purchase. The higher the number of periods (this can be days, months, *etc*.) the lower the recency variable of a customer will be, where five is the highest and one the lowest. ***Frequency*** refers to the number of purchases made in a given time period. The top quintile with the highest frequency is given the value of five. The ***monetary*** attribute is defined as the average amount of money spent during the analysed time. The more money spent on average the higher the monetary attribute will be. A combined RFM score can be calculated by adding the individual RFM attributes.

The RFM modelling technique is commonly applied in database marketing and a known tool to be used in developing marketing strategies [10]. This is because the RFM model can predict future purchases of customers. RFM is applied in most practical areas such as financial and government organisations, online and telecommunication companies as well as marketing and retail industries.

Apart from developing marketing strategies, the ***Customer Lifecycle Value*** (CLV) can be determined from the RFM analysis. The CLV provides insight of the profitability of the customer to the enterprise. This can give an expectation of nett future revenue of the company [17]. RFM analysis is also used for clustering customers into segments based on similar traits.

RFM can be used in this initiative to create provisional customer clusters for new participating members.

2.3.2 Market Basket Analysis

Market basket analysis (MBA) is a technique used for discovering purchasing patterns of customers by identifying associations between products customers buy together [18], [19]. Association rule mining is a data mining tool with various algorithms to create association rules that are used in MBA. These rules are generated by analysing the transactional database containing customers' purchasing history. Given two products X and Y, an association rule in the form of $X \rightarrow Y$ indicates a buying pattern that if a customer purchases item X they also purchase item Y.

There are two measures needed to generate association rules, namely the *support* and the *confidence*. Support is the measure of how often the database contains both X and Y, where the confidence is a measure of the accuracy of the rule. This accuracy is defined by the ratio of the number of times both X and Y appear to the number of times only X appears in the transactional database. MBA associations between products can be used to identify cross-selling and upselling items to propose to customers. MBA results also help in marketing strategies where products having a strong association should not be placed on promotion at the same time.

According to [10] association models can be applied to selected levels of analysis. Transactional level refers to items bought at a single visit to the store. Aggregated information assesses products bought during a set time period by each customer and this is referred to as the customer level of analysis.

MBA can be used in this initiative to propose PDO to customers based on products they normally buy together. The following section provides the proposed model for this initiative.

3. PROPOSED MODEL

The proposed model gives an overall understanding of the case study that is addressed. In order to design the proposed model, the authors followed the Object-Process Methodology (OPM). OPM is an intuitive methodology that models the complex architecture of systems in a coherent way. Development and support are needed throughout the life cycle of artificial models. This calls for a comprehensive methodology that includes all challenging points in the evolution of a system [19]. Systems consist of three main aspects which are the *function*, the *structure* and the *behaviour* of the model. These aspects are alike for both artificial and natural systems, which makes OPM an unambiguous approach to gain a holistic view of a system. OPM is an ISO-certified methodology (ISO19450) which confirms that it is sufficient for practitioners to use OPM as a modelling paradigm to conceptualise systems in a varying amount of detail. The value of using OPM is within the visual graphics and semantics which make it easy to understand [20]. The proposed model created with OPM is visualised by the Object-Process Diagrams (OPD) that can be seen in Figure 1.

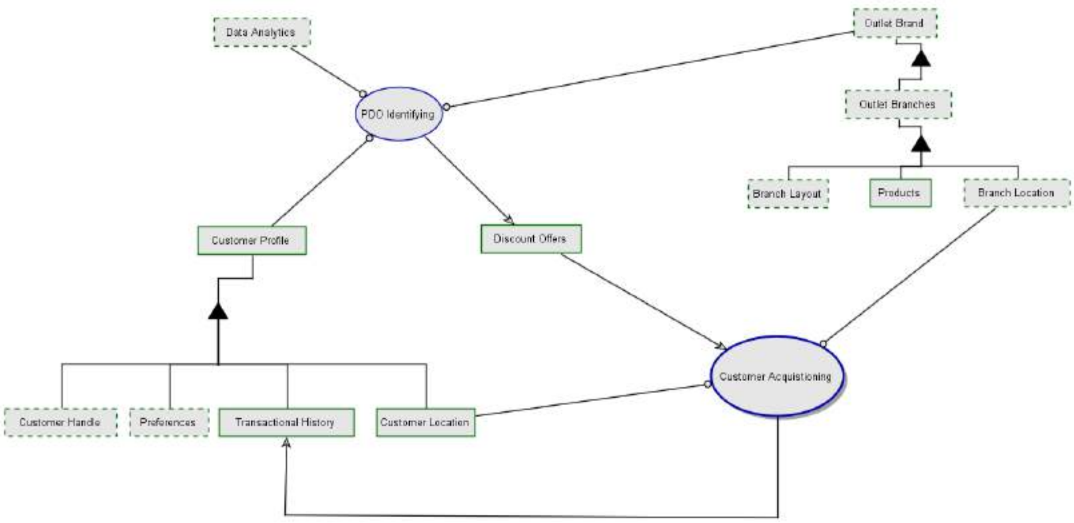


Figure 1: Architecture of the proposed model

The OPD is accompanied by the natural language description of the diagram, which makes OPM a desired approach. The natural language describing the proposed system is as follows:

Data Analytics is environmental.
Customer Profile consists of Customer Handle, Preferences, Transactional History, and Customer Location.
    Customer Handle is environmental.
    Preferences is environmental.
Outlet Brand is environmental.
Outlet Brand consists of Outlet Branches.
    Outlet Branches is environmental.
    Outlet Branches consists of Branch Layout, Products, and Branch Location.
        Branch Layout is environmental.
        Branch Location is environmental.
PDO Identifying requires Outlet Brand, Customer Profile, and Data Analytics.
PDO Identifying yields Discount Offers.
Customer Acquisitioning is physical.
Customer Acquisitioning requires Customer Location and Branch Location.
Customer Acquisitioning consumes Discount Offers.
Customer Acquisitioning yields Transactional History.

As stated previously, this proposed model is a very high-level view of the desired system. The processes are shown with elliptical symbols whereas the objects are represented by rectangles. In this proposed model the two processes are the *PDO identifying* and the *Customer acquisitioning*. Customer acquisitioning consists of two lower-level OPDs. For simplicity the lower-level OPDs are not included in this article.

The process of PDO identification requires the outlet brand, customer profile, and data analytics to make this process happen. The different requirements for PDO identification consist of other objects as well. This process yields discount offers, which is consumed by the customer acquisitioning process.

The outlet brand object consists of the participating branches and these branches include information such as branch layout, location and products. The customer profile object consists of the factual information of the customer and the preferences of the customer. This information is captured when the customer subscribes to the service and cannot be changed by the model itself. The customer profile also includes the customer location and transactional history of the customer.

The customer acquisitioning process is the second process in this top-level OPD. Customer acquisitioning requires the customer and branch location. As mentioned before, this process consumes the discount offers yielded by the PDO identifying process. The transactional history is yielded from the customer acquisitioning process. This creates a feedback to the customer profile which ensures the transactional history is updated with information regarding customer acquisitions from all participating outlets.

The next section explains a toy problem the authors designed to enlighten the reader on how the proposed model will function in a real-world situation.

4. TOY PROBLEM

The toy problem is modelled on a smaller scale of data. Figure 2 visualises how the real-world system would interact with the customer.

Retail enterprises subscribe to this system in order for customers to utilise this service within their respective stores. Initially, the customer subscribes to the services of the PDOs via a mobile app. In this subscription, the factual data of the customer is captured. Along with this, the customer is also allowed to enter preferences which would help to place the new customer in a provisional customer segment. Over time, the customer builds up transactional history, which the system uses along with data mining techniques to identify offers the customer would be most susceptible to. When the customer enters one of many participating retail outlets, the system is notified of the customer's location. If the moment in time relates to the time the system estimated the customer to be most susceptible to an offer, the customer receives a notification via the app.

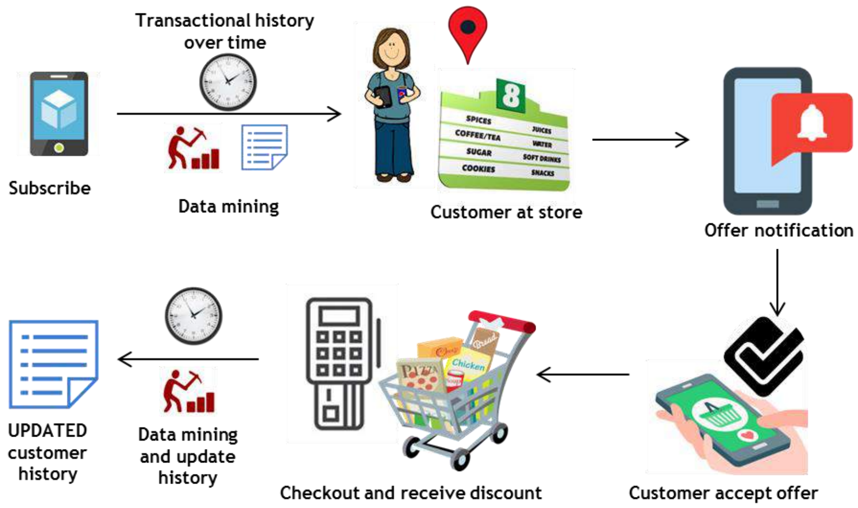


Figure 2: Schematic of the toy problem

The customer is presented with the personalised discount offer. This offer is personalised according to the purchasing behaviour of the specific customer. The offer is instant and temporary and will thus only be valid while the customer is in the store. The customer can either decide to accept the offer or to reject it. In the case of rejection, the system records the decision to improve the customer profile. If the customer decides to accept the offer, the discount is presented at the checkout point. Along with this the transactional history of the customer is updated. The customer can update the transactional history with every visit to a participating store by using the mobile app even though an offer is not presented. As time passes and the mobile app is used frequently, the customer profile will become more accurate and more appropriate offers can be identified.

The following section describes the simulator and demonstrator that was designed and developed by the authors in order to create the system that was described up until this point.

5. SIMULATOR AND DEMONSTRATOR

The proposed system consists of a:

1. simulator that creates initial customer historic data, without any PDO analysis.
2. demonstrator that uses the simulator in 1. to keep on creating customer purchases, but now the real-world process is emulated, as PDOs are identified and offered based on machine learning techniques. **Figure 3** summarises the differences between the simulator and the demonstrator in the proposed system.

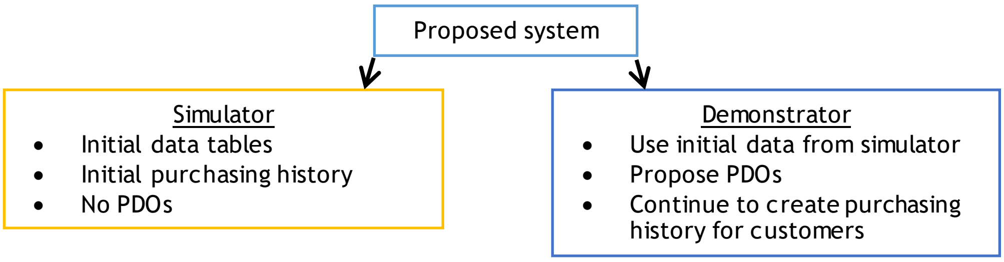


Figure 3: Summary of proposed system

5.1 Simulator

The authors developed the simulator using Matlab in order to simulate pseudo customer data that can be used for analysis purposes, thus creating historic data. For this system to be used in practice, enterprises will need to extract data in the correct format to be used in the demonstrator. The format in which the demonstrator requires the data is very specific and it is for this reason that the authors decided to simulate data in the correct format to be analysed. The simulator is not the main focus of this article and will thus not be explained in depth. The simulator simulated numerous different customers, branches, brands, products, preferences, *etc*. It was possible to create customers' transactional history using this initial information. Different statistical distributions were used to introduce variation within the initial simulated data.

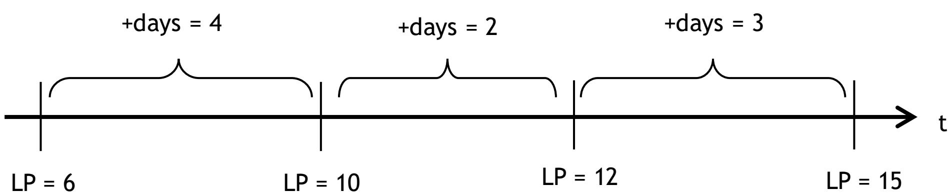
The initial orders or purchasing instances are contained within an orders table in a relational database. The orders table contains information regarding each instance a customer visits a participating store. Certain aspects need to be determined in order to populate this table for each simulated day. These aspects are:

1. The number of customers visiting stores each day.
2. The respective customers who visit the stores for the specific day.
3. The respective store each customer visits.
4. The time the customer visits a store.
5. To update the customers last purchase date.

These choices were based on using various statistical distributions in order to introduce variation within the purchasing history. The customers' last purchase dates are updated each time they visit a store again.

LastPurchase (LP)	...
~~06-01-2016~~	...
~~10-01-2016~~	...
~~12-01-2016~~	...
15-01-2016	...

Figure 4 explains visually how the last purchase date of customers is updated.



LastPurchase (LP)	...
~~06-01-2016~~	...
~~10-01-2016~~	...
~~12-01-2016~~	...
15-01-2016	...

Figure 4: Example of customer last purchase date update

Each purchasing instance that is recorded within the orders table is linked to information such as the products bought, the quantity of products and the unit price paid for them. This information is recorded in the transactional history table. These two tables are the most important tables in this study as they contain the purchasing behaviour of each customer and are used as input to the demonstrator.

All the initial information and purchasing behaviour generated by the simulator were saved in a Microsoft SQL Server database. The relevant information required for the demonstrator is extracted from this SQL database to Matlab. An ODBC data connection was created between SQL Server and Matlab using the Matlab Database Explorer Application. This connection is illustrated in Figure 5. The author used Matlab and SQL Server because it was available to the author. Alternative options can be used such as Python and MySQL.

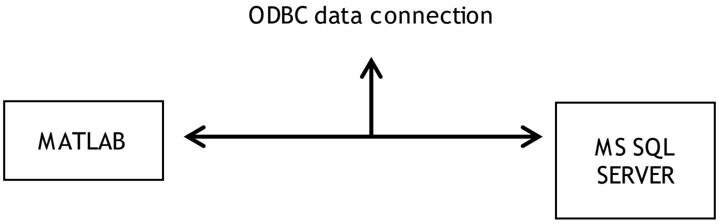


Figure 5: Data connection between Matlab and SQL Server

5.2 Demonstrator

The goal of the demonstrator is to identify patterns within the purchasing history of a customer. The demonstrator creates and presents PDOs to customers, based on the purchasing pattern of the specific customer. The demonstrator uses the simulated pseudo customer data created by the simulator explained in Subsection 5.1. The demonstrator is divided into two distinct parts or functionalities and is visualised by Figure 6.

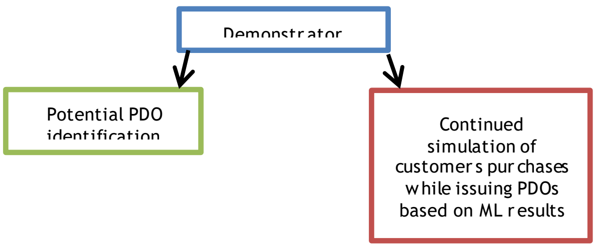


Figure 6: Summary of demonstrator functionalities

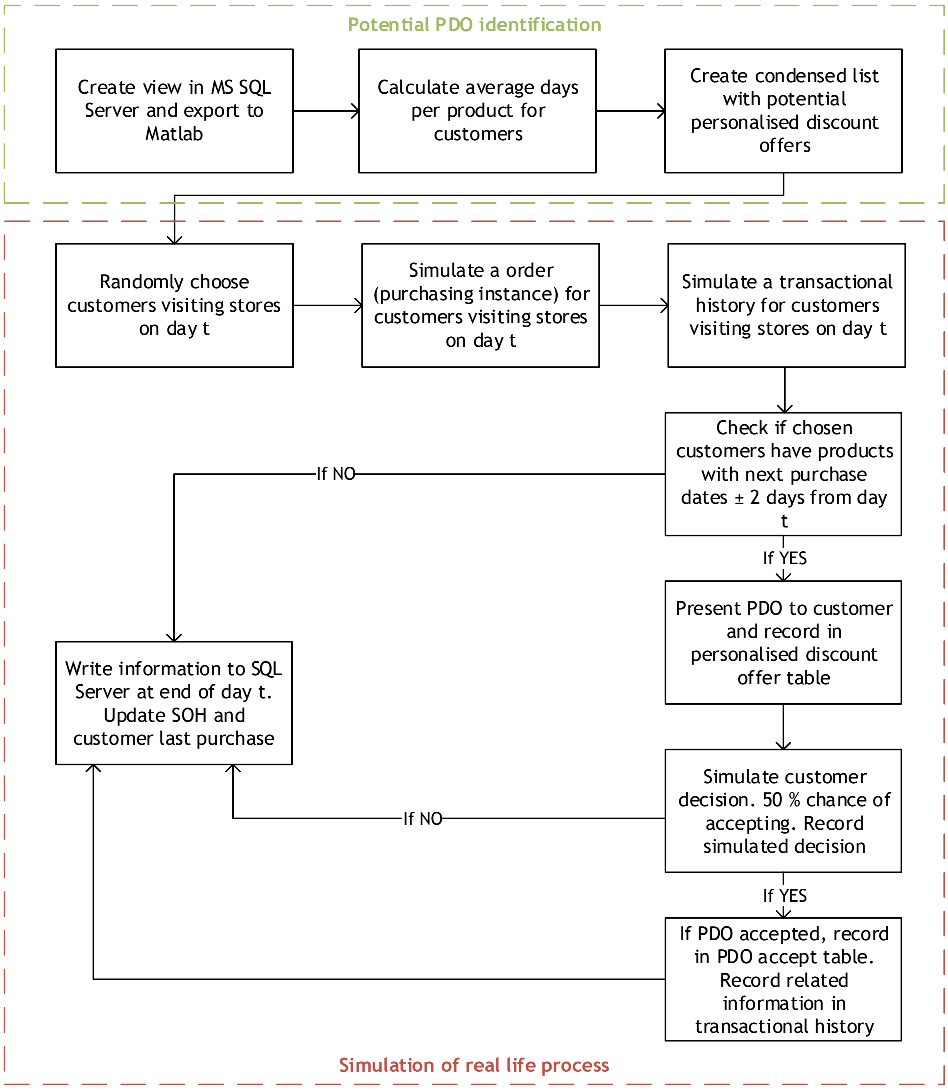


Figure 7: Conceptual schematic view of the demonstrator

The first part performs the identification of potential PDOs. The second part performs the continued simulation of the real world process where customers visit the stores, receive PDOs and continue to create a transactional history. Figure 7 is a schematic representation of the two parts.

To realise the first part of the demonstrator, the authors created a function that returns a table with products that may be potential discount items for customers. The function uses a view created in SQL Server and calculates an average days per product for products bought more than twice. This average is returned in a condensed structure along with the potential next purchase date of each product.

The average days per product was calculated taking into account the days between purchases of a specific item and the quantity that was bought. The data is updated in SQL at the end of each day to be ready for the calculation at the beginning of the next day. This explains the first part of the working of the demonstrator. Figure 8 visually explains how the function calculates the average days per product usage for example, Product X.

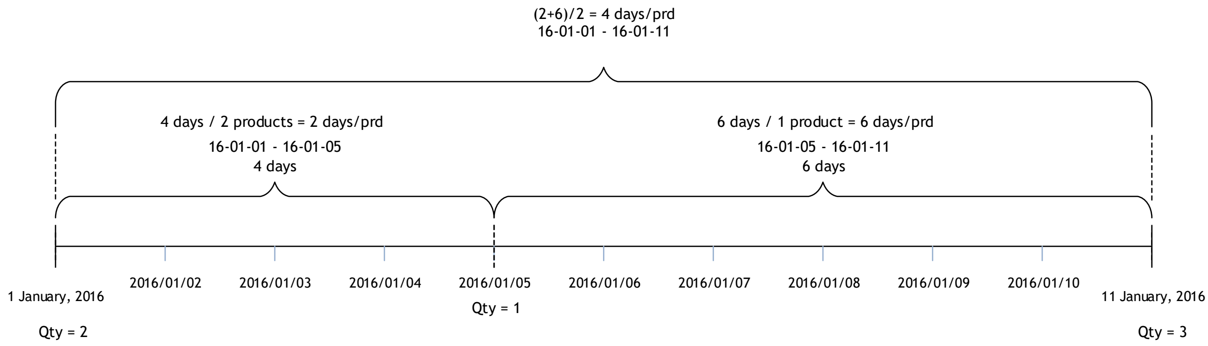


Figure 8: Calculations of average days per product

The average days per product usage for Product X was calculated as four days per product. Using the average days per product and quantity bought, the next potential purchase date for Product X is 23 Jan 2016.

The second part of the demonstrator is the simulation of the real world process of customers visiting retail stores and receiving potential PDOs. The condensed structure obtained by the Matlab function in the first part is used in this part of the demonstrator to evaluate whether the specific customers visiting the stores on the specific day qualify for a PDO. The demonstrator was developed so that it still created pseudo customer transactional data for those customers visiting the store on the specific day.

For each day, the demonstrator randomly chooses customers visiting stores that day. After the customers are chosen, the purchasing instance and transactional history are created as the simulator created it. Using the condensed list of potential PDOs, the demonstrator evaluates whether the customers entering a store that day have a product with a next purchase date ± two days from the current date. If any of the customers entering the store encompass products with a next purchase date within this date range, they will be presented with a PDO. The information regarding the specific PDO is recorded in a personalised discount offer table. This table also includes the acceptance or rejection of offers.

The demonstrator was developed to have an acceptance rate for PDOs of 50%. This was used as a starting point. It is inadequate to assume that the probability of a customer accepting offers based on their acceptance history. The offer could be a cross-sell or upsell offer and based on the customer's experience of the new product they might not accept it again. In practice this percentage will be determined by the customers' acceptance and rejection rate by analysing the customer historical data.

If the demonstrator simulates the customer to accept the PDO, the relevant information is documented in the acceptance table. The product bought, including the discount, is also recorded in the transactional history table along with the other products the customer bought during the particular purchasing instance.
At the end of each day, all relevant information recorded is exported to the SQL Server database. Except for the tables already mentioned, the stock on hand of products and the particular customers' last purchase dates are also updated during this event. The information regarding the purchasing history of customers is continuously updated by the feedback to the SQL Server database at the end of each day. This ensures that PDOs are estimated based on the current buying behaviour of customers.

The following section sheds light on the experiments and results that were obtained from the demonstrator.

6. RESULTS

The authors simulated a dataset containing historical purchasing data of customers using the simulator developed in Subsection 5.1. This simulated purchasing history was used as the initial input in the demonstrator in order to predict customers' PDOs.

The authors first simulated a smaller dataset to ensure the model is correct before attempting the simulation and analysis of a large dataset. After a validation and verification of the simulator and demonstrator, no evidence of errors was found in the development of the simulator and demonstrator. Any latent errors must still be observed within the model. The demonstrator presented PDOs to customers based on their historical purchasing behaviour. In order to show the result of this, the purchasing behaviour of Customer 2 toward Product 7 is used as an example.

Table 1: Purchase history of Customer 2 for Product 7

Purchase instance	Date of purchase	Quantity
1	06-03-2016	3
2	26-03-2016	2
3	15-04-2016	3

Based on the purchase history displayed in Table 1, the demonstrator predicted the next purchase dates for Product 7. Table 2 shows when Customer 2 bought Product 7. The table also includes the date on which Customer 2 was presented a PDO for Product 7 along with the customer decision of acceptance or rejection.

Table 2: Demonstrator predictions for Customer 2's Product 7

Purchase instance	Date of purchase	Quantity	Offered as PDO	Accepted/ Rejected	Expected NPD
4	01-07-2016	2	No	-	27-07-2016
5	14-07-2016	2	No	-	31-07-2016
6	02-08-2016	1	No	-	11-08-2016
	12-08-2016	-	Yes	Rejected	13-08-2016
	14-08-2016	-	Yes	Rejected	17-08-2016
7	15-08-2016	1	No	-	18-08-2016
8	16-08-2016	1	Yes	Accepted	28-08-2016
	30-08-2016	-	Yes	Rejected	10-09-2016
9	21-09-2016	1	No	-	02-10-2016
	04-10-2016	-	Yes	Rejected	05-10-2016
	07-10-2016	-	Yes	Rejected	09-10-2016
10	08-10-2016	3	Yes	Accepted	22-11-2016
	24-11-2016		Yes	Rejected	25-11-2016
11	25-11-2016	1	Yes	Accepted	30-11-2016
12	29-11-2016	1	No	-	15-11-2016
13	13-12-2016	2	Yes	Accepted	06-01-2017
14	02-01-2017	2	No	-	27-01-2017

After each purchase instance of Product 7 by Customer 2 the next purchase date is updated as explained in Subsection 5.2. This example shows an extraction of the purchasing behaviour of Customer 2 in order to illustrate the working of the demonstrator. The authors are in the process of investigating using survival analysis for the prediction of the next purchase dates. The next section presents the business proposition for this initiative.

7. BUSINESS CASE

This section provides the business proposition this initiative has. A business model for this service was proposed by the authors by applying the Business Model Canvas designed by Osterwalder to explicitly state and refine the business model of this initiative [21]. shows the nine building blocks suggested by Osterwalder. The authors populated the nine building blocks with the focus on this project.

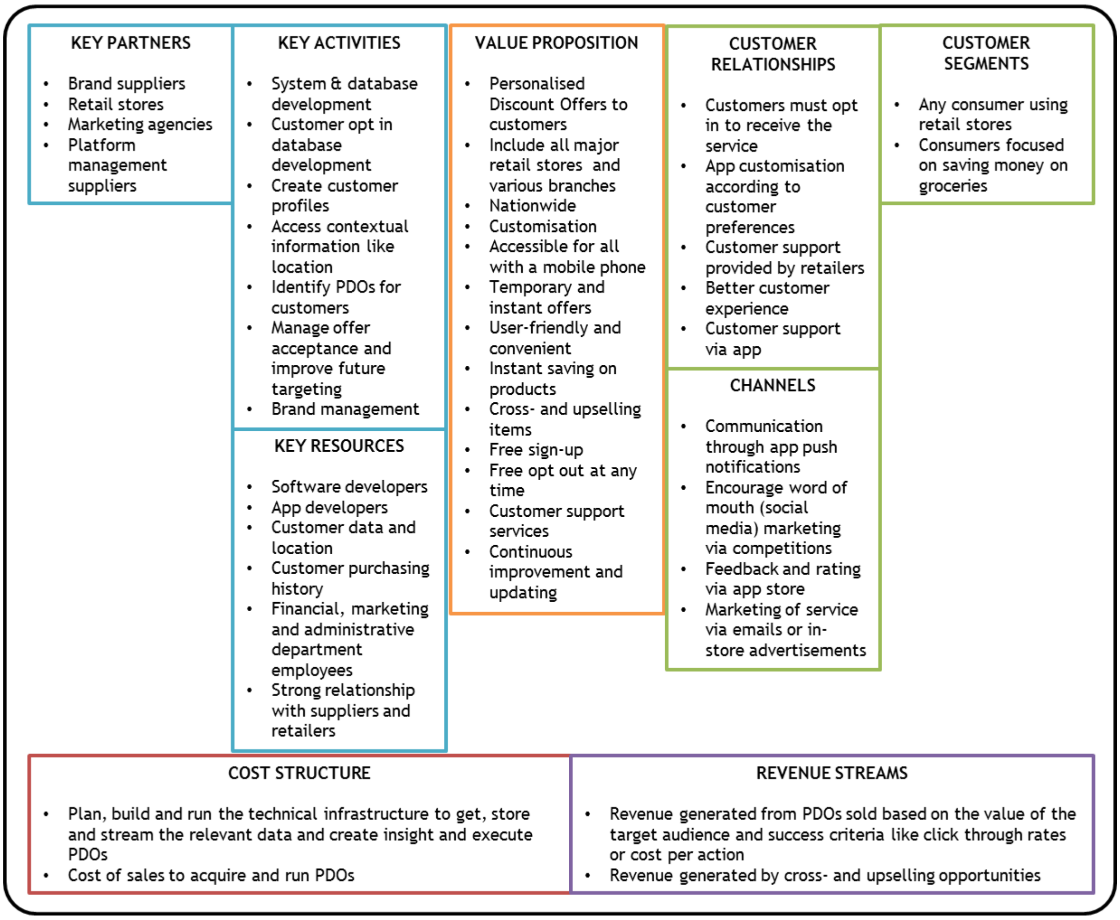


Figure 9: Business Model Canvas (adapted from [21])

This innovation suggests a new way of thinking and conducting business within the retail domain. The current relationship between retailers and suppliers in terms of creating promotions is a difficult and unpleasant task. With this new service, both retailers and suppliers can gain information regarding purchasing behaviour of potential customers along with ensuring low customer churn ratios. Promotions are personalised ensuring higher acceptance and including cross-selling and upselling offers based on purchasing behaviour, enterprises can expect an alternative revenue stream.

8. CONCLUSION

This article reflects on the work done by the authors as well as the mutual research done by the USMA Workgroup [22]. The literature study is considered fundamental to this study specifically. The literature review is followed by a proposed model. This article proposed a customer-centric marketing system in the context of data analytics where many customers with buying history can be targeted individually to enable cross-selling and upselling.

The proposed model visualises the architecture of the model and demonstrates how data mining will be used within the model. The study is explained by means of a toy problem. The toy problem is based on a smaller set of data and describes the interaction with the system. This illustrates precisely how the personalised discount offers are to be identified and how the customer will interact with the system.

The system was designed and developed in the form of a simulator and demonstrator. The simulator created pseudo customer historical data which was used as the input for the demonstrator. The demonstrator identified the personalised discount offers and proposed them to customers visiting the participating stores. Industrial engineers are system integrators who can see the detail but also the big picture. We are ideally positioned to engineer systems of the nature described in this article, i.e. systems that integrate to improve our lives, as dictated by the fourth industrial revolution.

REFERENCES

[1] **Gens, F.** 2013. The 3rd Platform: Enabling Digital Transformation, *IDC*, November, pp 1-13.

[2] **Xpertek Contact.** 2018. Aisle411. http://www.xpertekcontact.co.za/aisle-411/index.html?re=1&te=1. [Online Accessed: 04-09-2018]

[3] **Russom, P.** 2016. Big Data Analytics Guidebook, *TMForum*, May, pp 73-76.

[4] **Adomavicius, G. & Tuzhilin, A.** 2001. Using data mining methods to build customer profiles, *Computer*, 34(2), pp 74-81.

[5] **Bounsaythip, C. & Rinta-Runsala, E.** 2001. Overview of Data Mining for Customer Behavior Modeling, *VTT Information technology: Research Report: TTE1-2001-18*, pp 1-9.

[6] **King, N.J. & Jessen, P.W.** 2010. Profiling the mobile customer - Privacy concerns when behavioural advertisers target mobile phones - Part I, *Computer Law & Security*, 26(5), pp 455-478.

[7] **Lanjewar, R. & Yadav, O.P.** 2013. Understanding of Customer Profiling and Segmentation Using K-Means Clustering Method for Raipur Sahkari Dugdh Sangh Milk Products, *International Journal of Research in Computer and Communication Technology*, 2(3), pp 103-107.

[8] **Fan, S., Lau, R.Y.K. & Zhao, J.L.** 2015. Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix, *Big Data Research*, 2(1), pp 28-32.

[9] **Jansen, S.M.H.** 2007. Customer segmentation for a mobile telecommunications company based on service usage behavior, University of Maastricht.

[10] **Tsiptsis, K. & Chorianopoulos, A.,** 2010. *Data Mining Techniques in CRM: Inside Customer Segmentation*, 1st Edition, John Wiley & Sons.

[11] **Kubiak, B.P. & Weichbroth, P.** 2015. Cross-and up-selling techniques in e-commerce activities, *Journal of Internet Banking and Commerce*, 15(3), pp 1-7.

[12] **Schiffman, S.** 2005. *UPSELLING TECHNIQUES (That Really Work!)*, 1st Edition, Adams Media.

[13] **Salazar, M.T., Harrison, T. & Ansell, J.** 2007. An approach for the identification of cross-sell and up-sell opportunities using a financial services customer database, *Journal of Financial Services Marketing*, 12(2), pp 115-131.

[14] **David, F. R.** 2005. E-Crm: From a Supply Chain Management Perspective, *Information Systems Management*, 22(1), pp 37-44.

[15] **Dornheim, S., Groß, K., Debus, M., Achmann, F. & Hirschle, J.** 2015. *Managing the PSTN Transformation: A Blueprint for a Successful Migration to IP-Based Networks*, 20141201, CRC Press.

[16] **Wei, J., Lin, S., & Wu, H.** 2010. A review of the application of RFM model, *African Journal of Business Management*, 4(19). pp 4199-4206.

[17] **Krishna, G. J., & Ravi, V.** 2016. Evolutionary computing applied to customer relationship management: A survey, *Engineering Applications of Artificial Intelligence*, 56, pp. 30–59.

[18] **Giudici, P., & Passerone, G.** 2002. Data mining of association structures to model consumer behaviour, *Computational Statistics & Data Analysis*, 38(4), pp. 533–541.

[19] **Chen, Y. L., Tang, K., Shen, R. J., & Hu, Y. H.** 2005. Market basket analysis in a multiple store environment, *Decision Support Systems*, 40(2), pp 339–354.

[20] **Dori, D.** 2002. *Object-Process Methodology: A Holistic Systems Paradigm*, Springer.

[21] **Osterwalder, A. & Pigneur, Y.** 2013. *Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers*. John Wiley & Sons.

[22] **USMA Workgroup.** 2017. University of Stellenbosch.

DEVELOPMENT AND DEMONSTRATION OF A CUSTOMER SUPER-PROFILING TOOL TO ENABLE EFFICIENT TARGETING IN MARKETING CAMPAIGNS

M. Walters[1*] & J. Bekker[2]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
17618142@sun.ac.za

[2]Department of Industrial Engineering
Stellenbosch University, South Africa
jb2@sun.ac.za

ABSTRACT

Being part of a competitive generation demands having good marketing policies to attract new customers as well as to retain existing customers. This research outlines a general methodology for segmentation of customers by using the model of Recency, Frequency and Monetary (RFM) to identify types of customers, and then predict their customer profiles, based on demographic and behavioural features. A few previous studies dealt with the question using non-aggregate customer data. We, however, also address the problem by using decision trees, something which has rarely been done before. We applied and demonstrated this tool on a large customer dataset and found useful results.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

The research work reported on in this paper is the continuation of previous work published in [1] in the domain of data analytics and deals with specific aspects of customer profiling. In today's fast-moving world of marketing from product-orientation to customer-orientation, the management of customer treatment can be seen as a key to achieving revenue growth and profitability [2]. To gain more insights into customer behaviour, customer profiles should be constructed. Customer profiles are not the same as demographic information. Demographics usually provide the key dimensions that advertisers seek (age, gender, *etc*.), whereas profiling groups these dimensions along with other elements (behaviour) in creating the ideal customer profile [3].

This paper offers an approach to build customer profiles through data mining tools (*i.e.* supervised and unsupervised learning) and techniques (*i.e.* classification, regression, clustering *etc.*) when having a customer dataset with typical monetary transactional data, demographic data and extra value adding customer attributes; which include mobile phone type, medical aid *etc*. Data mining is referred to as a technique used to extract knowledge from information [5]. The goal of data mining differs from one area to another. When applying data mining to analyse data and create customer profiles, it will help to discover hidden knowledge in datasets to better understand customer behaviour and needs [4]. Thus, one can define data mining, with respect to customer profiling, as being the technology that allows the building of customer profiles (among other functions), where each profile describes the specific habits, needs and behaviour of a customer group. Therefore, developing customer profiles is an important step for targeted marketing campaigns, for it not only classifies new customers, but also provides information on current customers.

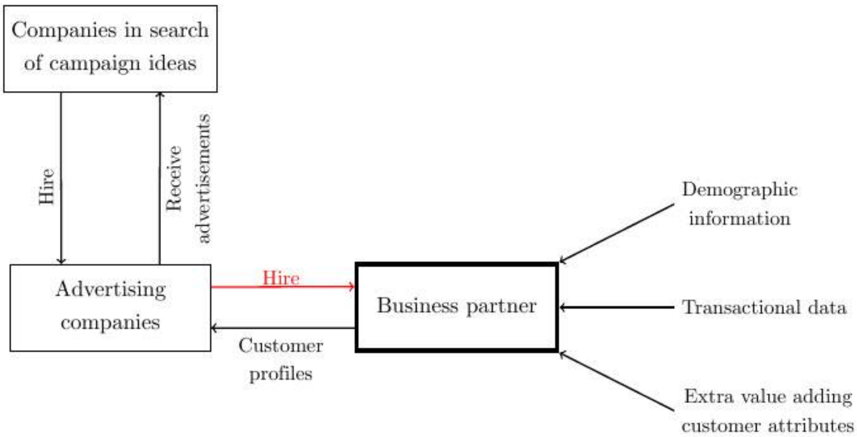


Figure 1: Illustrating the use of customer super-profiling

Figure 1 shows the purpose of the proposed *Customer Super-Profiling* (CSP) tool. It functions as a super-profiling analytics tool that receives various customer attributes as input to create customer super-profiles [1]. The customer attributes include demographic information (age, gender, ethnicity, *etc*.), transactional data, as well as extra value-adding attributes (transportation type, mobile phone, *etc*.). Companies in search of campaign ideas appoint advertising companies to assist them with marketing campaigns. Conversely, advertising companies may be in search of companies/developers that possess a profiling tool to provide them with reliable customer profiles for targeted marketing campaigns. These advertising companies are the value-creation partners: when they collaborate with the business partner they provide a revenue stream. The value that the advertising companies receive is knowledge about current and/or potential customers: who they are, what their behaviour and interests are and where to find them. This information provides the companies with insights in order to target suitable customers.

It is the authors' purpose to develop a CSP *tool* that has the ability to analyse a large dataset by utilising various big data analytics tools and techniques. To conduct this research, big datasets were necessary. In order to determine the structure and content of the data, a data simulator was developed to provide the super-profiling tool with data. Eventually, a different user of this super-profiling analytics tool could provide their own data, as long as the data have the same format and structure.

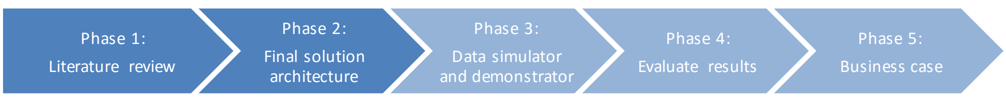


Figure 2: Project methodology

Figure 2 indicates the structure of this paper. Phase 1 of this project refers to the literature review that will be presented in *section 2*, which is brief, as the literature necessary to perform this work was presented previously in [1]. *Section 3* corresponds with Phase 2 in the figure, and presents the final solution architecture for the CSP tool. Phases 3 to 5 form the 'new' work, indicated as future work in [1]. *Section 4* contains the development of the data simulator and demonstrator, as well as evaluating the results. *Section 5* documents the business case applicable to this demonstration tool, while the last section gives the concluding remarks.

2. LITERATURE REVIEW

This section will provide a literature summary of specific data mining tools and techniques utilised to create the proposed customer super-profiling demonstrator. To initialise the literature review, a brief summary regarding segmentation will be provided.

2.1 Segmentation

Segmentation is often used in conjunction with profiling. However, segmentation is a term used to describe the process of dividing customers into homogeneous groups based on shared or common attributes, *e.g.* habits, tastes, *etc.*, while customer profiling describes the customers by using demographic information, *e.g.* age, gender, *etc.* as well as the behaviour of their homogeneous group.

Segmentation is performed on an *unordered* customer dataset and is the process of separating markets into groups of potential customers with similar needs and/or characteristics, who are likely to exhibit similar purchasing behaviour [6]. Segmentation of a customer dataset provides a more targeted communication with the customers, because the characteristics of a certain group of customers, or a cluster, are known [7].

Literature does not provide a distinguishable difference between market segmentation and customer segmentation, therefore the authors adopted the view that market segmentation is generally used for high-level strategy, whereas customer segmentation provides a more detailed view. Next to be discussed is the well-known heuristic approach called Recency, Frequency and Monetary (RFM) analysis.

2.2 RFM analysis

According to [8], most marketers experience difficulties in identifying the right customers to engage in successful campaigns. Thus far, customer segmentation is a popular method that is used for selecting appropriate customers for a campaign. Subsequently, [9] mentioned that for product advertising and promotions, there are mainly two approaches that are used in practice; *mass marketing* and *direct marketing*. Mass marketing targets large groups of customers; it does not distinguish between individual customers within a cluster/group and the information delivered to customers is uniform, whereas direct marketing targets individuals or households. Different customers receive different marketing information.

To achieve business success, engaging in effective campaigns is a key task for marketers. Traditionally, marketers first segment the market, and then target profitable customers. However, this process produces problems, as the correlation between customer segments and a campaign is neglected. Therefore, as stated by [10], it is necessary to consider significant campaign-dependent variables of customer targeting in customer segmentation. An approach to combine customer segmentation and customer targeting for campaign strategies was defined by [11]. The investigation identified customer behaviour, using the well-known *Recency*, *Frequency* and *Monetary* (RFM) analytical model ([2], [11], [12]). The detailed definition for each RFM parameter is as follows [13]:

- Recency (R): Represents the duration of time between the last purchase date/time and the date/time the 'survey' took place. The shorter the interval, the higher the recency value of a customer.
- Frequency (F): Represents the total number of purchases during the specific period (survey). The higher the number of purchases in an interval, the higher the frequency value of a customer.
- Monetary (M): Represents the monetary value of the purchases in the time interval considered. The higher the amount spent in an interval, the higher the monetary value of a customer.

After applying the RFM model to represent the customers' behaviour, the data is coded (encoded) into five categories. This is seen as one of the traditional applications of the RFM model and is called '*the customer quintile method*'. By coding, each customer is compared with all the others, depending on the variables used [11], [14]. If the value lies between 100% and 80%, the categorical value is set to 5; if between 80% and 60%, the value is set to 4, *etc.* In this way, the database is divided into 125 ($5 \times 5 \times 5$) equal clusters. The customers who obtain the highest RFM scores are generally the company's most profitable customers.

The purpose behind utilising the widely used behavioural-based method, RFM, is to analyse customers' behaviour and then make predictions based on the behaviour in the database [15]. The model is used in various research areas, which defines valuable customers as those simultaneously having high *recency*, *frequency* and *monetary* values. According to [2], one of the most effective customer segmentation models, based on customer value, is the RFM model. By adopting the RFM model, decision-makers can identify valuable customers and then develop effective marketing strategies [15].

The RFM model has been widely applied in many practical areas and its indicators are adaptable to measure customer value and to segment customers in different service areas ([14], [15], [16]). Next, a clustering technique (unsupervised learning), that can be performed on values obtained from the RFM analysis, will be discussed. This approach is explored because many researchers have considered the RFM variables when developing clustering models.

2.3 Clustering — *k*-means

The clustering technique that will be discussed is called, *k*-means, which forms part of the partitioning (non-hierarchical) method. The authors decided on using *k*-means as it is the most frequently used clustering technique, whilst it provides a good foundation for understanding clustering and is a simple and elegant approach to portioning a dataset into *k* distinct clusters.

The motive for using a clustering technique within this demonstrator is mainly to group together customers with similar purchasing or transactional patterns. When performing RFM analysis prior to the clustering process, it is possible to cluster customers based on the values obtained from the RFM analysis. The resulting clusters should have minimum dissimilarity within the cluster and maximum dissimilarity with other clusters [17].

It is important to understand the broad picture of how *k*-means performs clustering. After all the customer transactional data are transformed into RFM categories, the data can be grouped by using the *k*-means clustering technique. As mentioned, to be able to group the data into several clusters, certain steps need to be performed ([18], [19], [20]):

1. <u>**Determine the desired number of clusters**</u>: To find the optimal/desired number of clusters in a dataset, the *k*-means clustering algorithm needs to be performed for a range of *k* values and the results compared. In general, there is no method for determining the exact value for *k*, yet an accurate estimate can be obtained by using various techniques [19]. These techniques include: elbow plot, cross-validation, information criteria, the information theoretic jump method, the silhouette method, and the G-means algorithm. The *silhouette method* together with the squared *Euclidean distance* as distance metric will be used in this paper. Silhouette plot refers to a method of interpretation and validation of consistency within clusters of data. This technique provides a concise graphical representation of how well each object lies within its cluster (testing over the predefined range of *k* values). The silhouette method provides silhouette values for each data point by using the distance metric. The silhouette value is a measure of how similar an object is to its own cluster compared to the neighbouring cluster. The silhouette values range from -1 to +1 where a high value indicates that the object is well matched to its own cluster and poorly matched to the neighbouring cluster [20], [24].

2. <u>**Determine the centroids**</u>: The specialised algorithm called *k*-means++ algorithm uses a heuristic to find centroid seeds for *k*-means clustering. The *k*-means++ algorithm was proposed as a specific way of selecting centroids for the *k*-means clustering algorithm, instead of generating centroids randomly. It determines the initial centre points by calculating their squared distance from the closest centre already chosen.

3. <u>**Allocate data points to each cluster**</u>: The data points are assigned to its closest centroid. This is done by calculating the distance of each data point with regards to the centroid's position. The *objective function* that is employed by *k*-means is called the *Sum of Squared Errors* (SSE) or *Residual Sum of Squares* (RSS). The mathematical formula for SSE/RSS is
$$SSE\,(C) = \sum_{k=1}^{K} \sum_{x_i \epsilon C_k} \lVert x_i - c_k \rVert^2,$$
where c_k is the centroid of cluster C_k, and is denoted as
$$c_k = \frac{\sum_{x_i \epsilon C_k} x_i}{|C_k|}.$$

4. <u>**Recalculate the new cluster centroids**</u>: After all the data points are assigned to the closest cluster, compute the average of the data points in each cluster to obtain *k* new centroid locations.

5. <u>**Repeat steps 2—4**</u>: After obtaining the new centroids for each cluster, repeat steps 2 through 4 until cluster assignments do not change, or the maximum number of iterations is reached, because *k*-means is an *iterative algorithm*

This concludes the discussion regarding *k*-means for the purpose of this paper. A supervised learning technique that can be applied on clustered data, discovered by a clustering algorithm, will be discussed next.

2.4 Predictive model(s)

Classification algorithms are used to derive rules from the clustered results, obtained from the previous sections' results. These classification or decision rules are useful for identifying each and every customer from their purchasing patterns (RFM information) [17], [21]. There are various techniques for classification, *e.g.* decision trees, neural networks, *etc*.

The classification technique that will be utilised in this paper is decision trees. Once a decision tree or decision rule solution is generated from data, it can be used for estimating or predicting the response or class variable for a new case. The application of a decision tree is a straightforward top-down decision process, controlled by evaluating the tests and taking the appropriate branch, beginning at the root and terminating when a leaf node is reached. Decision tree and decision rule solutions offer a level of interpretability that is unique to symbolic models. This makes these solutions easily understandable for non-technical end users and makes this technique very appealing in decision support related data mining activities where insight and explanations are of critical importance. This approach is technically viable because most modern symbolic modelling methodologies succeed in formulating solutions that are also competitive in predictive accuracy, compared to non-intuitive or quantitative techniques, such as neural networks. This is an important reason for making use of decision rule modelling techniques to generate rules directly from data [21].

Estimating the true accuracy of a decision tree or rule model is an important aspect of the modelling process. A solution generated from a set of training examples will almost always be highly accurate on the same dataset, but far less accurate on new data. A two-fold strategy will be applied in this paper, where the first step involves generating the model from training data, and the second step involves testing the proposed solution on independent cases as part of future work [21], [24].

After completing the literature review, the authors noticed that very little has been done to apply unsupervised learning and supervised learning in conjunction, at least in the Industrial Engineering domain. This offers an opportunity to investigate the integrated use of both data analytics tools. This concludes the literature review for the paper. Next to be discussed is the final architecture for the super-profiling tool.

3. FINAL ARCHITECTURE FOR THE CUSTOMER SUPER-PROFILING TOOL

The proposed data simulator and super-profiling tool were developed in precursory work of [1], using the standard Object Process Methodology (OPM) approach [22]. Revisions have been made and the final solution architecture can be seen in Figure 3.

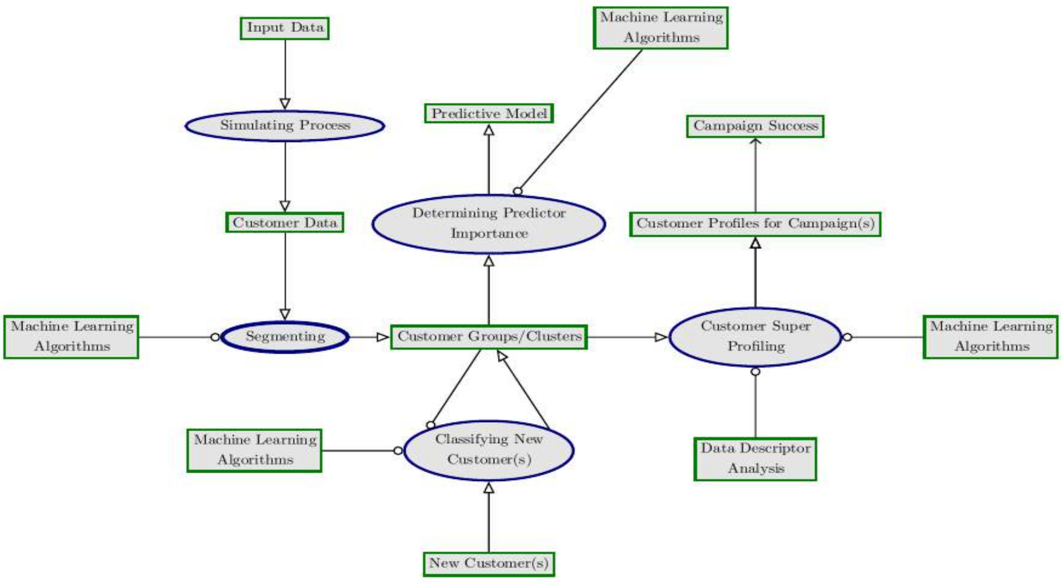


Figure 3: System diagram (SD) representing the working of the final proposed data simulator and CSP tool (Revised from [1]).

Object Process Language (OPL) is the semantic counterpart of the graphic OPM system specifications. The OPL is automatically generated as a textual description of the system in a subset natural English. Following the OPM guidelines, the OPL for Figure 3 is:

Customer Profiles for Campaign(s) relates to Campaign Success.
Customer Super Profiling requires Statistical Analysis and Machine Leaning Algorithms.
Customer Super Profiling consumes Customer Groups/Clusters.
Customer Super Profiling yields Customer Profiles for Campaign(s).
Segmenting requires Machine Leaning Algorithms.
Segmenting consumes Customer Data.
Segmenting yields Customer Groups/Clusters.
Simulating Process consumes Input Data.
Classifying New Customers requires Customer Groups/Clusters and Machine Leaning Algorithms.
Classifying New Customers consumes New Customer(s).

Classifying New Customers yields Customer Groups/Clusters.
Determining Predictor Importance requires Machine Leaning Algorithms.
Determining Predictor Importance consumes Customer Groups/Clusters.
Determining Predictor Importance yields Predictive model.

OPM is powerful since it presents a system architecture in visual and textual format.

4. DATA SIMULATION AND CUSTOMER SUPER-PROFILING TOOL

The previous section proposed the final system's architecture. This architecture aids in reaching the goal of the research, which is to develop a tool that contains a suite of Big Data Analytics tools and techniques which will allow for customer super-profiling.

The need for a data simulator that creates datasets with *specific properties* was identified in the previous sections. This section presents the process followed to create the specific datasets. The datasets will be used by the tool to illustrate the concept of customer super-profiling. The demonstrator is designed to contain the specific simulated datasets as input data, and if an enterprise wants to utilise the demonstrator, customer data with the same format and structure would need to be extracted.

4.1 Customer data

This section will provide more insight into the database that was constructed for this research. Firstly, the various customer attributes that provide more knowledge about a customer will be mentioned. These attributes will form the tables which constitute a database. The authors decided on utilising various customer characteristics (*e.g.* demographic and behavioural attributes). A few of the attributes are basic demographic information, while other attributes are more specific to customer behaviour (*e.g.* mobile phone type). Table 1 and Table 2 describe the available customer demographic data and customer behavioural features — those related to customer purchasing behaviour.

Table 1: Customer features used in the study

Variable name	Explanation	Scaling
Gender	Male or Female	Categorical
Ethnicity	Black African, Coloured, Indian/Asian or White	Categorical
Province	Eastern Cape, Free State, Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West or Western Cape	Categorical
Age	15-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 or 80+	Categorical
Education	Less than Gr.12 and no other qualification, Less than Gr.12 and with diploma or certificate, Gr.12, Gr.12 with diploma or certificate, Degree or post graduate degree or Honours degree or higher	Categorical
Employment	Employed, Unemployed or Not economically active	Categorical
Annual income	R0-R12 000, R12 001-R54 000, R54 001-R192 000, R192 001-R360 000 or More than R360 001	Categorical
Relationship status	Married or domestic partner, Never married or single, Widowed or Divorced	Categorical
Children status	Yes or No	Categorical
Household size	1, 2, 3…,10+	Categorical
Medical aid	Yes or No	Categorical
Housing ownership	Rented, Owned (not fully), Owned (fully), Occupied rent free or Other	Categorical
Housing type	Cluster house in complex, Flat or apartment in flat block, House or brick structure on yard or stand, House, flat or room in backyard, Informal – shack in backyard, Informal – shack not backyard, Other, Room or granny flat or large dwelling, Semi-detached house, Townhouse, Traditional dwelling – hut or Overcrowding	Categorical
Transportation	Train, Bus, Taxi, Car, Walk/Cycle or Other	Categorical
Mobile phone	Samsung, Other, Apple, Huawei, Nokia, Blackberry, Sony, LG, HTC, Motorola or Siemens	Categorical
Mobile contract	Prepaid or Contract	Categorical

Table 2: Customer purchasing (transactional) behaviour features

Variable name	Explanation	Scaling
Retail shop names (anonymised)	ShopWrong, Select&Debt, RetailA, Nylonworths, WePay, Kliks, ThisKem, RetailB, JetPlane, Cokcor, VosGroup, MrsFee, RetailC, Inspectets, WoolOn, RetailD, Poems, Kara, MarkHim, Retail E	Categorical
Activities/ Transactions	Retail shop	Categorical
	Transaction date	Date
	Amount spent	Numeric

The authors decided on simulating the datasets according to South African demographics [23]. The datasets contained in the different tables have different distributions so the data are random and more realistic. The authors started the data simulation with a set of assumptions derived from the real world (deductive), and produced simulation-based data that can be analysed (inductive). These assumptions include: (1) customers that are still in school will have an employment status of not economically active and (2) customers below the age range of 25-29 will not be able to have an educational level higher than a first degree or diploma qualification.

Two software packages were used to support this project, namely Matlab® and Microsoft® (MS) SQL Server®. Matlab is a high-level language and interactive environment for numerical computation, visualisation and programming and has the ability to access a database server and then perform data manipulation. To access data from Matlab a data source and connection to MS SQL Server database is necessary. The Database Explorer app (in Matlab) accesses the Microsoft ODBC Data Source Administrator automatically when configuring an ODBC data source. Figure 4 conceptualises the connection between Matlab and MS SQL Server, also indicating that an ODBC connection is created.

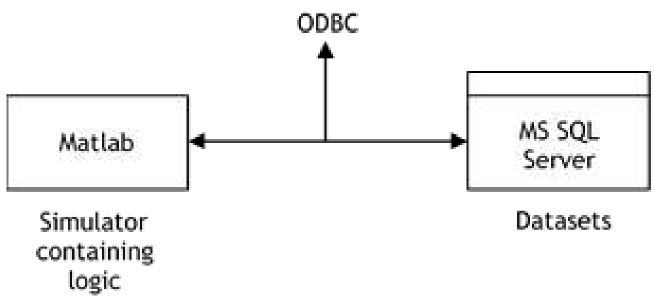


Figure 4: Conceptual illustration of the connection between Matlab and MS SQL Server

4.2 Application of proposed tool

The goal of this research is to develop a customer super-profiling demonstrator. In order to create this demonstrator, a simulator was developed which creates various datasets. These datasets contain demographic, typical transactional and personal preference information, of 50 000 customers. The demonstrator will make use of various techniques to analyse the data in its control. The techniques as well as the results of each technique will subsequently be discussed.

Figure 5 illustrates the broad outline of the CSP tool which will be developed throughout this section. The main steps within this outline include:

1. *Select* data.
2. *Do* RFM analysis.
3. *Do* clustering.
4. *Develop* a predictive model.

The subsections to follow will visit these steps individually, explain what each of them means and what needs to be done, as well as document relevant results retrieved during the steps.

4.2.1 Select data

The first step (1) in Figure 5 indicates that data needs to be selected. Large amounts of information, if used correctly, can help generate important patterns and trends. These patterns provide useful insights into customer purchasing behaviour, and when used in combination with customer demographic information, even more insights can be generated, and powerful customer profiles created.

The simulated customer data is stored in MS SQL Server, which functions as the database, and can be accessed in Matlab by using a *selectquery* command. Relevant datasets are created along various dimensions and imported into Matlab for analysis. The data must first be 'cleaned'. Next, the RFM analysis will be performed.

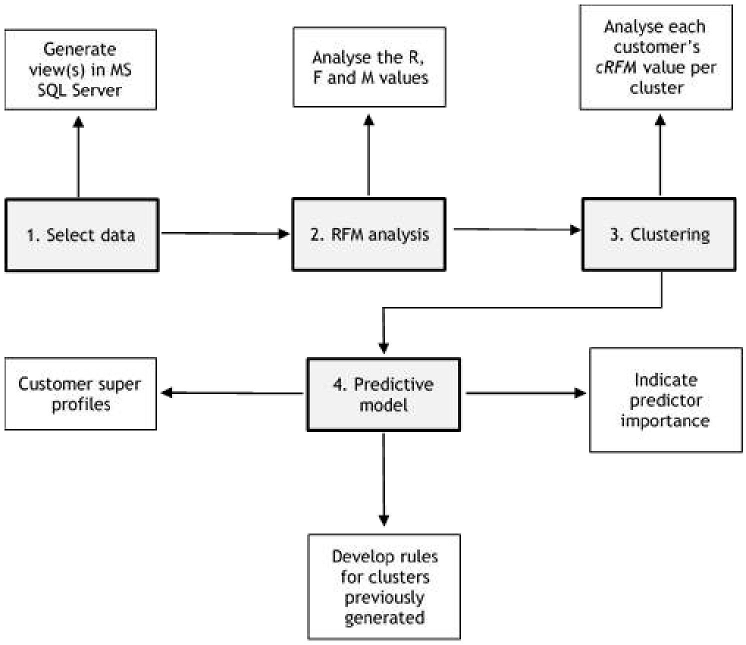


Figure 5: Schematic representing the CSP tool

4.2.2 RFM analysis

After selecting the appropriate customer information, the second step includes the *RFM analysis*. The RFM model is one of the best-known customer value analysis methods, which extracts characteristics of customers using fewer criteria as clustering attributes to reduce the complexity of the model. The RFM analysis will be performed on the dataset created in the previous step (1): this dataset includes all *retail shops* each customer visited, will not focus on only one shop or products purchased. The values that will be obtained from this analysis will provide insights into the individual R, F and M values of each customer taking into consideration all the retail shop visits of each customer, and will be used to fulfil the next step, namely *clustering*.

The transaction data, which forms part of the behavioural feature called activities (Table 2) was simulated to vary from 01/01/2015 to 31/12/2016 (two years). The RFM method was implemented as follows:

- *Recency (R):* It represents the interval between the customer's latest active date and the date selected as the last date (31/12/2016). The older the active date, the lower the recency category of that customer.
- *Frequency (F):* It represents the number of times a customer was active during the specified period for this study. The higher the number of transactions in an interval, the higher the frequency category
- *Monetary(M):* It represents the monetary value of the purchases in the specified period for this study. The higher the amount spent by a customer, the higher the monetary category. The average amount spent by each customer is calculated by adding all the money spent by a customer and dividing that amount by their frequency value. This average amount spent is used to allocate a monetary value to each customer.

Figure 6 schematically represents the R, F and M category range values.

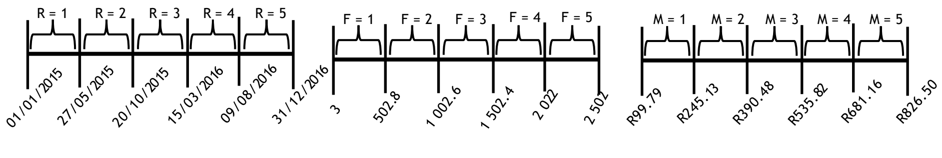


Figure 6: Schematic representing the R, F and M category range values

Table 3 indicates the transformation of the data after each customer has been assigned their R, F and M category value of the transactional data. After assigning each customer their own R, F and M value, it can be stated that more than 80% of the customers have a high R-value; the majority of customer have an F-value equal to 3, and lastly more than 80% of the customers have a low M-value. This data is now ready for the next step, which is clustering.

Table 3: Top five rows of customer RFM matrix

Customer_ID	*R*	*F*	*M*
2	5	3	1
3	5	4	1
4	5	3	1
5	5	3	1
7	5	4	1

4.2.3 Clustering

Clustering is the process of grouping similar objects. Clustering is performed on the RFM dataset created in the previous step (2). The *k*-means clustering method will be applied to this dataset. The best number of clusters is determined using Matlab to yield silhouette plots. The number of clusters associated with the highest silhouette value is seen as the best number of clusters. However, as indicated by literature, the selection of the best number of clusters is subjective. The silhouette values provide a reasonable indication of the cluster structure; the higher the silhouette value, the better the cluster structure. Figure 7 indicates the silhouette criterion value for each number of clusters tested. The best number of clusters suggested by this function is 10, with a silhouette value of 0.957. As mentioned, this is only a suggestion, and the authors decided to select two clusters for this dataset (silhouette value of 0.84). Various aspects were considered for not selecting the best number of clusters, but rather two clusters, such as marketing cost, the more clusters the more marketing efforts need to be funded and smaller clusters lead to more analysis possibilities. The number of customers that need to be clustered is 33 510, and with ten clusters, each cluster would possess a small group of customers. Therefore, two clusters is a reasonable number to perform analysis on, and the clusters will be divided more equally.

In total 50 000 customers were simulated, and only 33 510 customers participated in the RFM analysis. The remaining 16 490 customers did not visit a retail shop (those involved in this study) and will form by default their own cluster. Figure 8 illustrates a scatter plot of the two clusters that were formed when applying the *k*-means clustering method. This figure indicates that if only the recency and frequency values were considered, the centroids would lie very close to each other (almost overlap), therefore a third dimension, the monetary values, contributes to creating a suitable cluster structure for this dataset, dividing the two centroids along the vertical axis. The results received from Table 4 (percentage customers in each RFM category per cluster) are consistent with the structure and positioning shown in Figure 8.

As indicated by Table 4, cluster 1 has the majority observations, followed by cluster 3; which is the non-shoppers, followed by cluster 2. The biggest difference between (the customers of) cluster 1 and cluster 2 lies in the frequency parameter. This means that the customers allocated to cluster 1 are more frequent customers and could be seen as more loyal customers, as opposed to the customers allocated to cluster 2. ("Loyal" in this context is towards the retail stores involved in this study.)

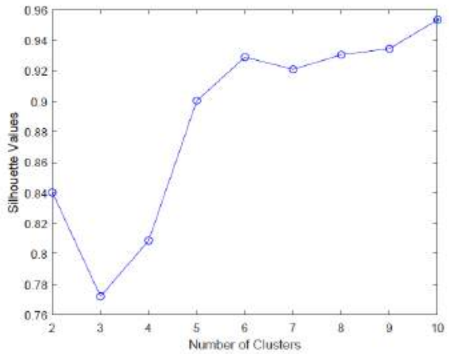


Figure 7: Plot of the silhouette criterion values for each number of clusters tested

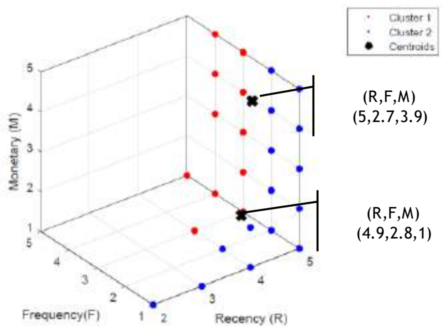


Figure 8: Scatter plot representing the two clusters

Table 4: Summary of the customers in each RFM category per cluster

	Cluster 1			Cluster 2			Cluster 3
	R	F	M	R	F	M	Not applicable
1	0	0	90.46%	0	4.27%	81.44%	
2	0	0	1.44%	0.02%	95.73%	10.84%	
3	0	70.77%	1.18%	0.05%	0	1.59%	
4	0.28%	28.38%	6.40%	33.74%	0	4.87%	
5	99.72%	0.85%	0.51%	66.19%	0	1.26%	
	21 671			11 839			16 490

For deeper insights into the customers' value and purchasing behaviour, we defined a *combined* RFM for each customer, calling it cRFM. This value is determined by adding each customer's R, F and M category value and then dividing the total by 3, as follows:

$$cRFM/customer = \frac{R+F+M}{3}.$$

Each customer has their own cRFM category value, and knowing this value provides a different perspective of the customers. The customers can easily be compared with each other when assigning a combined (cRFM) value to them. With a big customer dataset it is necessary to be able to compare customers and draw conclusions based on the comparisons. It is, however, still possible to interpret each customer's RFM category values separately, if needed. Many decision-making domains use some form of scoring with a single value to distinguish between alternatives/candidates.

Table 5 represents a summary of the percentage customers in each cRFM category, for clusters 1 and 2. The majority of customers in cluster 1 have high cRFM values, whereas the majority of customers in cluster 2 have lower, more distributed cRFM values. These two clusters sufficiently separate the two types of customers present in the dataset. Cluster 1, which contains the more *'loyal'* customers is the bigger cluster, whereas cluster 2 which contains the less *'loyal'* customers is smaller.

Table 5: Percentage customers in each cRFM category for clusters 1 and 2

	Cluster 1 (21 671)	Cluster 2 (11 839)
cRFM = 1	0%	0.068%
cRFM = 2	0%	35.73%
cRFM = 3	61.50%	56.51%
cRFM = 4	31.56%	6.48%
cRFM = 5	6.94%	1.22%

Now, the researcher will stray from the classical application of the RFM analysis to achieve marketing intelligence. With the help of the two clusters and the customers' cRFM values, it is possible to discover various present *types of customers* in the dataset.

By grouping (clustering) and 'ranking (scoring)' (cRFM category values) customers, the decision-maker can differentiate between types of customers and target them based on predefined and justified values instead of blindly reaching out to every customer. The researcher considered both the cRFM category values as well as the individual RFM category values to determine the association between the various types of customers and cRFM category values. It is possible to consider the individual RFM category values when working with the cRFM category values, for they consist of the RFM category values. A study conducted by [25] made use of a *'RFM pattern'* to indicate how the RFM category values of each customer segment differ from the 'original' RFM category values (Table 4) in the dataset.

The authors decided to adapt this RFM pattern technique by applying it to the cRFM category values, illustrating how each cRFM category (consisting of R, F and M parameters) differs from the original RFM category values in each cluster. The total average RFM category values in both clusters are compared with the average RFM category values which constitute the cRFM category values. If the average R (F, M) category value present within each cRFM category (*e.g.*the RFM category values which are combined to form the cRFM categories equal to 1, 2, …, 5) exceeds the total average R (F, M), then an upward arrow (↑) is shown; otherwise, a downward arrow (↓) is shown.

The five customer types that were decided on are listed in Table 6, together with various customer characteristics' explanations, adapted from research conducted by [26], the customers' identification traits (cRFM category values) and the RFM patterns of each cRFM category.

Table 6: Types of customers in dataset

Type of customer	Customer characteristics:	Cluster 1	Cluster 2
(New) low spenders	These customers have made significant low purchases on their (first) buying experience.	cRFM = 3 R↓F↓M↓	cRFM = 2 R↓F↓M↓
(New) big spenders	These customers, as opposed to the new low spenders, have made significant high purchases on their (first) buying experience. These customers are wealthy and will spend their money over a lifetime of their relationship with a brand(s). They usually content themselves with a few big purchases, or a few small ones.	cRFM = 5 R↑F↓M↑	cRFM = 4 R↑F↑M↑ cRFM = 5 R↑F↑M↑
Low loyal customers	These customers buy often but are not able to spend more than they can afford or more than they think something should cost. These customers make	cRFM = 4 R↑F↑M↓	cRFM = 3 R↑F↑M↓

	purchases carefully but trust the retail groups that they support.		
Churned cheap customers	These customers spend as little as possible, buy very few goods and their purchase history is from a long time ago. It is extremely unlikely that these customers are a source of repeat purchases. Marketers believe that these customers are not worth time and trouble.		cRFM = 1 R$\downarrow$F$\downarrow$M$\downarrow$
Prospects	No transactions are registered in the database; only demographic information is available.	Only cluster 3.	

Next, a prediction model that includes the customer demographic information will be developed, according to the various types of customers.

4.2.4 Predictive model

A customer segment, or cluster, is not sufficient to identify and then ultimately predict a customer's behaviour. Many researchers believe that the RFM values of customers are generally associated with customer profiling [17]. Integrating the RFM analysis with both clustering (step 3) and classification provides useful information for current and new customers and more behavioural knowledge of customers is attained; as opposed to other independent clustering and classification techniques.

According to [27] using a decision tree in conjunction with other data mining techniques, such as unsupervised learning (*k*-means) which determines whether concept structures exist within the dataset, would provide a good, if not complete implementation of a data mining process.

For these reasons, decision rules were discovered using the customers' demographic and extra value adding features (age, gender, province, medical aid, mobile phone type, *etc*.), as seen in Table 1, to generate customer super-profiles for the various types of customers (Table 6). The cRFM values allocated to each type of customer refine the customer profiles, forming the *super-profiles*. Decision rules are extracted from generated decision trees. This is called an indirect method of creating decision rules. The decision rules may not be mutually exclusive, meaning more than one rule may cover the same instance.

In total 18 decision rules were generated for this study to identify a type of customer. Table 7 shows a selection of four decision rules that are utilised when predicting a customer's type (*i.e.* (new) low spenders, low loyal customer, (new) big spender, or prospect). This set of rules can provide (1) customer super-profiles for each type of customer and (2) classify new/future customers.

Table 7: Selection of decision rules to identify the type of customer

Rule 2:	**if** *Province* = Eastern Cape or Free State and *Age* = 20-24, 30-34, 35-39, 40-45, 50-54, 65-69 or 80+ and *HouseholdSize* = 1, 2, 4, 6, 7 or 10+ and *Education* = Less than Gr.12 and with diploma or certificate or Degree or post graduate degree and *MobilePhoneType* = Apple, Nokia, Blackberry, HTC or Siemens **then** Low loyal customer.
Rule 3:	**if** *Province* = Eastern Cape or Free State and *Age* = 20-24, 30-34, 35-39, 40-45, 50-54, 65-69 or 80+ and *HouseholdSize* = 1, 2, 4, 6, 7 or 10+ and *Education* = Less than Gr.12 and with diploma or certificate or Degree or post graduate degree and *MobilePhoneType* = Samsung, Other, Huawei, Sony, LG or Motorola **then** Prospect.
Rule 8:	**if** *Province* = Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West or Western Cape and *EmploymentStatus* = Employed or Unemployed and *HouseholdSize* = 1, 2, 3, 4, 5, 6 or 9 and *HousingType* = Townhouse or Traditional dwelling—hut and *MobilePhoneType* = Samsung, Other, Huawei, Nokia, Blackberry, LG or HTC **then** (New) low spender.
Rule 14:	**if** *Province* = Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West or Western Cape and *EmploymentStatus* = Not economically active and *Age* = 30-34, 35-39 or 40-44 and *HouseholdSize* = 2, 7, 8 or 9 and *MobilePhoneType* = Motorola **then** (New) big spender.

The decision rules are developed by determining the most *distinguishing* customer feature within the dataset, for example, this feature would be *province* for the decision rules shown in Table 7; then a rule is *formulated* to 'divide' the dataset into various groups (in this case provinces). The customers are classified as either belonging to 'Eastern Cape or Free State' (Rules 2 and 3) or to 'Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape, North West or Western Cape' (Rules 8 and 14). Those customers that belong to the first group of provinces have a different second customer feature that are used to distinguish then further, namely *age*, as opposed to the other group of customers (*employment status)*. This process is repeated until the rules contain all the customer features, no distinguishing customer features are present, or it is preferred that the rules only contain certain customer features.

Not all types of customers are restricted to one cluster. Therefore, after the customer type is known (Table 7), the cluster to which that customer belongs can be determined. A set of decision rules for each type of customer is constructed. These rules can be used to (1) predict to which cluster a specific type of customer belongs *via* a customer super-profile as well as (2) provide customer super-profiles for targeted marketing campaigns, when the type of customer (*e.g.* low loyal customer) is known. As noted before, the cRFM values allocated to each type of customer refines the customer profiles, forming the *super-profiles*.

For *(new) low spenders*, only customers that belong to the specific groups indicated in Table 6, columns three and four were selected. Again, only a selection of four rules are shown, as seen in Table 8. The rules that are reported have a misclassification rate of 9.3 percent. This means that, for instance, at least 90.7 percent of customers following rule 1 are in cluster 1. The most distinguishing customer feature for this set of rules is the *employment status*.

Table 8: Selection of decision rules to identify (new) low spenders

Rule 1:	**if** *EmploymentStatus*= Employed or Unemployed **then** Cluster 1.
Rule 3:	**if** *EmploymentStatus*= Not economically active and *Province* = Eastern Cape or Free State and *Age*= 30-34, 35-39, 65-69 or 75-79 and *MobilePhoneType* = Samsung, Other, Huawei, Nokia, Blackberry, Sony, LG or Motorola and *RelationshipStatus* = Married/domestic partner, Never married/single or Widowed **then** Cluster 1.
Rule 4:	**if** *EmploymentStatus*= Not economically active and *Province* = Eastern Cape or Free State and *Age*= 30-34, 35-39, 65-69 or 75-79 and *MobilePhoneType* = Samsung, Other, Huawei, Nokia, Blackberry, Sony, LG or Motorola and *RelationshipStatus* = Divorced **then** Cluster 2.
Rule 5:	**if** *EmploymentStatus*= Not economically active and *Province* = Eastern Cape or Free State and *Age*= 30-34, 35-39, 65-69 or 75-79 and *MobilePhoneType* = Apple **then** Cluster 2.

Table 9 shows four selected decision rules generated to identify (new) big spenders and the cluster to which they belong. The misclassification error for this set of rules is 10.9 percent, with the most distinguishing customer feature as the *age* category.

Table 9: Selection of decision rules to identify (new) big spenders

Rule 1:	**if** *Age*= 15-19, 20-24, 25-29, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 or 80+and *Gender* = Male and *ChildrenStatus* = Yes and *Age* = 20-24, 45-49, 65-69, 75-79 or 80+ **then** Cluster 1.
Rule 3:	**if** *Age*= 15-19, 20-24, 25-29, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 or 80+ and *Gender* = Male and *ChildrenStatus* = Yes and *Age* = 15-19, 25-29, 50-54, 55-59, 60-64 or 70-74 and *HousingType* = House, flat or room in backyard, Informal—Shack in backyard or Room, granny flat or large dwelling **then** Cluster 2.
Rule 6:	**if** *Age*= 15-19, 20-24, 25-29, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79 or 80+ and *Gender* = Male and *ChildrenStatus* = No and *HousingOwnership*= Owned (not fully), Owned (fully), Occupied rent free or Other **then** Cluster 1.
Rule 10:	**if** *Age*= 1 30-34, 35-39 or 40-44 and *Education* = Less than Gr.12 and no other qualification, Less than Gr.12 and with diploma or certificate, Gr.12 or Gr.12 with diploma or certificate and *Age* = 35-39 or 40-44 and *HouseholdSize* = 3, 4, 7, 8 or 10+ and *Province* = Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, Northern Cape or North West **then** Cluster 2.

Next, the decision rules for low loyal customers are shown in Table 10. The misclassification rate for this decision tree is 9.8 percent. The distinguishable customer feature is the customer's *province*.

Table 10: Selection of decision rules to identify low loyal customers

Rule 2:	**if** *Province* = Eastern Cape or Free State and *AnnualIncome* = R12 001-R54 000 or R54 001-R192 000 and *HousingOwnership* = Owned (not fully) and *HousingType* = Cluster house in complex, Flat or apartment in flat block, House or brick structure on yard or strand, Informal—Shack in backyard, Informal—Shack not backyard, Other, Room, granny flat or large dwelling, Semi-detached house, Townhouse, Traditional dwelling—hut or Overcrowding **then** Cluster 2.
Rule 5:	**if** *Province* = Eastern Cape or Free State and *AnnualIncome* = R12 001-R54 000 or R54 001-R192 000 and *HousingOwnership* = Rented, Owned (fully), Occupied rent free or Other and *RelationshipStatus* = Widowed or Divorced and *Age* = 60-64 or 75-79 **then** Cluster 1.
Rule 8:	**if** *Province* = Eastern Cape or Free State and *AnnualIncome* = R0-R12 000, R192 001-R360 000 or More than R360 000 and *HousingType* = House or brick structure on yard or strand, House, flat or room in backyard, Informal—Shack in backyard, Informal—Shack not backyard, Other, Townhouse, Traditional dwelling—hut or Overcrowding and *HousingOwnership* = Other and *Age* = 15-19, 20-24, 25-29, 30-34, 40-44, 50-54, 60-64, 65-69 or 70-74 **then** Cluster 2.
Rule 10:	**if** *Province* = Gauteng, KwaZulu-Natal, Limpopo, Mpumalanga, North West, Northern Cape or Western Cape and *EmploymentStatus* = Employed or Unemployed **then** Cluster 1.

As indicated in Table 6, the churned cheap customers are not of interest to marketers. However, in this customer dataset, churned cheap customers only belong to cluster 2, have a cRFM value of 1, and only make up 0.068 percent of the customers within cluster 2. No decision rules are necessary to profile this type of customer. The last type of customer is the *prospects*. These customers only belong to cluster 3, for no transactional information is registered in the database for such customers, only demographic and extra value adding features. Therefore, no additional decision rules are necessary.

Comparing the individual decision rules for the type of customers (Table 8—Table 10), it is found that the rules used to identify (new) low spenders (Table 8) have the lowest misclassification rate. This can be as a result of the (new) low spenders being the biggest customer dataset, thus having more training and testing data.

This concludes the discussion regarding the development of the predictive model. The purpose of this article was to develop a tool that demonstrates the generating of customer super-profiles given a specific dataset, through performing the steps illustrated in Figure 5. The predictive model for this study, which forms part of the CSP tool, includes various sets of decision rules, all leading to creating customer super-profiles for the five customer types present within the dataset. Marketers using this tool will have more knowledge of their customers especially when they know which type of customer they are interested in.

5. BUSINESS CASE

The previous sections indicated how to develop a customer *super-profile* to enable efficient targeting in marketing campaigns. The business case to illustrate the value added by this project is as follows.

Being able to identify *who* to target, as well as *where* and *how* to advertise marketing campaigns is an important task. The proposed demonstrator has the ability to run a deterministic audience discovery to reveal customer profiles for the marketers. These profiles contain demographic information, typical transactional data as well as customer preferences. The demonstrator can be used when marketing a product for a certain target group is necessary. This target group can be 'found' by the demonstrator and will provide all relevant information regarding that group, *e.g.* demographic information, transportation type, mobile phone ownership and activities as well as RFM values. This type of information will provide more insight into the customers and will decrease the frustration experienced by marketers when performing 'tossing a coin' type of targeting.

6. CONCLUSIONS

Customer super-profiling consists of a large set of analysis models that could be used for predicting the behaviour and characteristics of current and/or new customers, which enables efficient targeting in marketing campaigns. This paper presents a proposed customer profiling demonstrator by incorporating the RFM analysis into data mining techniques to provide marketing intelligence. Initialising the study by performing the RFM analysis draws attention to the importance and advantages of this analysis. In order to evaluate the proposed demonstrator and show the benefit of using it in direct marketing, a customer dataset was simulated containing 50 000 customers, with purchasing behaviour over a two-year period. The results received consider several parameters together, such as the customer clusters, the *cRFM* values of the customers, as well as potential future customer behaviour. Industrial engineers, with their understanding of systems and system integration as well as analytical knowledge, should find these challenges exciting and relevant in our modern world. With the fourth industrial revolution upon us, who better to lead it than the adaptable and flexible industrial engineers.

REFERENCES

[1] **Walters, M. and Bekker, J.** 2017. Customer super-profiling demonstrator to enable efficient targeting in marketing campaigns, *South African Journal of Industrial Engineering,* 28(3), pp. 113-127.

[2] **Hosseini, M. and Shabani, M.** 2015. New approach to customer segmentation based on changes in customer value, *Journal of Marketing Analytics,* 3(3), pp. 110-121.

[3] **Brown, O.** 2016. The Importance of Customer Profiling, http://www.hellostarling.com/the-importance-of-customer-profiling/. Accessed 21 May 2018.

[4] **Shaw, M., Subramaniam,W., Tan, W. and Welge, M.** 2001. Knowledge management and data mining for marketing, *Decision Support System,* 31(1), pp.127-137.

[5] **Chen, T. and Chen, C.** 2010. Application of data mining to the spatial heterogeneity of foreclosed mortgages, *Expert Systems with Applications,* 37(2), pp.993-997.

[6] **Weinstein, A.** 2013. *Handbook of Market Segmentation: Strategic Targeting for Business and Technology Firms*, 3rd edition, USA: Taylor & Francis.

[7] **Jansen, S.** 2007. Customer Segmentations and Customer Profiling for a Mobile Telecommunications Company Based on Usage Behaviour: A Vodafone Case Study. Masters thesis, Maastricht: University of Maastricht.

[8] **Chan, C.C.H.** 2008. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, *Expert systems with applications,* 34(4), pp.2754-2762.

[9] **Bose, I. and Chen, X.** 2009. Quantitative models for direct marketing: A review from systems perspective, *European Journal of Operational Research,* 195(1), pp.1-16.

[10] **Jonker, J.-J., Piersma, N. and Van den Poel, D.** 2004. Joint optimization of customer segmentation and marketing policy to maximize long-term profitability, *Expert Systems with Applications,* 27(2), pp.159-168.

[11] **Chan, C.** 2008. Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer, *Expert System Application*, 24(4), pp. 2754-2762.

[12] **Kim, S., Jung, T., Suh, E. and Hwang, H.** 2006. Customer segmentation and strategy development based on customer lifetime value: A case study, *Expert System Application*, 31(1), pp. 101-107.

[13] **Sarvari, P. A., Ustundag, A. and Takci, H.** 2016. Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis, *Kybernetes*, 45(7), pp. 1129-1157.

[14] **Dursun, A. and Caber, M.** 2016. Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis, *Tourism Management Perspectives*, 18(1), pp.153-160.

[15] **Wei, J.T., Lin, S.Y. and Wu, H.H.** 2010. A review of the application of RFM model, *African Journal of Business Management*, 4(19), pp.4199-4206.

[16] **Hsieh, N.C.,** 2004. An integrated data mining and behavioral scoring model for analyzing bank customers, *Expert systems with applications*, 27(4), pp.623-633.

[17] **Nimbalkar, D. D. and Shah, P.** 2013. Data mining using RFM Analysis, *International Journal of Scientific & Engineering Research*, 4(12), pp.940-943.

[18] **MathWorks**.2018. Kmeans, https://www.mathworks.com/help/stats/kmeans.html#bueftl4-1. The MathWorks, Inc. Accessed 24 January 2018.

[19] **Trevino, A.** 2016. Introduction to K-means Clustering. https://www.datascience.com/blog/k-means-clustering. Accessed 23 May 2018.

[20] **Kaufman, L.** and Rousseeuw, P.J. 2009. *Finding groups in data: an introduction to cluster analysis*, USA: John Wiley & Sons.

[21] **Aptè, C. and Weiss, S.**1997. Data mining with decision trees and decision rules, *Future Generation Computer Systems*, 13(219), pp. 197-210.

[22] **Dori, D.** 2002. *Object-process methodology: A holistic systems paradigm*. Germany: Springer Science & Business Media.

[23] **Statistics South Africa,** 2017. *Mid-year population estimates 2017*, http://www.statssa.gov.za/publications/P0302/P03022017.pdf . Accessed 1 October 2017.

[24] **Aggarwal,C.C. and Reddy, C.K.** 2014. *Data clustering: Algorithms and Applications*. USA: CRC press.

[25] **Shih, Y.-Y. and Liu, C.-Y.** 2003. A method for customer lifetime value ranking – Combining the analytic hierarchy process and clustering analysis, *Journal of Database Marketing & Customer Strategy Management*, 11(2), pp.159-172.

[26] **Evaldas, M.** 2017. *Practical use of RFM customer segmentation*, https://stacktome.com/blog/rfm-customer-segmentation . Accessed 14 May 2018.

[27] **Trewartha, D.** 2006. Investigating data mining in MATALAB. Masters thesis, Department of Science, Rhodes University, Grahamstown. http://pppj2012.ru.ac.za/g03t2052/CSHnsThesis.pdf

PRACTICAL APPLICATION OF AN OBJECTIVE METHODOLOGY TO EVALUATE INDUSTRIAL ENERGY EFFICIENCY MODELS FOR THE RSA SECTION 12L TAX INCENTIVE

L.A. Botes[1], W. Booysen[2], W. Hamer[3] & M. Kleingeld[4]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
lbotes@researchtoolbox.com

[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbooysen@researchtoolbox.com

[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
whamer@researchtoolbox.com

[4]North-West University, CRCED-Pretoria, Pretoria, South Africa
mkleingeld@researchtoolbox.com

ABSTRACT

Section 12L of the Income Tax Act (1962) allows a significant tax rebate for quantified energy efficiency (EE) savings resulting from an energy saving measure (ESM). However, EE savings cannot be directly measured since it refers to the absence of energy use. Therefore, models are used to quantify EE savings.

In order to have a high level of confidence in the reported EE savings, the development of multiple models is recommended. These various models should be objectively evaluated to ensure that the final selected model adheres to the multiple requirements associated with 12L. This paper provides a brief overview of an evaluation and model selection methodology. The paper primarily focusses on the practical application of the methodology on industrial ESMs.

The methodology is applied to three case studies: boilers, blast furnace and compressed air network ESMs. The evaluation results are used to rank the modelling options and recommend a final model. The results obtained are validated by comparing them to independent SANAS accredited measurement and verification results. The results show a 10% to 33% variance in the potential claim value. This significant variance highlights the importance of presenting a transparent and objective model selection process.

This work was sponsored by ETA Operations (Pty) Ltd.

[1] Ms. L.A. Botes holds a Master's of Engineering in Mechanical Engineering from the North-West University. She is currently enrolled for a PhD degree the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr. W. Booysen is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr Waldt Hamer is a holds a PhD in Mechanical Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[4] Prof M. Kleingeld is a registered professional engineer and holds a PhD in mechanical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre of Research and Continued Engineering Development (CRCED) in Pretoria.

1 INTRODUCTION

1.1 Section 12L of the Income Tax Act

Section 12L of the Income Tax Act (1962) was introduced by the National Treasury with the aim to incentivise tax payers to utilise energy efficiently [1]. This is achieved by allowing a tax deduction of 0.95R/kWh for measured and verified energy efficiency (EE) savings [2]. The procedure for applying for the allowance as well as the legal requirements is stipulated in the Regulations in terms of 12L of the Income Tax Act [3].

The Regulations stipulate that the quantification of the EE savings should be done in accordance with the South African National Standard (SANS) for the measurement and verification of energy savings (SANS 50010:2011 - referred to as the Standard) [4]. The Standard provides technical guidance in the form of multiple generic methodologies for the quantification of the EE savings.

In order to receive the 12L allowance the EE savings should be calculated in adherence to both the legal and technical requirements as stipulated in the Regulations and Standard. The guidelines provided by these legal documentations may, however, be implemented with different levels of rigour which would ultimately result in different EE savings results. Therefore, more research is required regarding the various options resulting from the implementation of these guidelines.

1.2 Model development

The EE savings resulting from an energy saving measure (ESM) is quantified by means of EE models. The key technical aspects of the 12L model development process were identified from relevant literature [4] [5]. These aspects will only be discussed briefly. The first key aspect is the selection of a measurement boundary. The Standard allows savings to be determined for a portion of a facility or the whole facility [4]. From the viable options the measurement boundary should be chosen in such a way that it encapsulates the effect of the ESM. The second key aspect is the assessment of the dataset used during model development. The selected measurement boundary determines the measurement points that can be used. The Standard allows either only the significant energy governing factors or all of the parameters associated with the energy system to be used [4]. In order to ensure accuracy of the dataset the Standard deems two sources of data as compliant, namely data obtained from invoices or measurements from calibrated meters.

The third and final key aspect is the development of the EE model. This is done by calculating the EE savings with a mathematical method based on the selected measurement boundary and dataset. Various methods are acceptable, however, the most commonly used methods in practice include energy intensity calculations and regression analyses [6] [7].

The available measurement boundaries, datasets and calculation methods can be paired in numerous ways during model development. Moreover, the current movement to Industry 4.0 will lead to the installation of more process meters and data collection methods. As a result more dataset and measurement boundary options will be available for the determination of EE savings. Previous studies that focussed on the practical application of 12L in industry recommend that multiple models be developed in order to holistically evaluate the EE savings [7] [6].

1.3 Model selection

There is no guidance available in literature on how to evaluate the various developed models or how to select the final model. This becomes critical when considering the direct impact of the chosen model on the monetary value associated with 12L. Since the chosen model should adhere to multiple legal and technical criteria, multi-criteria decision making (MCDM) methods will be used to aid in the selection process. A MCDM evaluates each of the available modelling options and ranks them according to the evaluation. Therefore the use of a MCDM addresses the problem of evaluating the multiple modelling options and selecting a final model for which there is no current literature available.

Various MCDM methods exist and have been used extensively in the energy application field. The most common methods used include the weighted sum method (WSM) [8], technique for the order of preference by similarity to the ideal solution (TOPSIS) [9], analytical hierarchy process (AHP) [10], *elimination et choix traduisant la realité* (ELECTRE) [11] and preference ranking organization method for enrichment evaluation (PROMETHEE) [8]. The most commonly used of the above-mentioned methods is the WSM. The WSM was selected as the preferred MCDM method in this study due to its transparency and simplicity. These are important factors to consider since the final results are generally presented to multiple stakeholders who don't necessarily have the expertise to understand the more sophisticated methods. Furthermore, comparative studies have shown similar results between the WSM and the more complex methods.

This paper is based on a methodology developed by the author [12] who makes use of the WSM to aid in the selection process of a final 12L model between multiple modelling options. The full documentation of the original study is available for detailed information. However, the purpose of this paper is to illustrate the practical application of the methodology on industrial energy saving measures (ESM).

2 OVERVIEW OF METHODOLOGY

This section provides a brief overview of the methodology used to evaluate different 12L energy efficiency models. The methodology consists of three steps. In the first step multiple modelling options are identified, where after each developed model is evaluated in the second step. In the third and final step each of the modelling options are ranked according to their evaluation scores and a final modelling option is selected. The methodology is depicted in Figure 2-1.

Figure 2-1: Three steps of methodology. Adapted from [12]

Identification of modelling options

In the first step of the methodology, all of the available information of the system under investigation is identified. Thereafter, this information is used to identify available modelling options in order to quantify the EE savings associated with an energy saving measure (ESM). Figure 2-2 illustrates the identification of various measurement boundaries and dataset options by the use of a process flow layout [6]. Various measurement points are identified on the layout, with the level of compliance with 12L Regulations indicated by a colour key.

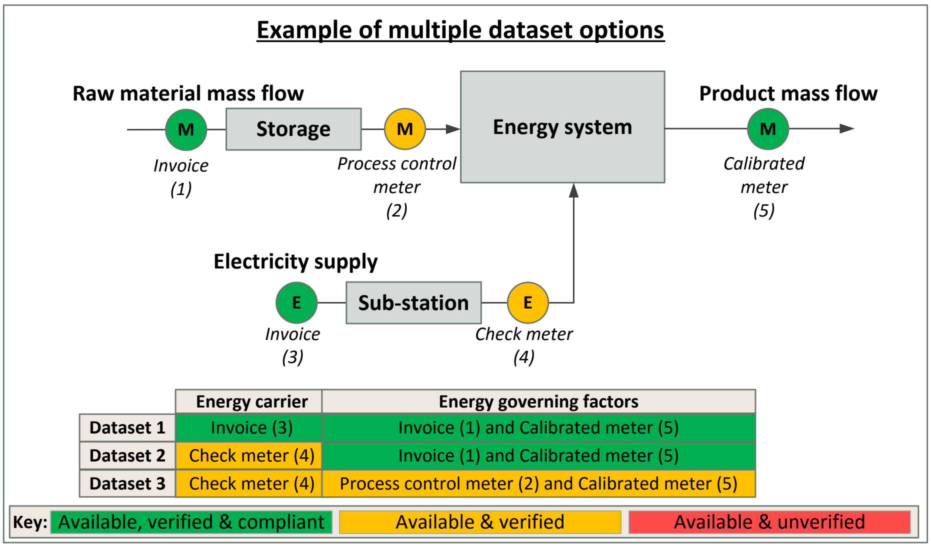


Figure 2-2: Illustration of process flow and multiple dataset options. Adapted from [6]

By the use of a layout, as illustrated in Figure 2-2, it is possible to identify the various possible measurement boundaries that would correctly encapsulate the effect of the ESM under investigation. All of the relevant energy streams and their respective measurement points associated with the process are also identified. This allows the identification of multiple points of measurement for a respective energy stream some of which may be available on different data resolutions (e.g. daily, weekly, monthly), various proximities to the boundary of focus and different levels of legal compliance.

All of the identified information can be used to develop multiple EE models. The models are developed by making use of well-established methodologies available from previous studies and published guidelines [6] [13] [4] [14]. The development of EE models have been extensively discussed in available literature. However, the focus of this study is the evaluation and selection process of the final modelling option, therefore the calculation process of EE savings and model development will not be covered in detail in this paper. The developed models may vary according to the selected measurement boundary, dataset and calculation method. Table *2-1* shows an example of summarising the different modelling options from which the final model should be selected.

Table 2-1: Summary of different modelling options

Model	Boundary	Data	Calculation Method
Model 1	Boundary 1	Dataset 1	Method 1
Model 2	-	-	-

…	…	…	…
Model m	Boundary n	Dataset x	Method z

Evaluation of modelling options
In this step of the methodology the various modelling options are evaluated in order to select a final modelling option which adheres to all of the requirements of 12L. This is done by first identifying the criteria that the modelling options need to adhere to. After a critical analysis of relevant literature, seven evaluation criteria were identified for this study [6] [4].

The first criterion is the compliance of the measurements used during the model development (C_1) while the second criterion considers the conservative nature of the quantified EE savings (C_2). The third and fourth criterion is based on the most prevalent statistical parameters used to evaluate regression models. This includes model correlation (C_3) and the root mean squared error (C_4).

The final three criteria evaluate whether the model is technically sound. The criteria thus include the significance of the savings (C_5), variance in the savings (C_6) and the fraction of energy the model accounts for (C_7). The significance of the savings is indicative that the correct measurement boundary was selected. The variance in savings between the available modelling options identifies extreme models. The fraction of energy the model accounts for is an indication that the relevant parameters are included in the model.

After identifying the evaluation criteria it is necessary to allocate weights to each criterion. The criteria weights represent the relative importance of each criterion. For this study, the criteria weights used were obtained through the use of pairwise comparison matrices as commonly recommended from literature [10] [15]. In order to test the applicability of the methodology ten individuals with experience in the measurement and verification field were asked to complete the pairwise matrices. This may, however, be refined, for example, by including more individuals in the questionnaire to establish the credibility of the results obtained. The weight obtained for each criterion is as follows:

- C_1: Compliance (0.213)
- C_2: Conservativeness (0.165)
- C_3: Model correlation (0.182)
- C_4: RMSE (0.112)
- C_5: Significance of quantified savings (0.126)
- C_6: Variance in savings (0.105)
- C_7: Fraction of energy accounted for (0.096)

After identifying the evaluation criteria and assigning weights to each criterion the modelling options are evaluated. This is done in three main steps. Firstly, a performance matrix is completed. In this matrix each modelling option is assessed according to each criterion.
Various units of measurement may be present in the performance matrix. Therefore, the scores in the performance matrix are normalised in the second step. This is done through the use of a relative strength of preference scale (rating from 0 and 1 in this study). The least and most preferred scenario in terms of each criterion is represented by the anchor points of the scale (0 and 1). This allows the performance scores to be linearly scaled which results in comparable score values between 0 and 1 for each model in terms of each criterion.

Thirdly, the weighted scores are calculated. This is done by multiplying the evaluation score (between 0 and 1) with the respective criteria weight, as shown in Equation 1.

$$weighted\ score = a_{ij}w_j$$

Equation 1: Calculation of weighted score

Where a_{ij} denotes the evaluation score of model i with respect to criterion j and w_j the weight of criteria j. Table 2-2 illustrates the weighted scoring matrix.

Table 2-2: Example of weighted scoring matrix

	C_1	…	C_j	…	C_7
model 1	$a_{11}w_1$	…	$a_{1j}w_j$	…	$a_{17}w_7$
model 2	$a_{21}w_1$	…	$a_{2j}w_j$	…	$a_{27}w_7$
. . .					

model i	$a_{i1}w_1$	…	$a_{ij}w_j$	…	$a_{i7}w_m$
model i +1	$a_{(i+1)1}w_1$	…	$a_{(i+1)j}w_j$	…	$a_{(i+1)7}w_m$
. . .					
model m	$a_{n1}w_1$	…	$a_{nj}w_j$	…	$a_{n7}w_m$

Ranking and recommendation of preferred model

In the last step of the methodology the overall score or multi-criteria score of each modelling option is calculated. This score takes all of the evaluation criteria into consideration. The overall score of each modelling option is equal to the sum of the respective weighted scores in terms of each criterion. The calculation is presented in Equation 2.

$$A_i = \sum_{j=1}^{n} a_{ij} w_j$$

Equation 2: Calculation of overall score

For $i = 1, 2, \ldots, m$ modelling options and $j = 1, 2, \ldots, n$ evaluation criteria. The overall scores are used to rank each modelling option. This allows the identification of the least and most preferred model between the various options according to the methodology. The methodology identifies the optimum solution from the modelling options as the one with the highest overall score.

The overall scores are thus ranked from highest to lowest. Figure 2-3 illustrates an example of four modelling options ranked according to their overall scores.

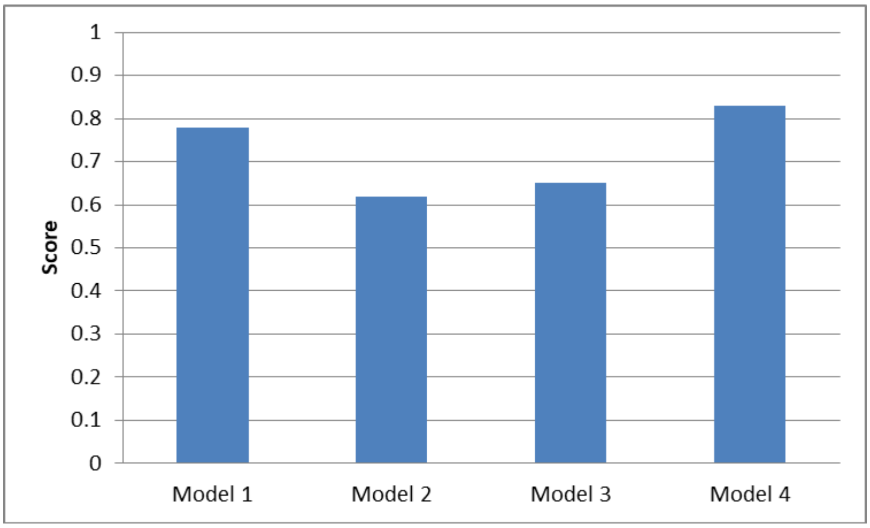


Figure 2-3: Example of ranking obtained from overall scores

From the example in Figure 2-3 it is seen that Model 4 obtained the highest overall score. This means that Model 4 adheres to all of the criteria more satisfactorily than the other available modelling options. The ranking results are thus used to recommend a preferred model for the specific I2L application under consideration. A summary of the three steps of the methodology is shown in Figure 2-4.

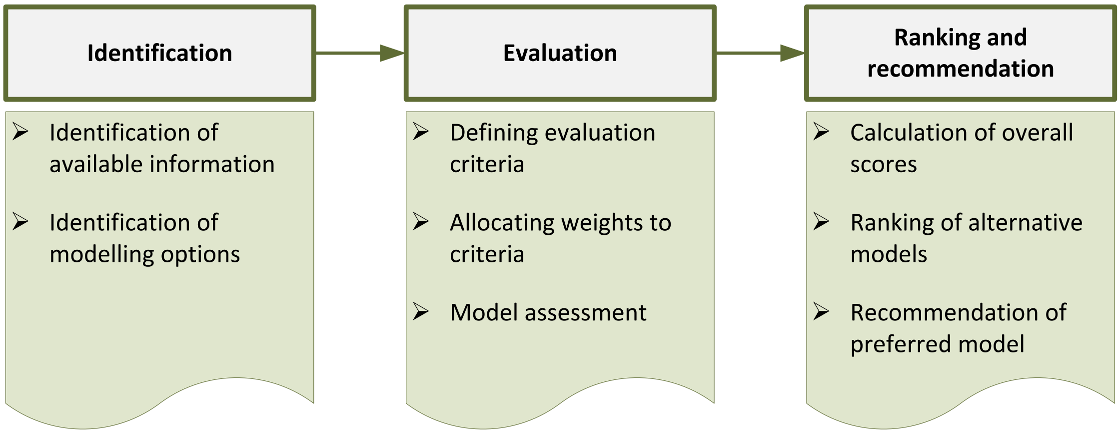


Figure 2-4: Extended steps of methodology. Adapted from [12]

3 CASE STUDY 1: BOILERS

3.1 Basic site description

The first case study focusses on boilers of a large industrial plant. The specific boundary consists of 15 coal fired boilers from which steam is generated. The steam is either used in downstream processes or for electricity generation. Eight of the boilers are situated in steam station 1 with the remaining seven in steam station 2. A basic layout of the boilers is depicted in Figure 3-1.

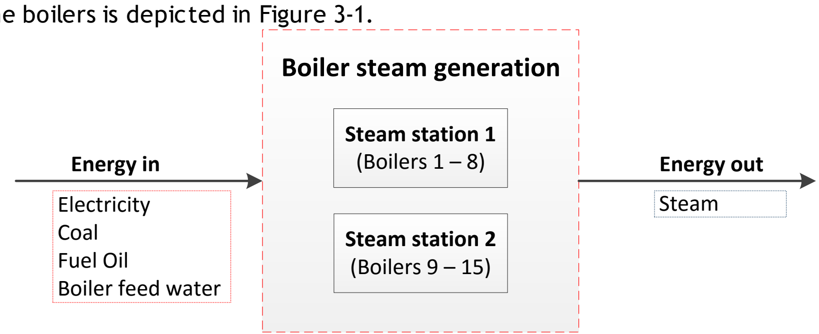


Figure 3-1: Basic layout (case study 1)

During the assessment period, the water and coal qualities were improved which resulted in measurable energy efficiency savings. This was done by repairing boiler tubes during general overhauls, focussing on the improvement of the water quality management and the assessment of coal received from mines.

3.2 Identification of modelling options

A detailed process flow layout was constructed in order to identify the available information regarding the boilers. The layout is presented in Figure 3-2. The relevant energy streams and measurement points are identified on the layout, while the compliance of each measurement is indicated by a colour key. Selected sections will be discussed in more detail.

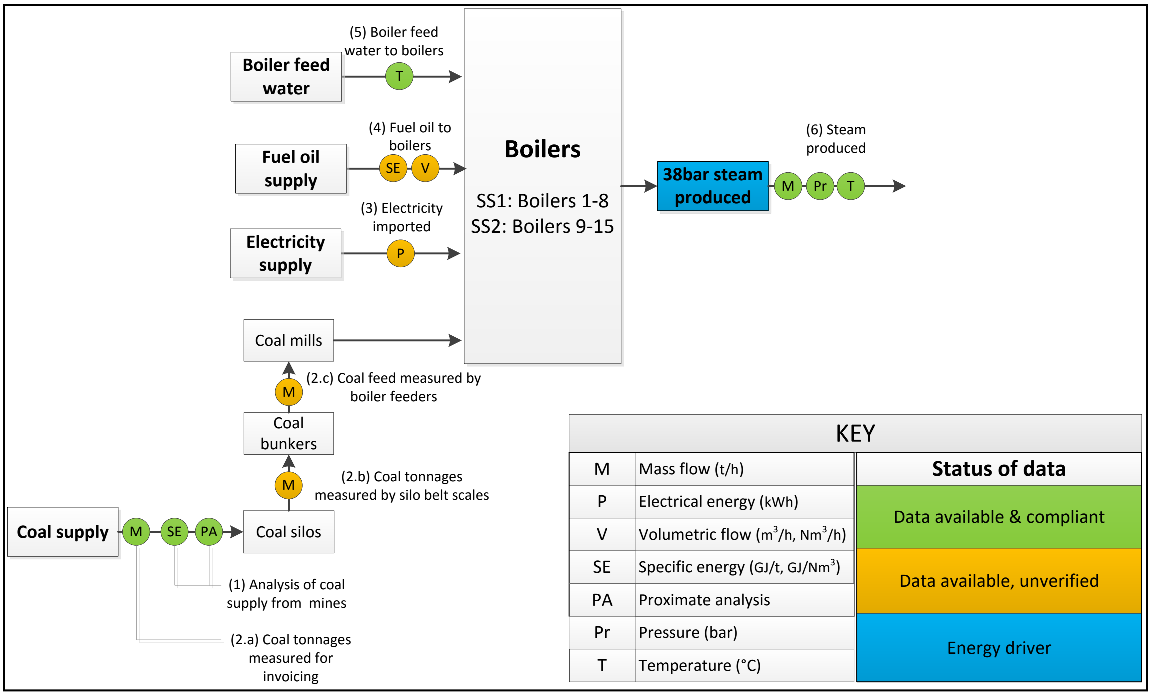


Figure 3-2: Overview of measurement points (case study 1)

Measurement points are represented by coloured circles. Mass measurements are indicated by M, electrical power meters by P, volumetric flow by V, specific energy by SE, proximate analysis by PA, pressure by Pr and temperature by T. The colour key indicates the compliance status of each measurement. Green represents available and 12L compliant data, while yellow represents data that is available but has unverified compliance status.

The coal supply to the boilers is measured at three different positions. The measurements taken in closer proximity to the boilers are available per boiler and on a high resolution (daily). Data for the supply of coal (furthest point from the boilers) are, however, only available on a monthly resolution as a total supply to the steam stations. The coal supply is deemed 12L compliant since it is based on invoices, while compliance could not be proven for the higher resolution measurement points.

Measurement points could be identified for the electricity consumption of each steam station and fuel oil feed to each boiler. However, compliance could not be proven for these measurements. Furthermore, relevant and compliant measurements could be identified for the steam production and boiler feed water. These measurements are available per boiler in daily resolutions.

The available information could be paired in several ways to develop a model to quantify the associated EE savings. Strategies available from previous literature and published guidelines were used to select a measurement boundary, evaluate the available datasets and ultimately develop EE models. This resulted in five different and feasible modelling options. The five models are summarised in Table *3-1* according to the various measurement boundaries, datasets and calculation methods used.

Table 3-1: Summary of different modelling options (case study 1)

Model	Boundary	Data	Calculation	Savings (GWh)
1	Total steam stations	Select parameter	Intensity	314.27
2	Total steam stations	Select parameter	Regression model	293.20
3	Total steam stations	All parameters	Regression model	251.40
4	Aggregate of per steam station	Select parameter	Intensity	324.82
5	Aggregate of individual boilers	Select parameter	Intensity	341.19

The measurement boundaries of the models varied between total steam stations, per steam station and individual boilers. This was mainly due to the higher resolution (daily values) but non-compliant coal measurements used. The data selection of the various models varied between select parameter and all parameter approaches. The select parameter models exclude boiler feed water as an energy carrier, while it is included in the all parameter approach model (Model 3).

Furthermore, the calculation methods used during model development were either an energy intensity calculation or a regression analysis. The savings quantified by using these models varied between 251.4 GWh and 341.2 GWh. In order to determine which modelling option should be selected as the final 12L modelling option each of the five developed models will be evaluated in the next section.

3.3 Evaluation of modelling options

After developing the various modelling options to assess the EE savings associated with the boilers, the models were evaluated. The first step of the evaluation phase is to complete a performance matrix. In the performance matrix each developed model was assessed according to its performance with respect to each evaluation criteria. Table 3-2 shows the performance of each of the five models with respect to the seven evaluation criteria.

Table 3-2: Performance matrix (Case study 1)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
	%	GWh	-	%	%	GWh	%
Model 1	50.00	314.27	-	-	2.89	9.29	87.60
Model 2	50.00	293.20	0.77	5.49	2.70	11.78	87.60
Model 3	60.00	251.40	0.75	6.49	2.31	53.58	100.00
Model 4	25.00	324.82	-	-	2.99	19.84	87.60
Model 5	33.33	341.19	-	-	3.14	36.22	86.48

Table 3-2 shows that none of the models are 100% compliant (C_1). This was due to the fact that compliance could not be proven for electricity and fuel oil measurements. It is noted that Model 4 and Model 5 received lower scores for compliance than the other models. The reason for this is that higher resolution (daily values) but non-compliant coal measurements were used in these models. The quantified savings of the five models ranged from 251.4 GWh to 341.2 GWh (C_2). Since Model 3 was the most conservative model (EE savings of 251.4 GWh), the model had the largest variance from the average quantified savings (C_6).

Only Model 2 and Model 3 obtained scores for the model correlation (C_3) and RMSE (C_4) criteria. This is due to Model 2 and Model 3 being regression analyses while the other three models are based on intensity calculations. The significance of the quantified savings for each of the five models was relatively the same (C_5).

Model 3 was based on an all parameter approach and therefore accounts for 100% of the energy within the selected measurement boundary (C_7). Model 5, however, accounted for the least amount of energy. This was due to electricity being excluded from the model since it was not available for the selected measurement boundary (per individual boiler). The rest of the models (Model 1, Model 2 and Model 3) were all based on a select parameter approach which excluded boiler feed water and thus accounted for the same amount of energy.

Next, the performance scores were converted to values between 0 and 1. This method was discussed in detail in the methodology section. After normalisation of the scores; the weighted scores were determined. This was done by multiplying the score values (between 0 and 1) with the respective criterion weight. The weighted scores for each of the five models are presented in Table 3-3.

Table 3-3: Weighted scoring matrix (Case study 1)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
Model 1	0.11	0.05	0.00	0.00	0.12	0.09	0.08
Model 2	0.11	0.09	0.14	0.07	0.11	0.08	0.08
Model 3	0.13	0.17	0.14	0.06	0.09	0.00	0.10
Model 4	0.05	0.03	0.00	0.00	0.12	0.07	0.08
Model 5	0.07	0.00	0.00	0.00	0.13	0.03	0.08

3.4 Ranking and recommendation of preferred model

In the final step of the methodology the overall scores for each model was calculated. This was determined as the sum of the seven weighted scores of each criterion. The overall scores of the five models are shown in Table 3-4.

Table 3-4: Overall score of each modelling option (Case study 1)

Model	Overall score
Model 1	0.179
Model 2	0.275
Model 3	0.276
Model 4	0.143
Model 5	0.127

The overall scores shown in Table 3-4 were used to rank the modelling options, which ultimately aids in the selection process of a final 12L modelling option. Figure 3-3 visually illustrates the overall scores of each of the five modelling options. Each model is ranked from worst to best according the overall scores and numbered from 1 to 5.

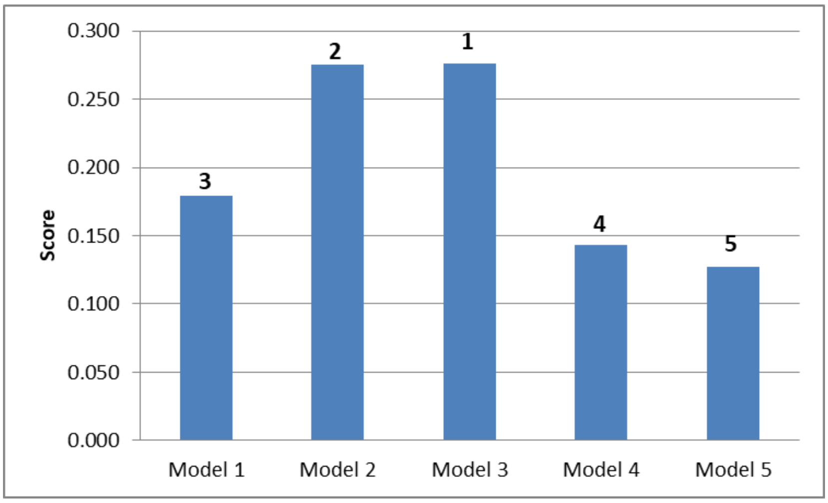


Figure 3-3: Summary of overall score of each modelling option (Case study 1)

Figure 3-3 shows that, although Model 2 and Model 3 obtained similar scores, Model 3 was ranked first place due to obtaining the highest overall score. Furthermore, it can be seen that Model 5 obtained the lowest overall score and was ranked in last place. Figure 3-4 presents more detail of the overall scores by considering each criterion's contribution to the final scores.

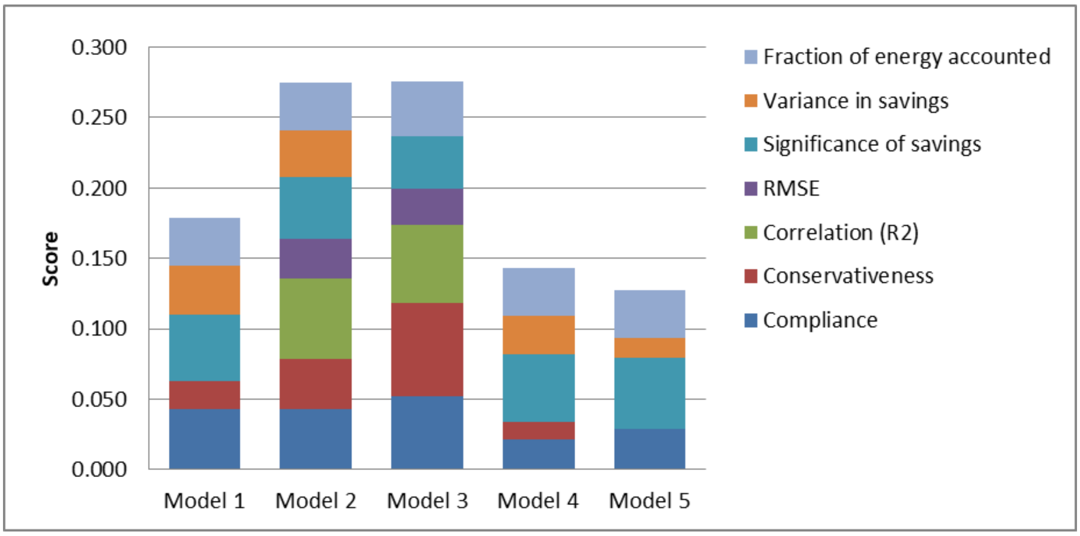


Figure 3-4: Summary of results (Case study 1)

It is observed from Figure 3-4 that Model 2 and Model 3 obtained higher scores than the other models. By evaluating the contribution of each criterion to the final scores it is visible that the higher scores were obtained due to being regression models. These two models could thus obtain scores with respect to the model correlation and RMSE evaluation criteria. The other models were however based on intensity calculations and could thus not receive a score with respect to C_3 and C_4. Furthermore, a large portion of the overall score of Model 3 may be attributed to its conservative nature. Model 3 is recommended as the preferred 12L modelling option according to the methodology due to obtaining the largest overall score.

4 CASE STUDY 2: BLAST FURNACE

4.1 Basic site description

The second case study is focussed on a blast furnace. Due to the relining and repair of the blast furnace, different modelling options are evaluated to quantify the EE savings. The blast furnace is used to produce liquid iron. The main energy sources consumed in this process are coke, pulverised coal injection (PCI), steam, natural gas, coke oven gas (COG) and electricity. During operation blast furnace gas is produced as a by-product which is either flared or re-used as a fuel gas in the process.

4.2 Identification of modelling options

A detailed process flow of the blast furnace indicating the relevant energy streams and measurements is depicted in Figure *4-1*. Compliance could only be proven for the steam consumption, coke supply and liquid iron production. A large portion of the energy consumption of the blast furnace consists of the other non-compliant energy streams. These streams were therefore still considered during modelling.

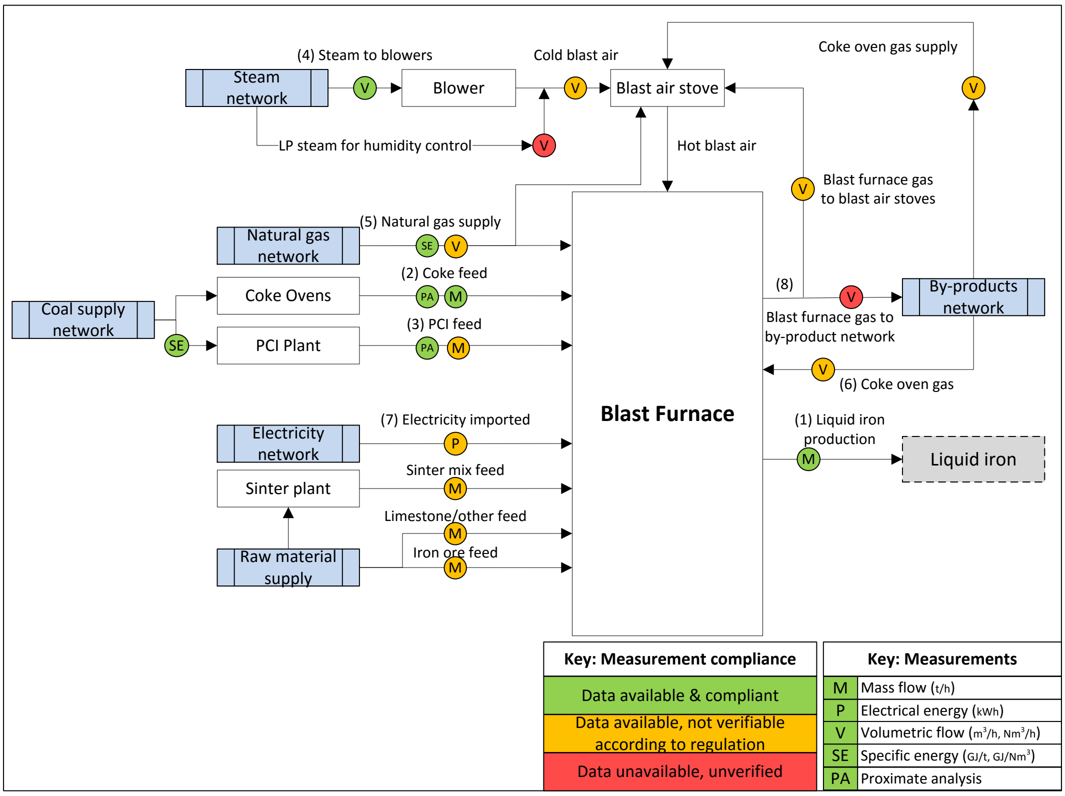


Figure 4-1: Overview of measurement points (case study 2)

During the relining and repair of the blast furnace no production took place for a period of eight months. Therefore, two baseline periods were considered in the modelling options. The first baseline (BL1) considered the 12 months prior to the assessment period. This baseline eliminates the unusable months due to the relining and repair and thus only consists of four months. The second baseline (BL2) considered the 12 months prior to the relining and repair while the assessment period remained the 12 months after the relining and repair.

Three modelling options were identified to evaluate the relining and repair of the blast furnace. The three options are summarised in Table 4-1. The dataset used in Model 1 and Model 2 is based on BL1, while Model 3 considers BL2. Intensity (Model 1) and regression analyses (Model 2 and Model 3) were considered in the various models. Furthermore, the quantified EE savings from the three modelling options vary between 234.17 GWh and 348.41 GWh.

Table 4-1: Summary of different modelling options (case study 2)

Model	Boundary	Data	Calculation	Savings (GWh)
1	Total blast furnace	BL1	Intensity	348.41
2	Total blast furnace	BL1	Regression	345.67
3	Total blast furnace	BL2	Regression	234.17

4.3 Evaluation of modelling options

In this section each of the three identified modelling options are evaluated. This was done according to their performance with respect to the seven criteria identified in the methodology section. The detailed procedure of the evaluation process is described in Section 2 (methodology). Table *4-2* presents the resulting weighted scoring matrix after evaluation.

Table 4-2: Weighted scoring matrix (Case study 2)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
Model 1	0.09	0.00	0.00	0.00	0.13	0.05	0.10
Model 2	0.09	0.00	0.18	0.11	0.13	0.05	0.10
Model 3	0.09	0.17	0.17	0.10	0.08	0.00	0.10

4.4 Ranking and recommendation of preferred model

In the last step the overall score of each modelling option was calculated. The overall scores were then used to rank the modelling options in order to recommend a final model. Figure 4-2 presents the overall score of each of the three modelling options as well as the contribution of each criterion to the final scores.

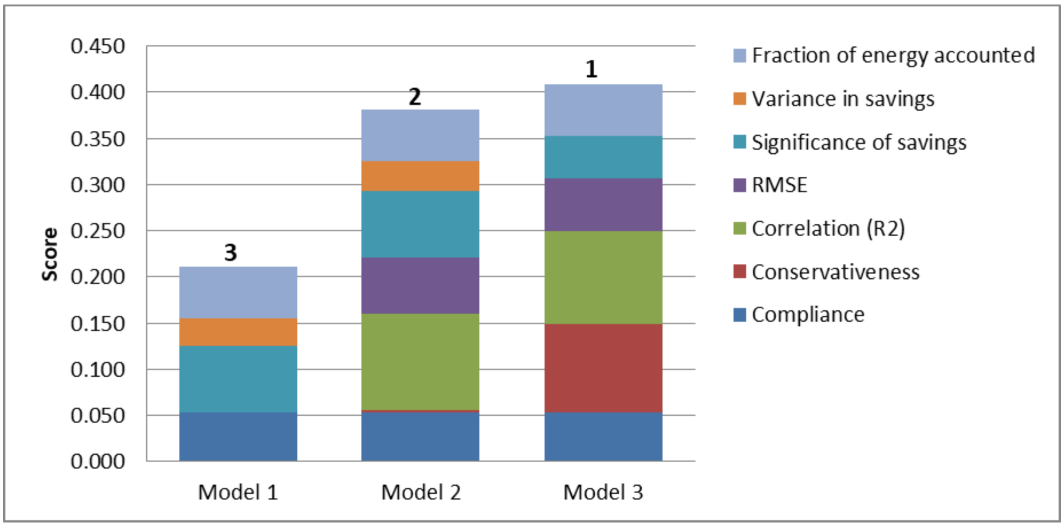


Figure 4-2: Summary of results (Case study 2)

From Figure 4-2 it can be seen that Model 3 obtained the highest overall score and was thus recommended as the preferred 12L modelling option. It is also observed that a large portion of the score attributed to Model 3 is due to its conservative nature. Furthermore, Model 1 received the lowest overall score. This may be attributed to the fact that Model 1 did not receive any score in terms of model correlation or RMSE, since it is based on an intensity calculation.

5 CASE STUDY 3: COMPRESSED AIR NETWORK

5.1 Basic site description

The third case study is focussed on the different modelling options available on a large industrial mine's compressed air network. The energy efficiency of the compressed air network was increased by optimising the compressor control philosophy. This was achieved by being able to reduce the compressed air supply during non-peak drilling periods.

Four large compressors are used to supply air to various pneumatic systems. The compressors rely on electrical energy for operation and have a total rated capacity of 17.5 MW.

5.2 Identification of modelling options

A detail layout of the compressed air network is depicted in Figure 5-1. From Figure 5-1 various measurement points are identified for the compressors electricity consumption. The various electricity measurements was deemed 12L compliant as monthly invoices are available for the electricity supply and the measurements from sub-meters are from calibrated meters.

Mass flow measurements are available for the compressed air flow supplied to main consumption levels Furthermore, the main production indicator of the mine is the tonnes of ore milled. Measurements of the ore production could also be identified.

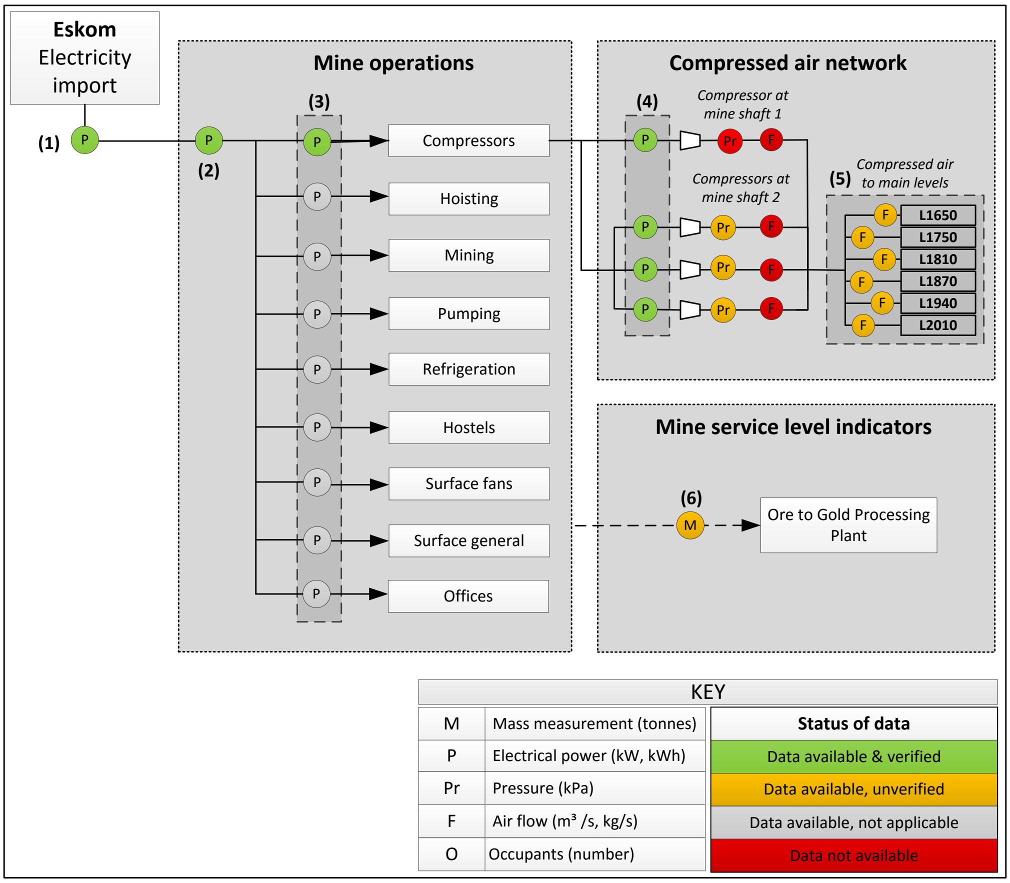


Figure 5-1: Overview of measurement points (case study 3)

Three different modelling options could be identified in order to evaluate the EE savings associated with the compressed air network. The three models are summarised in Table 5-1. In order to encapsulate the effect of the ESM, the measurement boundary for all three models was chosen as the compressed air network. The first model is based on a year-on-year energy reduction in the electricity consumption of the compressors. Model 2 and Model 3 are both based on regression analyses but vary according to the chosen energy driver of the system. Model 2 considers the peak drilling period compressed air as the energy driver, while the mine production is considered as an energy driver for Model 3. The savings of the three models vary between 6.17 GWh and 6.84 GWh.

Table 5-1: Summary of different modelling options (case study 3)

Model	Boundary	Data	Calculation	Savings (GWh)
1	Compressed air network	Electricity consumption	Unadjusted energy reduction	6.80
2	Compressed air network	Electricity consumption and peak period compressed air flow as energy driver	Regression	6.84
3	Compressed air network	Electricity consumption and mine production as energy driver	Regression	6.17

5.3 Evaluation of modelling options

The next step was to evaluate each of the three identified modelling options. Section 2 of this paper (overview of methodology) describes the evaluation process. Therefore, only the results from the evaluation process are presented in this section.

Table 5-2 presents the weighted scoring matrix.

Table 5-2: Weighted scoring matrix (Case study 3)

	C_1	C_2	C_3	C_4	C_5	C_6	C_7
Model 1	0.21	0.01	0.00	0.00	0.13	0.06	0.10
Model 2	0.11	0.00	0.15	0.07	0.13	0.05	0.10
Model 3	0.11	0.17	0.08	0.06	0.11	0.00	0.10

5.4 Ranking and recommendation of preferred model

In this section the overall scores of each of the three modelling options were calculated. Figure 5-2 presents the overall scores as well as the contribution of each criterion to the final scores for each modelling option. Furthermore, the models were ranked and numbered from first to third place according to the final scores.

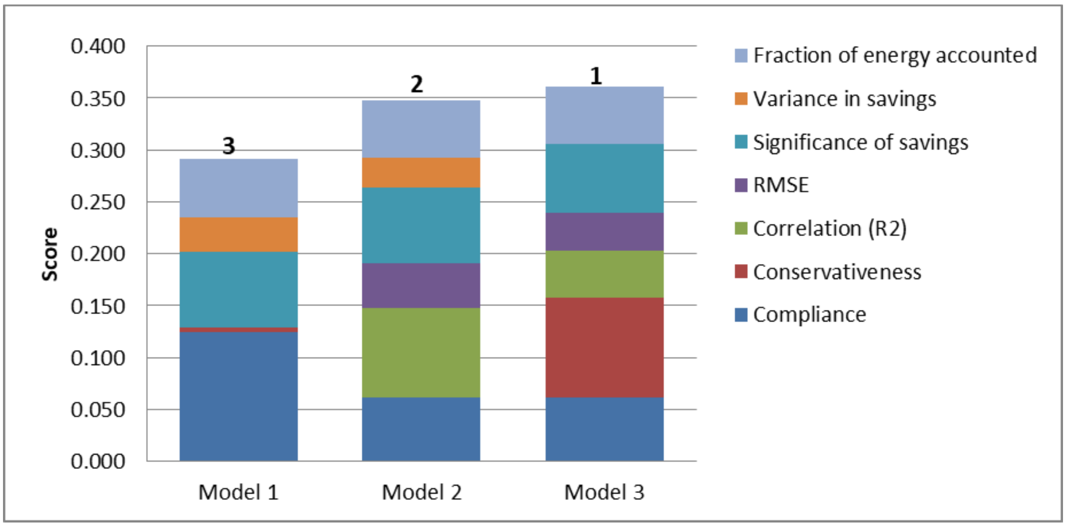


Figure 5-2: Summary of results (Case study 3)

A large portion of Model 1's score may be attributed to its compliance, as compliant electricity measurements were used. Although Model 1 was based on compliant data, it was still outranked by the other models. Since Model 1 was based on the unadjusted raw electricity consumption of the compressors, it did not receive any score with respect to model correlation and RMSE. Furthermore, Model 3 was the most conservative model and thus received a large score with regards to conservativeness. Model 3 obtained the highest overall score and was thus recommended as the preferred 12L model according to the methodology.

6 DISCUSSION OF RESULTS

The results obtained from the case studies are validated by comparing them to independent SANAS accredited M&V results from successful 12L applications. The comparisons are shown in Table 6-1, which validates that the methodology followed in this paper aligns with existing formal M&V methods. Effectively, the methodology is validated without disproving the official M&V work.

Table 6-1: Comparison of case studies' results with independent M&V results

Case study	Methodology	Independent SANAS accredited M&V team
Case study 1	Model 3	Model 3
Case study 2	Model 3	Model 3
Case study 3	Model 3	Model 3

Since the results summarised in Table 6-1 are comparable, there was no financial impact from the methodology results. However, the true benefit of the methodology lies in the objective, transparent and repeatable essence thereof.

All of the identified models for each case study showed comparable results and are suitable models to quantify the respective EE savings. However, each of these suitable models differed in the resulting EE savings. Table *6-2* illustrates the most and least conservative model results for each case study. Table 6-2 shows that the possible variance in the 12L certificate value from selecting either the least or most conservative modelling option could range between 10% and 33%. This validates the need for an objective method for the selection of a final modelling option from multiple options.

Table 6-2: Possible variance in 12L certificate value between modelling options

Case study	Quantified EE savings (GWh)		12L Certificate value (R millions)			
	Most conservative	Least conservative	Most conservative	Least conservative	Difference	
Case study 1	251.40	341.19	238.83	324.13	85.30	26%
Case study 2	234.17	348.41	222.46	330.99	108.53	33%
Case study 3	6.17	6.84	5.86	6.50	0.64	10%

A recommendation to possibly improve the results obtained in this study is to make use of more sophisticated and complex decision aid methods. This could be done to counter the weaknesses associated with the WSM; such as heavily conditioning scores towards criteria with extreme values, e.g. null values. An example of this weakness in this paper is highlighted in the fact that intensity calculations could not be evaluated according to model correlation and RMSE. The results obtained in this study are, however, still acceptable since regression models are the industry preferred method for quantifying EE savings.

The results obtained in this study illustrate that it is a challenging task to develop models which are completely based on compliant data sources. The non-compliant components of the final selected model could, however, be further assessed. This could be done to identify whether any possible influences of the non-compliant measurements could be mitigated in order to achieve conservative results. Once the necessary mitigations have been taken into consideration, the methodology could be repeated in order to re-evaluate the available modelling options.

7 CONCLUSION

In this paper it was highlighted that numerous models are developed in order to quantify the EE savings associated with an ESM for the 12L application process. There are, however, no guidelines available in literature on how to evaluate the multiple modelling options or how to select the final model. This paper therefore focussed on the practical application of a methodology which assists the 12L application process in the selection of a final modelling option.

The methodology was applied to three industrial case studies, namely steam stations, blast furnace and compressed air network ESMs. For each of the case studies multiple modelling options were evaluated and a final model recommended according to the methodology. The results were comparable to independent SANAS accredited M&V results. Furthermore, the possible 12L certificate value from the case studies' multiple modelling options varied between 10% and 33%. This verified the need for an objective and transparent methodology for the model evaluation and selection process, such as the one presented in this paper.

REFERENCES

[1] E. F. Du Toit, "Energy efficiency savings allowance in South Africa: An international comparison," Masters Dissertation, University of Pretoria, South Africa, 2011.

[2] Republic of South Africa, "Section 12L of the Income Tax Act (no. 58 of 1962) Deduction in respect of energy efficiency savings," South Africa, 2013.

[3] National Treasury, "Regulations in terms of section 12L of the income tax act, 1962, on the allowance for energy efficiency savings," Government Gazette, No. 37136, 2013.

[4] SABS, "SANS 50010:2011 Measurement and Verification of energy savings," 2011.

[5] Efficiency Valuation Organisation, "International Performance Measurement & Verification Protocol: Concepts and options for determining energy and water savings," 2012.

[6] W. Hamer, "A practical approach to quantify RSA Section 12L EE tax incentives for large industry," PhD Thesis, North-West University, South Africa, 2016.

[7] K. Campbell, W. Booysen and J. Vosloo, "Evaluating the feasibility of the 12L tax incentive for energy-intensive industries," *South African Journal of Industrial Engineering,* vol. 28, no. 3, pp. 15-28, 2017.

[8] J. R. S. C. Mateo, Multi Criteria Analysis in the Renewable Energy Industry, Berlin, Germany: Springer Science & Business Media, 2012.

[9] A. Kolios, V. Mytilinou, E. Lozano-Minguez and K. Salonitis, "A Comparative Study of Multiple-Criteria Decision-Making Methods under Stochastic Inputs," *energies,* vol. 9, no. 566, p. 21, 2016.

[10] T. L. Saaty, "Decision making with the analytical hierarchy process," *International Journal Services Science,* vol. 1, no. 1, pp. 83-98, 2008.

[11] N. J. G. Carrico, F. V. Goncalves, D. I. .. Covas, M. do Céu Almeida and H. Alegre, "Multi-criteria analysis for the selection of the best energy efficient option in urban water systems," *Procedia Engineering,* vol. 70, pp. 292-301, 2014.

[12] L. Botes, "Objective evaluation of industrial energy efficiency models for the RSA Section 12L tax incentive," Masters Dissertation, North-West University, South Africa, 2017.

[13] H. M. Janse Van Rensburg, "Structuring mining data for RSA Section 12L EE tax incentives," Masters Dissertation, North-West University, South Africa, 2015.

[14] A. G. S. Gous, W. Booysen and W. Hamer, "Data quality evaluation for measurement and verification processes," in *Industrial and Commercial Use of Energy*, 2016.

[15] J. R. Eastman, "Multi-criteria evaluation and GIS," in *Geographical information systems*, New York, Wiley, 1999, pp. Chapter 35, 493-502.

DEVELOPMENT OF AN ASSURANCE FRAMEWORK FOR SOUTH AFRICAN CARBON TAX

A.G.S. Gous[1], W. Booysen[2] & E.H. Mathews[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
agous@researchtoolbox.com
[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbooysen@researchtoolbox.com
[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
ehmathews@researchtoolbox.com

ABSTRACT

The Carbon Tax on greenhouse gas (GHG) emissions, set to help transition South Africa towards a low-carbon economy, induces several uncertainties. Liable parties, as well as national bodies (e.g. the South African Revenue Service, SARS), are at risk due to the lack of published guidance. Such uncertainties and risks can be managed or mitigated by introducing assurance mechanisms focused on the different key structures within the process.

This paper presents an assurance-centered reporting framework developed to document and support all decisions affecting the values submitted to SARS. The framework merges several international and national best practice auditing assurance standards together with the technical and legal requirements of carbon tax. Assurance is thereby provided by a transparent and traceable process presented in such a manner that it can be easily audited by independent parties.

The developed framework assures the (1) dataset constructed and used in (2) GHG quantification methods to (3) report the results for industrial case studies. Several practical assurance mechanisms are identified and applied within each structure to address the associated uncertainties. Aggregating these mechanisms ultimately provides holistic assured results that are likely to pass all levels of scrutiny.

This work was sponsored by ETA Operations (Pty) Ltd

[1] Dr Andries Gous holds a PhD in Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr Walter Booysen is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Prof Edward Mathews is a registered professional engineer and holds a PhD in mechanical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. SOUTH AFRICA'S CARBON TAX

Carbon pricing is applied in numerous countries as a strategy to mitigate greenhouse gas (GHG) emissions [1]. South Africa aims to transition to a low carbon economy with reduced emissions while promoting sustainable development [2]. In 2017, an updated "Draft Carbon Tax Bill" was published following the initial draft in 2015 [3], emphasising the government's approach to impose an environmental levy on GHG emissions, to be known as carbon tax.

The design of carbon tax is portrayed to be a simplistic self-assessment that levy CO_2 equivalent GHG emissions at R120 per tonne. Assessments are reported to the South African Revenue Service (SARS) and paid accordingly [3]. Presently, liable parties are only required to submit a single value (carbon tax amount payable) to SARS. Accurately reporting an entity's official carbon tax value to SARS is critically important. It is highly probable (as with all tax submissions) that some values may be audited from time to time.

The lack of assurance that the submitted value is correct presents a significant risk to both parties in the event of an audit [4]. SARS is at risk since results cannot be traced, audited or checked. The liable party is at risk for penalties, additional costs and the administrative burden of attempting to compile a report retrospectively for auditing purposes. It is therefore necessary to have a framework that can be used to document the process and support all values submitted to SARS [5].

Several areas of uncertainty exist within the design of the generic carbon tax process [4]. These uncertainties are related to the, (1) legislative compliance, (2) quantification of emissions, (3) potential mitigation of liability, (4) reporting the results, and (5) assurance that the submitted values are correct. It is therefore important to firstly understand the process and the uncertainties before assurance can be provided. Figure 1 illustrates the generic carbon tax process and highlights the uncertainties associated with the steps [4].

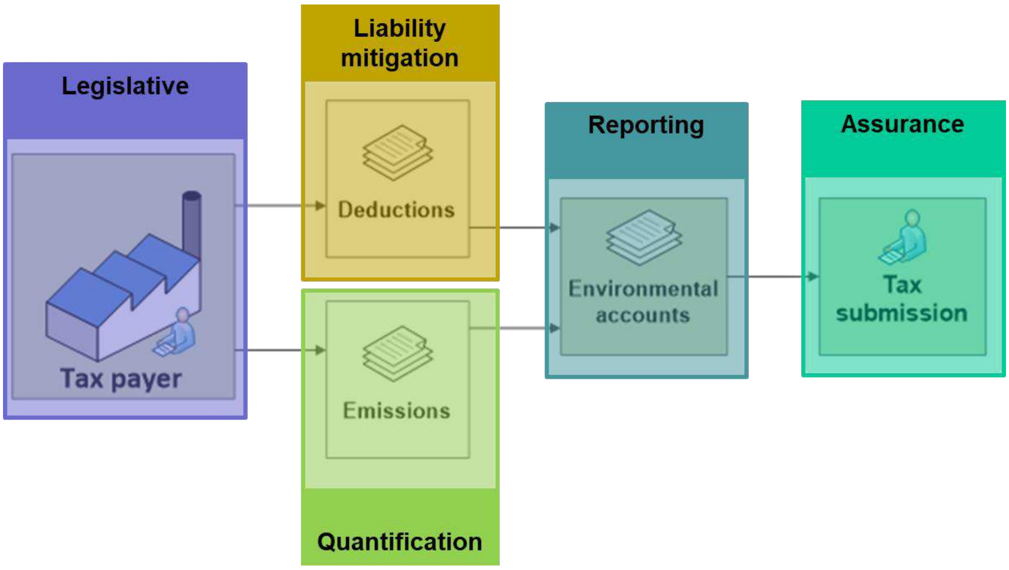


Figure 1: Generic carbon tax process and the associated uncertainties [4]

Before any steps are taken, industries and companies should know if they are required to submit and pay carbon tax. Liable parties are then required to quantify the emissions on which the tax should be paid. Deductions, such as offsets and other credits, need to be determined and subtracted from the liable emissions. The emissions and deductions should then be reported to SARS as part of the environmental accounts. After which the determined amount should be paid over to SARS.

Each area of uncertainty is broadly described with regards to the process steps.

Legislative uncertainty encapsulates the legal requirements that stakeholders should adhere to and provides answers to questions such as "who should pay the tax?" [4], [5].
Quantification uncertainty describes the uncertainty from the data and methodologies that are used to quantify the emissions. This highlights the underlying question of "how do you quantify the emissions?" [4]

The design of carbon tax allows for some *Liability mitigation* [3]. There are a vast number of deductions and allowances that can be claimed, but it is unknown for "what deductions do the liable party qualify?" [4]

Reporting uncertainty describes the queries experienced when compiling a report for relevant stakeholders [4]. This area of uncertainty stresses the question of "how should such a report look?"

Assurance uncertainty are provoked from possible errors and inaccuracies of the entire carbon tax process [4]. Matters associated with accuracy, correctness, and stakeholder confidence are accounted for in this area [5].

These uncertainties were identified by comparing the carbon tax process to the structure of the Section 12L tax incentive [4]. The Section 12L tax incentive provided a well-defined structure as reference point into handling similar issues [6]. However, there were a lack of clear reference documentation available that focus on carbon tax. Gous *et. al.* [4] presented a landscape of carbon tax specific references to assist in addressing each uncertainty. This landscape consists of references that are relevant to carbon tax and specific towards an area of uncertainty.

It is necessary to assess the landscape for each individual area of uncertainty to establish the core assurance for the final value. Analysis of the landscape would present multiple ways to assure the final results.

2. GENERAL ASSURANCE MECHANISMS

The approach to identifying the assurance applicable to each of the uncertainties is divided into two sections. The first identifies general mechanisms that are used to provide assurance. The second section identifies the assurance mechanisms that are applicable to the carbon tax process.

There are several fundamental requirements to obtain a high degree of the necessary assurance. These requirements are satisfied through general assurance mechanisms that are already available. Evaluation of such mechanisms will identify basic components that can be used within the assurance framework.

2.1 Requirements for assurance

The carbon tax landscape presents several documents capable of assisting with the areas of uncertainty. However, an analysis of the landscape is required to identify the requirements for providing assurance. Gous [5] critically analysed the landscape using a structured review to obtain assistance for assuring the different areas of uncertainty. The analysis identified general requirements for assurance and can be grouped under five topics [5]:

- accuracy and validation of results [7],
- addressing uncertainties [8], [9],
- independent assurance [10],
- traceability [6], and
- transparency [11], [12].

Accuracy and validation of results are assured by analyses and evaluations. These include historic, present and predictive processes. This evaluates the stability of results while identifying abnormal circumstances over a timeframe. Stable and representative results indicate a high level of assurance [7].

Addressing uncertainties similarly provides a high level of assurance. Uncertainties need to be known before they can be addressed [8]. A quantified uncertainty gives an indication of what impact it could have. Lessening this impact is done by addressing such uncertainties by obtaining additional knowledge or disclosing the associated information to stakeholders. Assurance is obtained by either the mitigation of uncertainties or the knowledge of its existence [9].
Independent assurance is often required when several stakeholders are involved [10]. The level of independence can differ but it is ultimately required that the entity that provides the assurance is not operationally involved [13]. Assurance is provided on the grounds that outcomes are checked and confirmed in an unbiased manner.

Traceability gives the pathway by which results were attained [6]. Knowing how and from what results are obtained provides assurance. Traceability mechanisms focus on the acquisition of materials that describes the flow of information from its origin to the reported result.

Being *transparent* with information gives a degree of certainty and assurance [11], [12]. Transparency can help stakeholders understand the context of results and aids in its interpretation [14]. Assurance is provided by increased acceptance of the results [14] through the disclosure of essential information [15].

2.2 Available assurance mechanisms

Several generic assurance mechanisms were identified from a detailed literature survey. Literature from several fields of study and countries were included to recognise mechanisms that are currently used. The identified mechanisms were further evaluated to understand their usability. Table 1 gives these mechanisms, a general description of the specific mechanism and the specific assurance it provides.

Table 1: Generic assurance mechanisms and assurance provided

Assurance mechanism	Mechanism description	Provided assurance	Ref.
Archived records	Storage of records for an extended period for revaluation.	Traceability: Historic results traceable to the original source document. Transparency: Information on process used during previous time periods.	[16], [17]
Availability of information	Ease of accessing information used during an evaluation. Ease of recreating the results from available information.	Transparency: Information is available within a predetermined domain and can be used to recreate results.	[10], [16]-[19]
Certified monitoring	Measurements and monitoring are certified to conform to technical requirements by an external party.	Accuracy and validation of results: Obtained information adheres to a set of technical requirements. Independent assurance: Certification is provided by external party.	[20]-[24]
Defined reporting structures	Designated platforms available for reporting information to stakeholders.	Transparency: Stakeholders obtain necessary information from an established platform.	[15], [17], [20], [23], [25]-[29]
Disclosure of methods and processes	Assumptions, methods and process used within to obtain the results are stated.	Transparency: Information used is known to stakeholders. Traceability: Information can be traced to the origin, be it an assumption or process.	[9], [10], [15], [17]-[20], [26], [30], [31]
Disclosure of uncertainty	Statements of known uncertainties with the potential to change the results.	Addressing uncertainties: Impact of uncertainty is known to stakeholders. Transparency: Uncertainties affecting the results are known for decision making.	[9], [20], [25], [32], [33]
Documented procedures	Documents that describe the procedure that is followed for a set activity or process.	Traceability: Outcomes can be traced to the source by evaluating the descriptions. Transparency: The descriptions provide insight into what is included or not.	[9], [20], [27], [31], [34]
Expert judgement	Opinion and/or recommendations from an individual or a team with specific technical knowledge or expertise.	Accuracy and validation of results: Assumptions and processes are based on information received from an expert in a specific field.	[19], [21], [35]-[40]
External evaluation	Evaluations performed on an activity by an entity which is not responsible for the operation thereof.	Accuracy and validation of results: Results are validated through additional evaluations and operational bias is reduced. Independent assurance: Evaluation by an external entity provides confidence in the results and lessens any bias.	[15], [18]-[21], [24], [25], [27], [33]
Historic-predictive analysis	Using historic analyses to predict and evaluate current assessments.	Accuracy and validation of results: Validation of results through the comparison between expected and actual results.	[6], [10], [41], [15], [20], [21], [23], [26], [29], [32], [37]

Assurance mechanism	Mechanism description	Provided assurance	Ref.
Independent M&V	Measurement and verification of activities performed by an independent entity.	Independent assurance: Independent confirmation that outcomes are correct and trustworthy.	[10], [15], [33], [37], [17], [20], [21], [24]-[27], [32]
Management of uncertainty	Processes or structures to mitigate and/or eliminate uncertainties.	Addressing uncertainties: Known uncertainties are addressed and the impact thereof is reduced.	[9], [15], [18], [20], [32], [42], [43]
Supporting documentation	A document that presents and confirms given information.	Traceability: Source documentation confirms stated information.	[6], [20], [43], [44]
Trend analysis based on key indicators	Using operational key indicators to trend and analyse outcomes.	Accuracy and validation of results: Outcomes are validated against trends of key operational indicators.	[23], [26], [29]
Uncertainty quantification	Quantified value of the potential impact that uncertainty may have on the outcome.	Addressing uncertainty: A quantified value that indicates the trustworthiness and potential risks of the results. Accuracy and validation of results: A quantified margin of accuracy of the results.	[9], [15], [18], [20], [32], [37], [43]-[45]

The results summarised in Table 1 show that each mechanism is designed differently and achieves a specific outcome in a unique manner. These designs include strategies, processes and calculations that are not universally applicable to all areas of uncertainty. However, the different designs create the opportunity to combine and adapt several mechanisms to address specific needs.

2.3 Combining mechanisms to deliver holistic assurance

The assurance requirements can be reduced to a generic set of mechanisms/topics. However, the evaluation of available mechanisms shows that the universal implementation of specific mechanisms is not always possible. In these situations, alternatives mechanisms should be combined to allow for a framework that provides assurance. Several of the mechanisms presented in Table 1 provide the same assurance, but through alternative means. It is therefore possible to select the more practical mechanism to achieve a required assurance. However, it is necessary to know what assurance is required before an alternative mechanism can be selected.

3. ASSURANCE APPLICABLE TO CARBON TAX

Assurance is required to address the uncertainties associated with the carbon tax process. Such assurance can be provided by the generic mechanisms. However, these mechanisms need to be matched to the relevant areas of the carbon tax process. Therefore, the relationship between the assurance mechanisms and carbon tax needs to be established.

3.1 Framework development process

The framework, that establishes a relationship between carbon tax and assurance, expands on existing literature. Acting on shortcomings that are still present within the carbon tax process. Figure 2 illustrates the expansion on literature from uncertainty identification to the assurance framework for the final results.

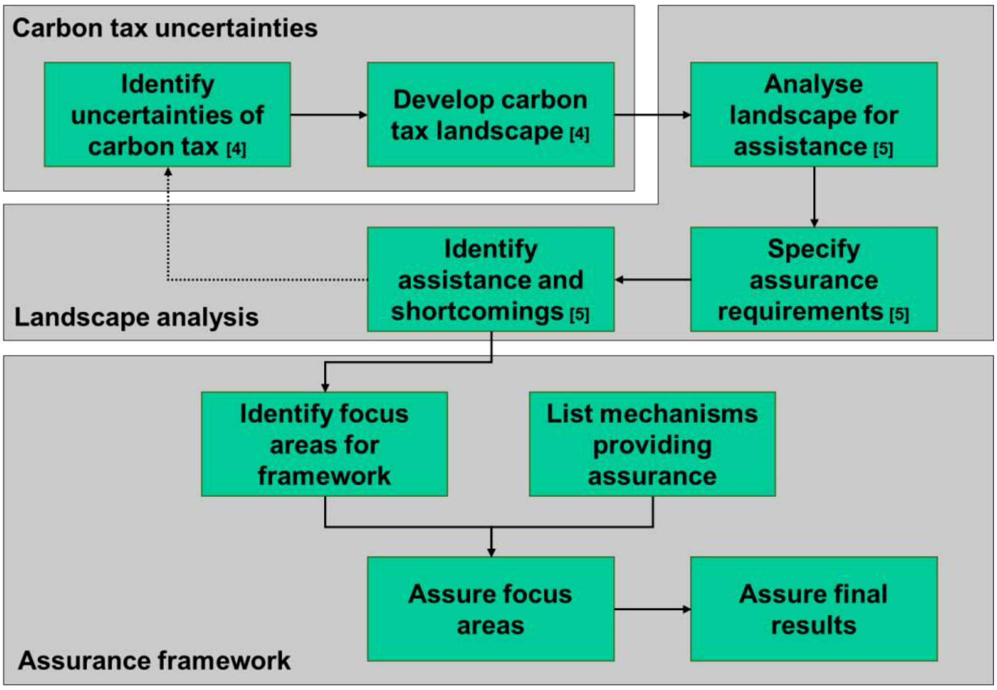


Figure 2: Development process of the assurance framework

Uncertainties for the carbon tax process was identified and a resource landscape was developed to provide focussed assistance. These resources were analysed to obtain the necessary assistance through assurance requirements. Adherence to the requirements provided sufficient assistance for some areas of uncertainty. However, several shortcomings were identified that could not be addressed from the landscape.

Shortcomings are used to identify focus areas for an assurance framework. Additional mechanisms that provide assurance are identified to resolve such shortcomings. The available mechanisms are applied to the focus areas to provide specific assurance.

Specific assurance for each focus area aggregates into the final results. The final results are therefore holistically assured by the confidence in the underlying focus areas. Which ultimately provides guidance for the uncertainties from the carbon tax process that could not be addressed by the landscape.

3.2 Matching assurance mechanisms to areas of uncertainty

The carbon tax landscape does not only provide numerous generic assurance mechanisms but also assistance for some of the areas of uncertainty. Specific resources provide the necessary clarification and support to address the respective uncertainties.

Two areas of uncertainty, legislative and liability mitigation, have detailed resources to help stakeholders manage the unknown. These resources include legal documents that can be queried relating to specific matters. This consequentially provides assurance for these areas of uncertainty within the carbon tax process. The assurance framework therefore focusses to assure the uncertainties where shortcomings are currently not addressed.

Shortcomings were found with regards to quantification (including the data and methods), reporting and overall assurance [5]. The quantification sections will have a significant effect on the final quantified emissions. However, accurate and true reporting is not guaranteed by delivering assurance on the quantification results alone [15]. Items such as review processes and undisclosed information can potentially influence how the reported values are observed and interpreted by stakeholders. Reports that are deemed as misleading will decrease stakeholder confidence and could lead to legal consequences, such as penalties and imprisonment [17]. Assurance of the report and associated processes is therefore necessary to provide stakeholders with a high degree of confidence [27].

Presenting an overall assurance for the combined sections is complex and not easily performed due to their interrelated nature [39], [46], [47]. It is therefore necessary to separate the concepts and assure each uncertainty individually in order to simplify the process [46].

The framework therefore focusses on three specific areas to obtain the individual assurance. Focussing on assuring the (1) data and (2) methods that are used during quantification of GHG emissions and the (3) reporting thereof. These focus areas were identified from the shortcomings that were observed for the areas of uncertainty.

This approach requires the various assurance focus areas (data, calculation methods and reported results) to be matched to the available assurance mechanisms, linking the areas of uncertainty, respective shortcomings and the mechanisms together to meet the assurance requirements.

3.3 Mechanisms applicable to specific focus areas

Three focus areas (data, calculation methods and reported results) were identified from the uncertainties that could not be sufficiently addressed by the carbon tax landscape. The quantification data focus area assures that the uncertainty induced from the dataset are managed. Which can then be used to quantify emissions. However, using good data does not constitute good quantified results.

Emission quantification methods are therefore assured to provide confidence that the results were obtained in a suitable manner. The third focus area is centred around the way these results are reported. Assurance of each respective focus area will ultimately provide overall assurance.

Several mechanisms (Table 1) can potentially be applied to each focus area to attain the desired assurance. However, not all mechanisms are always practical/applicable for a specific scenario [48]. It is therefore necessary to evaluate the available mechanisms and understand what is practically applicable for carbon tax.

3.3.1 Data assurance:

Constructing a dataset is one of the first phases towards quantifying carbon tax. Table 2 provides a summary of assurance requirements and the mechanisms for the quantification data that can be found within the landscape.

Table 2: Mechanisms providing assurance for quantification data

Carbon tax focus area	Assurance requirements	Applicable mechanisms
Quantification data	Accuracy and validation of results Addressing uncertainty Independent assurance Traceability	Archived records Certified monitoring Disclosure of methods and processes Disclosure of uncertainty Documented procedures Expert judgement External evaluation Historic-predictive analysis Independent M&V Management of uncertainty Supporting documentation Trend analysis based on key indicators Uncertainty quantification

Datasets should firstly be accurate, representative and valid. The accuracy of data can have a significant influence on the results [6], [7], [49] and therefore requires assurance. Uncertainties need to be addressed to ensure that the impact of data uncertainty is minimised [47]. One of the main strategies to assure a dataset is through independent assurance that confirms the properties of the data [10]. There is a normal variance within data and this contributes to data uncertainty [50]. It is thus required that data should be traceable to the measurement point before assurance can be confirmed [6].

Assurance on the dataset can be conferred through the applicable mechanism. Administrative and calculation mechanisms, such as record keeping and analyses, form the basis for dataset assurance. These are combined with unbiased evaluations to adhere to the requirements. The practicality of this combination causes concern due to the difficulty of unbiased evaluations.

Independent evaluators are either limited to a specific field or expensive due to their expertise [51]. An alternative is to make use of certified monitoring. Equipment and measurement procedures need to be periodically evaluated to confirm that they are within specifications before they are certified. Such certification can only be provided by an individual with the necessary competency and accreditations [22]. Similar to independent evaluators, certification providers are also limited or expensive [52]. Implementing as well as maintaining these mechanisms can therefore place a financial burden on an entity.

Several carbon-intensive industries are already under financial pressure and would not be able to use these mechanisms [53]. Practical implementation of these mechanisms is thus required to attain assurance of a dataset.

3.3.2 Quantification methods:

Methods are used to interpret the dataset and determine the emission quantities. In Table 3, the mechanisms and assurance requirements are listed for the quantification methods.

Table 3: Mechanisms providing assurance for quantification methods

Carbon tax focus area	Assurance requirements	Applicable mechanisms
Emission quantification methods	Accuracy and validation of results Addressing uncertainty Independent assurance Transparency	Archived records Availability of information Certified monitoring Defined reporting structures Disclosure of methods and processes Disclosure of uncertainty Documented procedures Expert judgement External evaluation Historic-predictive analysis Independent M&V Management of uncertainty Trend analysis based on key indicators Uncertainty quantification

Three of the requirements for assurance are the same for the data and the method focus areas. Transparency replaces traceability when assuring the methods, indicating that the methods used should be presented to stakeholders in an understandable manner.

No method will be 100% accurate due to its sensitivity to the provided inputs [8]. This introduces uncertainty to the outcome and needs to be addressed before it is reported [54]. Assurance is therefore required on the accuracy and relevance of the methods, while uncertainty is addressed. Such assurance is commonly provided by adhering to the industry best practices [9].

Best practices are available through guidelines and standardised processes, thus also providing assurance that the methods would be able to obtain independent assurance [55]. Independent assurance is provided against a set list of criteria (best practices) that are available in standards, such as the SANS 14064 standards [10], [18], [19].

Independent and external evaluations are required to confirm that the calculated results are correct and true. However, the competency requirements for quantification evaluations are extremely specialised. Verification specialists require expertise and experience in general M&V, GHG quantification methods as well as the relevant production process [13]. The specialist nature of the work leads to a significant shortage of verification specialists capable of providing assurance.

Transparency is important for processes where other forms of assurance are either not available or impractical [56]. Assurance on transparency is provided by using clearly defined structures for reporting, while disclosing details of the process and providing stakeholders with access to the information.

Providing stakeholders with the relevant information will additionally address the uncertainty of the methods [6]. Uncertainties can be mitigated and managed if stakeholders are kept aware of all relevant information. Although assurance on the methods can be provided through the different mechanisms, some form of adaption might be required to practically implement the mechanisms.

3.3.3 Reporting results:

Data and methods are used to calculate results that have to be presented to various stakeholders. The requirements and assurance mechanisms for reported results are given in Table 4.

Table 4: Mechanisms providing assurance for reporting results

Carbon tax focus area	Assurance requirements	Applicable mechanisms
Reporting results	Addressing uncertainty Traceability Transparency	Archived records Availability of information Defined reporting structures Disclosure of methods and processes Disclosure of uncertainty Documented procedures Management of uncertainty Supporting documentation Uncertainty quantification

When reporting results, assurance is required for traceability, transparency and uncertainty. These requirements are essential for stakeholder confidence, since the results are provided at a low level of detail [57]. Stakeholders therefore need to know and understand how the results were obtained. Supporting documentation allows stakeholders to trace results back to the source, while definite structures describe what was specifically done to obtain the results.

Uncertainty can be addressed with the applicable mechanisms, but also by managing the potential risks that it induces. Statements and proof that uncertainty is addressed, managed and quantified will provide stakeholders with assurance. Assurance for the reported results are thus provided twofold, for the actual submitted value as well as understanding the accompanied risks.

4. RESULTS AND FINDINGS

The assurance framework was applied to an industrial case study to show the practical implementation thereof. The case study focusses on the three focus areas within the framework. Assuring each individual focus area would provide the necessary support to present the overall deliverable with a high degree of confidence.

Different elements (data variable, calculation strategies and results) that are used to present the final outcome are evaluated for the level of assurance. Assurance was judged sufficient when one or more of the applicable mechanisms were present for the evaluated element. Results from the assurance presented by the framework are finally compared to a published report from industry.

4.1 Assurance of the constructed dataset

The data used to quantify the carbon tax value for an industrial processing plant was evaluated to determine whether assurance can be provided for it. The assurance results are displayed in Table 5. Assurance was evaluated against all five criteria even though transparency is not a specific requirement for data. Each variable was evaluated for its respective level of assurance.

Table 5: Assurance of quantification data

Framework structure	Calculation element	Carbon tax assurance requirements: Accuracy and validation	Addressing uncertainty	Independent assurance	Traceability	Transparency	Carbon tax assurance
Data assurance	Fuel to boilers	✓	✓	✓	✓	✓	✓
	Gas 1 to turbines	✓	✓	✓	✓	✓	✓
	Gas 2 to turbines	✓	✓	✓	✓	✓	✓
	Fuel oil	✓	✓	✗	✓	✓	(✓)
	Fuel gas	✓	✓	✗	✓	✓	(✓)
	Production CO_2	✓	✓	✓	✓	✓	✓
	Cleaning CO_2	✓	✓	✓	✓	✓	✓
	Gas flares	✓	✓	✗	✓	✓	(✓)
	Off-gas flares	✓	✓	✗	✓	✓	(✓)
	Product flares	✓	✓	✗	✓	✓	(✓)
	Process gas flares	✓	✓	✗	✓	✓	(✓)
	Fuel gas distribution	✓	✓	(✓)	✓	✓	(✓)
	Gas feed distribution	✓	✓	✓	✓	✓	✓
	Gas1 distribution	✓	✓	✓	✓	✓	✓
	Gas 2 distribution	✓	✓	✓	✓	✓	✓
	Fuel handling	✓	✓	✓	✓	✓	✓
	Product transfer	✓	✓	✓	✓	✓	✓
Quantification assurance	Emission factor	✓	✓	✓	✓	✓	✓
	Direct measurement	✓	✓	✓	✓	✓	✓
	Mass balance	✓	✓	✓	✓	✓	✓
Reported results assurance	Results	✓	✓	✓	✓	✓	✓

✓ denotes sufficient assurance; (✓), plausible assurance but not confirmed; ✗, assurance not available.

The results show that the data used has an acceptable level of assurance associated with it. Ten data sources were sufficiently assured, while the remaining seven sources did not obtain a perfect score. Assurance was not deemed sufficient, since not all requirements were satisfied or even potentially satisfied. However, assurance was still deemed plausible due to the majority of requirements being satisfied.

It is seen that the independent assurance requirement was not achieved for all data sources. Further evaluation showed that the sources that do not meet the requirement are related to either flares or auxiliary streams (fuel gas and oil). These sources are not fundamentally important to the operation of the plant and would thus explain the lack of checks in an independent manner. However, these sources do play a role towards the calculated carbon tax and were therefore highlighted as potential risks.

Assurance that was obtained for the data sources was provided by several mechanisms. Trend analysis, certified monitoring and external evaluation were the most common mechanisms to assure the accuracy and validation of results. Uncertainty was managed for most sources through onsite processes that monitor the specific data. Detailed documents are available to describe the uncertainty management and provide detailed information on the production process. These supporting documents assured a clear traceability pathway for the data from the different sources.

4.2 Assurance of GHG quantification strategies

GHG emissions for the processing plant were quantified using three different strategies and methodologies. The methodologies were evaluated to see whether mechanisms were utilised or specifically applied during the quantification process. The results are shown in Table 5.
Results show that the quantification strategies have a high level of assurance for all of the requirements. The most prominent mechanisms that provided assurance were: trend analysis, disclosure of uncertainty, external evaluation and documented procedures. These mechanisms were present within all three of the strategies that were applied. This showed that it is possible to sufficiently assure the quantification strategies using the generic mechanisms.

4.3 Assurance of reported results

The reported results are difficult to evaluate as a single component due to the limited information reported. Assurance on the reported results are also presented in Table 5. The assurance requirements for the reported results were confirmed by the presence of archived records and uncertainties that were disclosed. Reported results with a high level of assurance provides a high degree of confidence to all stakeholders.

Incorporating several mechanisms into the process will present a high level of confidence to all parties involved during different stages of an audit or evaluation.

4.4 Holistic assurance of carbon tax

The previous sections focussed on the individual focus areas within the framework and how the assurance aggregates to assure the final reported results. Results from the framework were compared to the annual reports published by the processing plant. Table 6 indicates the assurance provided by the annual report and the framework.

Table 6: Assurance comparison between the developed framework and independent results

Evaluation criteria		Annual report	Assurance framework
Dataset assurance	Accuracy and validation of results	✅	✅
	Addressing uncertainty	❌	✅
	Independent assurance	✅	🟠
	Traceability	🟠	✅
	Overall assurance	❌	✅
Quantification assurance	Accuracy and validation of results	✅	✅
	Addressing uncertainty	🟠	✅
	Independent assurance	✅	✅
	Transparency	🟠	✅
	Overall assurance	🟠	✅
Reported results assurance	Addressing uncertainty	🟠	✅
	Traceability	✅	✅
	Transparency	✅	✅
	Overall assurance	🟠	✅

✅ denotes sufficient assurance; 🟠, plausible assurance but not confirmed; ❌, assurance not available.

Assurance within the annual report could not necessarily be linked to specific mechanisms. The report was therefore surveyed for either descriptions, values or figures that would be able to provide assurance for the different requirements.

It is important to note that the processing plant is in compliance with legislative requirements, as well as a forerunner regarding GHG management and public participation into mitigating climate change. However, from the results it is evident that assurance was not sufficiently provided by the annual report when compared to the developed framework. It is possible to improve the assurance by incorporating the developed framework into the current reporting structures, thereby improving stakeholder confidence that is already present.

5. THE DEVELOPED ASSURANCE FRAMEWORK

The developed framework combines available assurance mechanisms with specific focus areas that were identified from the carbon tax uncertainties. The three focus areas are combined to deliver a framework to assure carbon tax. Figure 3 illustrates the framework and the assurance requirements that are given for each area.

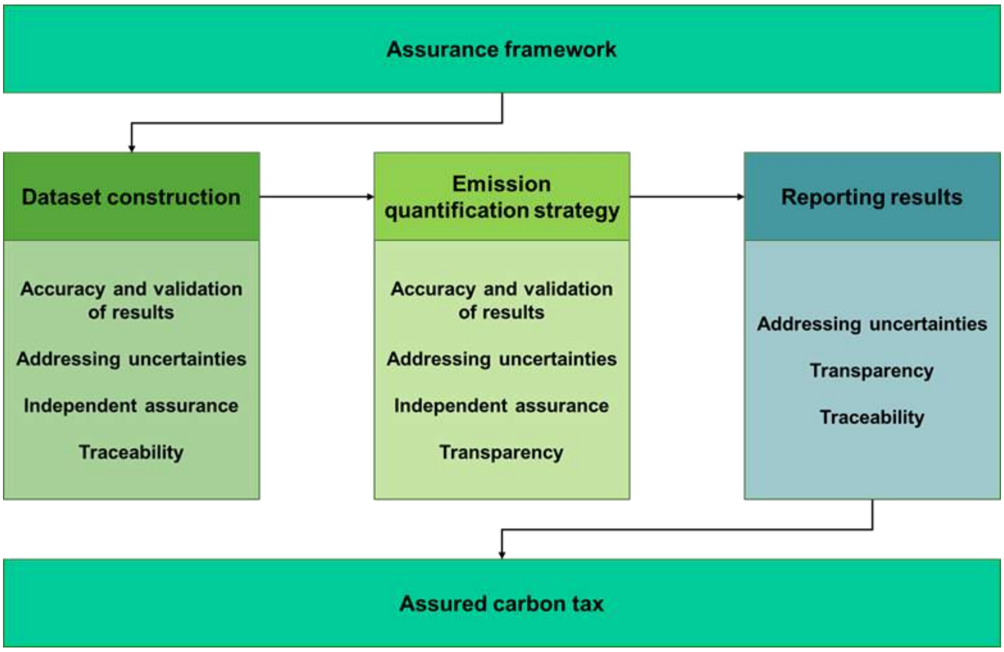


Figure 3: Assurance framework for carbon tax

The framework focusses firstly on constructing a dataset that is accurate and traceable, which can then be used to calculate the taxable emissions according to international best practise and in a transparent manner for stakeholders to understand. The outcome of these components is presented in a report that addresses uncertainties and provides stakeholders with confidence in the results.

Assurance provided for the dataset manifests in the outcome due to confidence in the reliability of quality of the data. This allows the relevant party to use quantification methods that are already widely available without significant risks. Potential risks are continuously communicated or disclosed through different assurance mechanisms, thus allowing informed decision making based on the results.

Reported results are compiled from information that has been exposed to different levels of scrutiny. Reports can furthermore be adapted to provide additional assurance to all relevant stakeholder. The reported results are thus highly resilient against uncertainties, placing a high level of assurance on the value submitted to SARS and would provide confidence to all stakeholders in the event of an audit.

6. CONCLUSION

Carbon tax is one of the strategies aimed at helping South Africa transition towards a low carbon economy. The carbon tax process has several inherent uncertainties within the process that decrease stakeholder confidence in the payable amount. By looking at general assurance mechanisms and linking them to carbon tax areas of uncertainty, an assurance framework was developed.

Results from applying the framework to an industrial case study highlights how assurance can be provided for the dataset used, the quantification strategies and the reported results. It also showed that it is possible to increase the assurance presented in public reports. The reported tax amount is thus resilient against scrutiny during different stages of an audit through assurance.

7. REFERENCES

[1] **S. Vera and E. Sauma**, "Does a carbon tax make sense in countries with still a high potential for energy efficiency? Comparison between the reducing-emissions effects of carbon tax and energy efficiency measures in the Chilean case," *Energy*, vol. 88, pp. 478–488, 2015.

[2] **RSA National Treasury**, "Reducing Greenhouse Gas Emissions: The Carbon Tax Option." Pretoria, p. 75, 2010.

[3] **RSA National Treasury**, "Draft Carbon Tax Bill." pp. 1–65, 2017.

[4] **A. G. S. Gous, W. Booysen, and E. H. Mathews**, "Addressing uncertainties in the South African Carbon Tax landscape," in *Industrial & Commercial use of Energy (ICUE)*, 2017, pp. 204–211.

[5] **A. G. S. Gous**, "A practical approach to quantify carbon tax in South Africa," PhD Thesis, North-West University, 2018.

[6] **W. Hamer**, "A practical approach to quantify RSA Section 12L EE tax incentives for large industry," PhD

thesis, North-West University, 2016.

[7] **A. G. S. Gous, W. Booysen, and W. Hamer**, "Data quality evaluation for measurement and verification processes," in *Industrial & Commercial use of Energy (ICUE)*, 2016, pp. 9–15.

[8] **C. Sweetapple, G. Fu, and D. Butler**, "Identifying key sources of uncertainty in the modelling of greenhouse gas emissions from wastewater treatment," *Water Res.*, vol. 47, no. 13, pp. 4652–4665, 2013.

[9] **Intergovernmental Panel on Climate Change (IPCC)**, "General guidance and reporting," *2006 IPCC Guidelines for National Greenhouse Gas Inventories*, vol. 1. pp. 1–309, 2006.

[10] **SABS**, "SANS 14064-3: Specification with guidance for the validation and verification of greenhouse gas assertions." SABS Standards Division, p. 35, 2006.

[11] **A. R. Brandt, Y. Sun, and K. Vafi**, "Uncertainty in regional-average petroleum GHG intensities: Countering information gaps with targeted data gathering," *Environ. Sci. Technol.*, vol. 49, no. 1, pp. 679–686, 2015.

[12] **C. Daniel Mullins *et al.***, "Guidance document: Global pharmacoeconomic model adaption strategies," *Value Heal. Reg. Issues*, vol. 5, pp. 7–13, Dec. 2014.

[13] **SABS**, "SANS 14066: Greenhouse gases — Competence requirements for greenhouse gas validation teams and verification teams." SABS Standards Division, pp. 1–35, 2012.

[14] **R. Grieve, J. Hutton, and C. Green**, "Selecting methods for the prediction of future events in cost-effectiveness models: A decision-framework and example from the cardiovascular field," *Health Policy (New. York).*, vol. 64, no. 3, pp. 311–324, 2003.

[15] **World Business Council for Sustainable Development (WBCSD) and World Resources Institute (WRI)**, "A Corporate Accounting and Reporting Standard," *Greenh. Gas Protoc.*, pp. 1–116, 2004.

[16] **Republic of South Africa**, "Promotion of access to information Act," *Gov. Gaz.*, vol. 416, no. 20852, pp. 1–45, 2000.

[17] **RSA Department of Environmental Affairs**, "National Greenhouse Gas Emission Reporting Regulations," *Government Gazette*, vol. 622, no. 40762. pp. 4–24, 2017.

[18] **SABS**, "SANS 14064-1: Specification with guidance at the organization level for quantification and reporting of greenhouse gas emissions and removals." SABS Standards Division, pp. 1–31, 2006.

[19] **SABS**, "SANS 14064-2: Specification with guidance at the project level for quantification, monitoring and reporting of greenhouse gas emission reductions or removal enhancements." SABS Standards Division, p. 29, 2006.

[20] **SABS**, "SANS 50010: Measurement and verification of energy savings." SABS Standards Division, pp. 1–25, 2011.

[21] **United Nations Framework Convention on Climate Change**, "CDM validation and verification standard," 2014.

[22] **Republic of South Africa**, "Accreditation for Conformity Assessment, Calibration and Good Laboratory Practice Act," *Gov. Gaz.*, vol. 501, no. 29712, pp. 1–13, 2007.

[23] **SABS**, "SANS 14001: Environmental management systems — Requirements with guidance for use." SABS Standards Division, pp. 1–47, 2015.

[24] **RSA National Treasury**, "Regulations in terms of section 12L of the South African Income Tax Act, 1962, on the allowance for energy savings," *Gov. Gaz.*, vol. 37136, no. R.971, pp. 3–10, 2013.

[25] **Global Sustainability Standards Board**, "Consolidated set of GRI sustainability reporting standards." p. 443, 2016.

[26] **SABS**, "SANS 50001: Energy Management systems - Requirements with guidance for use." SABS Standards Division, pp. 1–31, 2011.

[27] **CDP**, "Guidance for companies reporting on climate change on behalf of investors & supply chain members 2016," no. 1122330, pp. 1–182, 2014.

[28] **RSA Department of Energy**, "Draft Regulations regarding Registration, Reporting on Energy Management and Submission of Energy Management Plans," *Gov. Gaz.*, vol. 38614, no. R. 259, pp. 5–48, 2015.

[29] **RSA Department of Environmental Affairs**, "National Pollution Prevention Plan Regulations," *Gov. Gaz.*, vol. 712, no. 40996, pp. 81–87, 2017.

[30] **United Nations Framework Convention on Climate Change**, "CDM Project standard." pp. 1–76, 2015.

[31] **C. Thomas, T. Tennant, and J. Rolls**, "The GHG Indicator: UNEP Guidelines for Calculating Greenhouse Gas Emissions for Businesses and Non-Commercial Organisations," 2000.

[32] **VCS Association**, "Verified Carbon Standard." pp. 1–53, 2016.

[33] **RSA Department of Environmental Affairs**, "Technical Guidelines for Monitoring, Reporting and Verification of Greenhouse Gas Emissions by Industry," vol. TG-2016.1, p. 197, 2017.

[34] **RSA Department of Energy**, "Draft Post-2015 National Energy Efficiency Strategy," *Gov. Gaz.*, no. 40515, pp. 426–471, 2016.

[35] **D. G. Cacuci**, *Sensitivity & Uncertainty Analysis, Volume I: Theory*. 2003.

[36] **SABS**, "SANS 31010: Risk management — Risk assessment techniques." pp. 1–93, 2010.

[37] **The Gold Standard**, "The Gold Standard Requirements." pp. 1–53, 2012.

[38] **SABS**, "SANS 14065: Greenhouse gases - Requirements for greenhouse gas validation and verification bodies for use in accreditation or other forms of recognition." SABS Standards Division, pp. 1–39, 2016.

[39] **D. R. Johnson, H. H. Willis, A. E. Curtright, C. Samaras, and T. Skone**, "Incorporating uncertainty analysis into life cycle estimates of greenhouse gas emissions from biomass production," *Biomass and*

Bioenergy, vol. 35, no. 7, pp. 2619–2626, 2011.

[40] **L. I. Tong, C. W. Chang, S. E. Jin, and R. Saminathan**, "Quantifying uncertainty of emission estimates in National Greenhouse Gas Inventories using bootstrap confidence intervals," *Atmos. Environ.*, vol. 56, pp. 80–87, 2012.

[41] **M. Banks and R. Eaton**, "The WWF Climate Savers Program: Greenhouse Gas Reduction Strategies at Work," *ACEEE Summer Study Energy Effic. Ind.*, pp. 1–7, 2003.

[42] **Department for Business Energy and Industrial Strategy**, "Methodology Paper for Emission Factors," 2016.

[43] **SABS**, "SANS 9001: Quality management systems — Requirements." SABS Standards Division, pp. 1–43, 2015.

[44] **U.S. Environmental Protection Agency**, "Greenhouse Gas Reporting Program, Emission Calculation Methodologies," 2013.

[45] **Climate Action Reserve**, "Climate Action Reserve program manual." Climate Action Reserve, pp. 1–48, 2015.

[46] **M. Zehetmeier, M. Gandorfer, H. Hoffmann, U. K. Muller, I. J. M. De Boer, and A. Heibenhuber**, "The impact of uncertainties on predicted greenhouse gas emissions of dairy cow production systems," *J. Clean. Prod.*, vol. 73, pp. 116–124, 2014.

[47] **H. Yoshida, J. Clavreul, C. Scheutz, and T. H. Christensen**, "Influence of data collection schemes on the Life Cycle Assessment of a municipal wastewater treatment plant," *Water Res.*, vol. 56, pp. 292–303, 2014.

[48] **J. Hong, G. Q. Shen, Y. Peng, Y. Feng, and C. Mao**, "Uncertainty analysis for measuring greenhouse gas emissions in the building construction phase: A case study in China," *J. Clean. Prod.*, vol. 129, pp. 183–195, 2016.

[49] **W. Booysen**, "Measurement and verification of industrial DSM projects," PhD thesis, North West University, 2014.

[50] **Joint Committee for Guides in Metrology**, "Evaluation of measurement data — Guide to the expression of uncertainty in measurement," *JCGM*, vol. 100. pp. 1–134, 2008.

[51] **K. Campbell**, "Evaluating the feasibility of the 12L tax incentive for energy intensive industries," M.Eng dissertation, North West University, 2016.

[52] **A. Schoeman**, "Omskakel van CO2 in krediete kos tot R1 m.," *Netwerk24*. 2016.

[53] **G. H. Popescu, E. Nica, R. O. Ştefănescu-Mihăilă, and G. Lăzăroiu**, "The United States (U.S.) steel import crisis and the global production overcapacity till 2016," *Metalurgija*, vol. 55, no. 3, pp. 538–540, 2016.

[54] **L. Uusitalo, A. Lehikoinen, I. Helle, and K. Myrberg**, "Environmental Modelling & Software An overview of methods to evaluate uncertainty of deterministic models in decision support," *Environ. Model. Softw.*, vol. 63, pp. 24–31, 2015.

[55] **International Organization for Standardization**, "ISO and IEC International Standards for policy makers," 2015.

[56] **Z. Song-Li and W. Wen-Tao**, "Development of National Greenhouse Gas Emissions Inventories in the Context of International Climate Negotiations," *Adv. Clim. Chang. Res.*, vol. 4, no. 1, pp. 60–68, 2013.

[57] **Intergovernmental Panel on Climate Change (IPCC)**, "Reporting instructions." pp. 5–7, 1996.

TOWARDS AN INTERNET-OF-THINGS FRAMEWORK FOR ASSISTING QUALITY-CONTROLLED-LOGISTICS DECISION MAKING WITHIN THE FRESH PRODUCE SUPPLY CHAIN.

L. Joubert[1*] & L. Louw[2]

[1,2]Department of Industrial Engineering
University of Stellenbosch, South Africa
[1]17518113@sun.ac.za, [2]louisl@sun.ac.za

ABSTRACT

Fresh food is mainly wasted due to overproduction and the natural decay of food quality which cannot be prevented. Hence, actors in the fresh food supply chain are responsible to monitor and control activities that influences the quality of fresh food. The emergence of new technologies such as Internet-of-Things (IoT) creates the opportunity to collect real-time food quality information, which may be used to assist and adapt logistic activities to ensure that food quality remains in the accepted quality limits. The aim of this paper is to identify current knowledge on quality-controlled logistics (QCL) in the fresh produce industry, and to identify whether there are opportunities to implement IoT-technologies, from the perspective of experts working in the fresh produce industry.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, University of Stellenbosch, South Africa
[2]Corresponding author

1. INTRODUCTION

Roughly one third of edible fresh food is wasted because the quality has dropped below acceptable limits [1]. Food waste is therefore a global issue. In South Africa it is estimated that the fruit and vegetable commodity group have a significant contribution to food waste, as it contributes 44% to the total food waste [2]. The World-Wide Fund [3] estimated that 50% of fruit and vegetables are wasted during the post-harvest stage, 25% is wasted during process and packaging, and 20% is wasted during the distribution and retailing stages. Two main factors that contribute to food waste are overproduction of produce and the natural decay of food quality which cannot be prevented but is accelerated by poor supply chain management [4].

Food waste could be reduced by implementing quality control in food logistics, meaning that information about attributes which affect food quality, could be used to make better logistic decisions. Recent technological developments make it easier to collect the relevant information regarding food quality attributes. Researchers are investigating the use of Internet-of-Things (IoT) technologies within the food industry, as they believe that IoT will create numerous benefits for the fresh food supply chain [5]. Sundmaeker et al. [6] mention that IoT in the supply chain will create the following capabilities:

1. Better sensing and monitoring of production, crop development, and food processing;
2. Better understanding of environmental conditions, and to gain knowledge about appropriate management actions;
3. More sophisticated processing and logistics operations by actuators and robots;
4. Improving food quality monitoring and traceability by remote controlling the location and conditions of shipment and products; and
5. Increasing consumers' awareness of sustainability and health issues by personalised nutrition and wearables.

Although researchers believe that the use of IoT-technology provides several benefits, there is still uncertainty whether stakeholders in the fresh food supply chain will reap the full benefit and value from IoT technologies, as Sundmaeker et al. [6] state that there are technical and non-technical challenges that need to be addressed.

The aim of this paper is to identify current knowledge on quality-controlled logistics (QCL) in the fresh produce industry, and to identify whether there are opportunities to implement IoT-technologies, from the perspective of experts working in the fresh produce industry. Section 2 provides a summary of literature reviewed on fresh food supply chains, food quality, quality-controlled logistics, and Internet-of-Things within food logistics. Section 3 discusses the data collection and feedback received from experts within the fresh produce industry, and Section 4 concludes this paper.

2. LITERATURE REVIEW

The following section provides a summary of literature reviewed on fresh food supply chains and the use Internet-of-Things within the food industry.

2.1 The complexity of fresh food supply chains

A supply chain consists of actors (companies) that collaborate to put products in the market [7]. Each actor has a series of activities and processes to fulfil to add value to the product so that the consumers are willing to buy the product [8]. Fresh food supply chains usually consist of four stages, as shown in Figure 1. The first stage is called harvesting, where farmers are responsible for the production and harvest of fresh food. After harvesting, food is transported to the second stage, called processing and packaging. Here the fresh food is washed, packaged and stored until it needs to be transported to a distribution centre, which is the third stage. At the distribution centre, fresh food is stored until it must be shipped to the retailer. The last stage is the retailing where the retailer is responsible for selling the fresh food to the consumers.

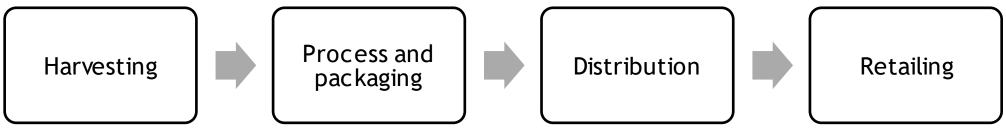


Figure 1: The four stages in a fresh food supply chain.

Fresh food supply chains are classified complex due to various reasons. Although the fresh food supply chain normally consists of four stages, it does not necessarily mean that there are only four actors. The number of actors involved depend on the size and strategy of the supply chain. Van der Vorst et al. [9] explains that actors in fresh food supply chains usually participate in different supply chain processes which means that that they may collaborate with partners that are competitors in other chains. Jonkman et al. [10] further explains that fresh food supply chains are complex because the supply design and performance measurements are highly dependent on product integrity. The reason is that fresh food is perishable by nature, meaning that the products deteriorate rapidly once they have been harvested [11].

Perishability influences the value and the quality of the products [11], and according to Aramyan et al. [12] food quality should be considered as a key performance measurement since it is one of the main characteristics that makes the fresh food supply chain complex.

2.2 Food quality

Food quality is becoming increasingly important to measure and monitor throughout the fresh food supply chain. This is in part due to consumers' expectations on food quality which are becoming a key influenc e during their purchase decision [8]. It is difficult to measure food quality because (i) each actor in the supply chain has their own perception of quality; and (ii) there are various product and environmental factors involved that influences food quality [13] [14].

To address the challenge of measuring and monitoring food quality, researchers in food science developed the concept of shelf-life. Shelf-life is defined by Jedermann et al. [15] as the 'time span for which fresh food can be stored at a certain reference temperature until it is no longer suitable for human consumption or when the food quality does not meet the freshness requirements of consumers. Shelf-life models can be used to predict the time span that is left in total for transport, storage and display in the shop as a function of the environmental conditions to which fresh products may have been exposed to, if such information is available [4]. The accuracy of the shelf-life model depends on the number of quality attributes that are available to monitor and measure, but Jedermann et al. [15] mention that even the simplest shelf-life model provides great insights on product quality and estimated remaining shelf-life.

2.3 Quality Controlled Logistics

Dani [16] defines logistics as 'the operational component of supply chain management, including the quantification, procurement, inventory management, transportation management, and data collection and monitoring'. In fresh food supply chain management, one can then argue that logistics is the movement of fresh food through the supply chain until it reaches the consumer. What makes fresh food logistics more challenging, is that the movement of products must be in adequate environments to ensure that food quality is retained [17]. Van der Vorst et al. [8] introduced a concept called "*Quality Controlled Logistics*", where they suggest that product quality should be considered to determine the required logistic strategies to implement throughout the supply chain. They proposed to view product quality as a dynamic issue and that one should use time dependent quality information to adapt logistics activities. Van der Vorst [18] continued to improve on this concept and developed key features and the necessary requirements that need to be considered when implementing Quality Controlled Logistics, as shown in Figure 2.

Consumer preference

- Understanding consumer preference of quality attributes.
- Get insigths in product acceptance period for all market segments.

Critical control points (CCP) for quality and logistics

- Get insight in dynamic behaviour of product quality attributes.
- Systematic assessment of CCP that have major impact on product quality and availability in retail outlets.

Product quality measurement and prediction

- Have ability to measure and analyse quality at CCP.
- Have automated quality measuring at CCPs
- Have prediction models to calculate real product quality.

Data logging and exchange of information

- Select local target for quality attributes, reduce variability internally and manipulate product quality in line with market demands.
- Apply quality driven inventory management principles.
- Sort homogeneous small batches, have quality based picking and transport.

Local dynamic or adaptive logistics and quality control

- Use advanced data loggers and techniques to capture relevant information real-time and reliable
- Exchange quality, supply and demand information in chain.

Supply chain management

- Establish full chain collaboration to create demand-driven chains.
- Direct goods to highest value markets and apply efficient replenishment in JIT principles for responsive distribution.
- Use temperature controlled containers and packaging.

Figure 2: Requirements to implement Quality Controlled Logistics (reproduced from [18]).

2.4 Internet-of-Things in food logistics

There is a growing interest in using Internet-of-Things (IoT) technologies in various industries, as it is an emerging technology that is expected to offer promising solutions such as tracking, monitoring and data accessibility to transform the operations within these industries [5]. IoT can be defined as ‘a dynamic global network infrastructure with self-configuring capabilities based on standards and interoperable communication protocols where physical and virtual “things” have identities, physical attributes, and virtual personalities and use intelligent interfaces, and are seamlessly integrated into an information network” [19]. Hence, IoT is a paradigm where objects can interact and cooperate with one another through wired and wireless connection, as shown in Figure 3. The main goal of IoT is to interconnect unique, addressable things to generate and share information across diverse platforms and applications, considering security and privacy issues [19] [20].

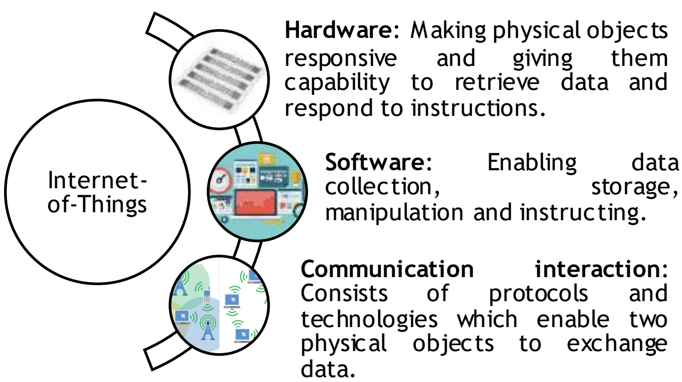


Figure 3: Components of Internet-of-Things.

As mentioned previously, fresh food supply chains can become complex due to the number of actors involved, product quality, and various logistics activities [21]. Several authors studied the potentials of using IoT technologies within the fresh food industry. Da Xu et al. [22] concluded that IoT provides the possibility to improve the traceability, visibility and controllability of food quality and safety throughout the entire supply chain. It can also be used to capture large amounts of data that can be further analysed and used to improve business processes and support decision-making. Sundmaeker et al. [6] mentioned that the use sensor technologies such as humidity, light and ethylene indicators are increasingly used to manage food quality, and temperature sensors are used to monitor the conditions in packing facilities, cooling storages and transportation. Implementing IoT technology in the fresh food supply chain creates the opportunity to base logistics decisions on the dynamic change of food quality over time.

Researchers estimate that the implementation of IoT technologies creates the opportunity to tackle food waste as well. Smart sensors can be used to monitor temperature, freshness, respiration and ethylene gas to detect food spoilage. This information then can be used to determine accurate shelf-life and redirect products to closer locations depending on the remaining shelf-life. Thus, it reduces the chance that food is wasted before it reaches the consumer [23]. Additionally, Vasseur [24] explained that IoT sensor data and analytics is a promising start to gain insights on identifying elements that contributes to food waste. It allows supply chain actors to take the necessary action to control or mitigate the elements to reduce food waste. He also discussed that linking IoT monitoring with task management encourage behavioural change and efficiency to minimize food waste. Integrating IoT monitoring with mobile devices and apps, allows staff to take just-in-time actions to protect food and minimize waste. Notifications and alerts via email, SMS, or automated phone calls can be used to take notify staff about anomalies in the supply chain.

From the literature that has been reviewed, it was identified that fresh food supply chains are defined as unique and complex especially since food quality should be considered when planning the logistic activities. Researchers developed the concept of Quality Controlled Logistics to incorporate food quality with the logistic activities. Furthermore, researchers believe that IoT-technologies can be used to collect data on food quality that can provide more accurate information to assist Quality Controlled Logistics. In the next section the paper will discuss opinions and feedback from experts working in the fresh produce industry to acknowledge whether they are familiar with Quality Controlled Logistics as well to receive their opinion on the use of IoT-technologies in the fresh produce industry.

3. EXPERT INSIGHTS REGARDING QUALITY CONTROLLED LOGISTICS AND INTERNET-OF-THINGS

The following section describes the approach used to receive feedback from experts within the fresh produce industry as well as an analysis and discussion on the feedback they provided.

3.1 Data collection technique and approach

Selecting the best technique to collect data often depends on the (i) purpose of data collection; (ii) type of information required; (iii) resources available; and (iv) evaluation of the collected data [25]. The purpose for data collection within this paper was to gain relevant knowledge on the following topics:

- identifying essential logistics decision regarding fresh produce quality and the requirements or inputs necessary for these decisions;
- establishing ways to use IoT-technologies to enhance Quality Controlled Logistics; and
- identifying essential requirements or considerations to implement IoT in the fresh produce industry.

An interviewer questionnaire was designed with the purpose to ask semi-structured questions to experts working in the fresh produce industry. Interviewer questionnaires refer to those questionnaires where the interviewer physically meet participants and ask the questions face to face. Semi-structured questions are a mix of open and closed questions, meaning the participant has the opportunity to provide feedback based on a predetermined selection of answers, as well as providing answers based on the participant's freedom.

3.2 Setting and sampling

FRUIT LOGISTICA is a fresh produce exhibition, annually in February in Berlin, the capital of Germany. It covers every sector of the fresh produce industry, provides the latest innovations and offers networking and contact opportunities to global companies. From the 7th until the 9th of February 2018, companies from multiple countries presented their products and services such as (i) multiple fresh products; (ii) technical systems; (iii) logistics; and (iv) other services. Further information can be found on the FRUIT LOGISTICA website: https://www.fruitlogistica.de/en.

During the exhibition, companies were asked to complete the questionnaire. Fifteen companies were willing to participate. The participants who completed the questionnaires were from companies that provided (i) logistic solutions for transportation and warehousing management; and (ii) technical solutions that can be integrated with transport, warehouse and quality management, as shown in Figure 4 on the left side. The participants who partook, were either marketing and sales agents, managers within their field of expertise, or the company CEO, as illustrated in Figure 4 on the right side.

Figure 4: Participants feedback profile in terms of company type (left) and role within the company (right).

3.3 Feedback

The questionnaire was designed to collect feedback regarding two themes: (i) Quality Controlled Logistics; and (ii) Internet-of-Things (IoT). The first set of questions were designed to identify whether companies consider fresh produce quality within their logistic activities in the supply chain, and which logistic activities influence produce quality before it reaches the end-consumer. These questions were designed to receive feedback based on predetermined answers, meaning the participants were able to select the answers that they believed were most suited for the question. Each time a specific answer was selected, it was counted to determine the most popular answers.

The second set of question were designed to determine whether companies within the fresh produce industry believe that IoT-technologies can create value within the supply chain, and to determine what they believe are requirements and challenges to implement IoT-technologies throughout the supply chain. These questions were designed to receive feedback based on the participant's knowledge and opinion on the topic. Similar feedback from the participants were clustered together and the most frequent responses were included in the feedback analysis.

3.3.1 Feedback regarding Quality Controlled Logistics

The first question related to identify general types of logistic decisions that need to be considered in the fresh produce industry. From the feedback shown in Figure 5, temperature management, and transport modes and route scheduling are general logistics decisions to consider within the fresh produce industry since more than half of the participants responded that they consider it as general logistic decisions. Managing inventory and stock levels and considering storage practices can also be noted as general logistic decisions, as more than five participants responded that they also focus on these logistic decisions within their companies. Less than five participants said that they consider order picking, shelf space allocation, layout design, and replenishment policies are general logistics decisions within their company.

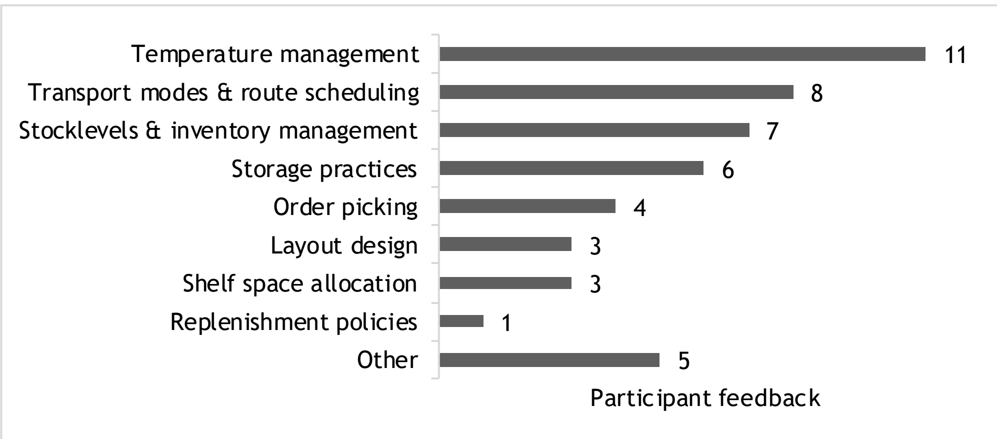


Figure 5: General logistics decisions to consider within the participant's company.

Questions two and three were asked to identify key logistic decisions that influences produce quality after harvest, as well as acknowledging critical quality control points provided by the participants feedback. From Figure 6, the feedback provided by the participants are relatively equal. This suggests that all four decisions mentioned in the question, have influence on produce quality once it has been harvested. Figure 7: Critical quality control points. Figure 7 illustrates that the participants believe that the conditions during transportation as well as storage conditions in either warehouses or distribution centres are critical quality control points in the fresh produce supply chain.

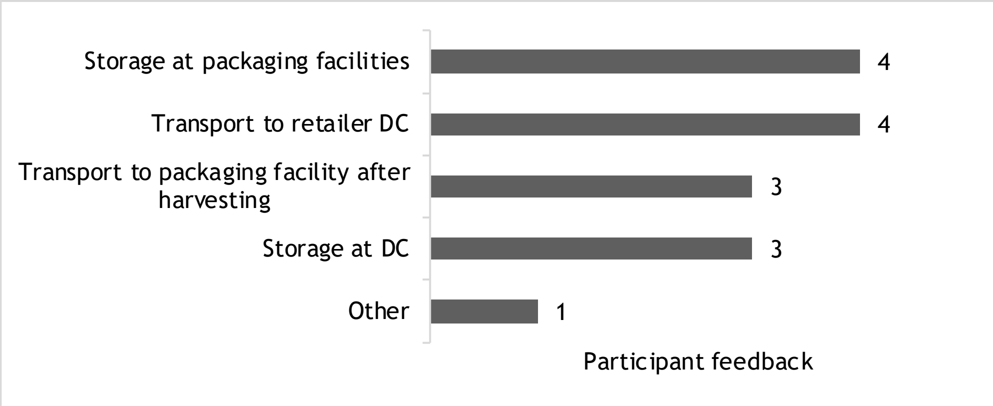


Figure 6: Logistic decisions with the most influence on quality after harvesting.

Figure 7: Critical quality control points.

The purpose of questions four and five were to determine what environmental conditions influence produce quality, and which of these environmental conditions were considered as important conditions to monitor throughout the supply chain. Figure 8 shows the environmental conditions that the participants suggest are important to measure throughout the supply chain. Temperature and relative-humidity are considered as extremely important environmental conditions to monitor, as stated by at least 13 participants. Ethylene is also considered as a condition to monitor, since four participants said it is extremely important to monitor it, and seven participants said it is moderately important. Other environmental conditions such as the lighting, shock and vibration during transportation, as well as controlled atmosphere during transportation and storage were mentioned by the participants.

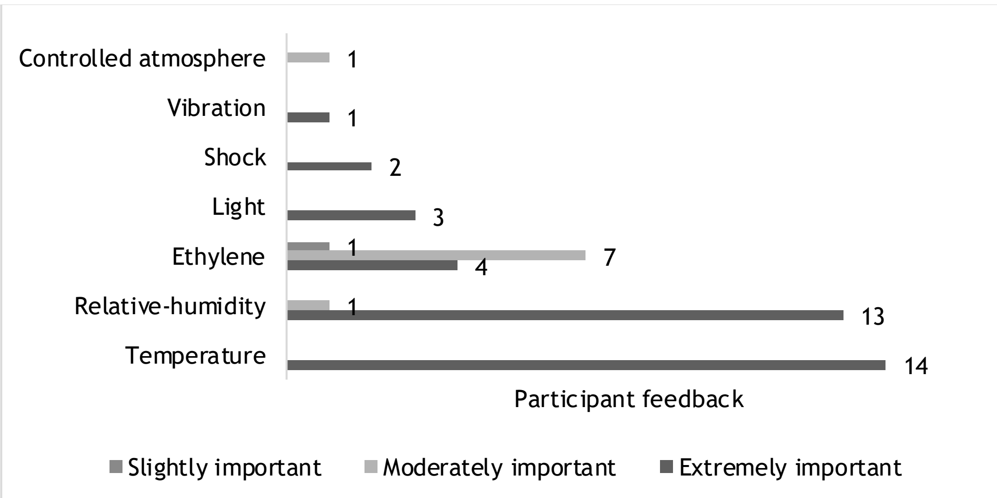


Figure 8: Importance of monitoring the defined environmental conditions.

The next question related to the level of influence predetermined logistic activities have on fresh produce quality. A list of logistic activities was given to the participants and they had to select the level of influence each activity has on produce quality. From the feedback shown in Figure 9, it is noted that majority of the participants mentioned that transport scheduling and routing have an extreme influence on produce quality. Activities such as transport mode selection, distribution network design, and storage mode and capacity were also highlighted as activities that have extreme influence on produce quality. Participants also mentioned that activities such as loading capacity and order picking planning had moderate influence on produce quality.

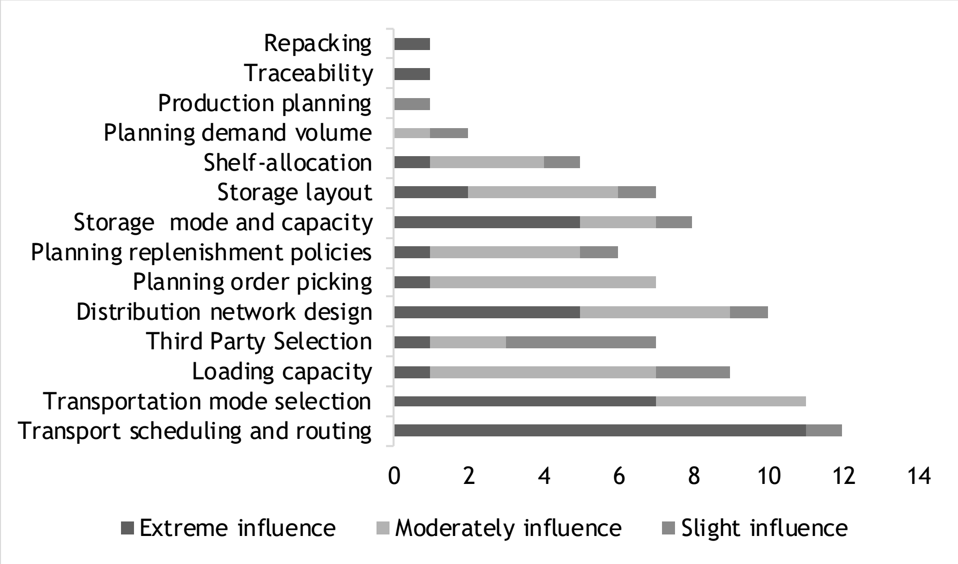


Figure 9: Different levels of influences each logistic activity has on quality.

The last question focusing on Quality Controlled Logistics, was asked to receive feedback on how produce quality information can add value to current logistics activities in the fresh produce industry. The participants shared their feedback and it was then clustered into groups with similar response. As shown in Figure 10, monitoring environmental conditions, transportation management, and supply chain management were mentioned as the most common activities where produce quality related information can create additional value. Four participants mentioned that it will create additional value to improve transparency and visibility through the supply chain, and three participants mentioned it will create additional value to inventory management and quality management.

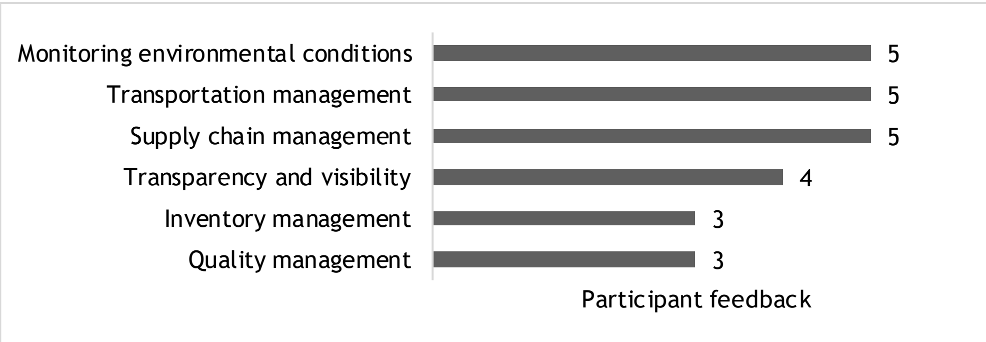


Figure 10: Logistic activities where quality information may add value.

3.3.2 Feedback regarding Internet-of-Things

The following set of questions were asked to the participants to identify their knowledge on the concept of IoT, and whether they believe that IoT-technologies can contribute to assist logistic activities and decision making in the fresh produce industry. The first question gave participants the opportunity to provide some benefits when implementing IoT-technologies in the supply chain. Their feedback was clustered into eight groups as shown in Figure 11. From the feedback, eight participants mentioned that IoT-technologies will assist with the monitoring of environmental conditions, and six participants said it will provide quick access to real-time data and route scheduling. It was mentioned five times that it will provide visible product quality through the supply chain. Other benefits that were mentioned, include optimising supply chain, processes, cost reduction, supplier trust and increased supplier performance, and reducing human error.

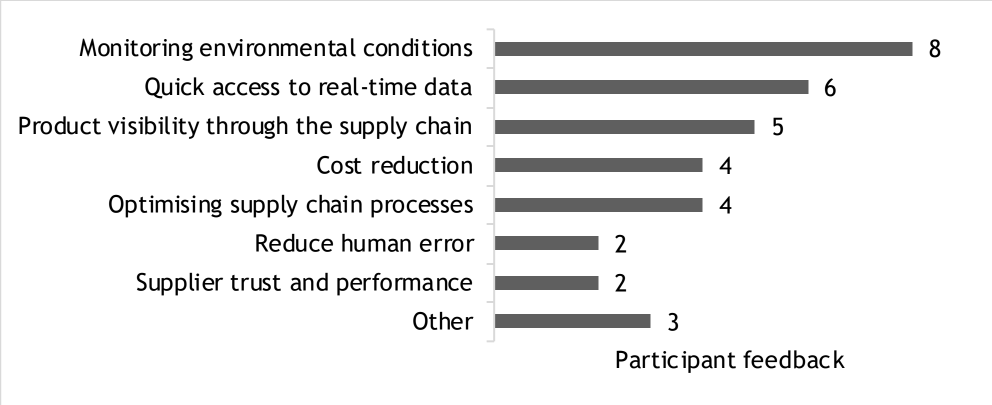


Figure 11: Benefits implementing IoT-technologies in the produce industry.

Next, the participants were asked to provide feedback on how IoT-technologies can add extra value to logistic decisions within the produce industry. Feedback was once again clustered into similar responses provided by the participants. Figure 12 shows the different value creation opportunities that were identified by the participants. Participants mentioned that IoT-technologies can add value to strategic supply chain planning by creating more flexible and adaptive planning, and using data to identify separate market requirements, and developing new business models. Another popular response is that IoT-technologies have the capabilities to make logistic activities more visible and transparent throughout the supply chain. An interesting response that was mentioned six times, is that participant believe that IoT technologies will enhance stakeholder relationships and trust throughout the supply chain.

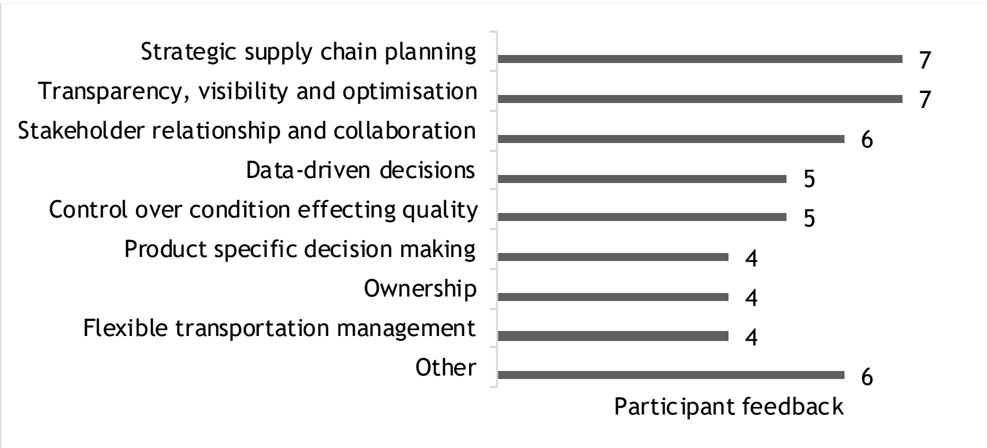


Figure 12: Value creation in the fresh produce supply chain.

Figure 13 shows the feedback participants provided to the question related to identify technical requirements for the implementation of IoT-technologies in the fresh produce industry. Popular feedback from the participants were the kind of network connectivity required, the type of application devices as well as the user interfaces required, and the type of sensing devices and data loggers to capture relevant data. Only two participants mentioned that data security is an important requirement for IoT implementations.

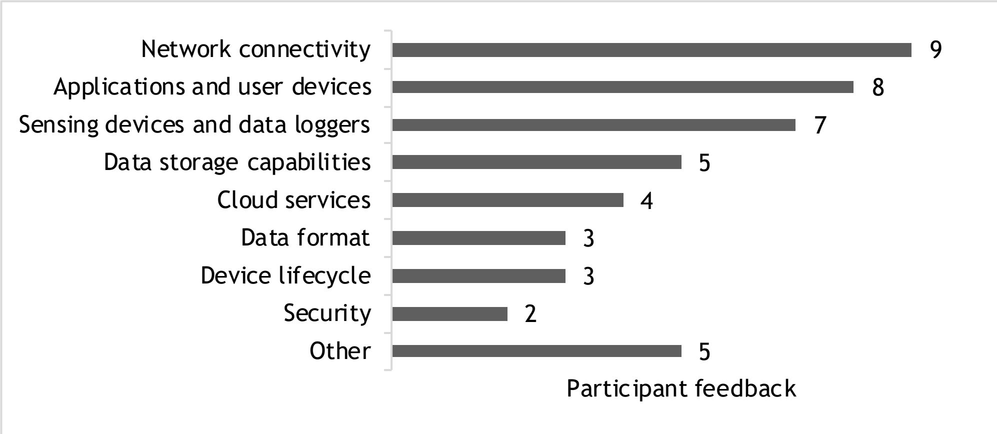


Figure 13: Technical requirements for IoT implementation.

The aim of the last question was to determine what type of challenges the participants believe there are to implement IoT-technologies in the produce industry. From Figure 14, it is noted that acceptance and collaboration to implement IoT-technologies, as well as integrating these technologies with current systems, were the most common challenges mentioned by the participants. The cost of implementation was also mentioned by four participants, but other participants believed that the associated costs to IoT-technologies are not a significant challenge. Other challenges that were mentioned by the participants are the physical installation of the IoT-system, training the users, and receiving valuable information from the collected data.

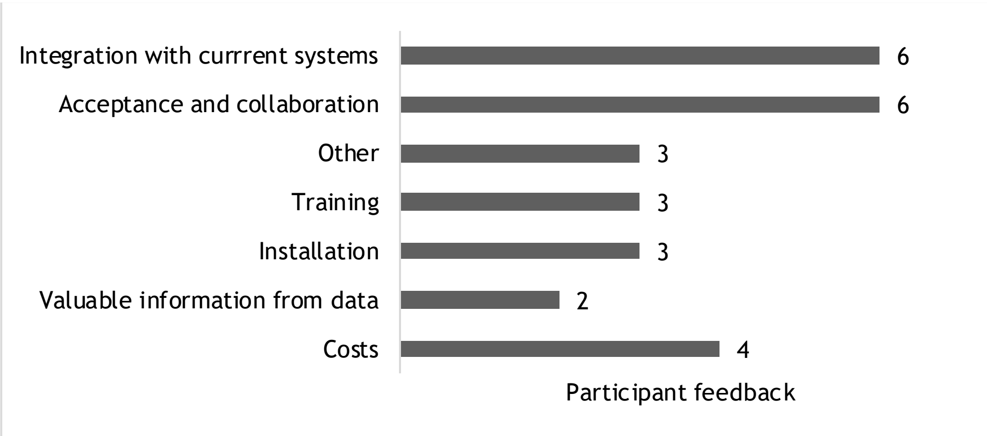


Figure 14: Challenges to implement IoT-technologies.

3.4 Discussion

By analysing the feedback from experts in the fresh produce industry, there seemed to be a positive attitude towards the implementation of IoT-technologies to assist several logistic activities in the fresh produce supply chain. The use of IoT-technologies in the fresh produce industry is not new, since there already exist solutions to monitor environmental conditions. Although it is widely used during transportation management, it is believed that these technologies can create further benefits in the fresh produce industry and will be discussed in the points below.

3.4.1 Temperature management and transportation

From the feedback given by the participants, temperature management, otherwise known as "cold-chain monitoring" during transportation, plays an important role within fresh produce supply chain logistics. Stakeholders in the fresh produce industry realise that inappropriate temperature-management is one of the key reasons why produce are wasted throughout the supply chain. Therefore, transportation companies realise the value of reliable temperature monitoring within different transportation modes. Transport companies have started to invest in solutions such as (i) refrigerated and deep-freeze trucks; (ii) trucks with isolation barriers at the doors; and (iii) temperature recording systems within the trucks, to monitor the temperature conditions while transporting fresh produce to their required destinations. The use of these solutions makes it easier to ensure that temperature stays within acceptable limit, and provides more accurate temperature reading, since temperature logging was done manually in the past.

3.4.2 Monitoring other influences on produce quality

Produce quality depends on both intrinsic and extrinsic attributes. From the feedback, it is realised that companies invest in multiple solutions to monitor extrinsic attributes such as temperature, relative-humidity within transportation and storage infrastructure, and lighting, shock, and vibration during transportation. Although many participants agreed that influences such as ethylene have an influence on produce quality, there have not been many solutions implemented to monitor and track these intrinsic attributes throughout the supply chain. Monitoring intrinsic attributes can provide meaningful information on produce quality, especially when these attributes can be incorporated into shelf-life or prediction models to assist logistic decisions.

3.4.3 Importance of shelf-life models

Participants were asked what type of methods they use to estimate remaining shelf-life of produce, and majority responded that best-before-dates are used. The problem with this method, is that it is a static method and ignores temperature exposure history as well as other conditions affecting shelf-life. To develop more accurate and dynamic shelf-life models, it is required to monitor attributes that influence the produce quality such as temperature, relative-humidity, and ethylene release, depending on the produce characteristics.

Developing accurate quality related shelf-life models can be a game-changer in the fresh produce industry, as it has capabilities to become a useful tool to assist decisions regarding several logistic activities in the fresh produce supply chain. Most warehouses and distribution centres use a "First-In-First-Out" inventory management approach. The problem with this approach is that it assumes the quality of produce arriving the same day at warehouses or distribution centres have equal shelf-life, which is not true due to different conditions they have been exposed to. Knowing the predicted shelf-life based on quality, products can be distributed via a First Expiry-First-Out (FEFO) approach. Another benefit using quality related shelf-life models and the FEFO inventory management approach, is that produce shipment can be distributed more "intelligently" by matching the remaining shelf-life to the required lead time until the shipment reaches the desired location. Not only can it reduce food waste during transportation, but it ensures that remaining shelf-life in a shipment is uniform, meaning the variation of quality is less.

3.4.4 Need for accurate visual inspection

Visual inspection is a fast and common method used to assist quality management, and to decide whether the fresh produce should be accepted or rejected. The problem with visual inspection is that it only reflects visible deterioration and do not consider other influences that reduces quality. The challenge with perishable products is that they usually look fresh until just before they are about to get spoiled. It makes it difficult to distribute correct quality fruit to the various market segments, as good-quality produce may be rejected, or bad-quality produce being accepted. Fortunately, from the feedback provided, participants realise this challenge, and agreed that there is a need to make intrinsic quality attributes more visible and transparent through the supply chain to improve and assist logistic activities such as quality management, better supply chain management, and monitoring environmental conditions to ensure optimal produce quality for end-consumers. The benefits of making produce quality information more visible are numerous since it might (i) improve quality consistency; (ii) reduced waste due to better logistic decisions; and (iii) a higher delivery yield, since produce will be accepted based on more accurate quality information.

3.4.5 Need for pallet-level monitoring

Many companies in the fresh produce industry realise the importance of monitoring environmental conditions, since it influences the quality of produce to some extent, and multiple solutions can be implemented to assist companies. Only a few companies realise that pallet-to-pallet variations exists, hence they realise that environmental monitoring only provide limited produce quality information. To address this problem, new solutions must be created to monitor and act upon these variations, and one method is implementing pallet-level monitoring. Pallets are an effective position to monitor, as it spends majority of its life-cycle with the produce self. Pallet-level monitoring is capable to capture the actual conditions of the produce and provide more accurate data to develop shelf-life prediction models.

Another important benefit of implementing pallet-level monitoring is that it can act upon quality variation and reduce losses. For example, Jedermann et al. [15] mention in a case study that the temperature within pallets during transit is not the same. This result that some pallets may experience quality degradation faster than others. Knowing the quality of produce inside pallets, can change the way several logistics activities such as inventory management, and distribution are executed, and enhance shelf-life optimisation.

4. CONCLUSION

Food waste is a significant issue globally and many initiatives are being researched and implemented to reduce food waste. Concepts such as quality-controlled logistics are promoted within the fresh produce supply chain. It is believed that it can contribute to the reduction of food waste when produce quality is considered whilst determining logistic strategies. Researchers further believe that the IoT-technologies create the opportunity to assist quality control by monitoring the conditions fresh produce are exposed to.

The goals of the questionnaire were to determine whether experts working in the fresh produce industry were familiar with quality-controlled logistics as well to receive their opinion on the use of IoT-technologies in the industry. From the feedback it was identified that most experts have already implemented quality-control activities such as temperature management, since inappropriate temperature management is one of the key reasons why produce is wasted in the supply chain. Experts are also becoming more interested to implement IoT-technologies in the industry. They realise that it creates the ability to collect product specific quality data that can be used to determine the quality of fresh produce more accurately. Furthermore, they believe that IoT-technologies can provide insights to enhance supply chain planning, such as implementing shelf-life models to adapt inventory and transportation management.

There is still limited literature available to gain sufficient understanding on how IoT-technologies and quality-controlled logistics can be combined to potentially reduce food waste along the fresh produce supply chain and to improve overall produce quality. Also, to the researcher's knowledge there are currently no definite framework available to assist fresh produce companies with the implementation of IoT-technologies to improve quality-controlled logistics.

Hence, this paper forms part of an ongoing research towards the development of a framework to assist quality-controlled logistics by using IoT -technology. Further research is currently being done to compare the impact of traditional logistic activities versus quality controlled logistic activities on food waste, which is caused by shelf-life loss. Thereafter a proposed IoT solution will be developed that collects available real-time data that can be used as input to determine dynamic shelf-life estimations of fresh produce. It is believed that the IoT solution will assist quality-controlled logistics, which will contribute to potentially reduce food waste along the fresh produce supply chain.

5. REFERENCES

[1] **Gustavsson, J., Cederberg, C., Sonesson, U., Van Otterdijk, R. & Meybeck, A.** 2011. Global food losses and food waste. *Food and Agriculture Organization of the United Nations, Rom*.

[2] **Oelofse, S.H.H. & Nahman, A.** 2013. Estimating the magnitude of food waste generated in South Africa. *Waste Management & Research*.

[3] **WWF.** 2017. Food loss and waste: Facts and futures.

[4] **Sciortino, R., Micale, R., Enea, M. & Scalia, G.L.** 2016. A webgis-based system for real time shelf life prediction. *Computers and Electronics in Agriculture*, 127, pp 451-459.

[5] **Zhou, L., Chong, A.Y.L. & Ngai, E.W.T.** 2015. Supply chain management in the era of the internet of things. *International Journal of Production Economics*, 159, pp 1-3.

[6] **Sundmaeker, H., Verdouw, C., Wolfert, S. & Freire, L.P.** 2016. Internet of food and farm 2020. Digitising the Industry-Internet of Things Connecting Physical, Digital and Virtual Worlds. River Publishers, Delft, pp 129-151.

[7] **Trienekens, J.H., Wognum, P.M., Beulens, A.J.M. & Van der Vorst, J.G.A.J.** 2012. Transparency in complex dynamic food supply chains. *Advanced Engineering Informatics*, 26, pp 55-65.

[8] **Van der Vorst, J.G.A.J., Beulaens, A.J.M. & Van Beek, P.** 2005. Innovations in Logistics and ICT in Food Supply Chain Networks. *Innovations in Agri-Food Systems*, pp 245-292.

[9] **Van der Vorst, J.G.A.J., Tromp, S.O. & Van der Zee, D.J.** (2009). Simulation modelling for food supply chain redesign: Integrated decision making on product quality, sustainability and logistics. *International Journal of Production Research*, 47, pp 6611-6631.

[10] **Jonkman, J., Bloemhof, J.M., Van der Vorst, J.G.A.J. & van der Padt, A.** 2017. Selecting food process designs from a supply chain perspective. *Journal of Food Engineering*, 195, pp 52-60.

[11] **Chen, H.K., Hsueh, C.F. & Chang, M.S.** 2009. Production scheduling and vehicle routing with time windows for perishable food products. *Computers & Operations Research*, 36, pp 2311-2319.

[12] **Aramyan, L.H., Ondersteijn, C.J.M., Van Kooten, O. & Lansink, A.O.** 2006. Performance indicators in agri-food production chains. In *Quantifying the agri-food supply chain*, pp 49-66, Springer.

[13] **Luning, P.A. & Marcelis, W.J.** 2007. A conceptual model of food quality management functions based on a techno-managerial approach. *Trends in Food Science & Technology*, 18, pp 159-166.

[14] **Heising, J.K., Claassen, G.D.H. & Dekker, M.** 2017. Options for reducing food waste by `Quality Controlled Logistics' using intelligent packaging along the supply chain. *Food Additives & Contaminants: Part A*.

[15] **Jedermann, R., Nicometo, M., Uysal, I. & Lang, W.** 2014. Reducing food losses by intelligent food logistics.

[16] **Dani, S.** 2015. *Food Supply Chain Management and Logistics: From Farm to Fork*. Kogan Page.

[17] **Macheka, L., Spelt, E., Van der Vorst, J.G.A.J. & Luning, P.A.** 2017. Exploration of logistics and quality control activities in view of context characteristics and postharvest losses in fresh produce chains: A case study for tomatoes. *Food Control*, 77, pp 221-234.

[18] **Van der Vorst, J.G.A.J., van Kooten, O. & Luning, P.A.** 2011. Towards a diagnostic instrument to identify improvement opportunities for quality-controlled logistics in agri-food supply chain networks. *International journal on food system dynamics*, 2, pp 94-105.

[19] **Sundmaeker, H., Guillemin, P., Friess, P. & Woelfflé, S.** 2010. Vision and challenges for realising the Internet of Things. *Cluster of European Research Projects on the Internet of Things*, European Commission.

[20] **Hartmann, M. & Halecker, B.** 2015. Management of Innovation in the Industrial Internet of Things. In *ISPIM Conference Proceedings, 1, The International Society for Professional Innovation Management (ISPIM)*.
[21] **Pang, Z., Chen, Q., Han, W. & Zheng, L.** 2015. Value-centric design of the internet-of-things solution for food supply chain: value creation, sensor portfolio and information fusion. *Information Systems Frontiers*, 17, pp 289-319.
[22] **Da Xu, L., He, W. & Li, S.** 2014. Internet of things in industries: A survey. *IEEE Transactions on Industrial Informatics*, 10, pp 2233-2243.
[23] **Shacklett, M.** 2018. Four Ways IoT Is Decreasing Commercial Food Waste, [accessed 21 May2018], available at: https://www.techrepublic.com/article/4-ways-iot-is-decreasing-commercial-food-waste/
[24] **Vasseur, T.** 2016. Can The Internet-of-Things Help Reduce Food Wastage, [accessed 21 May2018], available at: https://www.itproportal.com/2016/01/15/can-the-internet-of-things-help-reduce-food-wastage/
[25] **Saunders, M. and Lewis, P. & Thornhill, A.** 2009. *Research Methods for Business Students*. Prentice Hall.

THE ROLE GOVERNANCE STRUCTURE PLAYS BETWEEN THE MANUFACTURER AND DISTRIBUTOR DURING THE 4TH INDUSTRIAL REVOLUTION

L. Badenhorst[1*] & E. Van der Lingen[2]

[1]Graduate School of Technology Management
University of Pretoria
ludolf.badenhorst@luadia.com

[2]Graduate School of Technology Management
University of Pretoria
elma.vanderlingen@up.ac.za

ABSTRACT

A system that will be affected by Industry 4.0 is the value chain of manufacturers. Currently the relationship between manufacturers and distributors plays an important role in their success, but it will be tested vigorously while competing in the Industry 4.0 era. Governance structure between the manufacturer and distributor is a means to nurture their relationship, but will have to be tested whether it is sufficient to mitigate risks the distributor might face in the era of Industry 4.0. This study analysed the feedback from 98 South African distributors and found that there is a correlation between the current governance structure in place and distributors willingness to invest towards Industry 4.0, as well as the potential increase in transaction costs.

*Corresponding author

1. INTRODUCTION

Industry 4.0 will be of great economical value, since it is estimated that in 2020 the total global economic added value, purely based on the Internet of Things (IoT) market, will be around $1.9 trillion dollars and €140 billion in Europe alone [1]–[3]. This rapid growth of IoT devices raises great revenue potential for not only the manufacturing companies, but also the supply chains that need to move the product between the manufacturer and the end user [1]. However, the competition for this revenue will be tough, and in order for competing companies to survive in the modern economy they will have to adapt to the demands of Industry 4.0. The term Industry 4.0 has been defined in many different ways in the literature study, but in this study the term Industry 4.0 is a combination as it is used as an umbrella term for new technologies and concepts developed in new era for industrial production [4][5]. One system that will have to adapt to this new environment is the value chain network of manufacturers.

Having a good value chain or improving one's value chain is one of the most important challenges an organisation can face when developing a competitive edge [6]. The benefits of a value chain system can be seen in smaller inventories, better customer satisfaction, enhanced demand responsiveness, higher successful product commercialisation rate, and higher flow rate of knowledge and information [6], [7]. Manufacturers can take advantage of the value chain by relying on distributors to reduce the number of relationships and responsibilities with end customers, and therefore focus more on key accounts [8]. However, one of the biggest challenges distributors encountered, was the introduction to the internet in the 1990's that led to e-commerce. Distributors could no longer rely on product handling for success and had to change their role in the value chain, as the business customer could "with the click of a button" bypass them to procure products and obtain information from the manufacturer directly [8], [9].

It is clear from the literature study that manufacturers' distribution networks will again come under strain during Industry 4.0, and will still play a big role in their success in the new economy [1], [7]. Similar to manufacturers getting their IoT research & development (R&D) ready they should also get their distribution networks ready as this will play a key role in whether the end market will adopt their new technology [7].

The understanding on how Industry 4.0 will impact the distribution network could be enriched by investigating the current relationship between the manufacturer and the distributor, as well as the areas that could influence the relationship. To date, numerous literature have explored the issues of supply network relationships from the viewpoint of the manufacturer [10]–[12]. In contrast, little work has been reported on from the viewpoint of the distributor and their daily challenges, and how they react to these challenges to maintain a positive relationship with the manufacturer. Manufacturers can increase their distributor's relationship by having proper governance structure in place with distributors [13].

The objective of this study is to answer the proposed research question: "What role does existing governance structure play between manufacturers and distributors in lowering the distributor's risks during Industry 4.0?" This was done by investigating two identified financial risks and their dependence on governance structure.

Answering the research question is of importance as this study will contribute to the understanding of manufacturers regarding the current challenges distributors could face during Industry 4.0, as well as how governance can assist in minimising the potential negative impact on the manufacturer and distributor relationship.

2. LITERATURE REVIEW

2.1 The Importance of Distributors during Industry 4.0

PwC [1] released a study where a survey was conducted on 235 German industrial companies. The main areas of focus were on incoming supply chain, research and development (R&D), planning, manufacturing and service providing, rather than on their distribution networks, see Figure 1 [1]. This study shows that even though investments will be made in the distribution networks, companies will still rely heavily on their current distribution systems to deliver their products to the end consumer.

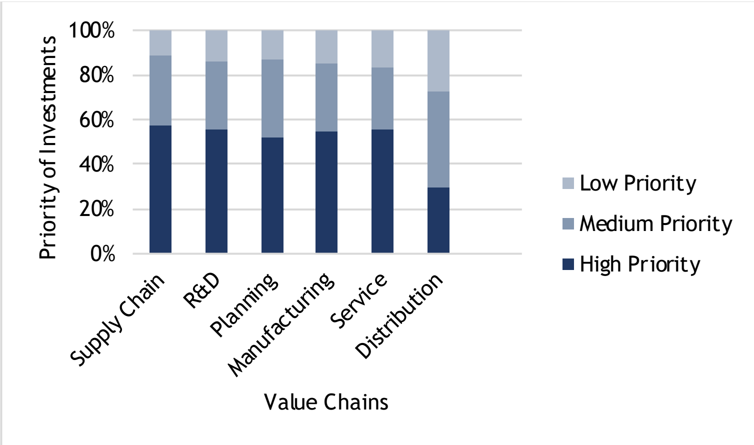


Figure 1: Industry 4.0 investments broken down according to the steps of the value chain (adapted from PwC [1])

PwC [1] concluded that co-operation in horizontal value chains will be of great importance to companies. Figure 2 shows that "Better satisfaction of customer requirements" is the main driving force for co-operation between value chain stakeholders followed by "Faster time to market". Both of these are directly influenced by the distribution network, as the distributor is one of the main sources for information flow to the manufacturer about customer needs, complaints, competitor technologies and major trends [8], [14]. "Faster time to market" will be influenced by the distribution network and in essence the efficiency of the distributor. These two reasons contribute to the important role manufacturers consider distributors to play in commercialising within the Industry 4.0 era.

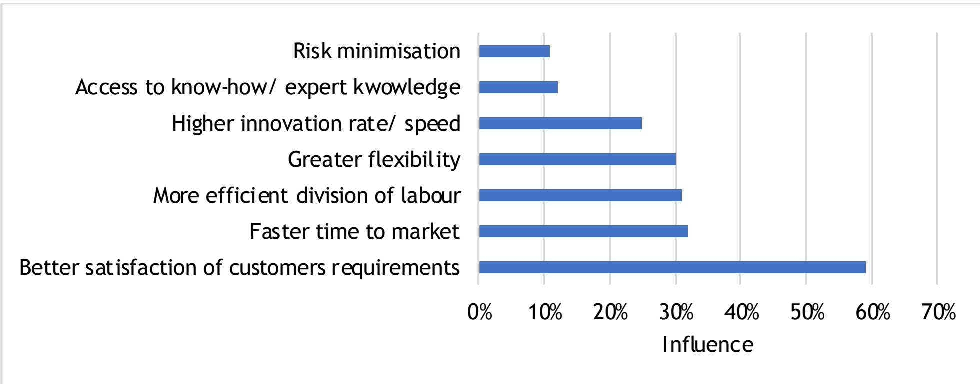


Figure 2: Reasons to co-operate in Industry 4.0 (adapted from PwC [1])

Another area where distributors will play an important part is to commercialise the concept of Industry 4.0 and the relevant technologies and concepts of the era moving forward. Geoffrey Moore has shown in his book "Crossing the Chasm" in 1991 that when commercialising a product to the mainstream market one first has to cross the "chasm" between the early market and the mainstream market. Moore [15] further indicated that the early market plays an important role in adopting the new technology, otherwise the pragmatist in the mainstream market will not accept the radical new innovation, due to their risk adverse attitudes leading to overall unsuccessful commercialisation. Industry 4.0 could follow the same technology adoption life cycle of Geoffrey Moore with rapid growth rates expected over the next few years. This, however, will only be possible if the early market have a positive attitude towards the industry change and in this case the early adopters can be seen as the distributors adding to the commercialisation process [7], [16]. Distributors will not just contribute by identifying potential users, but can have close relationships with end consumers. Aarikka-Stenroos & Sandberg [7] further indicated in their study that these close relationships make it easier to deliver on some of the main commercialisation activities such as: credibility establishment, trust creation, awareness building, distribution and customer education.

2.2 Governance Between the Distributor and the Manufacturer

According to Vázquez-Casielles et al. [13] governance structure consist of:

- Market governance: Transactional and coordination costs;
- Third-party enforcement: Governance enforced by contractual terms;

- Formal safeguards: Investments in non-recoverable expenditures; and
- Informal safeguards: Dependence and relational norms.

Governance between distributors and manufacturers poses as a safeguard towards their relationship, assisting both of them against opportunistic behaviour from one another, minimising transaction costs and increasing the flow of strategic information [13], [17]. Should the safeguards of governance structure not be in place the distributor and manufacturer might be reluctant to invest in resources and capabilities that could contribute to a more difficult commercialisation process [13].

2.3 Financial Factors Influencing Distributor Relationships

Moving from e-commerce to Industry 4.0, PwC [1] found in their study that 45% of the 235 manufacturers interviewed, indicated that they must change their business models to interact directly with end customers, which would result in an obsolete need for a distributor. 46% indicated that they will have to create digitalised value-added services, such as cell phone applications. This could make the distributor more dependent on the manufacturer for software licences and updates. 64% indicated that expanding on their digital services, this could lead to costly upfront investment by the distributor. These statistics makes it clear that manufacturers will change their business models to adapt to Industry 4.0, but that these changes might have an indirect impact on their distributors.

2.3.1. Distributors Willingness to Invest

Distributors could be faced with large transaction-specialised investments to support and distribute the manufacturers new technologies and concepts in this new era. These transaction-specialised investments could be anything ranging from marketing to maintenance tools and are often seen as non-recoverable expenditure [13]. Figure 3 shows the current challenges prohibiting the successful implementation of Industry 4.0 new technologies and concepts. The main challenges are "Unclear economic benefits and excessive investments" confirming that distributors are carrying significant risk with these kinds of upfront investments. Distributors are seen as critical players in the adoption networks and if they are not prepared to invest in an innovation, the end user will most likely also not [7], [16]. Thus, it is crucial that manufacturers have a supporting structure in place to secure trust with their distributors, even though economic benefits are still unclear. Some of these supporting structures could include sales training on measuring risks and rewards on IoT products, since IoT cost benefits have become a subject of great interest for companies [18].

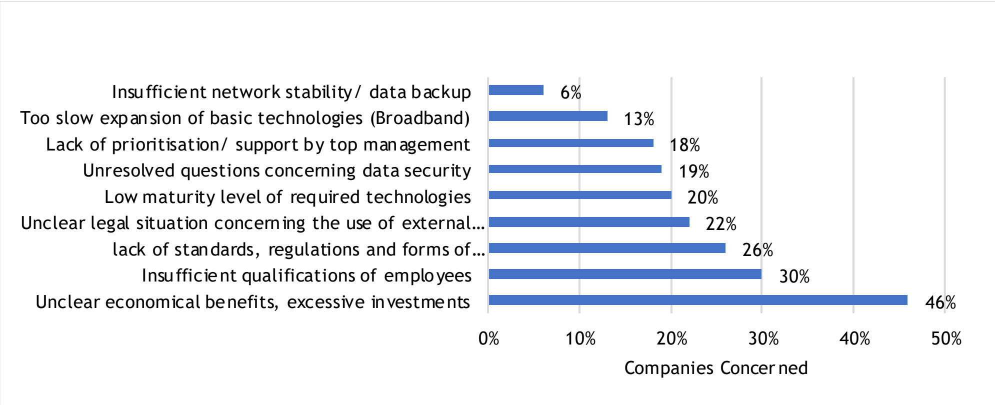


Figure 3: Challenges for successful implementation of Industry 4.0 (adapted from PwC [1])

2.3.2. Distributors Transaction Costs

Transaction costs can either be prior to, as well as after the commercial exchange. Transaction costs are defined as the perceived time and effort to communicate, negotiate and reach agreements in an ongoing exchange [13]. Vázquez-Casielles [13] further indicated that should these transaction costs become costly, they will reduce the willingness of the distributor sharing strategic innovation and collaborate on innovation activities with the manufacturer. An increase in dependency could ultimately add to the transactional costs as the distributor will become dependent on the manufacturer to resolve the client issues and ultimately delaying the commercialising process.

There are two areas that the distributor might become more dependent on the manufacturer. The first dependency is that products no longer will be sold but rather solutions and systems [1]. The flow of information is critical for distributors as it will help them not only to sell products, but rather tailor made systems to the needs of the end consumer [19]. This will become critical as PwC [1] indicated in their study that end consumers will be more systems orientated during Industry 4.0. Wang [19] has further indicated that this flow of information could assist the distributors by supplying more effective after sales services and technical support. Distributors have become more dependent on this information to provide better sales assistance, services and technical support. Further, they could become more dependent on manufacturers to share non-standard information on equipment.

Industry 4.0 is still in its immature phase and it is said that companies leading the race and implementing their technologies first will be the ones that will set the standards and governance rules [20]. With this race manufacturers are faced with a major problem of releasing immature or untested technologies into the commercial market [21]. This gave rise to the second dependency, namely that distributors carry the risk of major product recalls or software updates on these immature untested products that could lead to ultimate brand damage and negative transaction costs [22]. Some of these costly recalls over the past years include the recall of Samsung's Galaxy note 7 and Volkswagen AG Emissions Scandal [23].

3. Conceptual Model

The conceptual model that is derived from the literature study is shown in Figure 4. Figure 4 further shows the two financial risk factors investigated and their linkage to Industry 4.0 and governance structures.

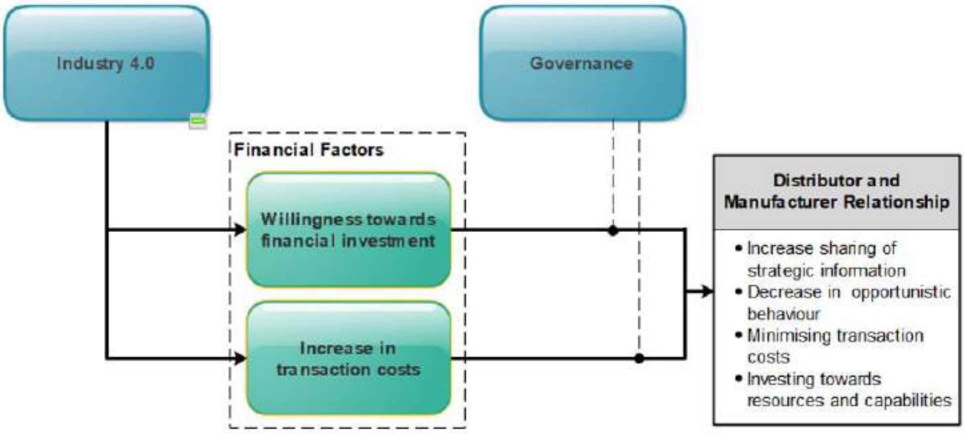


Figure 4: Proposed Conceptual Model

In order to achieve the objective of this study, the following two hypotheses were tested

$H1_1$: The likelihood of distributors investing in Industry 4.0 depends on the existing governance structure.

$H2_1$: Increase in transaction costs in Industry 4.0 depends on the existing governance structure.

4. RESEARCH METHODOLOGY

4.1 Research Design

This study utilised a mixed method approach between quantitative and qualitative methods to take advantage of the differences between the methods. First the quantitative method utilised close ended questionnaires in the form of a Likert scale to express the cause and effect between the identified issues and governance structure between the manufacturer and distributor [24], [25]. For the qualitative part this study utilised interviews to assist in understanding the reasons behind the feedback from the questionnaires, by obtaining rich detailed data from the interviewed distributor.

4.2 Research Setting

This study only focused on South African distributors and their relationships with either local or international manufacturers. The local distributors were identified through manufacturers that are currently members of the "Industrial internet consortium". Additional manufacturers actively advertising their involvement in Industry 4.0 were also identified, since many of the members of the "Industrial internet consortium" currently don't sell their products through the local distributors in South Africa. These distributors were then identified through the manufacturer's web sites and local directories on the internet. They were personally approached to participate in an electronic questionnaire. Random sampling was used to select distributors to participate in the in-depth interviews.

4.3 Sampling Criteria

The following specific criteria had to be met by the distributors in order for them to be included in the sampling for the questionnaire:

- The distributor had to have been actively involved in distributing the manufacturer's products for a period longer than one year. This requirement was designed to ensure that was a mature relationship between them;
- The distributor must distribute the manufacturer's products and not be merely a consultant, installer or repairer; and
- The participant of the distributor had to be at least one of the senior managers or owners of the company and not a general employee. The feedback had to be from employees involved in making strategic decisions within the company.

In total 120 questionnaires were received from the respondents. Of these, twenty-two were rejected for not complying with the sampling criteria, leaving a total of 98 questionnaires to be used in the study. Of the 98 distributors who participated in the questionnaire, only twelve indicated that they would be willing to participate in interviews and six followed through with the interviews.

4.4 Data Analysis

Data obtained through the questionnaires was analysed through the Chi-Square test of independence since the captured data was either a set of observed nominal or ordinal frequencies [26]. Once analysed, Cramer's V test was used as the statistical strength test. This is necessary to ensure the strength of the association, as well as that the significance between the variables is acceptable [26]. These tests were performed with the use of the software package Statistical Package for Social Sciences (SPSS).

The semi-structured interviews consisted of open-ended questions that were analysed through qualitative content analysis with the aim to quantify emerging characteristics and concepts.

5. RESULTS

5.1 Topology of Distributors

It was important to look at the topology of the distributors that participated in this study as each of the differences in the topology could potentially lead to a custom-made governance structure between the manufacturer and the distributor and leading to different standards while answering the questionnaire.

The degree on how much distributors are informed about Industry 4.0 was analysed, as this might have an impact on how they respond. Figure 5 below illustrates the topology of 98 distributors' knowledge on Industry 4.0 indicating that 78% of the participants have heard of the term, whether it be the new technologies and concepts developed or the new era for industrial production. The results were compared in this study with whether their knowledge level on Industry 4.0 made any difference towards both of the proposed hypotheses.

Figure 5: Topology of distributors' knowledge about Industry 4.0

Figure 6 indicates the distributors representing industries with the majority (40%) distributing industrial equipment, and the remaining 60% are divided between smart cities & factories, utilities, IT and security. Other industries that can also play a major role during Industry 4.0, as indicated by the industrial internet consortium, are healthcare, mining and transportation. However, although surveys were distributed to these industries, no credible responses were obtained to include in the study.

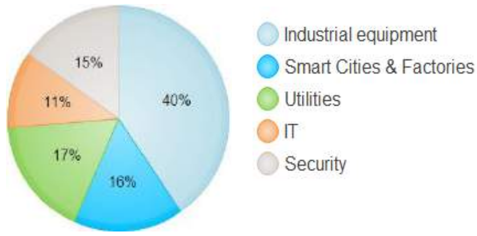


Figure 6: Distributors' representing industries

Figure 7 shows the governance structure between the manufacturer and the distributor. The majority of the distributors currently have a working governance structure in place, while the rest either don't have one in place or it is simply not being enforced. However, because of the low response rate in some areas the governance structure was adjusted to two sections "Working governance" (63%) and "Non-working governance" (37%). This was also done to solve the requirements of the Chi-square test that no cell must have an expectancy of less than five.

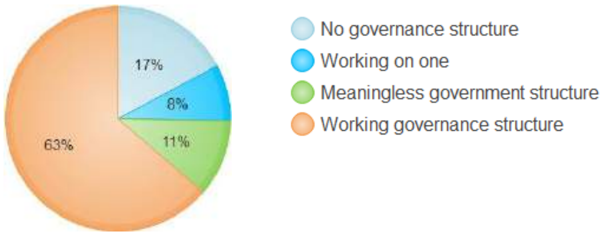


Figure 7: Governance structure between the manufacturer and the distributor

5.2 Distributors Willingness to Invest

Distributors are seen as critical players in the adoption networks and if they are not prepared to invest in a new innovation, similar to those new technologies and concepts of Industry 4.0 the end-user will most likely also not invest [7], [16]. For this reason, it was important to investigate distributors' willingness to invest, where the results can be seen in Figure 8. 83% of distributors are willing to invest towards Industry 4.0, compared to the 16% that won't make any investment. This high figure was not expected as PwC [1] indicated in the study that some of the main challenges towards Industry 4.0 are "Unclear economic benefits and excessive investments".

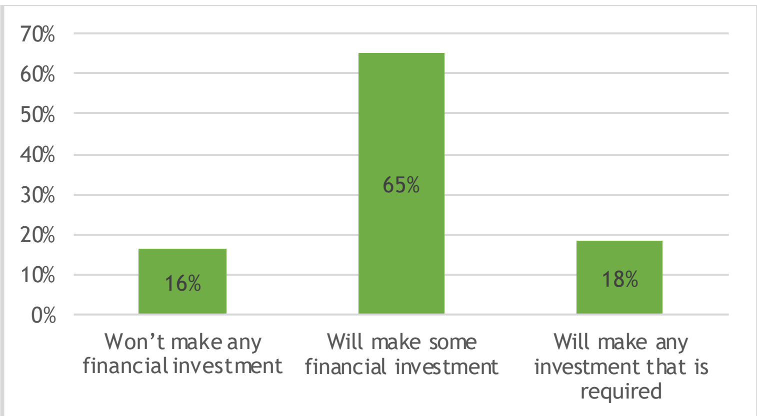


Figure 8: Willingness to make upfront investments

The low willingness to invest towards Industry 4.0 was investigated further by analysing the business models of those 76 distributors carrying knowledge of Industry 4.0 seen in Figure 9. These results create the opportunity for future research on why distributors are willing to invest towards Industry 4.0 even though they don't have a working business model in place leaving them with unclear economic benefits.

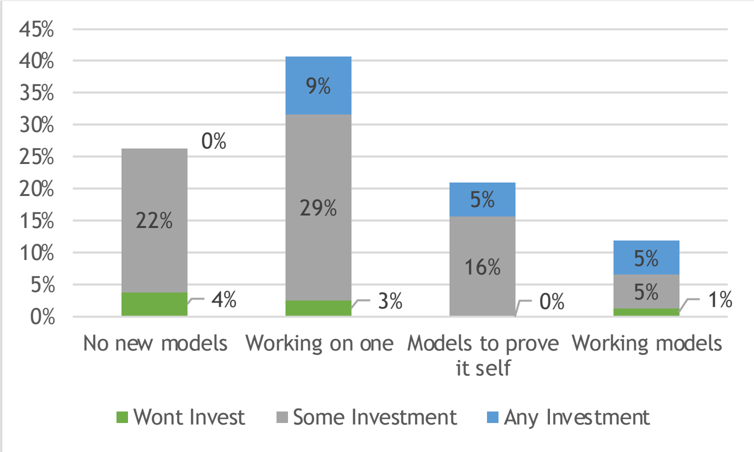


Figure 9: Current business model towards Industry 4.0

This willingness to invest was analysed through the Chi-square analysis to see if there was any correlation towards their current governance structure, see Table 1. Since it is observed that the probability value is less than the 0.05 significance level, it is concluded that the null hypothesis is rejected. Therefore, there is enough evidence to claim that the two variables are dependent.

Table 1: Results of statistical hypothesis 1 testing

Hypothesis	Chi-Square Test Result	Cramer's V Statistical Strength
The likelihood of distributors investing in Industry 4.0 depends on the existing governance structure.	P(x^2 > 14.333) = 0.0008 Accepted	0.382 Strong level of association

The concluded interviews revealed some correlation between the two variables. Three of the five distributors indicated that they will make investments towards Industry 4.0 with one of them having a non-working governance structure in place. The other two distributors with non-working governance structures said that they won't invest now as they don't see the benefits yet.

From the distributors that are willing to invest two of them indicated that they will make investments towards Industry 4.0, as both of them acknowledged that this new era will have a direct impact on distributors and by omitting to invest will cause falling behind. According to one distributor "The role distributors played 10 years ago is different to what they play now and definitely what they will play in 10 years' time. For this reason, I will invest in what we want in 10 years' time and not how it was 5 years ago".

Another observation was that three of the distributors indicated that their manufacturers have an overwhelming marketing campaign towards Industry 4.0 and are investing in promoting the concept to the market. However, all three indicated that the manufacturer has no real concrete strategy in place. Two indicated they won't invest towards Industry 4.0, but will invest once the manufacturer starts to make real concrete investments.

5.3 Distributors Transaction Cost

An increase in dependency could ultimately add to the transactional cost, as the distributor will become dependent on the manufacturer to resolve Industry 4.0 issues experienced by their clients, which will delay the commercialising process. The two dependencies investigated were System Design, and secondly Product Recalls or Software Updates.

The first dependency due to System Design questionnaire feedback is show in Figure 10 and indicates that the majority of distributors will become dependent on the distributor and also not just in the beginning stages but throughout the product life cycle.

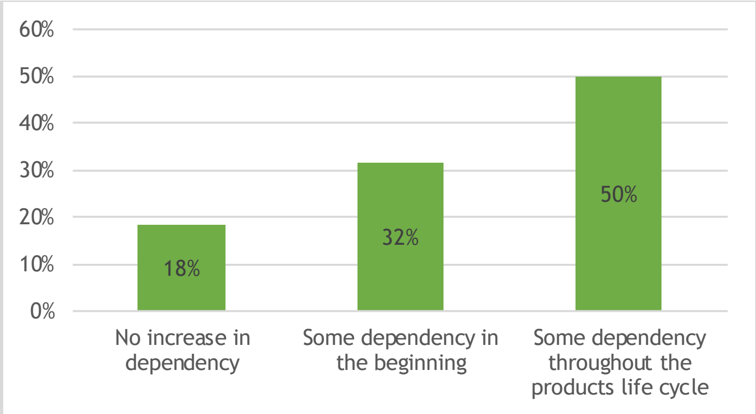


Figure 10: Dependency due to System Design

This dependency due to System Design was analysed through the Chi-square analysis to see if there was any correlation towards their current governance structure, see Table 2. Since it is observed that the probability value is less than the 0.05 significance level, it is then concluded that the null hypothesis is rejected. Therefore, there is enough evidence to claim that the two variables are dependent.

Table 2: Results of statistical hypothesis 2 (A) testing

Hypothesis	Chi-Square Test Result	Cramer's V Statistical Strength
Increase in dependency due to system design during Industry 4.0 depends on the existing governance structure.	$P(x^2 > 8.611) = 0.0135$ Accepted	0.296 Moderate level of association

The second dependency due to Software Updates or Product Recalls is show in Figure 11, where a similar trend can be seen as for the first dependency in Figure 10. This dependency was analysed through the Chi-square analysis to investigate if there was any correlation towards their current governance structure, see Table 3. Since it is observed that the probability value is less than the 0.05 significance level, it is then concluded that the null hypothesis is rejected. Therefore, there is enough evidence to claim that the two variables are dependent.

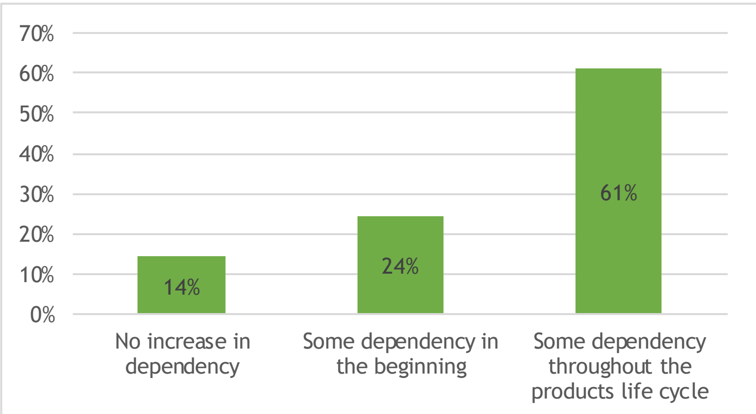


Figure 11: Dependency on Software Updates & Product Recalls

Table 3: Results of statistical hypothesis 2 (B) testing

Hypothesis	Chi-Square Test Result	Cramer's V Statistical Strength
Increase in dependency due to software updates and product recalls during Industry 4.0 depends on the existing governance structure.	$P(x^2 > 8.505) = 0.0142$ Accepted	0.294 Moderate level of association

When looking at the two dependencies in conjunction it is clear that the distributor will become dependent on the manufacturer to assist him or her in solving system design issues with hard to get information. Similar to that the distributor can become dependent on the manufacturer to resolve issues that can only be done by software updates from the manufacturers' team. Further, in in the first and second dependency 57% and 59% of the 98 distributors has working governance structures in place and indicated that they will become dependent on the manufacturer. This was not expected as the literature indicated that governance between distributors and manufacturers poses as a safeguard toward their relationship, assisting both of them by minimising transaction costs and increasing the flow of strategic information [13], [17].

The interviews revealed that for four distributors correlation exists between the System Design dependency and governance structure for those with "working governance in place". One distributor indicated that there is a "non-working governance in place" and that he will have some dependency on the manufacturer in the beginning.

Interesting to note is that all of those who indicated that they will be dependent or have some dependency in the beginning, mentioned the importance of training. Further, they will train themselves to become independent of the manufacturer. The other distributor also mentioned the importance of training but considers it the manufacturers' responsibility to makes sure the distributors are trained and knowledgeable regarding the new equipment.

A distributor with "no governance in place" and no dependency also indicated that their main focus areas is to train themselves upfront to become independent of the manufacturer. He did, however, indicate that they will be dependent on the manufacturer to take the lead in this new era and to show them what and when to focus on specific areas that they must train themselves in with regard to Industry 4.0, which is still not a very clearly layout subject.

It became clear that the distributors will invest in training to keep their dependency low towards the manufacturer where possible. This was seen in an additional question asked to the distributors in the questionnaire *"How likely will you become dependent on the OEM due to lack of IoT skills and tools in your company?"* and the higher independency rate is shown in Figure 12.

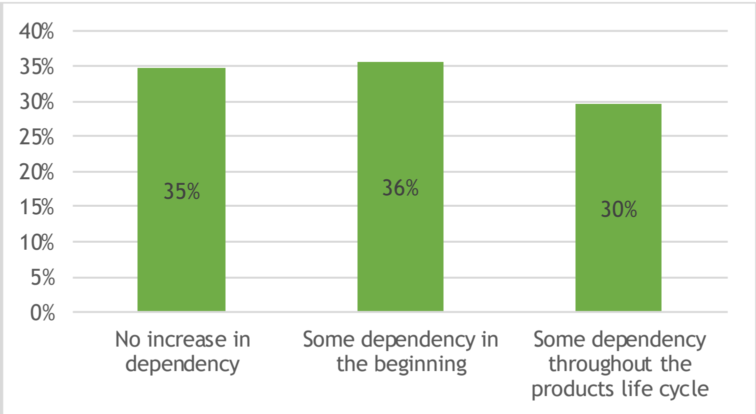


Figure 12: Dependency due to lack of IoT skills and tools

When analysing the second dependency based on Software Updates and Product Recalls no correlation could be found towards current governance structure, even though the quantitative study found that there was a correlation. Four of five distributors said that there will be no increase in dependency regardless of their governance structure in place. What was observed during the interviews was that these four distributors said that they have current working systems in place that they can rely on to sort out issues like this. One distributor said that they will not become dependent, as they have current systems in place where the manufacturer is replacing these parts or performing software updates before the customer or distributor even notice any problems with the equipment.

6. CONCLUSION AND RECOMMENDATION

This study tested the theory towards the impact governance structure has against the identified financial risks for distributors during Industry 4.0. The research question: "What role does existing governance structure play between manufacturers and distributors in lowering the distributor's risks during Industry 4.0?" was proposed and answered with the following two hypotheses being confirmed:

$H1_1$: The likelihood of distributors investing towards Industry 4.0 is dependent on the current governance structure in place.

$H2_1$: Increase in transaction costs during Industry 4.0 is dependent on the current governance structure in place.

From this study it was found that governance structure will have a direct impact on the financial risks distributors experience during Industry 4.0. However, it cannot be concluded that it will mainly have a positive impact, as those distributors with working governance structures in place indicated that they will become more dependent and hence have higher transaction costs.

6.1 Correlation between Willingness to Invest and Current Governance Structure

The found qualitative data between willingness towards investing or not investing were expected as the Chi-Square test has shown that there are a strong level of association between governance in place and willingness to make some investment towards Industry 4.0. This was expected as the literature indicated that governance between distributors and manufacturers poses as a safeguard toward their relationship [13], [17]. What was contradicted with this study was PwC [1] findings towards the main challenges during Industry 4.0 namely "Unclear economic benefits and excessive investments" as 83% of 76 distributors are willing to invest regardless whether they have a business model in place or not.

6.2 Correlation between Increased Transaction Cost and Current Governance Structure

This study indicated that there is a correlation between increased transaction costs and current governance structure. What was seen with both of the analysed dependencies was that the distributors that had a working governance structure in place will be much more dependent on the manufacturer than the other. This was not expected, as it is different to what was found in literature. According to the literature, governance between distributors and manufacturers poses as a safeguard towards their relationship, assisting both of them against opportunistic behaviour from each other, minimising transaction costs and increasing the flow of strategic information [13], [17].

6.3 Recommendations

Future research is necessary in order to determine:

- Why distributors are willing to make these upfront investments even though the risks are high for them?
- Identifying what are the distributors willing to invest towards, because those distributors interviewed had indicated the importance in getting themselves trained for the new industry.
- Why do distributors with a working governance structure between them and the distributor feel that they will be more dependent on the manufacturer due to system designs, software updates or product recalls?
- What role does training from the manufacturer and local facilities play in evening out the dependency distributors will have during Industry 4.0?

6.4 Shortcomings of the Study

With this study many additional factors came to light that has the potential to impact these identified risks for distributors during Industry 4.0. It is important to investigate all of these identified factors and their impact, as only governance structure in isolation will not give the manufacturer the full picture of how the distributor consider these identified issues and also the remedies to lower the risks towards them. Some of these identified factors to influence the results are shown in the results section 5.1 – "Topology of Distributors Participated" as each of the differences in the topology may lead to a custom-made governance structure between the manufacturer and the distributor. Other potential factors that could have influenced the distributor with answering the questionnaire, include:

- There were five industries that participated and each of them had their own set of local governing standards.
- When looking at IoT products there is a need to distinguish between off the shelf equipment or custom designed IoT engineering solutions or systems.

REFERENCES

[1] *PwC aktiengesellschaft wirtschaftsprüfungsgesellschaft*, "Industry 4.0 - Opportunities and challenges of the industrial internet" *PwC*, Germany, 2014. [Online]. Available: http:// https://www.pwc.nl/en/assets/documents/pwc-industrie-4-0.pdf [Accessed: 21-Sep-2018]

[2] R. M. Dijkman, B. Sprenkels, T. Peeters, and A. Janssen, "Business models for the Internet of Things," *Int. J. Inf. Manage.*, vol. 35, no. 6, pp. 672–678, 2015.

[3] B. D. Weinberg, G. R. Milne, Y. G. Andonova, and F. M. Hajjat, "Internet of Things: Convenience vs. privacy and secrecy," *Bus. Horiz.*, vol. 58, no. 6, pp. 615–624, 2015.

[4] H.-C. Pfohl, B. Yahsi, and T. Kuznaz, "The impact of Industry 4.0 on the Supply Chain," *Proc. Hambg. Int. Conf. Logist. (HICL)-20*, no. August, pp. 32–58, 2015.

[5] A. C. Pereira and F. Romero, "A review of the meanings and the implications of the Industry 4.0 concept," *Procedia Manuf.*, vol. 13, pp. 1206–1214, 2017.

[6] K.-M. Björk and C. Carlsson, "The Outcome of Flexible Lead Times on the Distributors," in *International Journal of Integrated Supply Management*, 2006, vol. 3, no. 1, pp. 86–102.

[7] L. Aarikka-Stenroos and B. Sandberg, "From new-product development to commercialization through networks," *J. Bus. Res.*, vol. 65, no. 2, pp. 198–206, 2012.

[8] S. Mudambi and R. Aggarwal, "Industrial distributors can they survive in the new economy?," *Industrial Marketing Management*, vol. 32, no. 4. pp. 317–325, 2003.

[9] R. Olsson, L. E. Gadde, and K. Hulthén, "The changing role of middlemen - strategic responses to distribution dynamics," *Ind. Mark. Manag.*, vol. 42, no. 7, pp. 1131–1140, 2013.

[10] M. Zineldin and P. Jonsson, "An examination of the main factors affecting trust/commitment in supplier-dealer relationships: an empirical study of the Swedish wood industry," *TQM Mag.*, vol. 12, no. 4, pp. 245–266, 2000.

[11] J. C. Lin and C. Chen, "Determinants of manufacturers' selection of distributors," *Supply Chain Manag. An Int. J.*, vol. 13, no. 5, pp. 356–365, 2008.

[12] P. M. Panayides and Y. H. Venus Lun, "The impact of trust on innovativeness and supply chain performance," *Int. J. Prod. Econ.*, vol. 122, no. 1, pp. 35–46, 2009.

[13] R. Vázquez-Casielles, V. Iglesias, and C. Varela-Neira, "Collaborative manufacturer-distributor relationships: the role of governance, information sharing and creativity," *J. Bus. Ind. Mark.*, vol. 28, no. 8, pp. 620–637, 2013.

[14] J. Hohenthal, J. Johanson, and M. Johanson, "Network knowledge and business-relationship value in the foreign market," *Int. Bus. Rev.*, vol. 23, no. 1, pp. 4–19, 2014.

[15] G. Moore, *Crossing the Chasm*, 1st ed. HarperCollins, 1991.

[16] V. Chiesa and F. Frattini, "Commercializing technological innovation: Learning from failures in high-tech markets," *J. Prod. Innov. Manag.*, vol. 28, no. 4, pp. 437–454, 2011.

[17] A. Rindfleisch *et al.*, "Transaction costs, opportunism, and governance: Contextual considerations and future research opportunities," *Mark. Lett.*, vol. 21, no. 3, pp. 211–222, Sep. 2010.

[18] I. Lee and K. Lee, "The Internet of Things (IoT): Applications, investments, and challenges for enterprises," *Bus. Horiz.*, vol. 58, no. 4, pp. 431–440, 2015.

[19] L. Wang, P. Iskanius, and P. Kess, "Improving the performance of manufacturer-distributor partnerships through knowledge transfer - A knowledge-based perspective," in *IEEM 2007: 2007 IEEE International Conference on Industrial Engineering and Engineering Management*, 2007, pp. 1634–1638.

[20] R. H. Weber, "Internet of things – Need for a new legal environment?," *Comput. Law Secur. Rev.*, vol. 25, no. 6, pp. 522–527, 2009.

[21] World Economic Forum and Accenture, "Industrial Internet of Things: Unleashing the Potential of Connected Products and Services," Geneva, Switzerland, 2015. [Online]. Available: http://www3.weforum.org/docs/WEFUSA_IndustrialInternet_Report2015.pdf. [Accessed: 21-Sep-2018]

[22] Y. Grégoire and R. J. Fisher, "Customer betrayal and retaliation: When your best customers become your worst enemies," *J. Acad. Mark. Sci.*, vol. 36, no. 2, pp. 247–261, 2008.

[23] Investopedia, "Samsung Galaxy 7 and the 5 Most Costly Product Recalls (SSNLF)," 2016. [Online]. Available: http://www.investopedia.com/slide-show/product-recalls/?article=1. [Accessed: 10-Sep-2017].

[24] D. R. Hittleman and A. J. Simon, *Interpreting Educational Research: An Introduction for Consumers of Research*, 2nd ed. Michigan: University of Michigan, 1997.

[25] L. J. Philip, "Combining Quantitative and Qualitative Approaches to Social Research in Human Geography—An Impossible Mixture?," *Environ. Plan. A* , vol. 30, no. 2, pp. 261–276, 1998.

[26] M. L. Mchugh, "Lessons in biostatistics The Chi-square test of independence," *Biochem. Medica*, vol. 23, no. 2, pp. 143–149, 2013.

[27] D. F. Polit and B. P. Hungler, *Nursing Research: Principles and Methods*, 3rd ed. Philladelphia: Lippincott, 1999.

INDUSTRY 4.0 IMPLICATIONS IN THE AUTOMOTIVE AFTERSALES SECTOR IN SOUTH AFRICA: A LITERATURE REVIEW

E. Davies[1*] & R. Coetzee[2]

[12]Department of Industrial Engineering
North West University, South Africa
[1]davies@audi.co.za, [2]Rojanette.Coetzee@nwu.ac.za

ABSTRACT

Industry 4.0 is currently a hot topic that has gained swift momentum across various industries and disciplines in recent years. It is therefore not surprising that several transformation frameworks have been developed for specific industries, to assist organisations to prepare and align themselves adequately for the change management required to usher them into the next era of Industrial Revolution. This paper provides a literature review of Industry 4.0 in general, where after the implications thereof in the automotive industry, and more specifically the automotive aftersales sector, are discussed. After considering existing Industry 4.0 transformation frameworks, the author proposes a transformation framework for the South African automotive aftersales sector.

*Corresponding author

1. INTRODUCTION

The fourth industrial revolution, also referred to as Industry 4.0, is a subject that has been gaining increasing focus in recent years. See Figure 1 for a Google search trends analysis for the terms "Industrie 4.0", "Industry 4.0" and "Fourth industrial revolution", which indicates a rapid increase in interest since 2013.

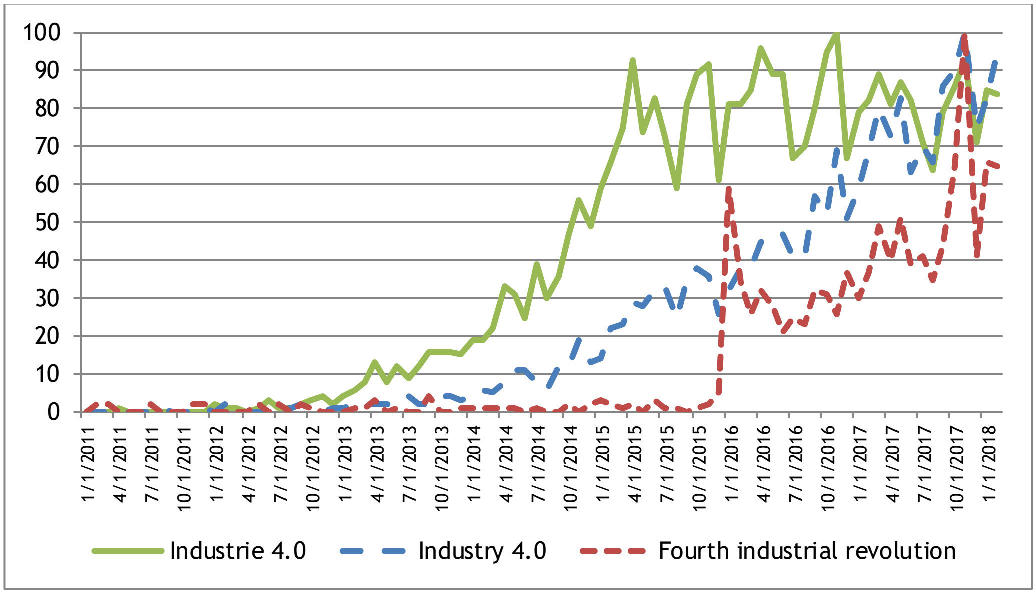


Figure 1: Google Trends progression for the terms "Industrie 4.0", "Industry 4.0" and "Fourth industrial revolution" for the period January 2011 to February 2018; a score of 100 relates to the peak in popularity for a specific term.

The preceding industrial revolutions marked periods in history where a significant technological advancement drastically altered the way in which we manufacture product, and ultimately the way we as humans do business. The first industrial revolution was ushered in with the advent of mechanisation mainly through steam power, which spanned the period from the mid-18th century to mid-19th century [1]. The revolution changed the economy from being agricultural and handcraft based, to an economy based on manufacturing and industry [1]. The second industrial revolution is attributed to mass production, assembly lines and electricity. This revolution spanned from about the mid-19th century to early 20th century. The third industrial revolution commenced as a result of automation and the development of computers, which started mid-20th century. The fourth industrial revolution, which is the focus of this paper, is currently gaining momentum across most industries; the implication of which will be discussed in more detail.

There is a general cautiousness and uncertainty in established organisations regarding Industry 4.0, which hinders investment and ultimately the transitioning into an Industry 4.0 organisation. This uncertainty stems from the high complexity of the topic, as well as a lack of guidance via roadmaps, frameworks or models [2][3]. This paper aims to shed some light on the complexities surrounding Industry 4.0, as well as provide a theoretical Industry 4.0 transformation framework for the South African automotive aftersales sector.

As we are currently in the transitionary period between the third and fourth industrial revolution, this paper will provide an overview of Industry 4.0 developments in context of the automotive sector, and also specifically examine the factors influencing the South African automotive aftersales sector. From the literature review, an Industry 4.0 transformation framework is formulated and proposed for the automotive aftersales sector in South Africa.

1. OVERVIEW OF INDUSTRY 4.0

As indicated in Figure 1, the term "Industrie 4.0" was the first to gain popularity. This makes sense as the term was first used in the Hannover Fair in 2011, whereafter the concept was finalised in 2013 by a German task team led by Siegfried Dais, from Robert Bosch GmbH, and Henning Kagerman, from the German Academy of Science and Technology [4]. The task team defined Industrie 4.0 as "a collective term for technologies and concepts of value chain organisation which draws together Cyber-Physical Systems, the Internet of Things, and the Internet of Services [4][5]". McKinsey's formulated a definition specifically related to manufacturing as "digitization of the manufacturing sector, with embedded sensors in virtually all product components and manufacturing equipment, ubiquitous Cyber-Physical systems, and analysis of all relevant data [6]".

McKinsey & Company have asserted that the imminent era of digital transformation will yield an assortment of new technologies that would change the way things are made. They believe that to gain most of Industry 4.0, organisations will have to invest in the following four dimensions; (1) Data, computational power, and connectivity; (2) Analytics and intelligence; (3) Human-machine interaction; and (4) Conversion of the digital to physical world. Even though it is becoming increasingly difficult to distinguish between these four dimensions due to interdisciplinary development, we will briefly look at these dimensions individually [7].

1.1 Data, Computational Power and Connectivity

Under the domain of *data, computational power, and connectivity,* technological developments such as big data, Internet of Things (IoT), open data and, cloud technology are featured. In simple terms *big data* refers to datasets that are too large for traditional data-processing, analytics programs and systems [8]. A new branch of science referred to as *data science* (including machine learning and data mining) was born as a result thereof. *Open data* on the other hand refers to data(sets) that is freely available online and which can be used for a host of applications, such as academic research, private sector use, etc., but ultimately for interoperability. According to the Open Data Handbook "Interoperability denotes the ability of diverse systems and organizations to work together (inter-operate). In this case, it is the ability to interoperate - or intermix - different datasets [9]". The Institute of Electrical and Electronics Engineers (IEEE) define the *IoT* on their website as:

"a self-configuring and adaptive system consisting of networks of sensors and smart objects whose purpose is to interconnect "all" things, including every day and industrial objects, in such a way as to make them intelligent, programmable and more capable of interacting with humans.[10][11]" *Cloud technology* is the delivery of Information Technology (IT) services over the internet rather than a direct connection to a server [12]. The internet in this sense is referred to as "the cloud". These services include storage of data, servers, networking, software, and analytics, to name but a few [12].

1.2 Analytics and intelligence

With the significant advances in artificial intelligence and robotics in recent years, automation of increasingly complex tasks has become possible. With the latest computers able to process larger datasets at a quicker rate than ever before, machine learning algorithms have been implemented to perform a myriad of everyday tasks [8]. Examples of applications where such algorithms are being used are; online personal retail recommendations (such as Amazon.com and Takealot.com), Google Maps route calculator, Uber, Spotify, Siri and Cortana.

1.3 Human-machine interaction

Human – machine interaction (HMI) refers to the interaction between the cyber and physical worlds. HMI have evolved from a point where it was initially primarily used in the industrial sector, to control machinery, robotics, etc., to a point where it is not uncommon to be used in our homes. Examples of HMI include; virtual reality, augmented reality and touch screen devices. An example of the advancement of HMI is the Festo Exohand [13], which is a pneumatic exoskeleton that can be worn similarly to a glove. The exoskeleton enhances the strength and movement of the human hand from the outside, and the possibilities of applications are vast; ranging from medical to industrial applications. As with Festo's Exohand, there exist a myriad of possible HMI applications which can enhance and simplify strenuous and hazardous tasks for humans.

1.4 Conversion of digital to physical world

In today's world there exist many interfaces between the physical world and digital domain. Examples of such interfaces are provided in Section 2.3. It addition to the increased interface to the digital world, it is now becoming progressively simpler to bring about changes to the physical world through digital means. Examples include Additive Manufacturing (AM), advanced robotics and, energy storage and harvesting [7]. In their research, Dimitrov and de Beer [14] found that 3-D printing is the most extensively used AM method in the professional world.

Another example of digital to physical conversion is advanced robotics whereby there is collaboration between robots and humans [7]. Examples include robotic surgery [15] and collaborative robots [16]. Collaborative robots can assist humans to perform strenuous or even hazardous tasks, which can reduce risk associated with health and safety, as well as increase the wellbeing of the human operator.

Energy harvesting techniques and the storage thereof is of particular importance to the IoT realm, as this provides a way of sustainably powering low-power IoT devices. Energy harvesting can be accomplished through the use of piezoelectric devices (where an applied stress will yield an electrical potential), thermoelectric devices (where a temperature difference will yield a related electrical potential), triboelectric devices (where an electrical charge is produced as a result of friction applied to the material), pyroelectric devices (where an electrical potential is generated by heating or cooling the device) and photovoltaic devices (where solar radiation is converted into direct current through the photovoltaic effect) [17]. Together with the development of light-weight efficient batteries, energy harvesting and storage systems, development and production of self-powered IoT devices are increasing.

An overview of these four dimensions relating to digital transformation provided a brief introduction to Indsutry 4.0 developments and the most prominent associated technologies. The next section will delve into Industry 4.0 developments in the Automotive industry, aptly termed Automotive 4.0.

2. INDUSTRY 4.0 IN AN AUTOMOTIVE CONTEXT

Section 2 provided an overview of Industry 4.0 in general. This section will discuss some of the implications of Industry 4.0 specific to the automotive sector. The automotive industry is one of the industries that have made significant strides in terms of implementing and developing Industry 4.0 technologies within manufacturing as well as in its products.

The World Economic Forum, in collaboration with Accenture [18], has identified three digital themes that will have an increasing influence and effect on the automotive industry. These three themes are the *connected traveller*, *autonomous driving* and *digital enterprise*. These themes, as they paint a holistic picture of Automotive 4.0 developments, are discussed in more detail below.

2.1 Connected traveller

The automobile has evolved from what was mostly a mechanical and analogue transportation machine into a complex computer operated vehicle which is becoming a mobile digital communications hub [18]. *In-vehicle Infotainment (IVI)* systems convey entertainment and information content to drivers and passengers within the vehicle. These systems are moving towards open-source systems that are more focussed on connectedness with mobile devices. IVIs are evolving into condition – and location based service, where customers will have access to services and systems on a need to have basis, similar to mobile applications today. *Usage based insurance* also becomes a reality where the increase of connected vehicles and telematics will enable insurers to monitor individual driver behaviour and adapt policies accordingly.

In Accenture’s World Economic Forum white paper, multimodal integration is comprehensively defined: “Multimodal connected transportation links all forms of road, rail and ferry travel, walking, cycling, all manner of automobile driving, public transit and the seamless connectivity among the modes. It brings together Original Equipment Manufacturers (OEMs), automotive and non-automotive suppliers, and government planning, tax and regulatory entities. Full-scale multimodal integration would create significant social and environmental benefits [18].” Multimodal integration is something that will not materialise in the short term, but rather the medium to long term future. The benefits of integration will lead to lower transportation costs for everyone, and ultimately lower cost of goods due to a more efficient supply-chain network.

2.2 Autonomous driving

In the truest sense of the word, *autonomous driving* refers to a vehicle being piloted by Artificial Intelligence (AI) without the need for human intervention. The level of autonomy of a vehicle is rated on a scale from 0 to 5, where 0 means the vehicle has no assistance systems in the form of automation, and 5 refers to vehicles that can navigate through all types of terrain and conditions, without the need for steering wheels or pedals [19]. Amongst companies currently testing fully autonomous vehicles in society is the company that was founded as a result of Google’s self-driving car project, Waymo [20-21]. Tesla states on their website that all their vehicles currently produced at their manufacturing facility have the required built-in hardware to enable the vehicle for full autonomous driving, at a more advanced safety level than that of a human driver [22]. Audi unveiled their A8 model in 2017 that will have the necessary hardware and AI ability to allow autonomous driving in slow-moving traffic conditions up to 60km/h [19]. Although the before-mentioned, as well as other major vehicle manufacturers either have the ability, or are busy developing the ability for full autonomous driving, legislation still needs to be developed for local authorities to deal with technicalities surrounding fully autonomous vehicles operating on public roads. One of the greatest threats of fully automated vehicles that are connected to the internet is cybersecurity [18]. Although fully autonomous vehicles are not yet that common in society, vehicles have been manufactured with increasing levels of assisted driving systems, such as adaptive cruise control, adaptive light control, automatic braking, automatic parking, blind-spot assist, collision avoidance systems, driver drowsiness detection, GPS navigation, hill descent control, tyre pressure monitoring and, lane departure warning systems.

2.3 Digital enterprise

Digitisation has driven significant improvements throughout existing value chains through optimised processes, increased efficiencies, reduced costs, increased collaboration and innovation [18]. With increased innovation and new ways of doing business, organisations will shift their strategies from simply selling products to customers to customer-centric value offerings. With increased levels of available customer data and ever evolving data science methods, businesses will be able to present customer specific offerings based on their purchasing preferences, lifestyle habits, etc. For the automotive industry to remain relevant during and after the Industry 4.0 transformation, some key areas in the automotive sector will see significant changes in the short to medium term.

A connected supply chain will lead to increased transparency and ultimately increased quality of products being produced. Increased data capturing through IoT devices will enable enhanced quality control which will in turn reduce defects, shorten the component design time, as well increase the efficiency of the manufacturing and logistics processes. In the longer term, 3D printing of parts on demand could significantly change the supply chain landscape, as the lead time will be reduced and the quality of the product will increase.

The automotive industry is one of the most aggressive industries when it comes to digitalisation and automation of processes. A combination of AI, robotics and IoT in manufacturing plants have led to increased flexibility in processes, shortened cycle times and increased productivity by reducing defects. The concept of a 'smart-factory' is considered an important outcome of Industry 4.0. Deloitte [23][24] defines a smart factory as "a flexible system that can self-optimize performance across a broader network, self-adapt to and learn from new conditions in real or near-real time, and autonomously run entire production processes".

The biggest cause for disrupted retail is customers whom are increasingly expecting a seamless customer experience through digital and physical touchpoints. With more and more customers thoroughly researching the products online before going to a bricks and mortar dealership, dealers of the future will most likely have smaller retail showrooms where the retail experience will be strengthened by augmented reality, virtual reality, digital showrooms, etc.

The traditional method of servicing and maintenance of a vehicle will also change as more vehicles become connected to the cloud. With the development of sophisticated in-vehicle diagnostic systems, smart components and ubiquitous vehicle connectivity, the vehicle will be able to proactively detect whether a component will require maintenance or replacement. With this level of predictive maintenance, bigger failures which result as a consequence of smaller component failures will drastically reduce. OEMs will also be able to utilise the connectivity of vehicles to send software updates directly to a specific vehicle, which will allow the owner to either accept or reject the update, as is currently the case with mobile devices, PCs and even smart televisions [18].

Vehicle lifespans are increasing as a result of increased manufacturing quality. Customers are therefore on average keeping their vehicles longer than in the past. As the automotive aftersales sector primarily earned revenue by selling parts and labour, increase digitisation, electrification and improved quality of vehicles will naturally lead to a significant decrease in revenue for traditional aftersales businesses. It is therefore imperative that this sector transforms into a digital aftermarket, where OEMs and aftersales providers will be able to offer software and hardware upgrades for infotainment and navigation systems [18].

Organisations in the automotive sector have and are collected large amounts of data on existing and past customers. This data is becoming valuable in its own right, as it can be analysed by data science methods to predict customer purchase preferences, behaviour, etc. This data is therefore valuable for the creation of new business models in an automotive data marketplace [18].

Another significant outcome of Industry 4.0 would be a connected infrastructure. Vehicle to infrastructure (V2I) and vehicle to vehicle (V2V) communication are key enablers towards intelligent transport. Through communication between connected devices such as vehicles, traffic lights, parking lots, RFID readers on road, and even bridges, an integrated communication network will increasingly be established that will continuously utilise data to increase the flow of traffic and increase road safety [18].

The first three sections of this paper provided an overview of Industry 4.0 and Automotive 4.0 developments and opportunities. It is clear from the complexity and extent of Industry 4.0 disruptions that it can present a daunting challenge to organisations to figure out how to ready itself for this revolution. Prior research has been done to formulate transformation frameworks for organisations to successfully navigate into this revolution. This will be the topic of the next session.

3. EXISTING INDUSTRY 4.0 RELATED TRANSFORMATION FRAMEWORKS

This section will discuss three frameworks from literature, focussing on Industry 4.0 and digital transformation. This section will conclude by looking at pros and cons of the presented frameworks.

In a report published by Massachussets Institute of Tehnology (MIT) and Capgemini Consuluting [25], *Digital Transformation* was defined as "the use of technology to radically improve performance or reach of enterprises". In their report they endeavoured to find out how senior executives could succesfully bring about digital transformation within their organisations. Their conclusion based on a study of 50 large traditional companies (exceeding annual turnovers of $1 Billion on average), which included 157 executives, was a digital transformation framework consisting of 9 elements. The nine elements of this framework are contained within 3 fundamental pillars, namely *customer experience*, *operational processes* and *business models* [25] as illustrated in Figure 2.

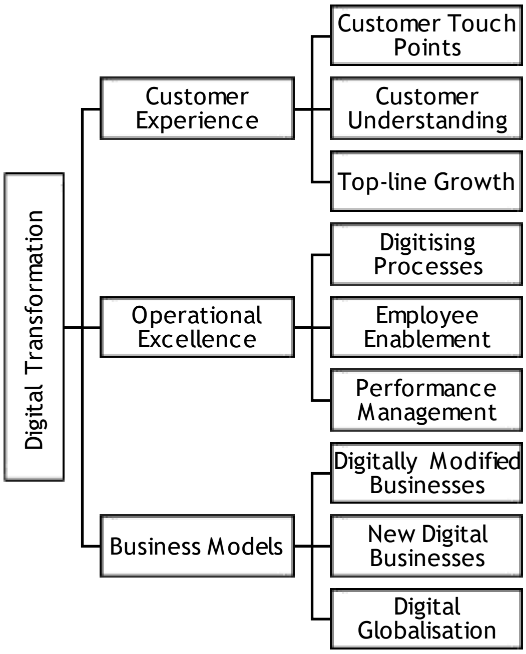


Figure 2: Graphical representation of MIT and Capgemini's digital transformation model

McKinsey and Company [6] also formulated a roadmap for the manufacturing industry to succesfully transition into the fourth industrial revolution. Their research comprised a survey of 300 participants from companies in Japan, USA and Germany. The surveys were then supplemented by industry interviews and further research with critically-acclaimed individuals in terms of the fourth industrial revolution. The resultant framework from their enquiry is shown in Figure 3 [6].

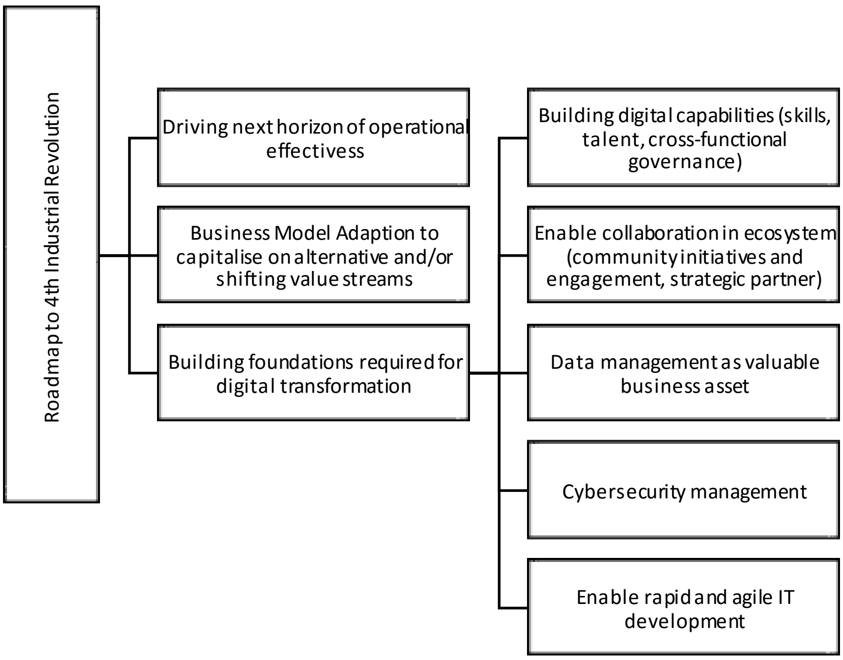


Figure 3: Graphical representation of McKinsey and Company's roadmap to the fourth industrial revolution

Erol, Schumacher and Sihn [26] proposes a three stage transformation model for a company to evolve into an Industry 4.0 ready one, as seen in Figure 4. The three stages are *Envision*, *Enable* and *Enact* [26]. Where the first two frameworks discussed in this section are focussed on organisations in general, and not a specific industry, this model rather focusses on the manufacturing indsutry. It is however considered in this paper due to the technical nature of the service dimension of the automotive aftersales industry. As the service department in the automotive aftersales business share many similiarties to a manufacturing environment, this model is deemed useful for the sake of completeness.

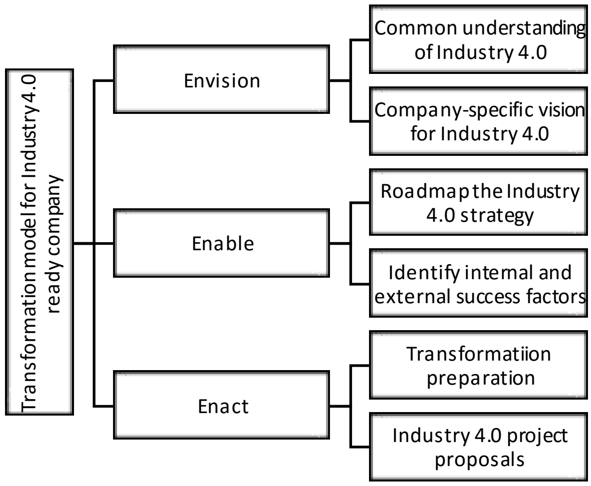


Figure 4: Graphical representation of Erol, Schumacher and Sihn's three stage Industry 4.0 transformation model

One of the biggest advantages of the discussed frameworks are the large aggregated amount of resourses in terms of research hours and budget to conduct research by means of surveys and, personal interviews with industry leaders and subject matter experts. It is also noteworthy that the individual resultant frameworks, although originating through different research design methodologies, share a large proportion of common elements and themes. In light of this study it however lacks considertation of the sector and country (or market) in question. Although the before mentioned frameworks are the results of extensive research conducted by experts in their respective fields, the multidisciplinary nature of the automotive aftersales sector demands a transformation framework that is distinctive and ordered to the sector's unique characteristics. It is therefore necessary to consider the South African landscape relative to the automotive aftersales sector to determine these distinctives, which is the focus of the next section.

4. SOUTH AFRICAN AUTOMOTIVE AFTERSALES SECTOR

Industry 4.0 brings a host of opportunities, albeit at times in the form of disruption, to almost every industry. It is important to consider the impact thereof on the automotive aftersales sector separately.

To visualise some of the future challenges and disruptions brought about as a result of Industry 4.0, in the context of a South African automotive dealership, an environmental scanning diagram is employed (see Figure 5). In the visual representation, the contextual environment includes the elements which are outside of the control or influence of the organisation, but which has an indirect impact on the organisation's goals and performance. The transactional environment on the other hand includes the elements which the organisation interacts with directly and which has a direct influence on the organisation's goals [27]. PESTLE analysis is the tool which was used to consider the influences in the contextual environment which include Political, Economic, Social, Technological, Legal and Environmental aspects [28]. The McKinsey 7Ss analysis [29] is used to consider the strategy, structure, systems, shared values, skills, styles and staff, which fall under the ambit of the transactional environment.

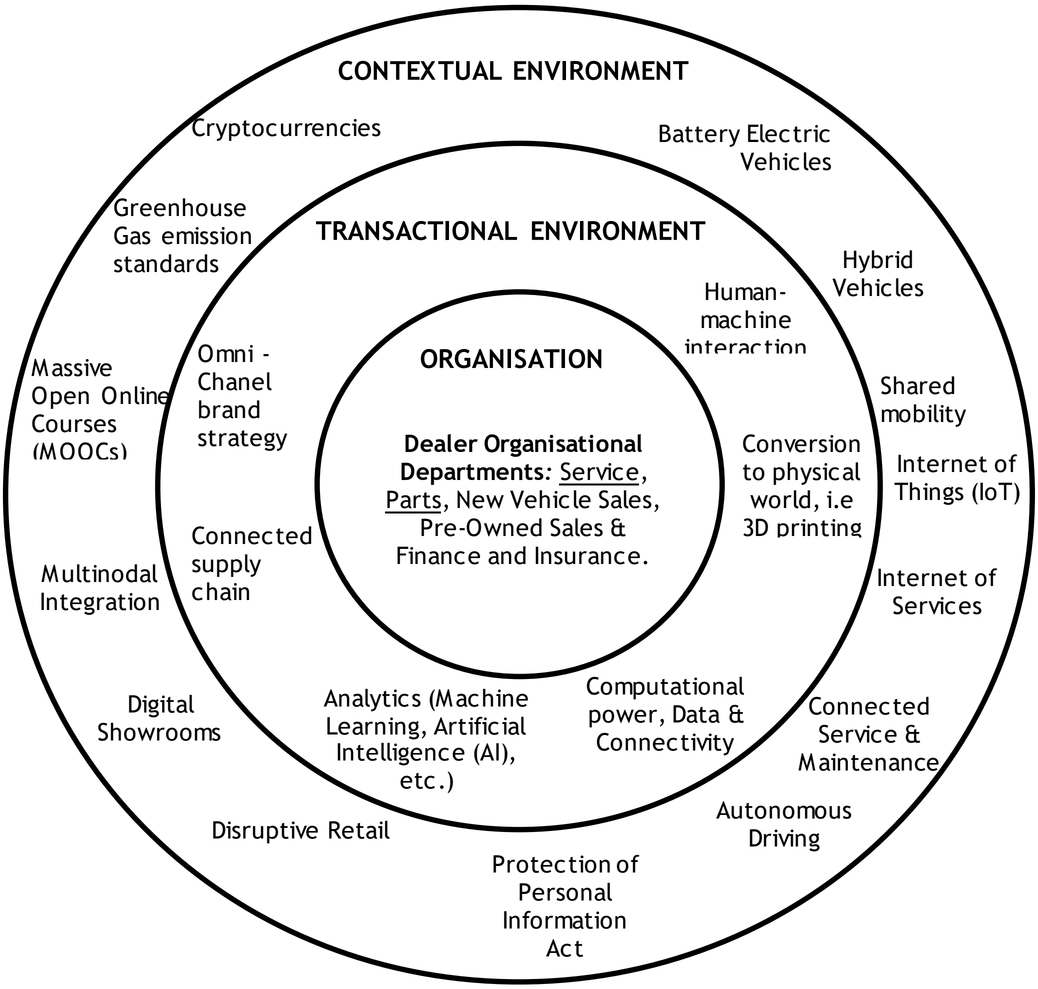


Figure 5: Environmental scanning of a generic South African automotive dealership in context of Industry 4.0

As seen in Figure 5, the traditional South African automotive dealership consists of various departments, of which the Parts and Service departments cumalitvely constitutes the Aftersales department. In this regard it is important to note that role and magnitude of the individual departments in the current model will differ drastically from the automotive dealership of the future, in terms of physical size, structure and profit contributions.

Not only does the autmotive aftersales sector have unique challenges to other industries, likewise South Africa is atypical to other countries in many other areas, which may present its own challenges for Industry 4.0 transformation. South Africa has an act protecting personal information of individuals called the Protection of Personal Information (POPI) act of 2013 [30]. This requires business to take special considertion when collecting, accessing and using personal information of individuals for the purposes of offering an omni-brand customer experience and customer specific offerings by using personal information in ML algorithms.

Some of the areas outside of the control of the automotive aftersales sector that are being adopted at a slower rate in South Africa than other markets include, electrical vehicles, hybrid vehicles, focus on reduction of GHG emissions, multinodal integration, digital showrooms and autonomous driving [31][32][33]. On the one hand this is not necessarily negative, as it provides the industry the opportunity to ensure it is ready for the Industry 4.0 and Automotive 4.0 expectations. To assist the automotive aftersales sector in assuring readiness, the next session will propose a transformational framework.

5. PROPOSED INDUSTRY 4.0 AUTOMOTIVE SERVICE SECTOR TRANSFORMATION FRAMEWORK

As the Google Scholar catchphrase asserts, “Stand on the shoulders of giants[13]”, similarly this paper draws on the expertise of thought leaders in their respective fields of research to address a certain gap in literature; both academic and industrial. Therefore in this section, the author considers the literature presented, as well as the

[3] https://scholar.google.co.za/

three existing Industry 4.0 transformation frameworks covered in this paper, to formulate a generic automotive aftersales sector transformation framework. The goal is for the proposed framework to be used as a generic theoretical framework for aftersales departments within the automotive sector, as well as a foundation for further research into Industry 4.0 transformation in this particular sector.

The three frameworks discussed in Section 4, namely (1) MIT and Capgemini's digital *transformation* model, (2) McKinsey and Company's roadmap to the fourth industrial revolution, and (3) Erol, Schumacher and Sihn's three stage Indsutry 4.0 transformation model, were used to formulate a base framework for the automotive service sector. The frameworks were broken down to its' individual elements and classified in tiers. The elements of each tier were individually considered in light of the literature review as well as the conditions of the environmental scanning diagram displayed in Figure 5. The elements which were not considered pertinent to the automotive aftersales sector were discarded, while the relevant elements were recorded. Similar elements in each tier were combined, whereafter the framework was formulated in a logical order of progression. The proposed framework, named the *Automotive Aftersales 4.0 Transformation Framework* is illustrated in Figure 6. The framework consists of three subdivisions, which are again broken down into individual framework steps. The three subdivisions are *customer centric experience*, *foundation for digital operations* and *digitally integrated business models*. These subdivisons are discussed in more detail below.

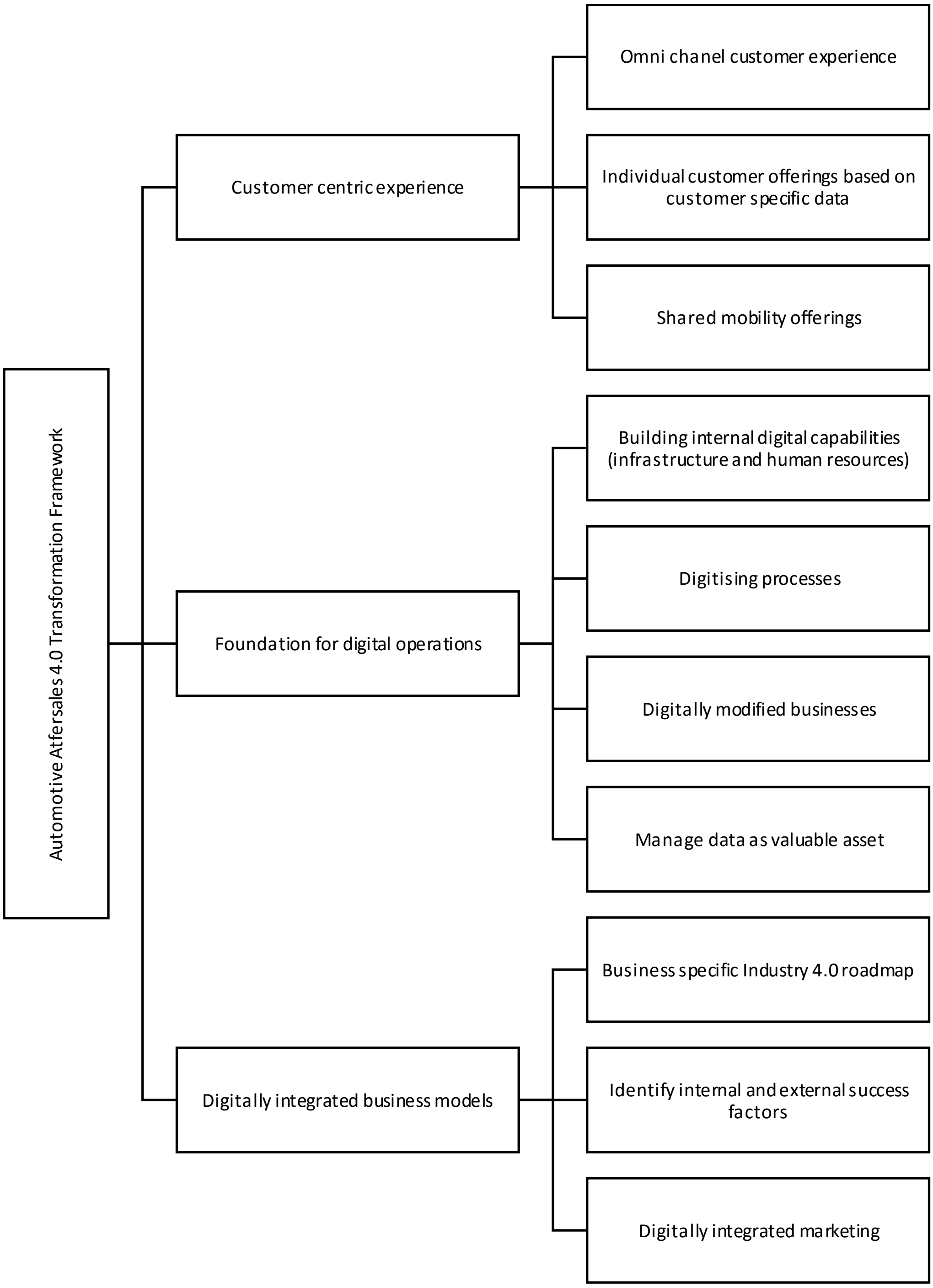


Figure 6: Proposed Industry 4.0 autmotive aftersales sector tranformation framework

5.1 Customer centric experience

The customer's service experience in the automotive aftersales department influences their loyalty to that department, dealership and brand, as well as their future purchasing decisions pertaining to the particular brand. Therefore, delivering customer centric experiences to each and every customer are of the utmost importance. With the wealth of available data at both OEM and dealer levels, it is possible to deliver customised service experiences to individual customers. As far as the digital customer is concerned, they want to experience a seamless, consistent and continuous process that transcends digital and physical platforms. The OEM and dealerships need to ensure they can provide this Omni-channel experience to customers. Another potential aftersales service that will enhance the customer centric experience is the offering of a shared mobility service to customers.

5.2 Foundation for digital operations

Before aftersales departments can digitise their business processes, it needs to ensure it has the requisite internal digital capabilities in terms of hardware and human resources. Once the aftersales department is digitally ready, the business can start integrating their internal and customer facing processes with digital platforms and devices. There exists great opportunity for elimination of wastes when migrating to digital solutions, which will lead to cost savings, as well as an improved and streamlined customer centric process. Once the digital foundation has been set in the aftersales department, areas such as augmented and virtual reality can be explored, which falls under the ambit of digitally modified businesses. Lastly, the fuel for the digital economy is data. It is therefore important for the business to be prepared to store large amounts of data, which are in formats which can easily be imported into data science tools, such as R, Python, SAS, SQL, Hadoop, etc.

5.3 Digitally integrated business models

As every organisation is unique, with different internal and external factors, each organisation will need to create its own Industry 4.0 specific roadmap. The organisation can consider their unique situations by utilising tools such as the environmental scanning, PESTLE analysis and McKinsey 7S analysis tools, as described in Section 5 in this paper. From these tools they will be able to determine their internal and external success factors. Once all the previously mentioned framework steps have been implemented, the business is digitally ready for effective digital marketing. Digital marketing can be pursued before the other framework steps are implemented, however it may not be as effective because the marketing strategy will be less driven by analytical decisions based on the available data, as it is by traditional marketing decisions.

6. CONCLUSION

This paper provided a literature overview pertaining to the foreseen implications of Industry 4.0 in general, as well as in the automotive industry in particular. The universal Industry 4.0 topics were discussed under four focus areas, namely *(1) data, computational power and connectivity, (2) analytics and intelligence, (3) HMI and (4) conversion of digital to physical world*. The three major themes discussed under Automotive 4.0 were the *(1) connected traveller, (2) autonomous driving and (3) digital enterprise*. Existing transformation frameworks relating to Industry 4.0 and the digital revolution were then presented. It was noted that these frameworks, however pertinent in a general sense, were not adequate to be used in the South African auomotive aftersales sector, which are affected by its own unique challenges. The author therefore proceeded to consider some of the topics in the South African environment that will influence the automotive aftersales sector and how it relates to Industry 4.0 and Automotive 4.0 developments.

The framework elements were then individually considered in light of the South African Automotive Aftersales sector and the beforementioned assessment thereof. This culminated into a theoretical Industry 4.0 transformation model, named the Automotive Aftersales 4.0 Transformation Framework.

REFERENCES

[1] **Groover, M.P.** 2010. *Fundamentals of Modern Manufacturing: Materials, Processes, and Systems. 4th Edition*. New Jersey: John Wiley & Sons, Inc.

[2] **Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. and Harnisch, M.** 2015. *Industry 4.0 - The Future of Productivity and Growth in Manufacturing Industries*. Boston Consulting Group.

[3] **Geissbauer, R., Schrauf, S., Koch, V. and Kuge, S.** 2014. *Industry 4.0 - Opportunities and Challenges of the Industrial Internet*. PWC – PricewaterhouseCoopers.

[4] **Kagermann, H., Wahlster, W. and Helbig, J.** 2013. *Recommendations for implementing the strategic initiative Industrie 4.0: Final report of the Industrie 4.0 Working Group*. Retrieved from http://www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Industrie_4.0/Final_report__Industrie_4.0_accessible.pdf. Accessed on 25 January 2018.

[5] **Hermann, M., Pentek, T. and Otto, B.** 2015. *Design principles for Industrie 4.0 scenarios — A literature review*. Technical University, Dortmund. Retrieved from http://www.snom.mb.tu dortmund.de/cms/de/forschung/Arbeitsberichte/Design-Principles-for-Industrie-4_0-Scenarios.pdf. Accessed on 24 January 2018.

[6] **McKinsey & Company.** 2015. *How to navigate digitization of the manufacturing sector*. Retrieved from http://worldmobilityleadershipforum.com/wp-content/uploads/2016/06/Industry-4.0-McKinsey-report.pdf. Accessed: 25 January 2018.

[7] **McKinsey & Company.** 2016. *Industry 4.0 at McKinsey's model factories*. Retrieved from https://capability-center.mckinsey.com/files/mccn/2017-03/digital_4.0_model_factories_brochure_2.pdf. Accessed: 27 January 2018.

[8] **Provost, F. and Fawcett, T.** 2013.Data science and its relationship to big data and data-driven decision making. Big Data,Vol.1, No. 1, pp. 51-59. Retrieved from https://www.liebertpub.com/doi/abs/10.1089/big.2013.1508. Accessed: 22 March 2018.

[9] **Open Knowledge International**. 2018. *Open Data Handbook*. Retrieved from http://opendatahandbook.org/guide/en/what-is-open-data/. Accessed: 22 March 2018.

[10] **Rutledge, R.L., Massey, A.K., Antón, I.E. and Swire, P.** 2014. *Defining the Internet of Devices: Privacy and Security Implications*. Georgia Institute of Technology Technical Report: GIT – GVU – 14 - 01. Retrieved from: https://smartech.gatech.edu/bitstream/handle/1853/52020/plsc2014-IoD.pdf?sequence=1&isAllowed=y. Accessed: 9 April 2018.

[11] **Institute of Electrical and Electronics Engineers (IEEE).** 2018. *About the IEEE Internet of Things (IoT) Initiative*. Retrieved from: https://iot.ieee.org/about.html. Accessed: 9 April 2018.

[12] **Mell, P. and Grance, T.** 2011. *The NIST Definition of Cloud Computing: Recommendations of the National Institute of Standards and Technology*. US Department of Commerce, National Institute of Standards and Technology. Retrieved from: https://www.researchgate.net/profile/Gillian_Oliver/publication/273692831_StoragS_is_a_Strategic_Issue_Digital_Preservation_in_the_Cloud/links/55de804c08ae7983879d1572/Storage-is-a-Strategic-Issue-Digital-Preservation-in-the-Cloud.pdf. Accessed 9 April 2018.

[13] **Festo.** 2018. *ExoHand*. Retrieved from: https://www.festo.com/group/en/cms/10233.htm. Accessed 11 June 2018.

[14] **Dimitrov, D.M. and de beer, N.** 2014. *Improvements in the capability profile of 3-D printing: An update.* The South African Journal of Industrial Engineering (25)2, pp. 1-12.

[15] **Davies, B.** 2015. *Robotic Surgery – A Personal View of the Past, Present and Future*. International Journal of Advanced Robotic Systems (12)5, pp. 1-6.

[16] **Maurtua, I., Ibarguren, A., Kidal, J., Susperregi, L. and Sierra, B.** 2017. *Human-robot collaboration in industrial applications; Safety, interaction and trust*. International Journal of Advanced Robotic Systems (14)4, pp. 1-10.

[17] **Lee, J., Kim, J., Hossain, M,S. and Kim, S.W.** 2016. *All-in-one energy harvesting and storage devices*. Journal of Materials Chemistry A, 4 (21), 7983-7999.

[18] **World Economic Forum (WEF)**. 2016. *World Economic Forum White Paper Digital Transformation of Industries: In collaboration with Accenture. Automotive Industry*. Retrieved from https://www.accenture.com/t20160505T044104__w__/us-en/_acnmedia/PDF-16/Accenture-wef-Dti-Automotive-2016.pdf. Accessed: 29 January 2018.

[19] **McGee, P. and Campbell, P.** 2017. *Financial Times: Audi launches most advanced self-driving car*. Retrieved from: https://www.ft.com/content/973f2f5c-6649-11e7-8526-7b38dcaef614. Accessed 10 April 2018.

[20] **Waymo.** 2018. *Waymo official website*. Retrieved from: https://waymo.com/. Accessed 10 April 2018.

[21] **Waters, R.** 2017. *Financial Times: Google owner tests first driverless car on city streets*. Retrieved from: https://www.ft.com/content/397134f8-c36c-11e7-a1d2-6786f39ef675. Accessed 10 April 2018.

[22] **Tesla.** 2018. *Tesla: Full Self-Driving Hardware on All Cars*. Retrieved from: https://www.tesla.com/autopilot?redirect=no. Accessed 10 April 2018.

[23] **Deloitte.** 2017. *The smart factory. Responsive, adaptive, connected manufacturing*. Deloiite University Press. Retrieved from: https://www2.deloitte.com/content/dam/insights/us/articles/4051_The-smart-factory/DUP_The-smart-factory.pdf. Accessed 7 June 2018.

[24] **Germany Trade and Invest.** 2017. *Smart factory*. Retrieved from: https://industrie4.0.gtai.de/INDUSTRIE40/Navigation/EN/Topics/Industrie-40/smart-factory.html. Accessed 18 August 2017.

[25] **Westerman, G., Calméjane, C., Bonnet, D., Ferraris, P. and McAfee, A.** 2011. *Digital Transformation: A Roadmap for Billion-Dollar Organizations*. MIT Center for Digital Business and Capgemini Consulting.

[26] **Erol, S., Schumacher, A. and Sihn, W.** 2016. *Strategic guidance toward Industry 4.0 – a three - stage process model*. International Conference on Competitive Manufacturing.

[27] **Nouwens, J.C.A.R.** 2004. *The outside in: questioning the use of electronic information services in organizations*. Alphen aan den Rijn: Kluwer.

[28] **Law, J.** 2009. *A Dictionary of Business and Management, 5th edition*. Oxford University Press. Retrieved from https://books.google.co.za/books?id=9gkkPwAACAAJ. Accessed: 8 June 2017.

[29] **Peters, T.J. and Waterman, R.H.** 2012. *In Search of Excellence: Lessons from America's Best-Run Companies*. HarperCollins Publishing: New York.

[30] **Government Gazette.** 2013. *Act No.4 of 2013: Protection of Personal Information Act, 2013*. Vol. 581 No. 37067. Cape Town, South Africa.

[31] **Price Waterhouse Coopers (PWC)**. 2016. *Industry 4.0: Building the digital enterprise*. Retrieved from https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf. Accessed: 28 January 2018.
[32] **Ernst & Young**. 2015. *Future of automotive retail: Shifting from transactional to customer-centric*. Retrieved from http://www.ey.com/Publication/vwLUAssets/EY-future-of-automotive-retail/$FILE/EY-future-of-automotive-retail.pdf. Accessed on: 4 February 2018.
[33] **Audi AG**. 2018. *Audi Media Info: Audi City*. Retreived from https://audi-illustrated.com/en/audi-city. Accessed on: 4 February 2018.

A DATA QUALITY EVALUATION FRAMEWORK FOR INDUSTRIAL ENERGY EFFICIENCY REPORTING

W. Hamer[1], W. Booysen[2] & E.H. Mathews[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
whamer@researchtoolbox.com

[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbooysen@researchtoolbox.com

[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
ehmathews@researchtoolbox.com

ABSTRACT

Data quality is fundamental in quantifying and reporting industrial-scale energy consumption. A narrow focus on isolated aspects of data quality can lead to inconsistent levels of rigour in dataset evaluation and selection. In this paper, a data quality framework is therefore developed to holistically evaluate if a dataset provides a fair representation of the underlying energy system.

The developed framework is based on three core aspects of data quality, namely accuracy, integrity and relevance. The framework emphasizes the use of traceability pathways to test data integrity and relevance. Quantitative and qualitative comparisons are proposed as practical options to test and evaluate multiple data sources. Based on the framework, a dataset can either be validated for reporting purposes or discarded based on the lack of data quality assurance.

The framework is applied to six isolated case studies. The results indicate that discrepancies relating to data integrity and relevance can significantly impact reporting functions. If left unchecked, these quantifiable discrepancies could result in data-based errors amounting to R1240 million (at R0.95/kWh) if viewed within the context of the Section 12L energy efficiency tax incentive. This highlights the role of holistic data quality evaluation to avoid propagation of erroneous data into reporting functions.

This work was sponsored by ETA Operations (Pty) Ltd

[1] Dr Waldt Hamer is a holds a PhD in Mechanical Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr Walter Booysen is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Prof Edward Mathews is a registered professional engineer and holds a PhD in mechanical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. BACKGROUND AND RELEVANCE

1.1 Data quality expectations for industrial energy efficiency reporting

When considering the rapid growth of information systems, data quality is a significant challenge in industrial scale operations [1], [2], [3]. Low quality data may lead to delays, wrongful decisions, inadequate prognoses, and troublesome handling and analysis of data [3]. In lieu, a high-quality dataset contains data that is "fit for use" in addition to being relevant, complete, error-free and representative [1]. Data quality is therefore a multidimensional challenge that is dependent on the specific use thereof [3].

Data quality is fundamental in quantifying and proving energy efficiency savings. Any lacking, irregular or questionable data will impact the assurance associated with subsequent calculations and observations made from a dataset. It is therefore critical to evaluate data quality as a first step of reporting processes. The importance of data quality becomes especially relevant when reporting functions have a monetary impact on performance contracting, employee incentive schemes, return on capital investments, and tax incentives.

In South Africa, the Section 12L tax incentive is applicable for measurable energy savings. Section 12L of the Income Tax Act (Act No. 58 of 1962) is an energy efficiency incentive taking form as a deduction of R0.95/kWh from taxable income per verified unit of savings achieved [4]. Compliance with the Section 12L regulations intends to ensure that reported energy efficiency savings are an accurate reflection of achieved savings [5]. Data quality is therefore a critical aspect of assurance to enable accurate reporting.

1.2 Review of present practice

Data quality is a fundamental requirement without which all subsequent findings are compromised. High quality data is important to provide an accurate representation of a system's operation [6]. Data quality for Measurement and Verification (M&V) of energy savings are defined by SANS 50010, the South African national standard which prescribes the minimum requirements of good M&V practice.

The standard necessitates supporting documents in the form of invoices and calibration records to prove data compliance. Compliance alone is, however, not sufficient to show the level of data quality. It is important to verify and validate collected data to determine a level of confidence at which the data can be used [7].

More specific guidance relating to M&V is provided by publications which are based on practical applications. This paper presents the results from a survey of specific M&V publications to evaluate how data quality is evaluated in practice. A wide range of M&V related publications (79 citations) were surveyed to determine existing M&V practices [8]. The following questions were therefore used for the survey:

- Which measurement option was specified?
- Has data compliance (official supporting references in terms of invoices or calibration records) been evaluated?
- Has data accuracy (measurement error or uncertainty) or integrity (traceability and relevance) been evaluated?

Seen in the results in Figure 1, it is evident that most of the case studies did not show an evaluation of data quality or disclose details thereof. It is therefore not possible to determine the level of assurance associated with the datasets used in most of the publications. Form the survey it was found that only 29% of the publications indicated an effort to evaluate data quality. Evaluation methods entailed the sporadic use of representative sampling methods, utility billing metering, calibration of measurement equipment, identification of interactive effects and the use precision and confidence limits.

The majority (80%) of the publications specified which measurement option was used. It is also noted that most of the measurement options entailed a key-parameter measurement option. This measurement option allows estimations when measured data is not available or inadequate. However, 72% of the publications linked to the key-parameter measurement option did not discuss the potential uncertainty caused by estimations.

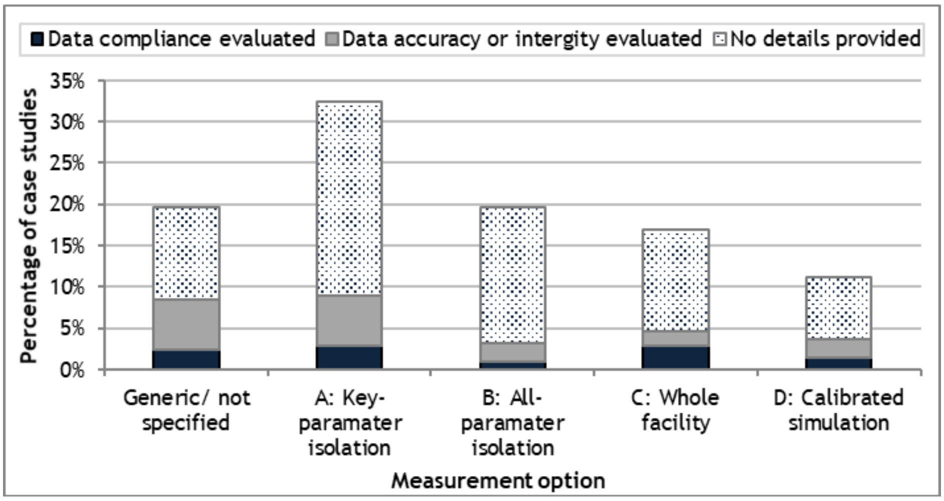


Figure 1: Critical analysis – Measurement option and data quality

The analysis showed that the use of the four standard measurement options, as specified by SANS 50010, is well-established in practice. However, data quality evaluation techniques are not commonly applied or disclosed. This highlights two major problems: unknown levels of data quality assurance and the inability to prove data compliance. Both aspects have a significant impact on the assurance provided by reporting functions.

A lack of consistent methodologies to evaluate datasets has also been noted in M&V literature [6], [9], [10], [11]. Investigation of the different sources of measurement error necessitates practical verification procedures to ensure consistent energy measurements [10]. Untraceable or unexplained assumptions cause up to 10% deviation between different datasets [11]. Additional verification methods are therefore required to ensure that the final dataset is "fit for use" [9]. This includes ensuring that the data is relevant, complete, error-free and representative [1].

1.3 Objective of the study

In this section the basic background and relevance to the topic of data quality in energy efficiency reporting was discussed. The critical analysis of available literature has indicated that several methods are used to ensure data quality has been identified in M&V case studies and literature. It is not common practice to indicate how overall data quality (accuracy, integrity and relevance) is evaluated. A narrow focus on an isolated aspect of data quality can therefore lead to inconsistent levels of rigour in dataset evaluation and selection. This highlights the need for consistent data evaluation methodologies specifically developed industrial energy efficiency reporting.

The remainder of this paper's objective is to review conventional data quality considerations, measurement and data handling practices (Section 2). A data quality evaluation framework is then developed to compile and evaluate multiple datasets that can be used in EE reporting (Section 3). The data quality evaluation framework presented in this paper forms part of a broader uncertainty management approach [12]. The framework was then applied to industrial case studies to test the approach. Isolated observations from six case studies are presented to highlight how data quality issues can be detected (Section 4). The potential impact of data quality mistakes is also quantified, before the final a conclusion is discussed (Section 5).

2. DATA MANAGEMENT METHODS

2.1 Data quality considerations

Data gathered during the M&V process is usually provided in the form of a spreadsheet [6], [9]. The data entries are, however, only a representation of measurements taken at an operational level. Data entries mostly result from time-series aggregated measurement readings, which are derived from instrument readings. Several problems can arise in this transfer from instrument to data source [6], [10]. It is therefore important to acknowledge several factors that can influence data quality.

In the field of M&V, data quality refers primarily to the management of uncertainty attributed to measurement, data sampling, estimation, modelling and interactive effects [13], [14], [15]. Uncertainty resulting from a measured dataset is mainly managed by investigating energy invoices, manufacturer's specified instrument accuracy and periodic calibration results [13]. These quality indicators are evaluated in different levels of rigour depending on the reporting criteria, allowed M&V costs and level of acceptable uncertainty [14]. The M&V process therefore always aims to balance uncertainty and cost [6], [14].

Data management in large industries tends to be prioritised to provide information for operational management and control [16]. This requires extensive time and resources to obtain acceptable data quality [3]. Certification in terms of quality (ISO 9001), energy (ISO 50001) and environmental (ISO 14001) management are also indications of investment in data quality [17], [18]. The focus for the M&V practitioner should therefore be to align data used for a specific M&V reporting function with the data management structure used within the existing operations.

Integration between M&V reporting requirements and existing industrial data management systems, can allow optimal usage of time and resources while ensuring alignment between different operational and reporting functions. This approach is therefore proposed in this study. The following sections discuss measurement and data handling practices before a new data quality evaluation framework is discussed in Section 3.

2.2 Measurement practice

2.2.1 Measurement boundary

Measurement boundary selection is a key deciding factor in establishing the scope of different reporting functions. Depending on the reporting function, boundaries can be selected to encapsulate a facility or isolated sections of a facility. A measurement boundary needs to have the following traits to support reporting functions:

- Encapsulation of the energy system targeted for reporting;
- Availability of energy consumption measurements;
- Availability of energy governing factor measurements;
- Allow statistically significant modelling of baseline and reporting period energy usage; and,
- Adherence with SANS 50010 measurement options and requirements.

The process of selecting an appropriate measurement boundary is well-established in practice and literature [6], [7], [9], [13], [14]. The convention is to select a single measurement boundary deemed appropriate for all subsequent data collection, calculation and reporting requirements. However, the proposed integration between existing industrial data management and the M&V process allows for multiple boundary options to be evaluated.

For example, a whole facility reporting boundary evaluation provides a holistic view of a facility's energy performance, while isolating sections thereof can provide more specific observations. In combination, it provides an opportunity to report on both top-down (integrated) and bottom-up (isolated) observations. Therefore, different viewpoints can be created by encapsulating the same energy systems by different measurement boundaries to improve understanding of the underlying energy system and its energy performance. In order to be able to identify and evaluate all the possible measurement boundaries, all the possible measurement points in an energy system need to be considered.

2.2.2 Measurement points

M&V includes the measurement of numerous variables which allow the quantification of energy performance. For each variable, a physical point of measurement with a specific measurement device or method is required. The types of variables can be classified into three basic groups:

- Energy carrier flow: Energy, mass or volumetric flow of energy carriers.
- Specific energy content: Calorific value or enthalpy of energy carriers.
- Energy governing factors: Service level indicators such as production volumes, product quality, operational set-points, temperature, occupancy, etc.

The listed variables are required for M&V of energy performance, but several are also measured for other operational purposes. If the M&V and operational measurement requirements do not align, additional measurement devices need to be installed and commissioned to enable data collection for reporting purposes. The following types of measurement devices and methods are typically used:

- Metering: Continuous or periodic measurement of energy and process flows that typically takes place in real-time.
- Laboratory analysis: Sampling of energy and process flows for compositional and quality analysis by means of laboratory procedures.
- Records: Logging of discrete data, including production loss incidents, occupancy, shipments, quantity surveys, etc.

The selection of measurement points will pinpoint the boundary options for energy reporting functions. It is therefore essential to identify all possible measurement points to assess the potential scope of reporting function. Figure 2 shows a simplified procedure to identify and classify measurement points. This can be used to simplify the boundary and dataset selection process.

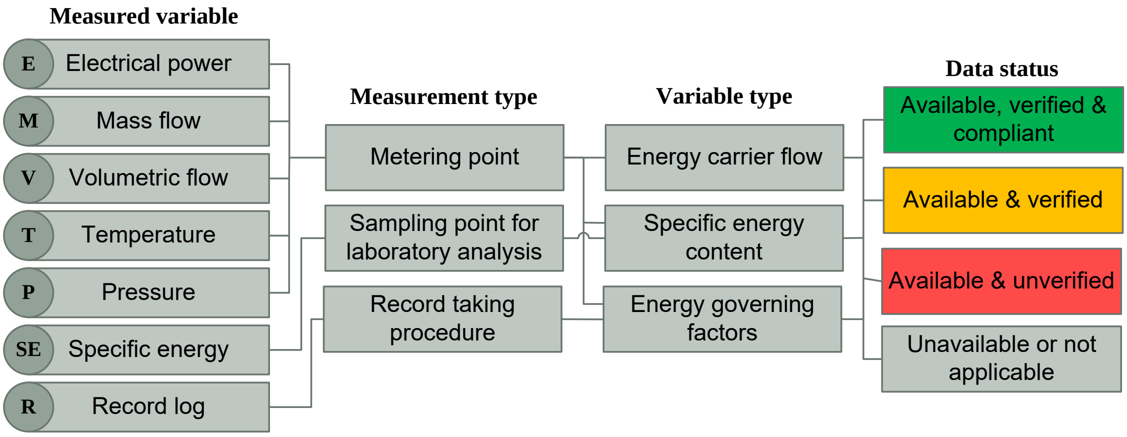
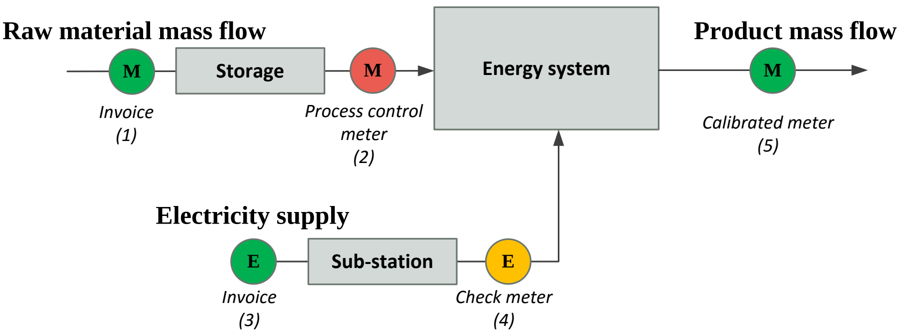


Figure 2: Management of measurement points for measurement boundary and dataset selection

Measurement points are classified according to the measured variable, measurement type, variable type and status of the data. Figure 3 illustrates an example in which the measurement point classification procedure is used to identify multiple dataset options for an energy system.



	Energy carrier	Energy governing factors
Dataset 1	Invoice (3)	Invoice (1) and Calibrated meter (5)
Dataset 2	Check meter (4)	Invoice (1) and Calibrated meter (5)
Dataset 3	Check meter (4)	Process control meter (2) and Calibrated meter (5)

Key:	Available, verified & compliant	Available & verified	Available & unverified

Figure 3: Example of multiple dataset options for measurement boundary and dataset selection

The example in Figure 3 indicates five different measurement points which can be used for M&V of the energy system. The convention would be to select only the compliant measurement points (1, 3 and 5) that allow a single dataset (Dataset 1). Albeit not necessarily verified or compliant, the use of the check meter (4) and process control meters (2) can provide additional datasets (Dataset 2 and 3, respectively). Additional datasets can provide additional insight into the system. For instance, an invoice is often limited to monthly billing period quantities, while meter data can be gathered in higher resolution. Three different datasets can therefore be compiled based on the measurement points in the example.

A good understanding of the possible measurement points is a key aspect. The use of existing measurements can expedite the M&V process. However, it is critical to evaluate if existing measurements and datasets are adequate. Data availability and accuracy can be limited due to lack of measurements. Estimations are allowed in these cases. However, the uncertainty caused by sub-standard measurements or estimated data need to be acknowledged in the development of reporting functions.

2.2.3 Measurement uncertainty

Data collected from measurements are never 100% error-free [14], [15]. Errors arise from accuracy of sensors, sampling rate, data integration, sensor drift, imprecise measurements and data capture errors [13], [19]. Each field of measurement and type of metrology has a set of standard requirements. This implies that management of measurement error is mostly based on the manufacturer's specifications, adherence to standard practice and management of periodic maintenance and calibration [13], [19].

Measurement uncertainty refers to the range within which the true measured value occurs. A typical uncertainty statement would entail the level of precision within a specified confidence interval [14], [15]. For example, the measured value can be expected to deviate by ±7.5% with 80% confidence [14]. SANS50010: 2018 requires the quantification of uncertainty, but good measurement practice is required to enable appropriate accuracy and repeatability of reported results [7].

The M&V standard allows two main sources of measured data, namely invoices of measured quantities and measurement equipment with appropriate calibration records [7]. The use of accredited calibration laboratories also provide assurance that certain minimum requirements are met [20]-[22]. Ideally, only measurement points that comply with the standard should be used. It is, however, not always practical since an increase in measurement accuracy will not necessarily justify the additional instrumentation and calibration costs.

A M&V plan is compiled to determine which data sources will be used for measurement of a specific energy system [23], [24]. This plan will indicate the extent to which measurements are required. Measurement accuracy is mostly emphasised in M&V literature, but data handling from measurement to point of use can also significantly affect data quality. The following section therefore reviews data handling practices.

2.3 Data handling

2.3.1 Data collection

Data collection for reporting functions is continuous for long periods. For the 12L tax incentive, datasets are required for a year of assessment as well as the preceding baseline year [5]. This necessitates data collection across a two-year period. Energy and energy governing factor data can originate from diverse sources based on identified measurement boundaries. Therefore, a high volume of data and supporting documents need to be gathered and stored. This is illustrated by a summary of possible data sources in Figure 4.

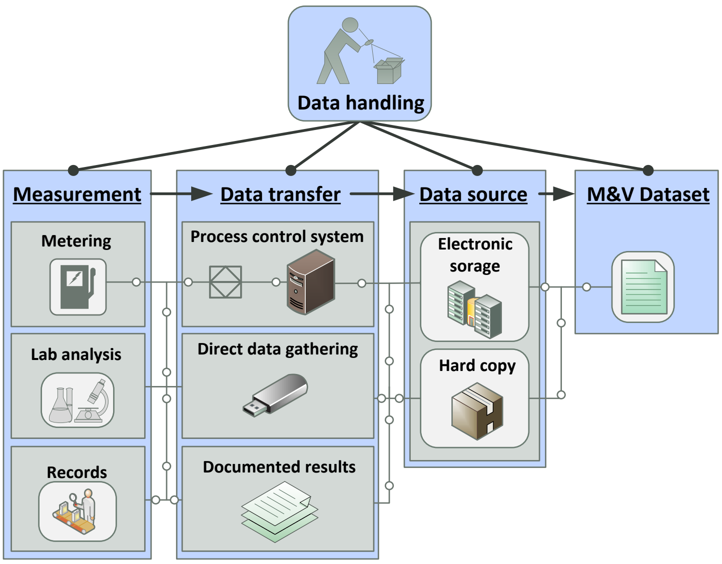


Figure 4: Overview of data handling across multiple data sources

Data can be gathered from process control historians, enterprise wide software and other electronic data management systems [25]. Documents such as invoices, calibration records and other physical records can be collected from the relevant departments. Data needs to be stored in an organised format with restricted access to restrict tampering [13]. In order to enable reproduction of the data it must be traceable from meter to measurement to data source. The relevant measuring instrument details (model, serial number, location, accuracy and calibration) must also be thoroughly documented.

In the conventional M&V processes, certain sources of error are not measurable and therefore remain unknown or unquantifiable [13]. These include poor meter selection, poor meter placement, inaccurate estimates and wrongful estimations of interactive effects [13]. This can lead to datasets that seem accurate, but inherently provide false observations. It is therefore important to test if a dataset is relevant to the intended use thereof.

2.3.2 Data inspection

Before the acquired data can be used it must be verified and validated to ensure a true representation of an energy system [6], [26]. A dataset with erroneous data will affect the reliability of calculations and observations. The data collection period must therefore include a full cycle of operation in order to accurately reflect the energy usage of operations [6], [13].

Reporting periods may vary depending on the required outcomes. SANS50010 specify that reporting periods should capture operational cycles and seasonal effects. However, abnormal operations across an extended period, such as strike periods or major overhauls, will limit the period of useable data. It is therefore necessary to evaluate data across the full period to identify abnormal occurrences and other possible discrepancies. It is also important to ensure that all energy streams are accounted for when considering a measurement boundary. Typical data inspection methods include the following:

- Long term trends and graphical representations [6], [27];
- Inspection of routine calibration records [14], [15], [27];
- Redundant measurements [6], [27]; and
- Error and precision of measurements [14], [15], [27].

Inspections are critical to determine if a dataset is fit for use and representative. Although inspections can be used to identify data discrepancies, dataset quality is largely fixed to the time of data acquisition and measurement [27]. This means data quality improvement in hindsight is limited. However, data filling or adjustment techniques can be used for periods in which data loss is noted [7], [19]. Such adjustments need to be documented in order to maintain repeatability of overall calculations. Untraceable alterations can affect the integrity of a dataset and influence observations made from the dataset.

Data integrity can be influenced intentionally or unintentionally by various data transfer activities. Intentional activities include data filling due to data loss, data corrections or formatting adjustments. Although these activities may not necessarily reduce data quality, they should be traceable to determine if they affect the reported results. Unintentional factors include mistakes in data transfer, aggregating and electronic storage software.

The critical analysis of M&V practice (section 1.2) indicated that data inspections or details thereof are normally not disclosed. Uncertainty then remains whether the used datasets are true representations of measurements and the actual energy system. Measurement accuracy and data compliance is usually emphasised in th e M&V process, but data integrity and its relevance to the actual energy system also needs to be evaluated. A holistic data quality evaluation is therefore required.

3. A DATA QUALITY EVALUATION FRAMEWORK

3.1 Preamble

The M&V process is reliant on data that allows an accurate analysis of an energy system. However, data quality can be a complex challenge in industrial energy systems. This is mainly due to the large number of measurement points and data sources that are available. It is therefore critical to evaluate the available dataset options.

The term data quality can be used to address various aspects of data use [1], [3], [14]. Data quality within the context of M&V and EE reporting is defined as: *The ability to prove that a dataset provides an accurate, traceable and relevant reflection of an energy system*. Four distinct aspects can be isolated from this definition, namely [12]:

- Accurate: Can measurement uncertainty be evaluated?
- Traceable: Can the dataset's integrity be tested from measurement point to data source?
- Relevant: Are all applicable energy and process indicators accounted for?
- Proof: Are supporting documents available?

The research methodology is based on the examination of these aspects of data quality and how it can be used to evaluate (section 3.2) and classify (section 3.3) datasets. The methods were compiled in a framework which was applied on a case study basis (section 4) to test the effect when used with real data observations from industry.

3.2 Dataset evaluation methodology

3.2.1 Data accuracy

The first aspect of data quality entails data accuracy which is primarily affected by measurement uncertainty. The common approach to address this uncertainty is to link the relevant measurement equipment to an invoice, calibration record or manufacturer's accuracy specification [7]. The methods used to evaluate measurement uncertainty are well-established in M&V guidelines and practice [13, 18]. This is, however, limited to a single aspect of data quality. The point of departure of this methodology is therefore to address the other decisive factors that can influence data quality, i.e. data integrity and relevance.

3.2.2 Data integrity

The terms data traceability and data integrity are often used interchangeably. A dataset is traceable if it can be tracked to a distinct data source, which in turn can be traced to one or more measurement points and instruments. By using this traceability pathway, it is possible to determine if data integrity has been compromised by any form of adjustment in the data transfer process. This is illustrated in Figure 5.

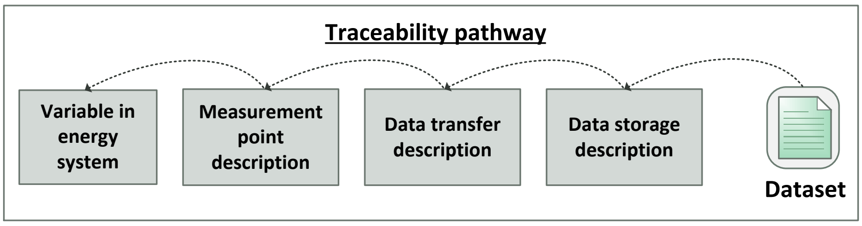


Figure 5: Traceability pathway concept to test data integrity

Testing data integrity can be a complex challenge due to the various activities across the data handling pathway. Therefore, a simplified data comparison method is proposed to identify data discrepancies. Data comparisons can be done by redundant metering or independent sources across the data traceability pathway. This is illustrated by an example in Figure 6 where three datasets of the same measured variable are compared.

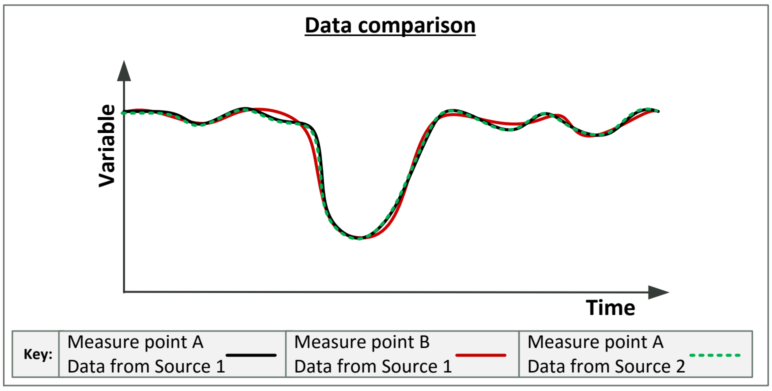


Figure 6: Example of visual data comparison

Figure 6 shows two types of data checks. Firstly, redundant measurements (Points A and B) are compared. Secondly, independent sources (Sources 1 and 2) of Measurement A are compared. Any deviations shown by these two comparisons will indicate a data discrepancy. These deviations need to be queried in order to determine if the dataset used fairly represents the measurements of the variable in the energy system.

Deviations can be identified by various methods. Analytical methods are commonly used to investigate the error between datasets. However, visual methods are useful to simplify identification of deviations [6]. Since equivalent datasets should match each other, it can be useful to plot a dataset comparison (i.e. y versus x). Deviations can then be identified both computationally and visually by establishing the relationship between datasets. This concept is illustrated in Figure 7 for both quantitative and qualitative comparisons.

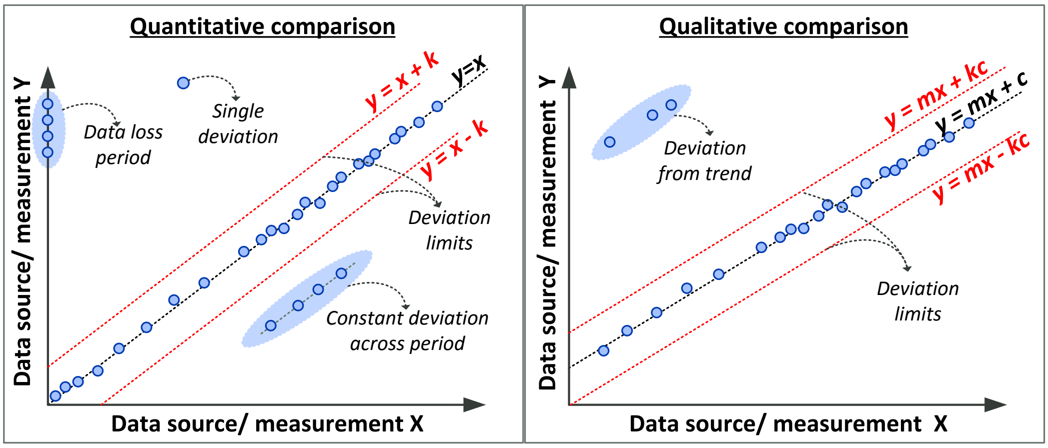


Figure 7: Example of quantitative and qualitative evaluations to test data integrity

Not all measurements or data sources provide a quantitative representation. For example, monitoring the status of a pump will only provide an indication of whether it is on or off. In practice, numerous such qualitative measurements can be collected and used to evaluate energy systems. Therefore, the data integrity test can be quantitative (y=x) or qualitative (y=mx+c) since the main concern is to identify deviations. As illustrated in

Figure 7, several deviations can be noted. These deviations are only considered as abnormal if they are beyond reasonable limits indicated by a constant (k).

Abnormalities can be linked to data loss, meter malfunctions, irregular operations or other explanations. If abnormalities cannot be linked to relevant explanations, they then create uncertainty on whether the dataset truly represents the energy system. Datasets will therefore either be validated or discarded based on the integrity test. The next evaluation tests the relevance of a dataset.

3.2.3 Dataset relevance

Data relevance is critical in the measurement boundary selection process. During this process the complex process flow of industrial systems is often reduced to enable a simplified overview [9]. This simplification allows the evaluation and selection of various measurement points. Certain measurements can also be excluded by a key-parameter isolation boundary (M&V options discussed in Section 1.2). However, the M&V exercise can be compromised if this simplification is flawed due to misinterpretations or lack of information. This can lead to measurement boundaries and/or datasets that do not fully represent the intended energy system.

The ideal solution would be to conduct a mass and energy balance across the energy system to allow full accountability of all energy streams [49]. This is, however, not practical due to the lack of measurements and the challenge in balancing dynamic state energy systems. Instead, this study proposes a simplified method to investigate energy usage trends. This is illustrated by a long-term energy intensity trend in Figure 8.

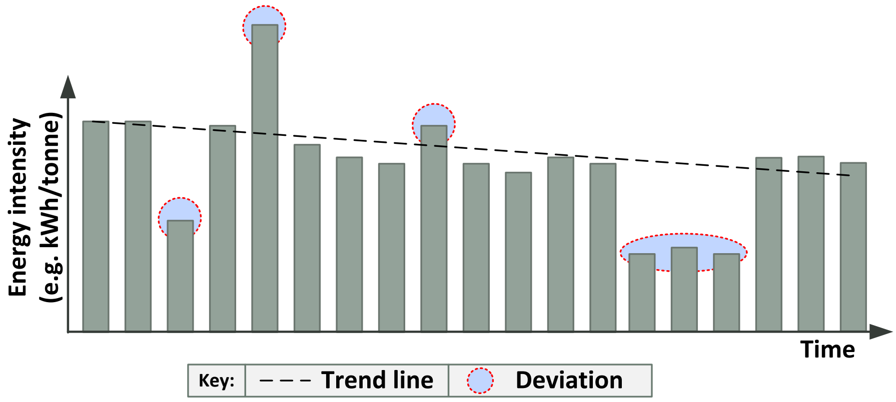


Figure 8: Example of long term intensity trend to evaluate data relevance

The example in Figure 8 shows a declining trend in energy intensity with certain deviations from this trend. These observations need to be linked to operational events in order to test the representativeness of the energy intensity data. For instance, the declining trend in energy intensity needs to be linked to a certain activity, such as a long-term energy saving initiative. Deviations from the trend also need to be linked to events, such as scheduled and unscheduled maintenance stops.

If the observations made from a dataset cannot be reconciled with actual operational events, it will lower confidence that the dataset is relevant. Conversely, correspondence between data-based observations and listed operational events will validate that the data is representative of actual operations. This will confirm the relevance of the dataset. Energy intensity trending is therefore a simple method to test the relevance of a dataset by identifying and explaining observations in the dataset.

3.3 Dataset classification

In the previous sections, dataset evaluation methods were discussed that assess data quality in terms of accuracy, integrity and relevance. By applying these methods, abnormalities can be identified and evaluated in order to determine if a dataset provides a true reflection of the actual energy system. Datasets will therefore be discarded if the data quality cannot be validated. Validated datasets will be classified as useable data.

Although validated datasets can be used for reporting purposes, only datasets with sufficient supporting documents (such as invoices and calibration records) can be proven to be official and compliant. Therefore, the last step in the data quality framework is to classify datasets as either, (1) official and validated, or (2) supporting and validated. Only a dataset with proof of compliance can be used for official purposes, such as quantifying 12L tax incentive reported EE savings. However, additional datasets can supplement this intent by avoiding an overreliance on a single dataset. This broadens the options and versatility of reporting functions.

The rationale for including multiple datasets (both official and supporting) is to allow multiple calculation and reporting options. Compliant data is often limited to monthly aggregated data from official documents, such as utility invoices and consignment certificates. This can restrict calculation and modelling options in terms of data resolution (e.g. daily or weekly resolutions). By expanding the M&V plan to include more datasets, more options are provided which can be used to evaluate specific data-based uncertainties.

A data quality evaluation framework is proposed based on the dataset classification concept. The framework is divided into two phases. The first phase indicates the conventional measurement boundary selection and data collection processes to determine which environment is being reported on. The processes required to determine available measurement boundaries and collect datasets are well-established. The focus of the framework is therefore in the second phase during which the collected datasets are evaluated. This is illustrated in Figure 9.

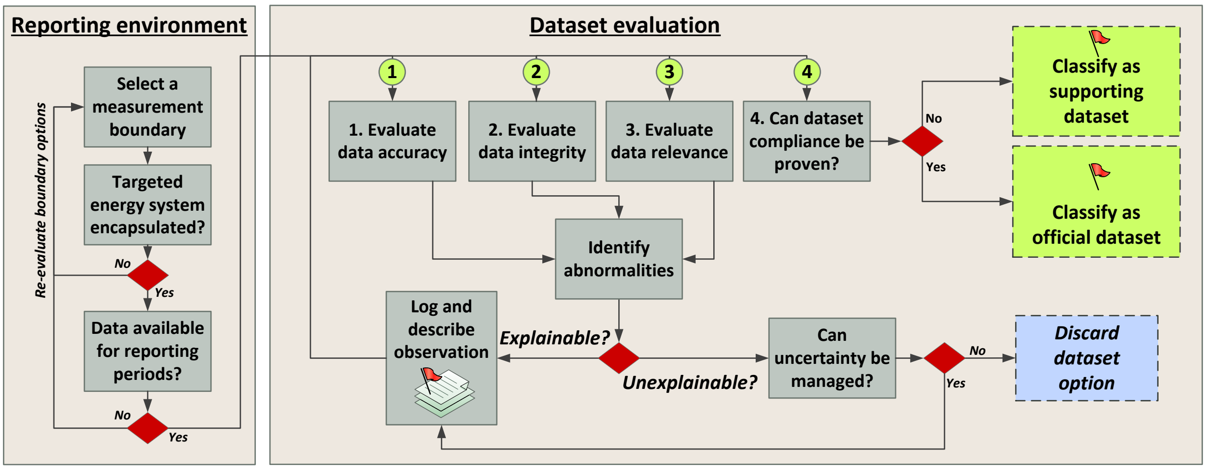


Figure 9: Data quality evaluation framework

Once a dataset has been acquired, three aspects are considered to evaluate the quality of data: accuracy, integrity and relevance (as discussed in Sections 3.2.1 to 3.2.3). A useable dataset is acquired once these three aspects are confirmed. If the dataset can be supported further by proof of compliance, then it is classified as an official dataset. A dataset is classified as a supporting dataset if proof of compliance is lacking while it can still be proven to be of representative quality by identifying and managing data abnormalities.

Verification of the developed data quality evaluation methodologies will now be discussed by case study observations.

4. RESULTS AND DISCUSSION

4.1 Overview of results presented

The data quality framework proposed by this paper was applied to several case studies based at industrial scale energy users in South Africa. Only the relevant observations from isolated case studies are presented in this section. Note that site specific details are omitted and normalised figures are presented due to the use of confidential information.

4.2 Case study 1 - Consistency of data source (Data integrity)

Case Study 1 is based on an industrial plant where several energy related datasets were gathered from various sources. Data quality evaluation was required to test if the company-wide aggregated datasets (high-level data) was traceable to actual measurement points at plant level (plant-level data). The main aspect of this data quality evaluation exercise was therefore to determine if the high-level dataset represented the actual energy system.

Intuitively, the datasets (plant-level and high-level data) should represent the same quantitative results since they are based on the same energy system. However, data integrity is tested since independent data sources are compared. Discrepancies are identified if the plant-level data mismatches the high-level data. For this specific case study, the production data was validated as illustrated in Figure 10 (A). However, a clear mismatch was identified between two independent data sources for coal consumption. This is illustrated by a quantitative comparison in Figure 10 (B).

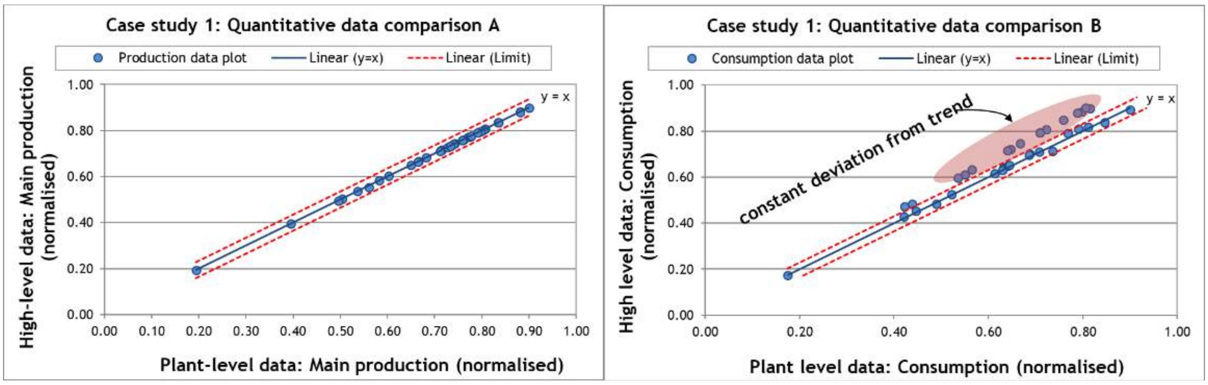


Figure 10: Case study 1 - Quantitative data integrity tests (A & B)

The mismatch identified by the data integrity test triggered an investigation into the coal measurement data. The discrepancy could be linked to inconsistent reporting of coal moisture conditions in the high-level dataset. The coal tonnages in certain instances were given as dry-base (moisture free) and others as-received (total moisture included) while it was assumed to be consistently at dry-base. Although this could be considered a trivial communication mistake, it could have resulted in significant errors if left unchecked.

In this case study a simple data integrity test was used to evaluate data quality. The production dataset was validated while the consumption dataset was invalidated due to a data transfer mistake. This mistake was identified prior to propagating it into energy efficiency reporting functions. Year-on-year, this mistake showed an under estimation of thermal energy usage by a total of 65 GWh. This shows a possible error worth R61 million if the current section 12L tax incentive rate (i.e. R0.95/kWh) is superimposed to this discrepancy.

4.3 Case study 2 - Representativeness of data source (Data integrity)

Case Study 2 evaluates the datasets gathered for a coal-fired boiler. The initial dataset was collected from a high-level data source. A data quality investigation was therefore conducted to evaluate if the high-level data was representative of operations at plant level. This case study focused on two measurements, namely steam production and coal consumption. The steam dataset was validated by the data integrity test as shown in in Figure 11 (A). However, numerous discrepancies were identified when comparing two independent coal datasets. The coal dataset evaluation is illustrated in Figure 11 (B).

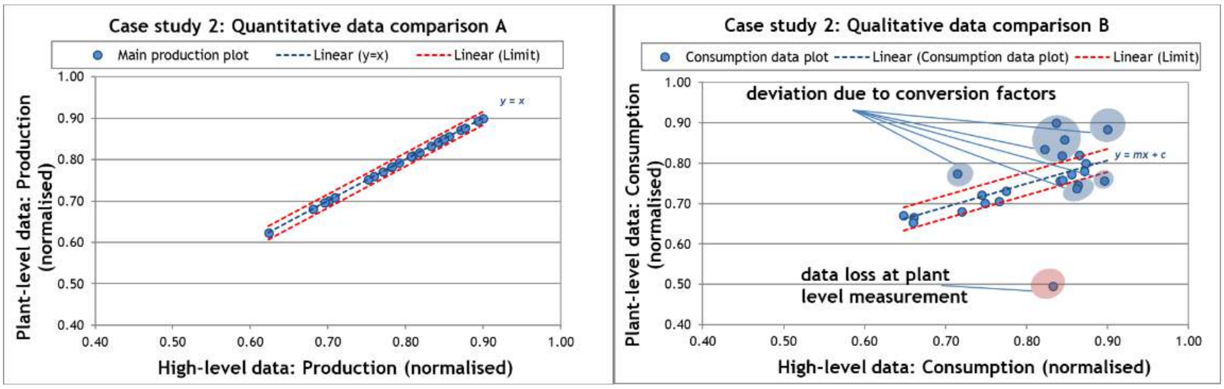


Figure 11: Case study 2 – Quantitative (A) and qualitative (B) data integrity tests

The comparison in Figure 11 (B) shows a qualitative correlation with several deviations. A single deviation was linked to a data loss period at the plant-level meter, while the remaining deviations were due to mismatched coal analyses which were used as conversion factors. This evaluation triggered an investigation which identified an unrepresentative data source which was subsequently discarded for reporting purposes. If the data quality was not checked, it would have resulted in significant errors if the invalidated data was propagated into reporting functions. Year-on-year this mistake displayed an over estimation of thermal energy usage by a total of 886 GWh with a superimposed tax incentive value of R840 million (R0.95/kWh).

4.4 Case study 3 - Representation of an operations (Dataset relevance)

Case study 3 evaluates the relevance of a dataset on an industrial production facility. Several datasets were collected and validated. Since these datasets were collected from various sources, it is important to check if the data is relevant to the operations. A long-term energy intensity trend was therefore plotted as illustrated by Figure 12.

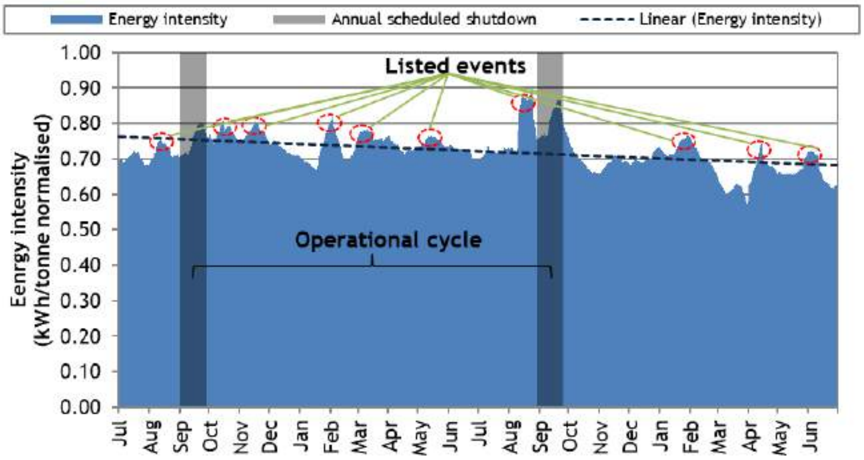


Figure 12: Case study 3 - Data relevance test

Figure 12 illustrates that in the long term, the energy intensity trend had decreased albeit with numerous fluctuations. The question that needs to be answered is whether the observations are a true reflection of the operations or due to data discrepancies. In this case study the observed trend (dotted line in Figure 12) was attributed to an energy management programme aimed at decreasing energy intensity. The significant fluctuations (encircled in Figure 12) were linked to officially recorded events.

The long-term trend also indicated that annual shutdowns (grey bars in Figure 12) are conducted as part of the normal operational cycle. This confirmed that a full operational cycle was included in the reporting period. The dataset was therefore validated since the data-based observations could be linked to independently recorded events. The evaluation increased confidence that the datasets were representative of the operations by reconciling operational events with data-based observations.

4.5 Case study 4 - Reliance on a single data source (Dataset relevance)

Case study 4 evaluates the relevance of a dataset on an industrial production facility. As with Case study 3, the goal is to reconcile operational events with data-based observations to test if the data is relevant. Figure 13 illustrates the long-term energy intensity trend used to evaluate the collected datasets.

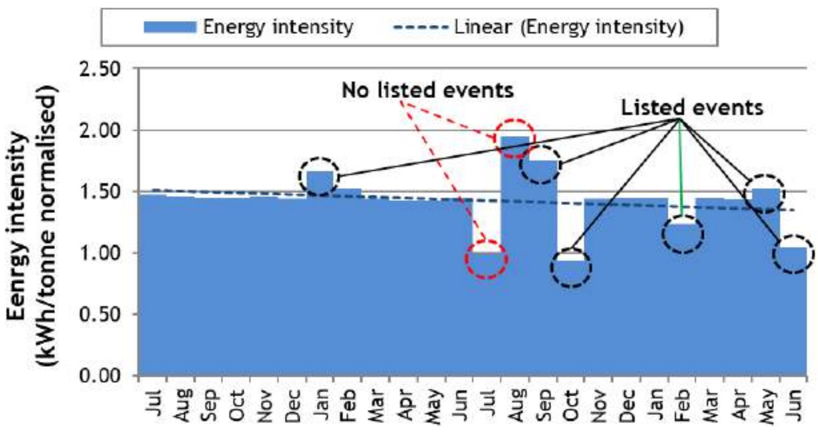


Figure 13: Case study 4 - Data relevance test

In Figure 13, a decreasing energy intensity trend (dotted line) and several fluctuations (encircled) are observed. Six of the fluctuations (encircled in black) were explainable by linking them to actual recorded events. However, two fluctuations (encircled in red) could not be explained and were flagged as discrepancies. After investigation, the discrepancies were linked to a billing recon in which invoiced quantities were consumed in the month preceding the invoice.

The mismatch caused by the billing recon will affect the correlation with energy governing factors, but will not affect total energy usage. This occurrence will only affect reporting functions if the recon took place across a baseline and assessment period, then it would unduly affect the reported change in energy usage. If this was the circumstance for this case study, a year-on-year under estimation of 356 GWh thermal energy usage would have been reported (i.e. superimposed value of R337 million if valued at R0.95/kWh).

In this case study, it is shown that even an official data source, i.e. invoice quantities, can cause unintentional errors. It is therefore critical to evaluate the relevance of a dataset by testing it with independent data sources. This will avoid an overreliance on a single dataset and increase confidence that the data is representative of the energy system.

4.6 Case study 5 - Electricity dataset options (Dataset classification)

Case Study 5 evaluates two datasets that are indicative of electricity usage. One dataset is based on the monthly invoices received from the utility provider. The other dataset is compiled from power meters that are installed on-site to monitor electricity usage. Only the first dataset is compliant since no calibration records were available for the on-site power meters.

Although the invoices are the only compliant data source, it is limited to monthly aggregated values. This will limit any further data handling to 12 data points. The power meter data is stored in half hourly intervals which can be aggregated to daily, weekly or monthly values. This will provide more data handling options since the main energy driver (production) is measured daily. This concept is illustrated in Figure 14 by trending (A) and plotting (B) the relevant datasets at different resolutions (i.e. daily and monthly).

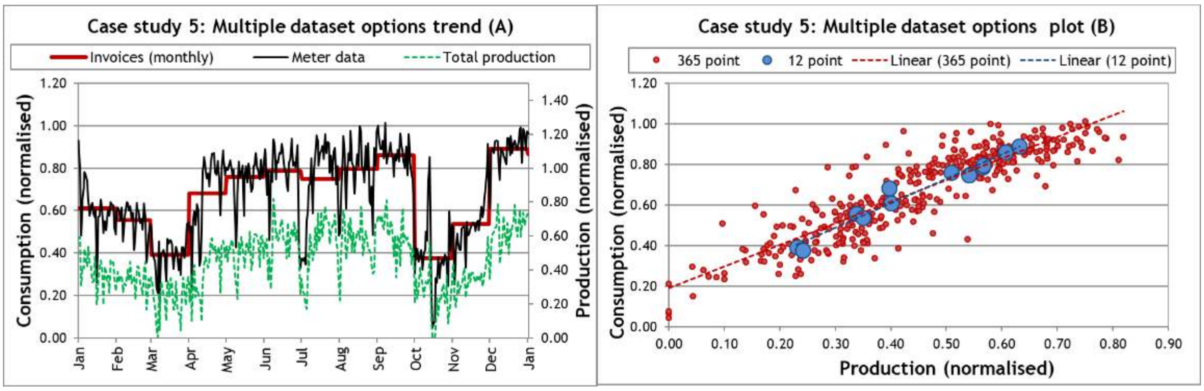


Figure 14: Case study 5 - Dataset classification trend (A) and plot (B)

In this case study, the convention would have been to only rely on the monthly dataset which is traceable to supporting documents in the form of invoices. This, however, limits reporting options (e.g. daily vs. monthly resolution reports). The dataset classification methodology doesn't quantitively impact energy reporting since both datasets present the same quantities. The impact is only gained by increasing the number of available reporting and M&V options. The methodology is therefore proposed to classify the monthly dataset as official, and the additional dataset as supporting since it has been validated with the official dataset.

4.7 Case study 6 - Coal dataset options (Dataset classification)

Case Study 6 classifies two datasets in a similar approach to that of the previous case study. Multiple data sources are available for the consumed coal quantities. However, the only compliant dataset is based on a measurement that can be traced to a monthly invoice (which in turn can be traced to a calibrated meter). This provides the most accurate quantitative indication of consumed quantities.

As with the previous case study, a monthly dataset based on invoices limits reporting and calculation options. In this case study, an additional disadvantage is noted since the invoiced values have an inherent latency due to storage buffers on the plant. An additional dataset is therefore included. This is based on the feeders which provide higher resolution data and a shorter latency, albeit not quantitatively accurate (uncalibrated). Since the feeders are not accurate, it must be normalised to the invoiced values quantitatively to provide an additional dataset. Figure 15 illustrates the two dataset options trended (A) and plotted (B) with the main energy governing factor (production).

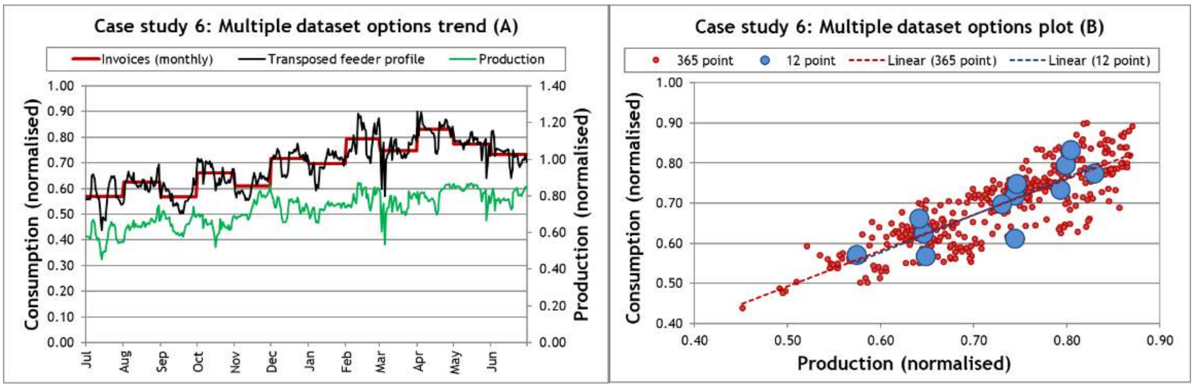


Figure 15: Case study 6 - Dataset classification trend (A) and plot (B)

The two different datasets plotted in Figure 15 have no quantitative difference, but it allows more data points that are more indicative of day-to-day operations. Therefore, by classifying both datasets it can be used to supplement and improve reporting functions.

4.8 Summary of findings and discussion

The data quality framework proposed by this paper was applied to several case studies based at industrial scale energy users in South Africa. Several highlights from data quality evaluations were presented. These are summarised in Table 1.

Table 1: Overview of results from case studies

Case study	*Quality aspect tested*	*Observations made*	*Impact noted*
1	Dataset integrity	Inconsistency of reporting basis	Quantifiable errors avoided
2	Dataset integrity	Undocumented measurement points and data loss	Quantifiable errors avoided
3	Dataset relevance	Data based observations validated	Dataset observations validated
4	Dataset relevance	Mismatch due to billing reconciliation	Quantifiable errors avoided
5	Dataset classification	Multiple resolutions reporting options enabled	Reporting functions expanded
6	Dataset classification	Multiple resolutions reporting options enabled	Reporting functions expanded

Three case studies showed quantifiable discrepancies which could have resulted in data-based errors amounting to R1 240 million (at R0.95/kWh) when considered in terms of the monetary value of the section 12L tax incentive. It is therefore evident that data quality must be evaluated holistically to avoid propagation of erroneous data into reporting functions.

It is expected that the majority of data discrepancies related to data integrity and data relevance occur due to the "size" of data in industry. A typical industrial scale facility employs hundreds of employees and sensors which continuously generate large volumes of data scattered across operations. Data generation has also increased drastically over the past few years owing to development of big data capabilities and Industry 4.0 initiatives [28]. Maintaining data quality and extracting value from growing data systems will therefore remain a relevant topic for future study.

This topic has been discussed for energy efficiency reporting in general and more specifically relating to the 12L tax inventive to highlight the possible monetary value associated with data quality. Since energy and greenhouse gas emissions are fundamentally linked in carbon-intensive industries, this study topic can also be expanded to other mechanisms that are included and related to South Africa climate change strategy [29], [30]. These include data management for greenhouse gas taxation, mitigation and mandatory reporting. This consolidation could provide a structured and integrated approach to sustainability management.

5. CONCLUSION

In this paper, the role of data management methods associated with energy efficiency reporting was reviewed. It showed that conventional methods tend to focus on data quality in terms of measurement uncertainty which is only a single aspect of data quality. Therefore, methodologies were proposed to guide data quality evaluation more holistically in terms of accuracy, integrity and relevance.

The framework sets out to compile various datasets. Each dataset is evaluated using the new methodology focusing on data accuracy, integrity and relevance. Data accuracy is widely discussed in the reviewed literature and is therefore only briefly presented in the methodology. Data integrity is evaluated using quantitative and qualitative comparisons to determine whether the dataset has been compromised along an established traceability pathway. The methodology finally conducts a long-term intensity trend and evaluation process to determine whether a dataset (in addition to being considered accurate and traceable) is a relevant representation of the system.

The dataset evaluation methodology aims to classify all available datasets into discarded, officially compliant or additional supporting data. Discarded datasets are excluded from all subsequent calculations to ensure that reporting functions are not compromised by erroneous data. The compliant datasets are used in all official calculations and the additional datasets are retained to supplement reporting functions. The new framework was applied to isolated case studies and highlighted the possible impact of discrepancies relating to data integrity and relevance. Recommendations are also made for further advancement of data quality assurance within the context of Industry 4.0 and sustainable development.

6. BIBLIOGRAPHY

[1] **R. Y. Wang and D. M. Strong.** 1996. "Beyond accuracy: What data quality means to data consumers," *Journal of Management Information Systems*, vol. 12, no. 4, pp. 5–33.

[2] **K. Vikhorev, R. Greenough, and N. Brown.** 2013. "An advanced energy management framework to promote energy awareness," *Journal of Cleaner Production*, vol. 43, pp. 103–112.

[3] **Hubauer, T., Lamparter, S., Roshchin, M., Solomakhina, N. and Watson S.** 2013. "Analysis of data quality issues in real-world industrial data," in *Annual Conference of the Prognostics and Health Management Society*.

[4] **Republic of South Africa**. 2013. "Section 12L of the Income Tax Act (no. 58 of 1962): Deduction in respect of energy efficiency savings", South Africa.

[5] **National Treasury.** 2015. "Regulations in terms of section 12L of the South African Income Tax Act, 1962, on the allowance for energy efficiency savings," *Government Gazette*, vol. 38541, no. R.186, pp. 1–4.

[6] **Booysen, W.** 2014. "Measurement and verification of industrial DSM projects," PhD Thesis, North-West University, South Africa.

[7] **SABS.** 2018. "SANS 50010: 2018 Measurement and Verification of energy savings," South Africa.

[8] **Hamer, W.** 2016. "A practical approach to quantify RSA Section 12L EE tax incentives for large industry," PhD Thesis, North-West University, South Africa.

[9] **Janse Van Rensburg, H. M.** 2015. "Structuring mining data for RSA Section 12L EE tax incentives," Masters Dissertation, North-West University, South Africa.

[10] **Meijsen,** L.A. 2014. "Verification procedures to ensure consistent energy measurements," Masters Dissertation, North-West University, South Africa.

[11] Van Niekerk, S.G.J. 2015. "Quantification of energy consumption and production drivers in steel manufacturing plants," Masters Dissertation, North-West University, South Africa.

[12] **Hamer, W., Booysen, W. and Mathews, E.H.** 2017., "A practical approach to manage uncertainty in the measurement and verification of energy efficiency savings," in *South African Journal of Industrial Engineering*, November 2017, vol. 28(3) Special Edition, pp 128-146.

[13] **Efficiency Valuation Organization**. 2012. "International Performance Measurement & Verification Protocol: Concepts and options for determining energy and water savings."

[14] **Clean Energy Ministerial.** 2014. "Global Superior Energy Performance (GSEP) Energy Performance Measurement and Verification: Guidance on Data Quality."

[15] **Efficiency Valuation Organization.** 2014. "Statistics and Uncertainty for International Performance Measurement & Verification Protocol."

[16] **Chen, Y.** 2016. "Industrial information integration - A literature review 2006-2015," *Journal of Industrial Information Integration*, vol. 2, pp. 30–64.

[17] **Meek, L.A.** 2013. "A systems engineering approach to improve the measurement and verification process of energy services companies," Masters Dissertation, North-West University, South Africa.

[18] **BSI Group.** 2015. "How sustainability standards can drive business performance." London, UK.

[19] **American Society of Heating Refrigeration and Air-conditioning**. 2002. "ASHRAE Guideline: Measurement of Energy and Demand Savings."

[20] **SANAS.** 2009. "TR15-06 Criteria for laboratory accreditation in the field of mass metrology."

[21] **SANAS.** 2013."R50-03 Estimation of the uncertainty of measurement by calibration laboratories and specification of calibration and measurement capability on schedules of accreditation."

[22] **SANAS.** 2015. "TR25-03 Criteria for performing calibration and intermediate checks on equipment used in accredited facilities."

[23] **Den Heijer, W. and Grobler, L.J.** 2010. "Measurement and Verification Guideline for Energy Efficiency and Demand-Side Management (EEDSM) Projects and Programmes," South Africa.

[24] **Efficiency Valuation Organization.** 2014. , "International Performance Measurement and Verification Protocol: Core Concepts."

[25] **Aspen Technology**. 2013. "Data Historians vs . DCS, SCADA, and PLC Systems." Burlington, MA, USA.
[26] **Gous, A. G. S., Booysen, W. and Hamer W.** 2016. "Data quality evaluation for measurement and verification processes," in *Industrial and Commercial Use of Energy*, Cape Town, South Africa.
[27] **Wilcox, S. M. and Mccormack, P.** 2011. "Implementing best practices for data quality assessment of the National Renewable Energy Laboratory's solar resource and meteorological assessment project," Raleigh, USA.
[28] **Koseleva, N. and Ropaite, G.** 2017. "Big data in building energy efficiency: understanding of big data and main challenges," *Procedia Engineering,* vol. 172, pp. 544 – 549.
[29] **Alton, T., Arndt, C., Davies, R., Hartley, F., Makrelov, K., Thurlow, J. and Ubogu, D.** 2014. "Introducing carbon taxes in South Africa," *Applied Energy*, vol. 116, pp. 344–354.
[30] **Department of Environmental Affairs.** 2018. "Climate Change Bill," *Government Gazette,* vol. 41689, no. 580, pp. 1–36.

CONSIDERING THE NEED FOR ALTERNATIVE INTERVENTION STRATEGIES FOR THE MANAGEMENT OF DIABETIC POLICY FORMULATION IN SOUTH AFRICA

V.V. Thomas[1*], I.H. De Kock[2] & L Bam[3]

[1]Department of Health Systems Engineering and Innovation Hub
University of Stellenbosch, South Africa
18287557@sun.ac.za

[2]Department of Health Systems Engineering and Innovation Hub
University of Stellenbosch, South Africa
imkedk@sun.ac.za

[3]Department of Health Systems Engineering and Innovation Hub
University of Stellenbosch, South Africa
louzanne@sun.ac.za

ABSTRACT

The increasing prevalence of diabetes in South Africa, alongside other non-communicable diseases, places a heavy burden on the health care system; especially when faced with the significant difference in quality of care between private and public health care, and the increased burden of disease.

This paper analyses various diabetic policies already implemented in South Africa, and considers the need to investigate alternative policies and intervention strategies to manage diabetes in South Africa. Due to the complex nature and non-linear interactions which exist within the health care system, a system dynamics-based approach is suggested as a useful analysis tool to evaluate and understand the dominant factors that influence the effective management of diabetes to potentially inform more effective and efficient diabetic policy formulation.

*Corresponding author

1. INTRODUCTION

The increasing prevalence of diabetes mellitus (commonly referred to as diabetes) in the world is a widespread concern [1]. According to predictions by the International Diabetes Federation, the prevalence of diabetes is expected to globally increase from 415 million in 2015 to 642 million in 2040 [2, 3]. While improvement has been made in the epidemiology and management of diabetes in the developed world [1], the same advances have not been made in sub-Saharan Africa. Sub-Saharan Africa, similarly to the rest of the world, is experiencing an increasing prevalence of diabetes alongside other non-communicable diseases [3]. In South Africa, this diabetic trend is still emergent in a region confronted with high rates of communicable diseases, such as HIV, as well as Tuberculosis [4].

Additionally, diabetes plays a significant role in contracting several other (often life threatening) diseases [4]. These diseases include both non-communicable diseases, such as cardiovascular disease and renal disease, and communicable diseases, such as pneumonia and tuberculosis, which have a considerable impact on morbidity and mortality in sub-Saharan Africa [4]. In addition, the prevention and treatment of diabetes is a significantly complex process, involving numerous role-players and stakeholders (i.e. government agencies, the healthcare system, communities and diabetic patients). It is, therefore, necessary to also consider diabetic health care in South Africa from a complex, system's perspective, with significant non-linear interactions. With the increasing burden of disease in South Africa, as well as limited resources and the complex, dynamic nature of the healthcare system, it is unsurprising that the prevalence for diabetes in South Africa continues to increase [1, 4].

In order to highlight the need to investigate alternative diabetic intervention and management strategies, this paper draws focus on (i) the South African health care system, (ii) diabetes in South Africa, (iii) policy and intervention strategies, and (iv) modelling techniques and approaches that could be utilised to model and subsequently evaluate complex systems. At the onset of the paper, an analysis of the South African health care system is presented to provide context for this paper, wherein the efficiency of the system, as well as the inequality between the public and private health care system and increased burden of disease, are discussed. The paper then focuses on specific aspects of diabetes in South Africa, such as the disease itself and the growing prevalence within the country, after which a discussion of the financial implications and management of the diabetes within South Africa highlights the need to address diabetes from the perspective of diabetic management intervention strategies.

The focus of this paper then shifts to a discussion of the approach used during policy analysis. An analysis of South African diabetic policy and intervention strategies is then presented, together with requirement specifications developed as an outcome of this analysis. These requirement specifications develop a need for a specific modelling technique, given the context, to analyse policy interventions. Thereafter, this paper introduces the concept of simulation modelling, where appropriate modelling approaches are discussed and evaluated so as to determine which approach most comprehensively meets the needs of the requirement specifications. Finally, this paper motivates the need to consider alternative intervention strategies for the management of diabetic policy formulation in South Africa by drawing on key points discussed throughout the paper.

2. SOUTH AFRICAN HEALTH CARE SYSTEM

During the past two decades, the South African government has aimed to improve the condition of the public health care system by outlining a clear model with a focus on primary health care (PHC) [5]. PHC, in the case of South Africa, refers to the first line of health care that a patient receives, at either a clinic, community health centre or district hospital, which may include the treatment of a disease, referral to more specialised care if required, and prevention through health education aimed at individuals, families, and communities [6]. The inadequacy of PHC available to the majority of the South African population during the apartheid era, however, led to a significant disproportionate excess of serious health problems and challenges, which was manifested in higher infant mortality rates, as well as lower life expectancies [5]. While the post-apartheid government has since developed a primary-centred health care model aimed at all South Africans, the quality of PHC in South Africa remains challenge-stricken [5].

In order to comprehend the current state of the South African health care system, it is necessary to investigate the efficiency of the system to provide health care for the South African public. Furthermore, the inequality between private and public health care, and the increased burden of disease experienced in South Africa, needs to be analysed to develop context for this paper, and is discussed in the sections below.

2.1 Efficiency of the health care system

In 2014, *The Economist* reported on a study that performed a 166-country health outcome report comparing the health care performance and spending patterns of various countries [7]. Figure 1 displays a plot that ranks the health outcomes of a country versus the ranking on healthcare spending for each country. The outcome measure was a combined function of (i) adult mortality in 2012, (ii) life expectancy at 60 years of age, (iii) disabilityadjusted life years, and (iv) health-adjusted life expectancy. This study found that health outcomes were directly correlated with healthcare spending [7]. According to the study, a country with a high ranking for healthcare spending should be expected to have a high ranking for healthcare outcomes. It is, therefore, contradicting to this trend that while South Africa ranks high in terms of spending, it is ranked significantly low in terms of healthcare outcomes.

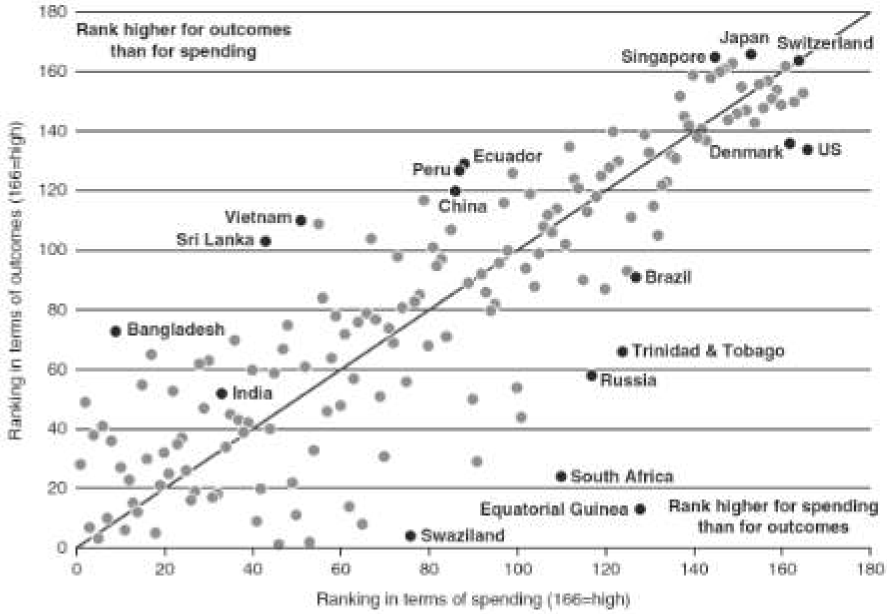


Figure 1: Health outcomes rank versus spending rank by country [7]

This result is echoed by the *Future Health Index* study, commissioned by Dutch technology company, Philips, which examined the current realities of how well the healthcare system of a country is set up for the future in order to quantify the readiness of health systems [7]. The results of the study ranked South Africa last among 19 nations in a global survey which measured healthcare system efficiency — the ability to deliver maximum results at the lowest possible cost. The study included countries such as China, France, the United States, Argentina, United Arab Emirates and Brazil, and whereas the group scored an average efficiency score of 10.5, South Africa achieved an efficiency score of 4.4 [7].

Maillacheruvu [5] argues that the quality of PHC in South Africa is challenged by two major issues: the inequality between private and public health care, and the increased burden of disease experienced in South Africa. Maillacheruvu [5] further argues that by addressing these issues, the health care outcomes of South Africa may be significantly improved. It is suggested that the inequality between public and private health care, as well as the increased burden of disease, is augmented by the historical struggles induced by the economic, political, and societal structure of the apartheid era [5]. These two issues are addressed in the respective sections to follow.

2.2 Inequality between private and public health care

Health care in South Africa varies from the most basic PHC offered by the state, to highly specialised health services available in the both public and private sector [8]. In 2015, the total health expenditure (THE) per capita was USD 570, where the THE formed 8.9% of the South African GDP [9]. According to the National Treasury's Fiscal Review [10], the South African GDP spent on health was split between the private health sector (48.5%) and the public sector (49.2%). The remaining 2.3% was donated and spent by NGOs. While the private and public sector both receive a similar share of the GDP for healthcare, the private healthcare sector only services an estimated 16% of the population, while the remaining 84% of the population relies on public healthcare services [10, 9].

In addition to the disparate patient populations between private and public health care, the number of health care providers within each system is also disproportionate. The World Health Organization estimates that only 30% of all South African physicians work in the public sector, despite the public health care sector serving 80% of the population [11]. The relatively lower number of health care providers, along with the increased number of patients in the public health care system, results in an overburdened public care system, when compared to the private sector. It is also argued that public health care workers are, in turn, overworked and thus, making it challenging to provide the same level of personalised services when compared to the private sector [12]. This is clearly evident in a study conducted at Tafelsig Clinic by Maillacheruvu [5], where clinicians were often overworked, and were unable to provide the health care needed to serve all the patient present at the clinic [5].

In addition to the disproportionate number of health care workers in the public and private health care sectors, the economic divide between the rich and the poor, also contributes to the persisting inequality between the two health care sectors. To further illustrate the extent of the inequality, it was reported in 2007 that South Africa had the world's 10th highest Gini index at 0.578, which is a measure of income inequity among a nation's population [13]. The economic divide in South Africa has created a separation within the national health care system, and, as a result, has developed a considerable discrepancy between the resources used by public health care and the private sector.

Another consequence of the economic divide between public and private health care is the poor usage of the health care referral system [14]. Referrals between public clinics, community health care clinics and district hospitals are standard practice [15]. The referral system begins at PHC clinics as the first step in the provision of health care [7]. If the clinic cannot assist, the patient will be referred to a community health center, as the second step in the referral system. The third step in the provision of health care is the district hospital. Thereafter, patients may be referred to secondary, tertiary, or quaternary level of care [7]. It was, however, noted by Mojaki [14] during a study conducted at the Dr JS Moroka District Hospital in the Free State, that the majority of patients seen in the outpatient department and casualty had bypassed the referral system by means of self-referral. More than 50% of these patients could have been managed at PHC clinics or community health centers, which is similar to study findings of Rutkove [16] at King Edward VIII Hospital in Durban. Long waiting times (close to 6 hours) in the hospital may have been reduced if patients were managed at their nearest PHC facility.

Mojaki [14] found that the primary reason for patients bypassing PHC facilities was their perception of the superior care and resource availability at hospitals. This perception may be rooted in the inadequacy of PHC available to non-white South Africans during the apartheid era. Other cited reasons included dysfunctional community health centers and lack of education about the referral system among patients and health professionals. In addition, less than 2% of patients were educated on the referral system, and none had been charged the bypass fees, despite a provincial policy. Mojaki [14] argues that the tendency of patients to bypass PHC facilities disrupts hospital's core functions and is linked to the overcrowded outpatient department in these hospitals.

It may, therefore, be seen from the studies at Tafelsig clinic and Dr JS Moroka Hospital, together with the plethora of other health care facilities facing almost identical situations throughout South Africa, that the overburdened public health care facilities have an insufficient number of health care providers needed to fulfil the personal, community-based values of the PHC model outlined in the National Health Bill.

2.3 Increased burden of disease

Along with the economic divide investigated in the previous section, another defining characteristic of the public PHC system of South Africa is the considerable array of long-term diseases that providers must treat. This, in turn, minimises the available time allocated to other aspects of PHC, which includes counselling and prevention through education [5]. Communicable diseases, such as HIV/AIDS and Tuberculosis (TB) are widespread in South Africa, but non-communicable diseases, including hypertension and diabetes, are growing in prevalence [17]. With more chronically-ill patients, public PHC facilities are under significant strain to dedicate sufficient resources to assist all patients.

The most considerable health problem that South Africa faces is the combined HIV and TB infection rate [18]. In 2011, the prevalence of HIV among South Africans aged 15-49 years was 17.3%, which is among the highest in the world [18], and, in 2014, TB was the leading cause of death in South Africa [18]. Additionally, individuals affected by HIV/AIDS are more susceptible to other infections, such as TB, due to their compromised immune systems. In fact, individuals who are HIV positive are ten times more likely to develop TB [19]. Treatment methods for TB exist and are readily available, but require a strict treatment regimen that may last up to six months, and may necessitate multiple visits to a clinic per week [19]. The prolonged period required to treat TB increases the burden of communicable diseases on PHC facilities, and also decreases the likelihood that patients will complete their recommended course of treatment [18].

Statistics South Africa [20] recently released a report on the top ten leading causes of death in South Africa, based on all death notification forms maintained by the Department of Home Affairs. Although South Africa is afflicted with a high HIV and TB rate, the data, provided by Statistics South Africa, found that 55.5% of all deaths were attributed to non-communicable diseases and that diabetes was the second leading cause of death in 2015 after TB [8]. Furthermore, it has been reported that diabetes is the number one killer of women living in the Western Cape Province [21]. Potential associations between diabetes and TB, as well as HIV, may also further complicate the pattern of increasing diabetes prevalence in South Africa and the challenges posed on resource-constrained health systems. A recent meta-analysis of thirteen studies found that individuals with diabetes were associated with a three-time elevated risk of tuberculosis [22]. Furthermore, the high prevalence of HIV, as well as antiretroviral therapy treatment for HIV, may increase the prevalence of diabetes risk factors and, consequently, diabetes incidence [5].

Although once regarded predominantly as a disease related to the developed world, it is clear that diabetes now exerts a significant burden in South Africa, which is expected to increase [4]. Many diabetic patients face significant challenges accessing diagnosis and treatment, further contributing to the high mortality and prevalence of complications observed [1].

3. DIABETES IN SOUTH AFRICA

Global projections by the International Diabetes Federation (IDF) has shown that diabetes prevalence is expected to double from 285 million in 2010 to 592 million in 2035, with the sub-Saharan African region bearing the burden of this increase, and South Africa at the forefront [24]. Of the 14.7 million people living with diabetes in Africa [2, 3], approximately 11 million are found in sub-Saharan Africa, and 2.3 million in South Africa. Until recently, diabetes was considered uncommon in these regions, but due to demographic and lifestyle changes, diabetes is increasingly identified as a prevalent health problem [25].

This section firstly provides a brief overview of the diabetes disease. Thereafter, the prevalence of diabetes in South Africa will be explored. Consequently, the increasing financial implications of the disease is discussed. Finally, the section concludes with an analysis of the management of diabetes in South Africa.

3.1 The diabetes disease

Diabetes mellitus[4], or diabetes, is a chronic and lifelong condition which may affect the human body's ability to use the energy from ingested food. There are three major types of diabetes[5]: type 1 diabetes, type 2 diabetes, and gestational diabetes [3, 26].

3.2 Diabetes prevalence in South Africa

Type 2 diabetes accounts for 90% of diabetes cases in sub-Saharan Africa, whilst type 1 diabetes and gestational diabetes constitute the remaining 10% [4]. The prevalence of type 2 diabetes has increased significantly from that recorded in pre-1985 surveys conducted within the region. These surveys found that the prevalence for diabetes in sub-Saharan Africa was typically below 1%, with the exception of studies in South Africa, where a 3.6% prevalence was observed [4].

Data from the IDF [2] estimates that, as of 2018, 7% of South Africans between the ages of 21 and 79 years have diabetes. Based on population estimates for South Africa, it is estimated that 2.29 million South Africans in the aforementioned age group have diabetes [6]. Of the 2.29 million people with diabetes, 1.4 million (61.1%) are undiagnosed [26]. Furthermore, it is estimated that an additional 5 million South Africans have pre-diabetes; a condition most likely caused by insulin resistance and results in blood glucose levels being higher than normal, but not significantly high enough to be classified as type 2 diabetes [3]. The highest prevalence of diabetes in South Africa is among the Indian population, with a prevalence of 11-13% [3]. This is followed by 8-10% in the coloured community, 5-8% among the black community, and 4% among the white community [3].

As diabetic symptoms may initially be extremely mild and develop gradually, combined with an ineffective PHC education intervention system in South Africa, many people fail to recognise symptoms as warning signs of diabetes [3]. In most cases, diabetic complications may have been avoided entirely by early diagnosis and proper treatment [3]. Due to the already considerable burden of disease in South Africa, however, the growing prevalence of diabetes may potentially be unavoidable and lead to increased strain on the already stressed health care system, as well as economic ramifications for South Africa.

[4] A variable disorder of carbohydrate metabolism caused by a combination of hereditary and environmental factors and is typically characterised by an insufficient secretion or utilisation of insulin, excessive urine production, significant amounts of sugar in the blood and urine, and by thirst, hunger, and loss of weight [3]. A number of medical risks are associated with diabetes, such as diabetic retinopathy, diabetic neuropathy, and diabetic nephropathy [26]. Furthermore, persons with diabetes have an increased risk of heart disease and stroke [26]. Treatment is required to maintain blood sugar levels within a target range, and includes taking several insulin injections every day or using an insulin pump, monitoring blood sugar levels and eating a healthy diet that spreads carbohydrate throughout the day [26].

[5] Type 1 diabetes is referred to as "insulin-dependent" diabetes and typically emerges during childhood. This variation of diabetes is an autoimmune condition, where the human body attacks its own pancreas with antibodies. After significant damage, the pancreas of a person with type 1 diabetes is be unable to produce insulin [4]. The most common form of diabetes is type 2 diabetes and accounts for 95% of diabetes cases in adults [3]. With the rise of obesity in children, however, type 2 diabetes is now being increasingly diagnosed in young people and teenagers [29]. In the case of type 2 diabetes, the pancreas is typically capable of producing some insulin. The insulin is, however, either insufficient for the needs of the body, or the cells of the body are resistant to the insulin. The final variation of diabetes is triggered by pregnancy, which is referred to as gestational diabetes, and occurs between 2% to 10% of pregnancies [26]. In contrast to type 1 and 2 diabetes, gestational diabetes typically resolves itself after pregnancy.

3.3 Economic implications of diabetes in South Africa

In 2010, the cost per person with diabetes in South Africa per annum was approximated by the Centre for Diabetes and Endocrinology (CDE) as USD 405.52 [28]. The Society for Endocrinology, Metabolism and Diabetes of South Africa (SEMDSA) estimated that the cost per person with diabetes in South Africa per annum in 2015 increased to USD 918.9 [28]. This is consistent with the observed increased prevalence of diabetes in South Africa. As discussed in section 2.2, the healthcare spending per capita was equivalent to USD 570, which is significantly lower than the cost per person with diabetes in South Africa estimated by SEMDSA.

Furthermore, the estimated cost per diabetic person in South Africa is likely to be significantly lower than the actual cost, due to factors such as undiagnoses, the cost for diabetes prevention programs, over-the-counter medications required for diabetes-related eye and dental problems, and the cost of reduced quality of life, pain and suffering which cannot be measured directly [29]. The economic implications further extend to the public health care system, which may continue to be over burded by the potencial increase of diabetcis to treat. With the increasing impact of diabetes that is expected to occur, future health care spending for diabetes is likely to increase.

3.4 Management of diabetes in South Africa

In order to address the increasing financial and economic strain of diabetes in South Africa, the management of the disease needs to be considered. In South Africa, there are a number of private and public agencies, with a wide spectrum of strategies, in attempt to manage diabetes.

In the private sector diabetes landscape, there exists the CDE - a diabetes management solutions enterprise in South Africa [28]. Their mandate is to improve the health and lives of diabetics by means of various formal diabetes management programmes, partnerships with medical aid schemes, and the education and accreditation of healthcare professionals in diabetes care principles [28]. Additionally, a non-profit organisation, Diabetes South Africa (DSA), which was founded to be a support and advocate for all people living with diabetes in South Africa [30]. The DSA primarily acts as an advocate for diabetics in South Africa by lobbying for better facilities, cheaper medication and better health care services, as well as promoting prevention through public awareness of diabetes, and its symptoms and risks [30].

While a plethora of private agencies play a role in the management of diabetes in South Africa, the most prominent is agency is SEMDSA – a scientific society that aims to further the clinical practice, as well as promote both clinical and scientific research and publication, into all branches of endocrinology, metabolism and diabetes [26]. This society strives to promote acceptable standards for training and the professional practice of endocrinology, metabolism and diabetes as well as to provide advice, where necessary, regarding the academic standard of individuals and training units [26]. SEMDSA also aims to promote access to the provision of health care services and adequate treatment for all affected diseases related to endocrinology, metabolism and diabetes, with particular focus on the poor and needy [26]. In 2017, SEMSDA released the *SEMDSA 2017 Guidelines for the Management of Type 2 diabetes mellitus* diabetic guideline to inform general patterns of care, to enhance diabetes prevention efforts and to reduce the burden of diabetes complications in people living with this disease, which is based on international best-practice [26]. The guideline addresses diabetes diagnosis, screening, diabetic lifestyle interventions, glucose management, comorbidities and complications, as well as focusing on special diabetic populations, such as children, adolescents, older persons, pregnant women [26]. It is, however, noted that this guideline only pertains to the care of adults with type 2 diabetes at primary care level [26].

In 2014, SEMDSA collaborated with the South African Department of Health (DoH) to produce the updated *Management of type 2 diabetes in adults at primary care level* policy guideline to manage diabetes from a public healthcare sector perspective [31]. The aim of implementing the updated guideline was to reduce diabetic complications, as well as to reduce premature mortality from diabetes. This formed an integral part of the Diabetes Implementation Strategy for South Africa, which was developed in response to the African Diabetes Declaration and Strategy of 2006 [31]. In this policy guideline, diabetes diagnosis, screening, glucose management, comorbidities and complications are addressed [31]. The policy, however, fails to address the treatment of diabetes at a level higher than PHC, or diabetic treatment for children.

In the 2013 *Strategic Plan for the Prevention and Control of Non-Communicable Diseases 2013 – 2017*, a listing of all the health care policies published by the DoH since 1998 are presented [32]. Prior to the 2014 *Management of type 2 diabetes in adults at primary care level* policy, only two other diabetic guidelines are listed – the 2005 *Management of diabetes type 1 and type 2 in adults at hospital level* and the 2008 *Guidelines for the management of type 1 diabetes in children* [32]. These two guidelines are, however, not publicly available. Kleinert [33] suggests that the content of South African health policy, as well as the poor documentation and availability thereof, is hampered by ineffective leadership, inexperienced and unaccountable managers, and a weak health system. In addition, both the 2005 *Management of diabetes type 1 and type 2 in adults at hospital level* and the 2008 *Guidelines for the management of type 1 diabetes in children* are only implemented as guidelines, as opposed to policy [32]. The 2014 *Management of type 2 diabetes in adults at primary care level* policy is, therefore, the most recent, publicly available diabetic policy implemented by the DoH at a national level and will be useful when considering the need to investigate alternative policies and intervention strategies to manage diabetes in South Africa. In addition, the *SEMDSA 2017 Guidelines for the Management of Type 2 diabetes mellitus,* while not published by the DoH, may prove to be useful in understanding the management of diabetes, as it is the most recently published diabetic guideline in South Africa [26].

In order to manage diabetes effectively in South Africa, the public health system should ensure that implemented intervention strategies and policies address the needs of diabetics at *all* levels of health care. This paper, therefore, argues that there is substantial need to investigate alternative policy and intervention strategies to inform diabetic policy formulation in South Africa.

4. POLICY AND INTERVENTION STATEGIES

According to Walt [34, 35], policy analysis is a multi-disciplinary approach to public policy, which aims to explain the interaction between the interests and ideas of the various stakeholders in the policy process, and may be useful a useful tool in understanding past policy failures and successes and planning future policy implementation.

As discussed in Section 3.4, it is crucial that policy and intervention strategies are developed, evaluated and implemented that ensures effective management of diabetic health care in South Africa. To further highlight this need, the methods of health policy analysis, as well as a brief overview of South African diabetic policy and intervention strategies, are discussed in the sections to follow.

4.1 Policy analysis approach

Walt [35] argues that health policy analysis typically focuses on the content of policy reform and neglects the actors involved in the reform, the processes required to develop and implement change, and the context within which the policy is developed. When a disproportionate amount of focus is placed on the content of a policy, attention is typically diverted away from understanding the processes, which substantiates why desired policy outcomes may fail to emerge [35]. Reich [36] argues that policy reform is generally a political process, which affects the inception, formulation and implementation of policy. Furthermore, policy-makers, whether politicians or bureaucrats, are acutely aware that reforms are often unpopular and may cause significant social instability [35].

From a policy-domain characterised primarily by consensus, health policy is increasingly subject to conflict and uncertainty. This encouraged Walt [35] to generate alternative ways of analysing policy. Walt argues that policy analysis from a systems perspective offers a more comprehensive framework for thinking about health reform, than approaches which concentrate on the technical features of the content of reform. Walt, therefore, suggests the use a simple analytical model, as shown in Figure 2, to conduct policy analysis. This model, commonly referred to as the policy analysis triangle, incorporates the concepts of context, process, content, and actors to allow policy-makers and researchers to understand the process of health policy reform better, and to plan for a more effective implementation.

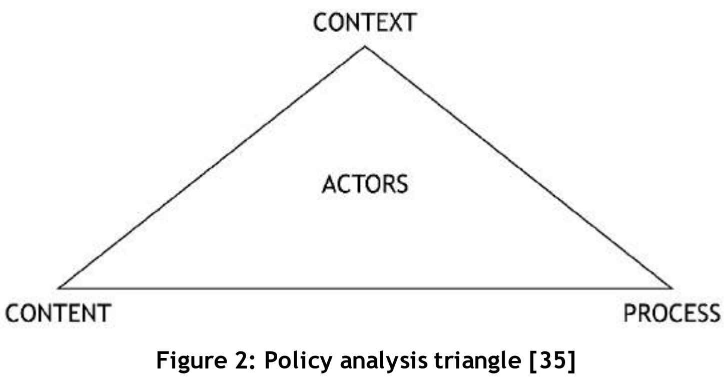


Figure 2: Policy analysis triangle [35]

The policy analysis triangle is a simplified model of a complex set of interrelationships. This model also considers the system as a whole, as opposed to each entity separately. This rationale may be observed as follows: *actors*, as individuals, members of interest groups, or professional associations, are influenced by the *context* within which they live and work. Context may be affected by many factors, such as instability or uncertainty created by changes in political regime, the economic standpoint of a country, social inequalities, historical experiences, or culture. The *process* of policy-making is in turn affected by actors, their position in power structures, their own values, and expectations of the policy. The *content* of policy may, therefore, reflect some or all of the above dimensions.

The traditional analysis focus on the content of policy, but neglects the other dimensions of process, actors and context, which may significantly affect the successful implementation of a policy [35]. In adopting the model, this paper argues that policy is not developed free of bias, but is rather the outcome of complex social, political and economic interactions, as well as non-interactions.

4.2 Analysis of South African diabetic policy and intervention strategies

As discussed in Section 3.4, two South African policy guidelines were identified to be in most prominant in managing diabetes in South Africa. The first of which is the *SEMDSA 2017 Guidelines for the Management of Type 2 diabetes mellitus* diabetic guideline — published by SEMDSA, a private organisation. The second is the 2014 *Management of type 2 diabetes in adults at primary care level* policy guideline — published by the South African Department of Health (DoH), a government entity. In order to understand these policy guidelines, a policy analysis is simultaneously conducted on both policies using the policy analysis triangle introduced in Section 4.1.

The policy analysis begins by considering the *actors* involved in both policies. In the context of the public policy of South Africa, it is observed that the policy was developed by the South African DoH with advice and direction provided by the SEMDSA Steering Committee and the Advisory Committee [26]. It is, however, noted that while these *actors* (the DoH and subject-matter experts) were acknowledged for their contribution to the policy, the consultation of diabetic patients, PHC clinicians, and community health workers were not mentioned. Public policy is, however, typically created in an open environment with free debate is subject to the force of law, whereas private organisations may develop their own policy based on the rules and regulations of their organisation without public accountability. This characteristic of private policy should to be considered when analysing the SEMDSA policy. While the *content* of the SEMDSA is based on international best-practice and the *actors* involved include the guideline committee (consisting of an extensive number of subject-matter experts), the public was not acknowledged as *actors* in this policy. Since there were, however, a significant number of subject-matter experts involved in the creation of both policies, as well as an external moderation process, the *content* of both policies may, therefore, be considered of a significantly high quality that meets international best-practice. It is, however, viewed that the *content* of both policies is directed at adults with type 2 diabetes at primary care level, and do not address higher levels of health care, or other types of diabetes.

While analysing the *actors* involved, it is seen that diabetic patients, PHC clinicians, or community health workers were not actively involved in the policy-making process. The knowledge of the involvement of actors is vital when the *context* of a policy is considered. As described in Section 2.2 and 2.3, there is a large inequality between the private and public sector, and an increased burden of disease in South Africa, respectively. This has placed a significant stress on the PHC clinicians and community health workers, and, therefore, increases the difficulty of diabetics assessing sufficient health care. In addition, the poor usage of the referral system in South Africa, as discussed in Section 2.3, is extensively relied on in the public policy. Although the *content* of both policies should equip PHC clinicians to treat diabetics effectively, the *context* of the current South African health care system prevents PHC clinicians from providing sufficient health care to diabetics. Finally, the *process* aspect of the policy analysis is considered. While the South African DoH released and implemented the *Management of type 2 diabetes in adults at primary care level* policy guideline in 2014 to manage diabetes in South Africa, the prevalence of the disease has significantly increased, as discussed in Section 3.2. This may be similarly found for the policy released by SEMDSA.

From the above policy analysis, the following requirement specifications may be developed to investigate alternative intervention strategies for the management of diabetic policy formulation in South Africa by means of a modelling approach:

i. **Problem identification**: Does the modelling technique assist the developer in identifying the intervention strategies and actors within the entire system?
ii. **Non-linear and dynamic intereactions**: Does the modelling technique accurately investigate the effect of intervention strategies on the heath care system?
iii. **Flexibility**: Does the model adapt well to varied input data and changes to interrelationships?
iv. **Outcome accuracy**: Do the results of the model justify the computational intensity?
v. **Indication of effect over time**: Does the model provide the expected outcome of the system over a time period specified in order to model future predictions?

These requirement specifications may, therefore, be utilised in identifying an appropriate modelling approach to investigate alternative intervention strategies for the management of diabetic policy formulation in South Africa.

5. THEORETICAL MODELLING OF DIABETIC INTERVENTION STRATEGIES IN THE SOUTH AFRICAN HEALTH CARE SYSTEM

Complex, dynamic systems typically presents multiple non-linear interactions, and may, therefore, also act as barriers to solving problems by means of analytical approaches. In order to address the issues that arise in complex systems, Banks [37] recommend that a mathematical computational method, based on iterative algorithms, should be employed, and identified simulation as the most appropriate method to analyse and understand such systems. Additionally, Banks [37], as well as Borshchev [38], noted that the use of simulation allows for the behaviour of a system to be evaluated and predicted beyond that of the available data for time or space. According to Banks and Borshchev, simulation is the only practical means to test complex system models, and suggest that without simulation, even the best conceptual models may only be tested and improved by relaying on the learning feedback through the real-world. This real-world feedback is, however, slow and often rendered ineffective by dynamic complexity, time delay, inadequate or ambiguous feedback, poor reasoning skills, or the cost of experimentation [37, 38]. In these circumstances, simulation becomes the only reliable way to test hypotheses and evaluate the likely effects of interventions, such as policy.

In the section to follow, simulation modelling approaches will be introduced and discussed. Thereafter, the simulation modelling methods will each be evaluted, based on their ability to meet the requirement specification developed in Section 4.2, in order to identify an appropriate modelling approach to investigate alternative intervention strategies for the management of diabetic policy formulation in South Africa.

5.1 Simulation modelling approaches

According to Borschev [38], the three most commonly used simulation modelling approaches are system dynamics modelling, agent-based modelling, and discrete-event modelling.

System dynamics is a simulation method, which was first developed in the 1950s by MIT Professor, Jay Forrester, to examine and characterise the dynamics of economic, as well as social structures [38, 39]. It is suggested by Borshchev [38], that when using a system dynamics approach, that the most important notion is to maintain a point of view from within the system, and that it should, therefore, be viewed as endogenous. This view is achieved by modelling the system as a casually closed structure which, in turn, defines the system's behaviour [38, 40]. The next step is to identify the various feedback loops. Feedback loops can be thought of as circular causality in that they essentially continually recalibrate the system based on the state of other parts within the structure [38]. Feedback loops are considered the crux of systems dynamics. It is suggested that all the concepts in the real system should be defined as continuous quantities, interconnected in loops of information feedback and circular causality. Furthermore, the system dynamics approach involves identifying the different accumulations, or stocks within the system, as well as the flows that affect them [38].

Agent-based modelling is a modelling approach developed in the early 2000s as a result of the rapid growth of available CPU power and memory. This is in the light of the fact that agent-based models are more computationally demanding of both when compared to system dynamics and discrete event models [37, 38]. In agent-based modelling, a system is modelled as a collection of independent decision-making entities, called agents [38]. In this modelling method, each individual agent assesses its situation and executes various decisions or actions based on a set of established rules [38]. Borshchev [38] suggests that agent-based modelling is ideal when the simulation developer possesses insight into how the objects in the system behave individually, rather than knowing how the system behaves as a whole. The model may, therefore, be constructed using the bottom-up modelling approach, commencing by identifying objects or agents and defining their behaviours. This modelling method differs from the systems dynamics approach, which develops a model through means of a top-down modelling approach [38].

Discrete-event systems are systems where specific state changes or events occur at discrete instances in time and that no state change takes place in the system between these events. Systems that are defined by occurrences discrete events may be modelled using discrete event modelling [38]. According to Borshchev [38], the level of abstraction suggested for discrete event modelling is significantly lower than that of system dynamics modelling. Abstraction refers to the complexity by which a system is observed [38], and is an important consideration depending on the type of problem at hand. The highest level of abstraction considers the entire system in low detail, whereas the lower levels of abstraction normally investigate smaller system components in a higher level of detail. In discrete event modelling, each object in the system is represented by an entity or a resource unit at a low level of abstraction, whereas the individual objects in a system dynamic model are aggregated, and are therefore modelled as a high abstraction [38].

5.2 Evaluation of simulation modelling approaches

When considering which simulation modelling method is most appropriate to understand the influencing factors of various diabetic intervention strategies for the management of diabetic policy formulation in South Africa, the decision is primarily based on the type of system being modelled and the purpose of this system [38]. Borshchev [38], however, also notes that the selection of a modelling approach is often largely influenced by the skill set or background of the simulation developer. Another consideration when selecting the modelling approach may be the level of *abstraction*.

In order to select the most appropriate modelling approach, discrete-event modelling, system dynamics modelling, and agent-based modelling will each be evaluated based on the ability of the approach to meet the requirement specifications developed in Section 4.2. As the selection of a modelling approach is often largely influenced by the skill set or background of the simulation developer, 'ease of creation' is added as an additional requirement specification. Table 1 contains a summarised description of the three modelling approaches and their capacity to meet the requirement specifications.

Table 1: Summary of simulation modelling techniques [37, 38, 39, 40]

Attributes	Modelling approaches		
	Discrete-event	System dynamics	Agent-based
Problem identification	Very poor	Excellent	Poor
Non-linear and dynamic intereactions	Very good	Very good	Excellent
Flexibility	Very good	Very good	Very good
Outcome accuracy	Poor	Excellent	Good
Indication of effect over time	Very good	Very good	Very good
Ease of creation	Poor	Good	Very poor

Discrete-event modelling is found to be an inadequate approach for the intended modelling purposes, as this approach is better suited to model queuing systems and supply chains [38], whereas the intervention strategies for a diabetic health care system may be viewed as a continuous flow of resources and information. Furthermore, the discrete-event approach is typically focused on the on the details of a system rather than from a holistic system perspective [38], and may, therefore, be unable to identify the effects of the various intervention strategies on the system. This may not be ideal in the context of national health care, as the problem is almost always viewed at a high-level perspective. An additional shortfall of the discrete-event approach is that ease of model creation is poor [38].

System dynamics modelling, and agent-based modeling are identified the most appropriate modelling techniques to understand the effects of various intervention strategies to manage diabetic policy formulation given the requirement specifications, as summarised in Table 1. Agent-based modelling is, however, rejected for the purposes of this study, as this approach is better suited to model a system at a low level of abstraction [37, 38]. Consequently, in the case of agent-based modelling, the 'Problem identification' aspect is scored 'poor'. Although agent-based modelling is excellent at incorporating non-linearity in a system, it is significantly criticised for its poor ease of creation [37].

System dynamics modelling is, therefore, selected as the appropriate modelling approach based on the identified requirement specifications. As the South African healthcare system consists of a multitude of stakeholders, the effects of diabetic policy interventions would be more effectively modelled from a systems perspective [39, 40], therefore, scoring 'excellent' for the 'problem identification' aspect, as well as 'very good' for the 'non-linear and dynamic interactions' category. Furthermore, the system dynamics approach adapts well to varied input data and changes to interrelations, and also has the capabilities to provide the expected outcome of the system over a time period specified in order to model future predictions [39, 40]. In conclusion, system dynamics modelling will also provide a more holistic solution to understand the influencing factors of various diabetic intervention strategies in the South African healthcare context for the management of policy formulation.

6. CONCLUSION

The increasing prevalence of diabetes in the world is a widespread concern. While improvements have been made in the epidemiology and management of diabetes in the developed world, the same advances have not been made in South Africa. In addition, it was noted by The Economist that while South Africa ranks high in terms of health care spending, it is ranked significantly low in terms of health care outcomes.

The ability of the South African health care system to provide sufficient PHC has been constrained by several factors, such as the increased stress on the public health system, caused by the inequality between public and private health care, and the increased burden of disease in South Africa. Communicable diseases, such as

HIV/AIDS and TB are widespread in South Africa, but non-communicable diseases, such as diabetes, are growing in prevalence. In fact, about 7% of South Africans have diabetes, and, according to the IDF, this percentage is only expected to increase. While public and private policies have been developed to address the growing prevalence of the disease, this paper argues the importance of implementing policy and intervention strategies which address the needs of diabetics, at all levels of health care, to effectively management diabetic health care in South Africa.

When a disproportionate amount of focus is placed on the content of a policy, attention is typically diverted away from understanding the processes, context, and actors of a policy, which explains why desired policy outcomes may fail to emerge. This paper argues that policy-making should not only focus on policy content, but also policy context, processes and actors, achieved in the triangle policy analysis. By applying this analysis tool to existing South African diabetic policies, requirement specifications were developed to identify a modelling approch to investigate alternative intervention strategies for the management of diabetic policy formulation in South Africa. These requirement specifications include (i) problem identification, (ii) non-linear and dynamic interations, (iii) flexibility, (iv) outcome accuracty, and (v) indication of effect over time.

This paper argues that the South African health care system, together with diabetic policy, is a complex system, with non-linear and dynamic interactions, and the approach to analysing such a system should take the form of a simulation model. In order to select the most appropriate simulation modelling approach, three of the most commonly used simulation modelling approaches were evaluated based on their ability to meet the developed requirement specifications. System dynamics modelling was deemed the most appropriate modeling tool to provide the holistic approach to investigate the influencing factors of various diabetic intervention strategies in the South African health care context for the purpose of informing policy formulation.

In conclusion, this paper argues that there is justified need to investigate alternative intervention strategies for the management of diabetic policy formulation in South Africa, and that the most appropiate tool for investigation would be by means of a system dynmics modelling approach.

7. REFERENCES

[1] **Idemyor, V.** 2010. Diabetes in sub-Saharan Africa: Health care perspectives, challenges, and the economic, *Journal of the National Medical Association*, vol. 102, nr. 7, pp. 650-653.

[2] **International Diabetes Federation.** 2017. Diabetes Atlas. [Online]. Available: http://www.diabetesatlas.org/across.

[3] **Ottermann, B.** 2017. Prevalence of diabetes in South Africa. [Online]. Available: https://www.health24.com/Medical/.

[4] **Hall, V., Thomsen, R., Henriksen, O. and Lohse N.** 2011. Diabetes in Sub Saharan Africa 1999-2011: Epidemiology and public health implications. A systematic review, *BMC public health*, vol. 11, nr. 1, pp. 564-576.

[5] **Maillacheruvu, P. And McDuff, E.** 2014. South Africa's return to primary care: The struggles and strides of the primary health care system, *Journal of Global Health*.

[6] **Dookie, S. and Singh, S.** 2012. Primary health services at district level in South Africa, *BMC Family Practice*, vol. 13, nr. 67.

[7] **Griffin, P.M., Nembhard, H.B., DeFlitch, C.J., Bastian, N.D., Kang, H. and Muñoz, D.A.** 2016. Healthcare Systems Engineering, Hoboken: John Wiley & Sons Inc.

[8] **Statistics South Africa**. 2017. Mortality and causes of death in South Africa. [Online]. Available: http://www.statssa..

[9] **Health Policy Project.** 2016. Health financing profile: South Africa. [Online]. Available: https://www.healthpolicyproject.com/pubs/7887/ SouthAfrica_HFP.pdf.

[10] **Statistics South Africa**. 2017. Mortality and causes of death in South Africa. [Online]. Available: http://www.statssa.gov.za/publications/P03093/P030932015.pdf.

[11] **Keeton, C.** 2010. Bridging the Gap in South Africa, *Bulletin of the World Health Organization,*, vol. 88, nr. 1, pp. 803-804.

[12] **Wadee, H.** 2003. Health Care Inequity in South Africa and the Public/Private Mix, *Unpublished*. School of Public Health, University of Witwatersrand.

[13] **Swanson, E.** 2010. World development indicators 2007, *The World Bank*.

[14] **Mojaki, M., Letskokgohka, M. and Govender, M.** 2011. Referral steps in district health system are side-stepped, *South African Medical Journal*, vol. 101, nr. 2, pp. 109-110.

[15] **Goudge, J., Gilson, L., Russell, S., Gumede, T. and Mills, A.** 2009. Affordability, availability and acceptability barriers to health care for the chronically ill: Longitudinal case studies from South Africa, *BMC health services research*, vol. 9, nr. 1, pp. 75-102.

[16] **Rutkove, S.** 1990. Patterns of care in an overburdened tertiary hospital outpatients department, *South African medical journal*, vol. 77, nr. 9, pp. 476-478.

[17] **Mayosi, B., Sitas, F., Tollman, S. and Bradshaw, D.** 2009. The burden of non-communicable diseases in South Africa, *The Lancet*, vol. 374, nr. 9693, pp. 934-947.

[18] **Khumalo, T.** 2014. South Africa: TB Is Number One Killer. [Online]. Available: https://www.voanews.com/a/tb-is-number-one-killer-in-south-africa/1876553.html.

[19] **South African Department of Health.** 2004. The South African National Tuberculosis Control Programme Practical Guidelines, *South African Department of Health*, Pretoria.

[20] **Business Tech**. 2017. 10 biggest killers in South Africa. [Online]. Available: https://businesstech.co.za/news/lifestyle/160957/10-biggest-killers-in-south-africa/.

[21] **News 24.** 2017. Diabetes is now number one killer in the Western Cape. [Online]. Available: https://www.news24.com/SouthAfrica/Local/City-Vision/diabetes-is-now-number-one-killer-in-the-western-cape-20170405

[22] **Jeon, C. and Murray, M.** 2008. Diabetes mellitus increases the risk of active tuberculosis: a systematic review of 13 observational studies, *PLoS Med*, vol. 5, nr. 7, p. 152.

[23] **Dooley, K. and Chaisson, R.** 2009. Tuberculosis and diabetes mellitus: convergence of two epidemics, *Lancet Infect Dis*, vol. 9, nr. 12, pp. 737-746.

[24] **Manyema, M.** 2015. Decreasing the burdenof type 2 diabetes in South Africa: The impact of taxing sugar-sweetened beverages, *PLoS One*, vol. 10, nr. 11, pp. 120-137.

[25] **Rotchford, A. and Rotchford, K.** 2002. Diabetes in rural South Africa - An assessment of care and complications, *South African Medical Journal*, vol. 92, nr. 7, pp. 536-541.

[26] **Web MD**. 2018. Types of diabetes mellitus. [Online]. Available: https://www.webmd.com/diabetes/guide/types-of-diabetes-mellitus#1.

[27] **The Society for Endocrinology, Metabolism and Diabetes of South Africa**. 2017. The 2017 SEMDSA Guideline for the Management of Type 2 Diabetes, *Journal of Endocrinology, Metabolism and Diabetes of South Africa*, vol. 21, nr. 1, pp. 1-196.

[28] **Centre for Diabetes and Endocrinology**. 2018. About CDE. [Online]. Available: http://www.cdediabetes.co.za/home/about-cde.html.

[29] **Murrel, D.** 2017. Type 2 diabetes statistics: Facts and trends. [Online]. Available: https://www.medicalnewstoday..

[30] **Diabetes South Africa**. 2018. About Diabetes South Africa. [Online]. Available: https://www.diabetessa.org.za/about-us/.

[31] **South African Department of Health**. 2014. Management of type 2 diabetes in adults at primary care level, *South African Department of Health*, Pretoria.

[32] **South African Department of Health**. 2013. Strategic plan for the prevention and control of non-communicable diseases 2013-17, *South African Department of Health*, Pretoria.

[33] **Kleinert, S. and Horton, R.** 2009. South Africa's health: departing for a better future?, *The Lancet* , vol. 374, nr. 9692, pp. 759-760.

[34] **Walt, G., Murray, S. and Gilson, L.** 2008. Doing health policy analysis: Methodological and conceptual, *Health policy and planning*, vol. 23, nr. 5, pp. 308-317.

[35] **Walt, G. and Gilson, L.** 1994. Reforming the health sector in developing countries: the central role of policy analysis, *Health policy and planning*, vol. 9, nr. 4, pp. 353-370.

[36] **Reich, M.** 1995. The politics of health sector reform in developing countries: three cases of pharmaceutical policy, *Health policy*, vol. 32, nr. 1-3, pp. 47-77, 1995.

[37] **Banks, J.** 1999. Introduction to simulation, in *Proceedings of the Winter Simulation Conference*, Phoenix.

[38] **Borshchev, A.** 2013. The big book of simulation modeling: Multimethod modeling with AnyLogic 6, Lisle: AnyLogic North America.

[39] **Maani, K. and Cavana, R.** 2007. Systems thinking, system dynamics: Managing change and complexity, *Pearson Education*.

[40] **Sterman, J.** 2000. Business dynamics: systems thinking and modeling for a complex world, *Mc Graw-Hill Companies, Inc.*, Boston.

DEVELOPING A BY-PRODUCT GAS SIMULATION MODEL TO ESTIMATE SYSTEM IMPROVEMENT INITIATIVES IN IRON AND STEEL MANUFACTURING

A. Ludick[1], J.H. Marais[2] & W.J.J. Breytenbach[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
aludick@researchtoolbox.com
[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
jhmarais@researchtoolbox.com
[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbreytenbach@researchtoolbox.com

ABSTRACT

High production costs and low steel demands recently placed the South African steelmaking industry under severe pressure. The industry is forced to evaluate and improve its current systems to improve profitability. The dated sector can relieve stress by implementing technologies of the present age. This paper proposes a simulation model that can be used to test system improvement initiatives on the gas network. The iron and steel making process produces combustible gases. These gases are recovered as by-products and used as fuel gas in combination with natural gas and Heavy Fuel Oil (HFO). By-product gases are used to generate heat typically for furnaces or to generate additional electricity through an on-site power generation plant. This gas distribution network is kept safe by flaring excess by-product gas. The proposed simulation model was verified on a case study highlighting the difficulties in predicting by-product gas behaviour as well as the benefit that a gas distribution simulation model can have.

This work was sponsored by ETA Operations (Pty) Ltd

[1] Alexander Ludick holds a Masters in Mechanical Engineering from the North-West University.

[2] Dr Johan Marais is a registered professional engineer and holds a PhD in Electrical Engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr Wynand Breytenbach holds a PhD in Engineering Development and Management from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

High product cost and low demands have placed the South African iron and steel manufacturing industry under severe pressure [1]. The industry is forced to re-evaluate current production strategies [2]. A failure in reducing production costs may force the South African industry to close down. The South African government supports the iron and steel industry by safeguarding the local steel market from cheaper imports from countries like China [3]. Governmental support can only help to a limited extent. It is the sector's responsibility to regain its position in the international market.

The steelmaking process starts with iron production in a blast furnace. Raw materials such as iron ore, coke and limestone are melted together with a constant supply of heated air within the blast furnace [4]. Ladle torpedoes transport liquid iron from the blast furnace to the steel plant after the tapping the blast furnace. Steel is produced from iron by either the Basic Oxygen Steelmaking (BOS) or Electric Arc Furnace (EAF) process [5]. In the BOS process, a Basic Oxygen Furnace (BOF) is used as opposed to an EAF [6]. The molten steel's composition is modified to the desired grade by stirring, a ladle furnace or ladle injection [7]-[9]. Finally, the steel is solidified in either slabs, blooms or billets by continues casters and rolled into different shapes at the mills [10].

The steelmaking process produces several off gases. Some of these gases are combustible and used as a by-product [11]. Throughout the steelmaking process, combusted by-product gas supplies heat where required. Three by-product gases are typically recovered and used in the industry. These gases include (1) Coke Oven Gas, (2) Blast Furnace Gas, and (3) Basic Oxygen Steelmaking Gas (BOSG). COG is a by-product from the coke making process, BFG from the iron making process and BOSG the steel making process [11]-[13]. Not all iron and steel manufacturing facilities reclaim these gases. Each facility's gas network is unique to the infrastructure. The cost of reclaiming and maintaining the by-product gas network needs to be less than then the potential profit.

Figure 1 illustrates the gas network on a typical iron and steel manufacturing facility. By-product gases are produced and recovered at their respective processes. Each by-product gas is generally integrated with a gas holder and flare stack. Gas holders provide the ability to store gas while maintaining a constant pressure in the gas network. Flare stacks relieve the gas network of pressure in the case of excess gas. An intricate distribution network connects the by-product gas to several plants which require gas for production. Excess by-product gas generates additional steam for electricity [12]. Natural gas and Heavy Fuel Oil (HFO) substitute by-product gases in case of by-product gas shortages [14], [15].

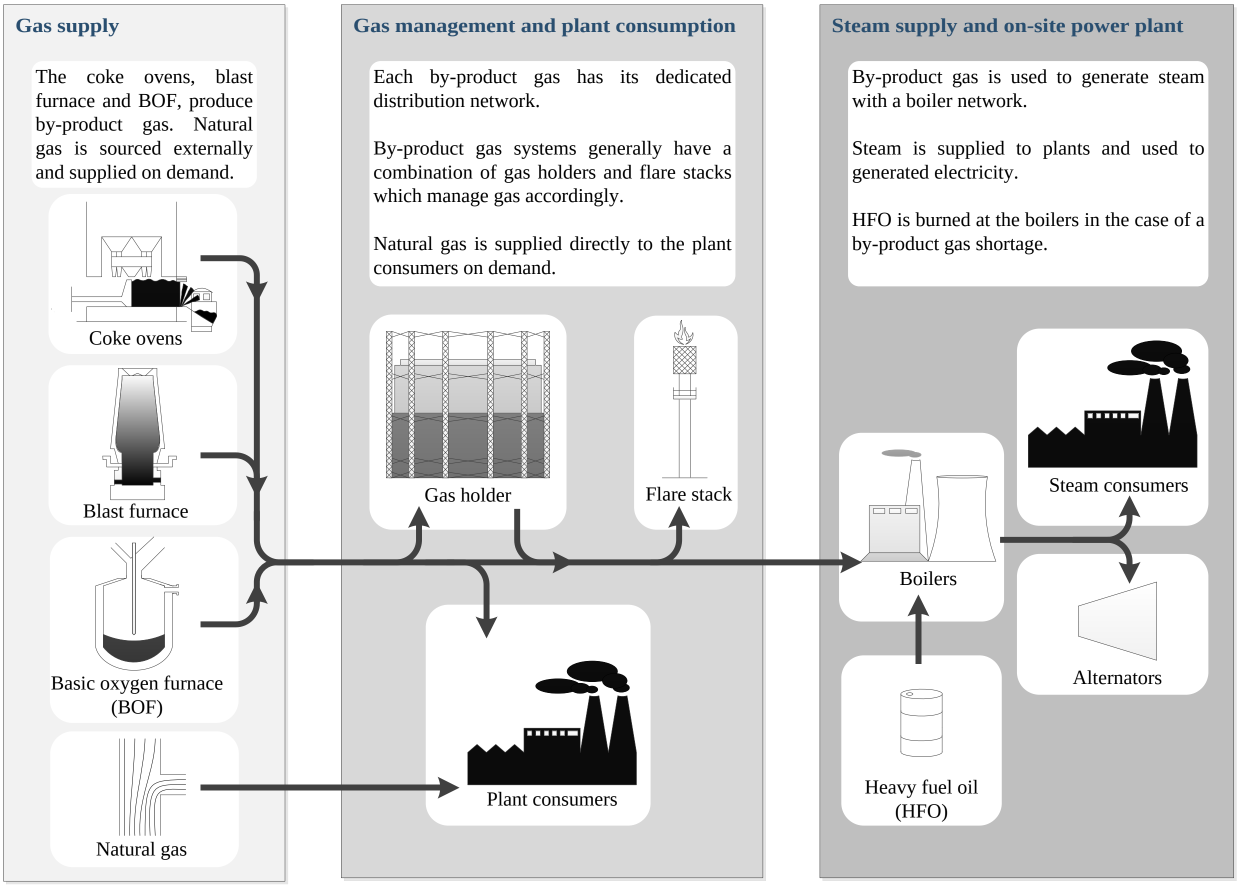


Figure 1: Iron and steel manufacturing holistic gas network

2. PROBLEM STATEMENT

By-product gases, if utilised efficiently, can save a steelmaking facility significantly. By-product gas utilisation reduces natural gas and HFO consumption as well as electricity purchased from an external supplier such as Eskom. Improved utilisation results in less gas flaring thus a reduction in CO_2 emissions which will soon be taxed by the South African government. It is difficult to utilise by-product gas effectively. Imbalances between by-product gas production and demand exist daily. These imbalances exist due to manufacturing variations, equipment failures, shutdowns and control discontinuities.

As an example, a steelmaking facility which produces steel using the BOS process with the manufacturing capacity of a million tonnes of steel annually can provide by-product gas of R60 000/h. The high value of by-product gases has led to much research in cost-saving initiatives. Table 1 is a summary of research related to system improvements on the gas control. The gas network is integrated with most of the plants on a steelmaking facility. System improvements on the gas network can therefore be done on most of the plants.

Table 1: Summary of work done of systems regarding gas

Section	Energy saving initiative	Study
Sinter and pelletizing	Use of waste fuels in the sinter plant, selective waste gas recycling and low emission and energy optimised	[16]-[18]
Coke making process	Programmed heating, automation and process control system and COG recovery	[16], [19], [20]
Blast furnace process	Recovery of BFG, hot blast stove automation	[16], [21], [22]
BOF process	BOFG recovery, BOFG sensible heat recovery	[19], [23], [24]
EAF process	EAF gas waste heat recovery	[25]
Casting process	N/A	-
Rolling process	Regenerative burners for reheating furnaces	[26], [27]
General	Energy monitoring and management system, high efficiency gas separation plant, gas turbine pilot fuel alternative technologies, by-product gas boiler power generation technology	[12], [28]-[31]

Most of the work presented in Table 1 require highly specialised skills and CAPEX. Both of which are currently lacking in the South African industry. The South African iron and steel industry has recently lost specialised knowledge due to a combination of high personnel turnover rates and inadequate and partial handovers. Furthermore, the industry has not seen much changes the last two decades which mean that the general South African steel manufacturer is running a dated facility. Dated machinery and equipment forces conservative operating conditions to avoid unnecessary strain.

The shortage of CAPEX, specialised skills and the deteriorated condition of overall plants place the industry in a difficult position to regain competitiveness. This paper focuses on a gas network simulation method which enables the steelmaking industry to simulate suggested system improvements on the gas network. The simulation method serves as a platform to test the effectiveness of system improvements and alterations without actual implementation. System upgrades can be reiterated, optimising the system upgrade before actual implemented to minimise implementation risk.

3. METHOD

There are many simulation software packages on the market as well as companies that specialises in plant simulations. It might be the preferred solution to contract the simulation model out to a specialised company. It is however, expensive and can be timely. Another disadvantage is that the plant personal does not gain a fundamental understanding and insight that comes with developing a simulation model. The simulation method in this paper will benefit the steelmaking facility optimally if the plant personal develops it. The model can be developed in Excel, VBA, Python, Matlab or any coding language with mathematical functions.

The simulation model consists of two sections. The first is gas production and consumptions predictions and the second is the distribution of by-product gas between these plants. The production and consumption prediction model provides insight on the by-product gas availability whereas the distribution model provides insight on utilisation. A production and consumption prediction model on its own can be used to predict by-product gas availability and schedule maintenance on the plants accordingly. A distribution model can be used to simulate isolated improvements or alterations on the distribution network.

These models were developed based on the current condition of the South African steelmaking industry. The industry currently functions on dated machinery which was modified over the years. Repairs, maintenance and modifications are typically done when plants are forced rather than preventive maintenance schedules. Machinery does not perform according to the commissioned specifications due to their dated and modified conditions.

The simulation model incorporates the dated condition of the machinery by using actual and historical plant data. This data provides an indication of the machinery's condition at the time that the data was captured. It supplies the model with actual availability and production insights.

4. SIMULATION MODEL

4.1 By-product gas availability prediction

COG, BFG and BOSG production are proportional to either coke, iron and steel production. The proportion is dependent on the plant's efficiency and the gas recovering system. By-product gases differ in quality which is dependent on how the relevant plants are producing their products. Gas quality is measured by energy density, Wobbe Index (WI), and how clean the gas is. Only the energy density is relevant for the prediction model. Wobbe analysers continually measure the WI of by-product gas. The by-product gas in energy is the product of the coke, iron or steel production, a scaling factor and the WI (Equation 1).

$$Q_x = A_x M_x WI_x \qquad \textbf{Equation 1}$$

$Q = gas\ energy\ [J]$
$A = scaling\ factor\ [m^3/tonne]$
$M = production\ (tonne)$
$WI = wobbe\ index\ (J/m^3)$
$x = iron, coke, steel$

Steel manufacturing facilities will typically have a planning schedule. Such a planning schedule will indicate the day to day production rates of each plant to reach the overall facility's production targets. There is generally a deviation between planned and actual production values. The difference between planned and actual production is typically constant from day to day. The proposed prediction model incorporates a regression between the planned production and the actual production. The regression will provide a function to predict the daily production.

The standard error of the sampling distribution of the regression model should be used as an indicator of the model's prediction reliability. The number of data points can adjust the fit of the regression model. A linear regression model should in most cases be sufficient. Alternatively, a combination of regression models that use daily, weekly, monthly and yearly values can be used. The interaction that these models have with one another can be used as a compounded result.

Predictions are based on the regression models in combination with the correlation between by-product gas and plant production. This model predicts how much by-product gas will be produced on a daily period. It is not useful if the by-product gas requirement of the plant is still unknown. The requirement can be calculated similarly. For example, by-product gas consumed by the reheating furnaces is proportional to the steel rolled at the mills. Similar to the by-product gas consumed at any other plant. Thus, the required by-product gas can be calculated using regression models between the plants' forecasts and actual production.

4.2 By-product gas distribution network

The by-product gas distribution network can be simplified to gas production, gas consumption, gas holders utilisation, gas flared and steam and electricity generation. By-product gas which is not consumed is stored in the gas holders. The gas is flared if the gas holder reaches the specified maximum limits. The interaction between by-product gas availability, gas holder utilisation, flaring and electricity generation is simulated in the distribution network simulation model.

Surplus by-product gas is stored in the gas holders. Once the gas holder reaches its specified maximum limit, it will automatically start to flare excess gas. The specified maximum gas holder level is ±85% to ensure safe operation. The gas holder level indicates by-product gas availability at any instance. If the gas holder level increases, then there is a surplus of by-product gas being produced and if the gas holder level decreases there is a shortage of by-product gas. These imbalances are the challenges that the operators face daily.

Figure 2 illustrates each component of the gas distribution simulation model. The simulation model starts by initialising the required variables. These variables include the system operating limits, gas energy density, gas rates, boilers and steam turbine and electricity generation efficiencies. The operating limits ensure that the system is operated in an appropriate and safe manner, for example, the gas holders' capacity and limits, electricity generation limits, steam turbine and generation ramp up and ramp down rates. Any parameter which ensures that the systems are operated in a safe condition.

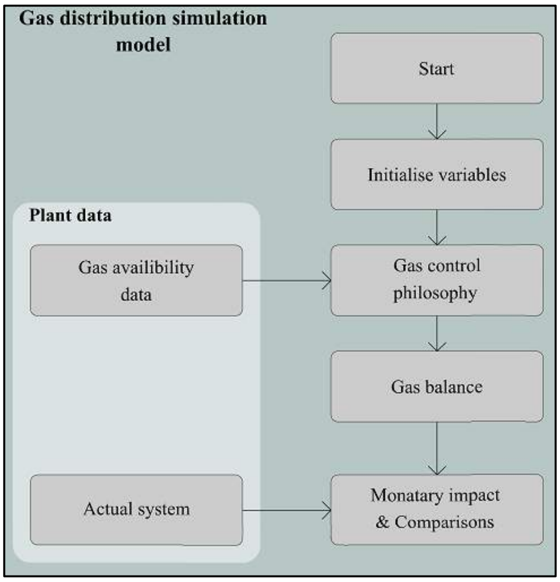


Figure 2: Gas distribution simulation model flow chart

The gas energy density measured by a Wobbe analyser should be as recent as possible due to the gas quality deviations which can have a significant influence on the electricity generation per volume gas. The gas rates provide the ability to measure the simulated impact on the facility relevant to the way the plant is controlled or different iteration of the gas control philosophy. The boilers' efficiency is used to determine steam generation from gas which is used to convert to electricity generation with the steam turbine and generator efficiency.

The initiated variables are fed with the actual gas availability data into the simulated control philosophy. Figure 3 illustrates a flow chart of the simulated time steps within the control philosophy component. For the first time step the initial conditions are used. The actual gas availability data of the following time step is added to the initial conditions. This data includes gas holder levels, gas flared, and electricity generated which are converted back to thermal energy.

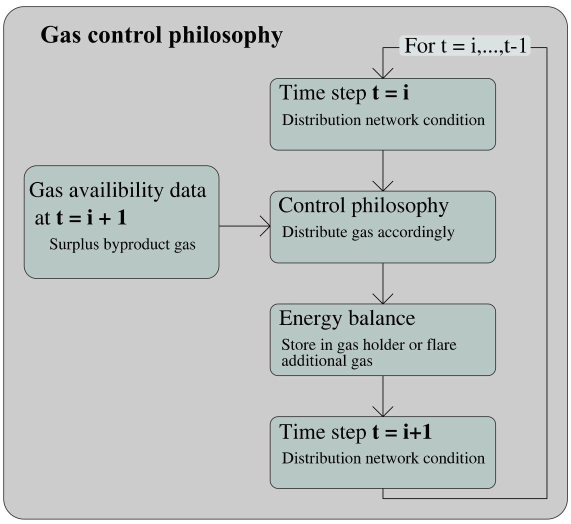


Figure 3: Gas control philosophy flow chart

Figure 4 illustrates the control philosophy component from Figure 3 in detail. The total thermal energy is calculated for each time step from the plant data. This includes the gasholder level, gas flared and electricity generation. The thermal energy is then distributed according to the control philosophy and the initial energy distribution. Once the energy is distributed then it is converted to either the gas holder level, gas flared or electricity generated.

An energy balance ensuring the conservation of energy is done within each simulated time step. This concludes the initial simulated time step. For the rest of the time steps the same process is repeated however the previous time step's distribution is used rather than the initial conditions. The control philosophy logic is the system improvement's impact simulated for. Different iterations and versions can be compared with one another using classes for version comparisons.

The time increments of the simulation model should be small enough to accurately capture the physics of the system. Thus, the driving factor for the time increments is the tempo by which by-product gas can be produced. This tempo is generally high and requires a time step of 30 seconds or less. Another gas balance needs to be completed after all the time steps are simulated. It is critical that no energy is lost, or additional energy gained within the simulated time. The sum of the energy flared, stored in the gas holder and electricity generated should be equal to energy into the system over the entire period.

The final step is to include a method to measure the impact of the simulated gas philosophy. This model can include the actual monetary impact using the gas rates. As long as a definite comparison can be made between the simulated control philosophy and either the actual data or a previous iteration of the gas control. This part allows for proving safe operation and a method to find the ideal control before actual implementation.

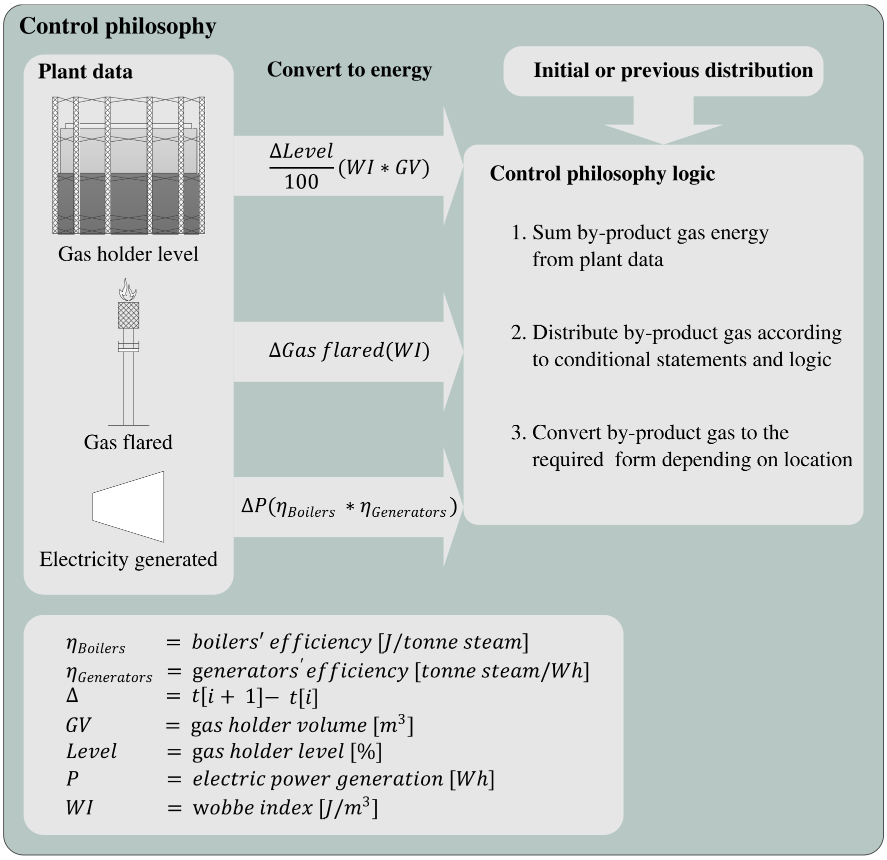


Figure 4: Control philosophy flow chart

5. CASE STUDY

5.1 Production and consumption prediction correlations

A case study was performed to illustrate the effectiveness of the proposed simulation model on a South African steelmaking facility. Both the production and consumption prediction models and by-product gas distribution model were implemented. The facility produces iron and steel with a blast furnace and BOF. On the facility, only COG and BFG is reclaimed and distributed to various consumers as by-product gas. The facility is extensively trying to reduce production costs to regain a competitive position in the market.

The model was constructed using rolling yearly, month and weekly linear regression models. A daily correlation model was also used. The model calculated a COG and BFG production prediction based on the planned iron and coke production. Similarly, the model estimated COG and BFG consumption prediction based on the expected production of the plants which consumes the gases for production. The surplus is the difference between the gas produced and consumed. Beneficial provisions can be made by knowing the surplus gas availability.

The COG and BFG surplus predictions are illustrated in Figure 5 and Figure 6. A forecast for one month was made based on the yearly, monthly, weekly and daily models. The actual surplus by-product gas was superimposed alongside the predictions for comparisons. The yearly prediction performed the best in both the COG and BFG cases. Similarly, the daily prediction model had the poorest performance in both cases. The prediction models struggled to predict the variances found in the actual data.

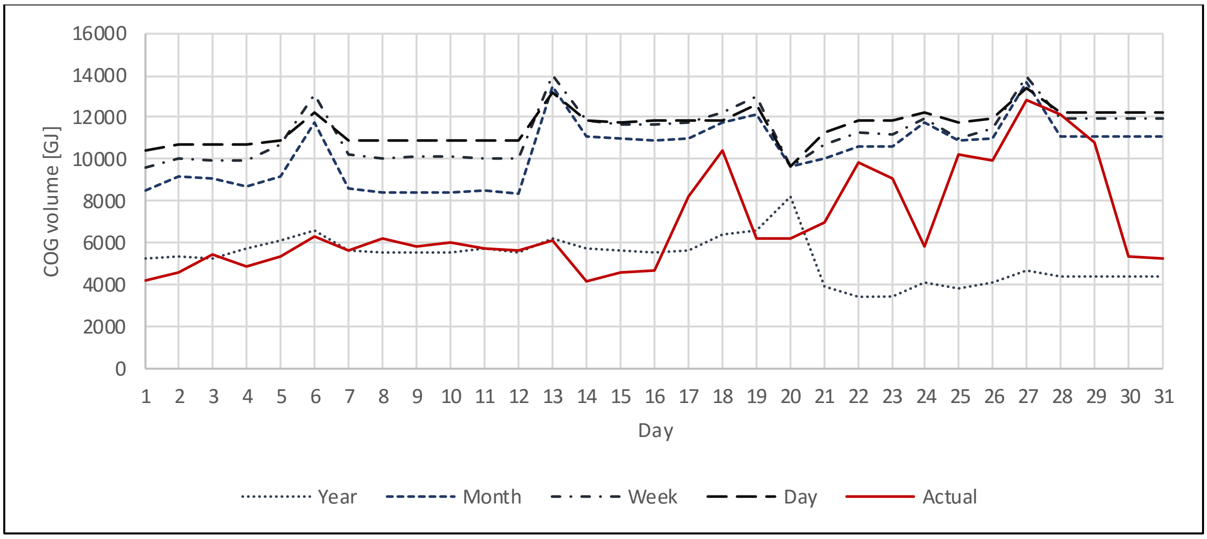


Figure 5: Yearly, monthly, weekly and daily surplus COG prediction vs actual

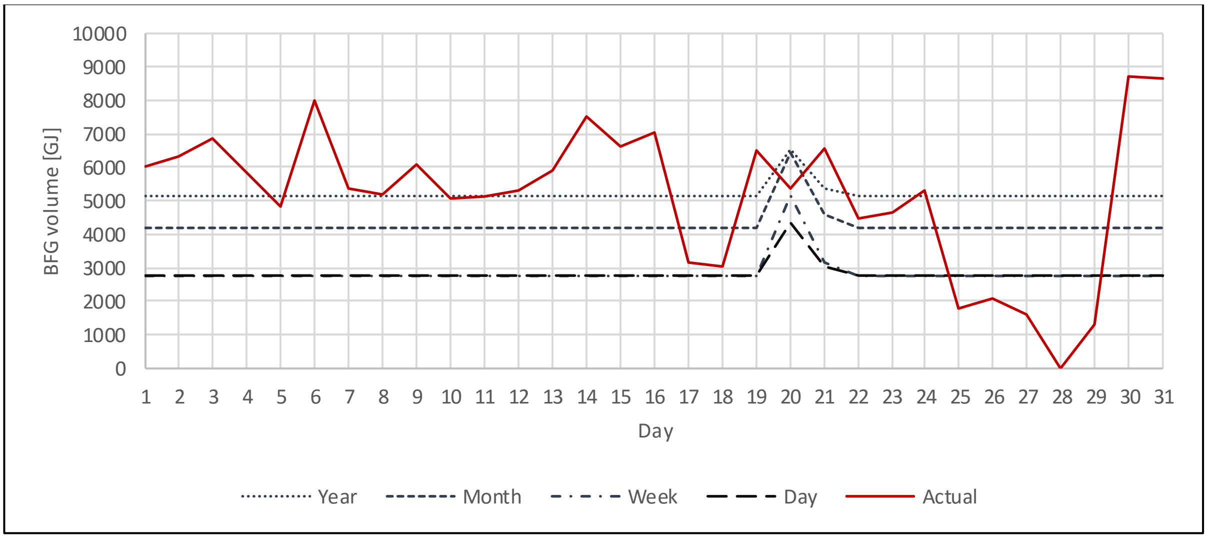


Figure 6: Yearly, monthly, weekly and daily surplus BFG prediction vs actual

The prediction models average daily error vs average daily Std Dev. are illustrated in Figure 7. The error was calculated as the percentage difference between the actual surplus gas and the prediction. A reasonable prediction would result in a low error and Std Dev. A high error indicates that the model is generally inaccurate whereas a high Std Dev suggests that the prediction model struggles to predict day to day variances.

The yearly COG prediction model had the lowest error and Std Dev. combination. The rest of the COG prediction models performed in the same regions with relatively high errors and Std Dev. In the BFG prediction models, the daily and weekly models performed the best. However, the other prediction models did not differ much regarding the error.

The Std Dev. were much more scattered than the COG. For example, the yearly BFG model's daily error was slightly lower than the daily and weekly's but had a scientifically higher Std Dev. The COG prediction varied more than the BFG predictions. However, the actual day to day variances was much more than both prediction models. This is highlighted in Std Dev. of the models.

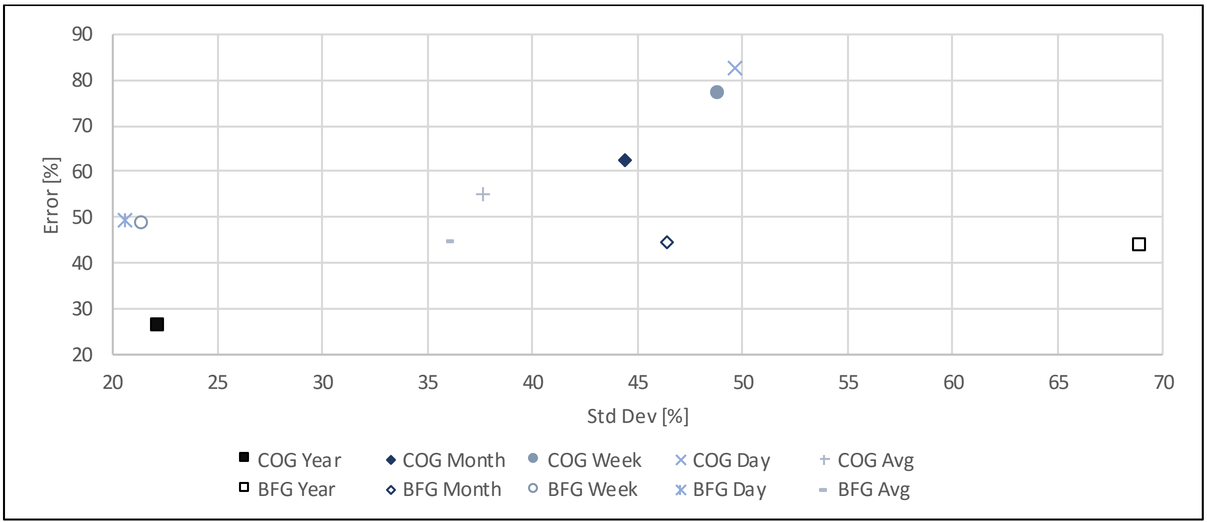


Figure 7: By-product gas prediction models' error vs Std Dev

5.2 By-product gas distribution system improvements

The gas distribution simulation model was implemented on the same steelmaking facility. The facility has two gas holders, one for the BFG and the other for COG. An improved BFG surplus gas control philosophy was developed in correspondence with the plant personnel. This control philosophy was designed with operation parameters and safety constraints. The plant did not have an automatic controller regulating the surplus by-product gas. Instead, excess by-product gas was controlled manually. The goal of the new control philosophy was to improve the facility's actual electricity generation by reducing the by-product gas flaring. A shift from passive to active control would allow for increased electricity generation.

The simulation was developed to test and possibly refine the controller on the actual system, without the danger of damaging any infrastructure. Actual system characteristics were incorporated into the simulation. The simulation consisted of four steps namely, variables initiation, gas control philosophy, gas balance and impact comparisons. The impact comparisons compared different iterations of the control to provide the safest control philosophy with the best improvement impact.

Figure 8, Figure 9 and Figure 10 illustrates how the simulated control philosophy performed against the manual control. In manual operation, the constant variances in the BFG availability caused the operator to generate electricity conservatively. This is to ensure that the gas holder remains at safe operating levels preventing the holder to drop too low too fast. The operators testified that they have too much to focus on to actively distribute surplus gas. They would instead generate less additional electricity which provides them with a buffer should the gas production unpredictably drop or consumption increase.

The simulated controller does control actively and not passively. Active control has the advantage of acting instantaneously providing safe control at a less conservative level. Any increase in the BFG availability is utilised by generating more steam for additional electricity. Similarly, the electricity generation would instantaneously decrease should the surplus gas decrease.

The simulation model displays the same dynamics of the actual system. Both the actual and simulation models start on the same instance, gas holder level, electricity generation and gas being flared. In the cases where the gas holder level increased so does the simulated electricity generation resulting in the simulated gas holder level increasing slower than the actual. In the cases where the actual gas holder reaches the flaring set point, the simulated gas holder generally reaches it later. It is due to the simulated control that generates more electricity causing the holder level to be operated on a lower average level.

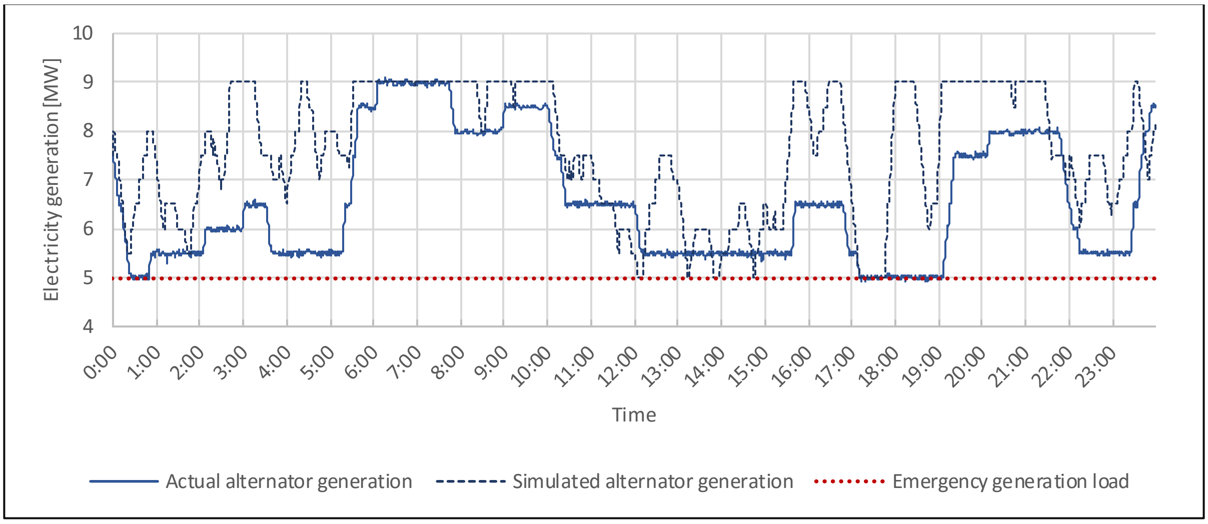


Figure 8: Improved simulated control philosophy vs actual electricity generated by facility

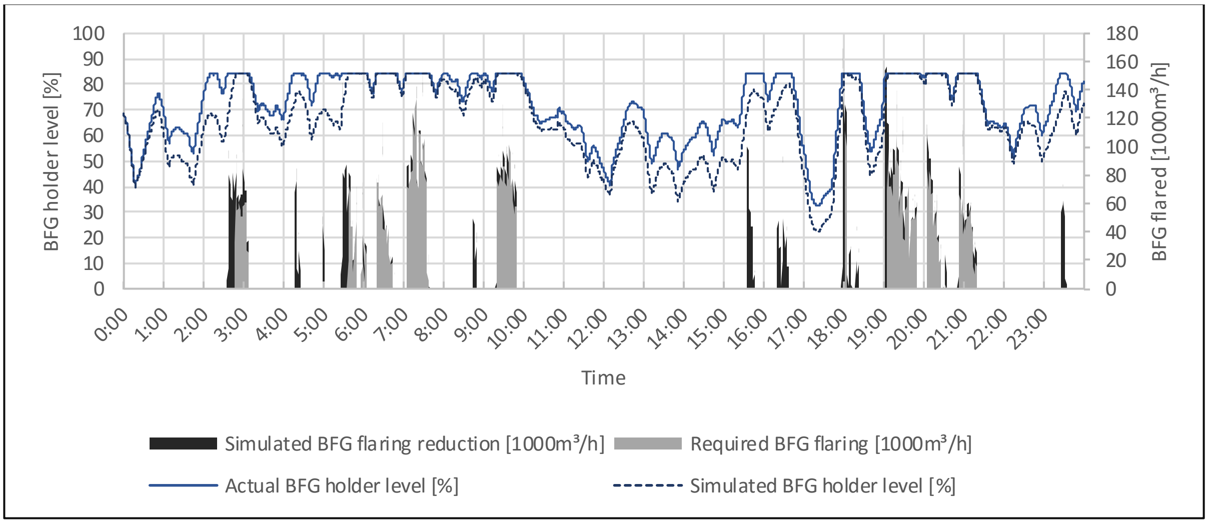


Figure 9: Improved simulated control philosophy vs actual gas holder utilisation and BFG gas flaring

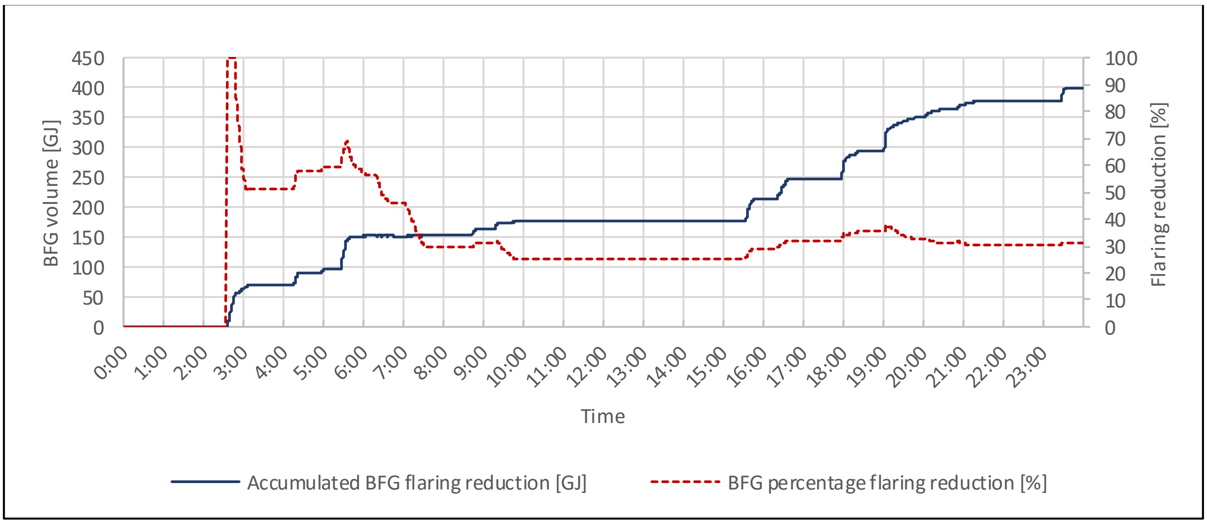


Figure 10: Improved simulated control philosophy flaring reduction compared to actual control

The simulation model proves that the prosed control can control the electricity generation, boilers and gas holders within safe operating parameters. The gas holder did not drop below 20 % or go above 85 %. Any additional gas was flared keeping the gas holder below 85 %. A minimum of 5 MW electricity was generated at all times which is the facility's emergency load. No more than the electricity generator's maximum capacity were reached. An energy balance between the actual and simulated model was done to prove that no energy was added or removed by the simulation model.

The steelmaking facility was satisfied with the simulated control philosophy. Forecasting proved possible flaring reductions of ± 400 GJ per day. Permission was granted to implement the control into SCADA system. The implementation period was relatively fast due to the control being simulated which required only a syntax change to the SCADA. Additionally, the new insights gained by the simulation model allowed the plant personal more insight in the gas holder utilisation and electricity generation dynamics.

6. RESULTS

The consumption and production prediction models proved that at best a crude prediction was made of the surplus by-product gas availability. Both the generally high average daily error and Std Dev. of the models indicated that there is room for improvements. Variations in the actual surplus gas that was missed in the prediction model can be the result of unreliable machinery. This is due to the model using planned production, the variations in the actual data mean that the planned production could not consistently be achieved.

The system of by-product gases in the steelmaking process is complicated. Both the enormity and deteriorated condition of the system are among the factors that make prediction models challenging to construct. The daily and weekly models will perhaps be more suitable for short-term predictions as it provides the most recent system condition. The model can be used as is for crude indications of surplus gas availability. For example to plan maintenance stops depending on gas availability potentially reducing natural gas consumption. Future work could entail combining the different prediction models for a more comprehensive forecast.

It was proved that the by-product gas distribution simulation model can be used to test system improvements before actual implementation. The dynamic behaviour of the actual system was accurately reproduced. Which allowed the facility to implement a system improvement knowing that it is safe and what the actual impact would be before implementation. The simulation proved that the system improvement has the potential to save in the order of 400 GJ of thermal energy per day. This translates to 25 MWh electricity which is approximately R20 000 cost reduction at a South African steelmaking facility.

Actual plant data was analysed to prove the effectiveness after the by-product gas distribution model was implemented. The model was challenging to measure due to many different systems influencing the plant gas consumption. These systems caused noise making it difficult to calculate the actual result that the by-product gas distribution model had. Thus, 5 different baseline models were constructed, 4 of the 5 models were in a band of R400 000 with two of the models, 1 and 4, in the middle of the band. Baseline model 1 and 4 proves a saving of R800 000 over 60 days which is R13 500 per day. Figure 11 illustrated the accumulated monetary benefit of each baseline model and the actual utilisation of the gas holder over 60 consecutive days.

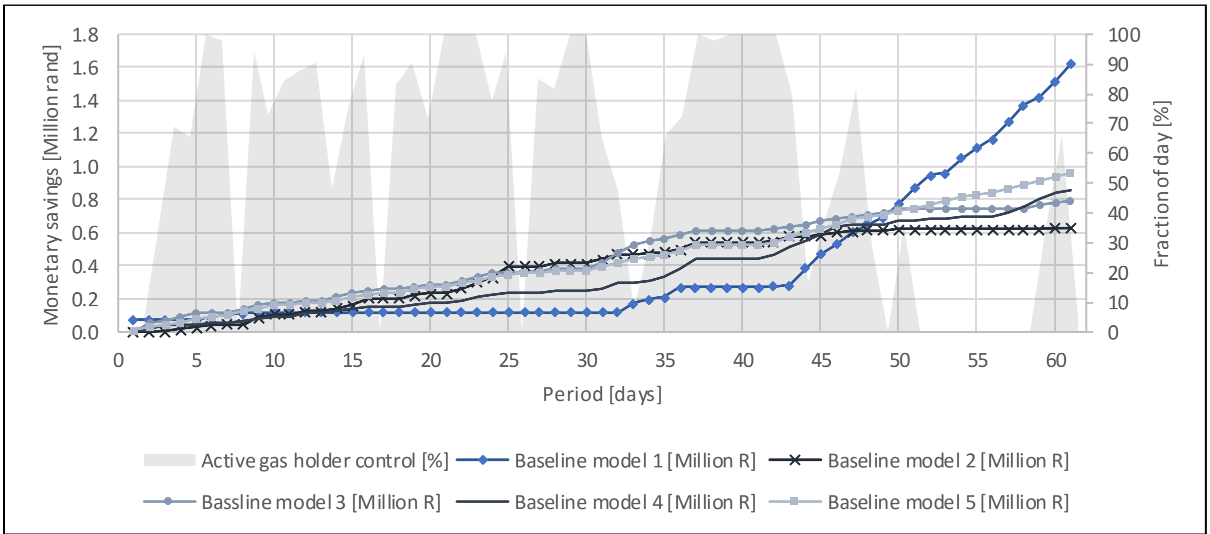


Figure 11: Accumulated monetary by-product gas distribution model benefit

A comprehensive simulation model can be developed by coupling the prediction and distribution model. The coupling would provide a predicted gas availability input for the distribution model. This type of model would be beneficial for forecasting monthly gas utilisation. Therefore, allowing predictive gas distribution modifications for optimal energy utilisation. The prediction model needs to be more reliable and have a higher prediction resolution.

7. CONCLUSION

Simulating by-product gas behaviour in the South African steelmaking industry is complicated due to the enormity of the systems and the dated condition of the equipment. A gas network simulation model can provide steelmaking facilities with a platform to develop, test and expand system improvements without putting fragile and dated machinery at risk. The proposed model proved that predicting by-product gas surplus availabilities are difficult and complicated. Results from the case study demonstrated that system improvements on the gas distribution network could save the facility around R20 000 a day. Actual results proved a saving of R13 500 per day thus justifying the need for an accurate and reliable simulation platform to test, promote and validate similar system improvements.

8. REFERENCES

[1] K. Cloete, "South Africa not surrendering its steel industry, says Davies," *Engineering news*, 2016. [Online]. Available: http://www.engineeringnews.co.za/article/south-africa-not-surrendering-its-steel-industry-says-davies-2016-08-23. [Accessed: 05-Oct-2017].

[2] L. Steyn, "SA's steel industry on brink of collapse," *Mail & Guardian*, 2016. [Online]. Available: https://mg.co.za/article/2016-04-07-sas-steel-industry-on-brink-of-collapse. [Accessed: 12-Sep-2017].

[3] S. Njobeni, "Duty to protect SA steel industry," *Business Report*, 2017. [Online]. Available: https://www.iol.co.za/business-report/companies/duty-to-protect-sa-steel-industry-9021440. [Accessed: 04-Oct-2017].

[4] D. Fu, G. Tang, Y. Zhao, J. D'Alessio, and C. Q. Zhou, "Modeling of iron ore reactions in blast furnace," *Int. J. Heat Mass Transf.*, vol. 103, pp. 77–86, 2016.

[5] R. Hu and C. Zhang, "Discussion on energy conservation strategies for steel industry: Based on a Chinese firm," *J. Clean. Prod.*, vol. 166, pp. 66–80, 2017.

[6] P. J. W. K. De Buzin, N. C. Heck, and A. C. F. Vilela, "EAF dust: An overview on the influences of physical, chemical and mineral features in its recycling and waste incorporation routes," *J. Mater. Res. Technol.*, vol. 6, no. 2, pp. 194–202, 2017.

[7] I. Ahmad, M. Kano, S. Hasebe, H. Kitada, and N. Murata, "Gray-box modeling for prediction and control of molten steel temperature in tundish," *J. Process Control*, vol. 24, no. 4, pp. 375–382, 2014.

[8] X. Wang, M. You, Z. Mao, and P. Yuan, "Tree-structure ensemble general regression neural networks applied to predict the molten steel temperature in Ladle Furnace," *Adv. Eng. Informatics*, vol. 30, no. 3, pp. 368–375, 2016.

[9] T. Kulju, S. Ollila, R. L. Keiski, and E. Muurinen, "Validation of three-phase CAS-OB CFD-model," *IFAC-PapersOnLine*, vol. 48, no. 17, pp. 1–5, 2015.

[10] S. K. Choudhary, S. Ganguly, A. Sengupta, and V. Sharma, "Solidification morphology and segregation in continuously cast steel slab," *J. Mater. Process. Technol.*, vol. 243, pp. 312–321, 2017.

[11] S. Nomura and T. Nakagawa, "The yield of coke oven gas from hard and semi-soft coking coals," *Int. J. Coal Geol.*, vol. 168, pp. 179–185, 2016.

[12] V. B. de Oliveira Junior, J. G. C. Pena, and J. L. F. Salles, "An improved plant-wide multiperiod optimization model of a byproduct gas supply system in the iron and steel-making process," *Appl. Energy*, vol. 164, pp. 462–474, 2016.

[13] J. Yang, J. Cai, W. Sun, and J. Huang, "Optimal allocation of surplus gas and suitable capacity for buffer users in steel plant," *Appl. Therm. Eng.*, vol. 115, pp. 586–596, 2017.

[14] X. Zhao, H. Bai, Q. Shi, X. Lu, and Z. Zhang, "Optimal scheduling of a byproduct gas system in a steel plant considering time-of-use electricity pricing," *Appl. Energy*, vol. 195, pp. 100–113, 2017.

[15] M. Kovačič and B. Šarler, "Genetic programming prediction of the natural gas consumption in a steel plant," *Energy*, vol. 66, pp. 273–284, 2014.

[16] United States Environmental Protection Agency, "Available and emerging technologies for reducing greenhouse gas emissions from the iron and steel industry," 2012.

[17] J. Reidetschlaege, H. Stiasny, S. Hoetzinger, and A. Fulgencio, "Selective waste gas recirculation system for sintering plants," *Stahl und Eisen*, vol. 132, pp. 25–30, 2012.

[18] R. Remus, S. Roudier, M. a. Aguado Monsonet, and L. D. Sancho, "JRC reference report: Best Available Techniques (BAT) reference document for iron and steel production," 2013.

[19] Lawrence Berkeley National Laboratory, *The State-of-the-Art Clean Technologies (SOACT) for steelmaking handbook*. Washington, DC, USA: American Iron and Steel Institute, 2010.

[20] International Energy Agency, *Tracking industrial energy efficiency and CO2 emissions*, vol. 30, no. 10. Paris, France: IEA, 2007.

[21] A. Sahin and M. Morari, "Optimizing control of hot blast stoves in staggered parallel operation," in *IFAC Proceedings Volumes (IFAC-PapersOnline)*, 2008, pp. 1982–1987.

[22] J. Zetterholm, X. Ji, B. Sundelin, P. M. Martin, and C. Wang, "Dynamic modelling for the hot blast stove," *Appl. Energy*, vol. 185, pp. 2142–2150, 2017.

[23] W. Uribe-Soto, J. F. Portha, J. M. Commenge, and L. Falk, "A review of thermochemical processes and technologies to use steelworks off-gases," *Renew. Sustain. Energy Rev.*, vol. 74, pp. 809–823, 2017.

[24] M. McBrien, A. C. Serrenho, and J. M. Allwood, "Potential for energy savings by heat recovery in an integrated steel supply chain," *Appl. Therm. Eng.*, vol. 103, pp. 592–606, 2016.

[25] X. Lin, Z. Peng, J. Yan, Z. Li, J. Y. Hwang, Y. Zhang, G. Li, and T. Jiang, "Pyrometallurgical recycling of electric arc furnace dust," *J. Clean. Prod.*, vol. 149, pp. 1079–1100, 2017.

[26] S. M. Thai, S. J. Wilcox, C. K. Tan, A. Z. S. Chong, J. Ward, and G. Andrews, "Monitoring regenerative steel reheating burners using an intelligent flame diagnostic system," *J. Energy Inst.*, vol. 87, pp. 48–58, 2014.

[27] X. Shanqing and W. Daohong, "Design features of air and gas double preheating regenerative burner reheating furnace," *Phys. Procedia*, vol. 66, pp. 189–192, 2015.

[28] F. Marchiori, A. Belloni, M. Benini, S. Cateni, V. Colla, A. Ebel, M. Lupinelli, G. Nastasi, M. Neuer, C. Pietrosanti, and A. Vignali, "Integrated dynamic energy management for steel production," *Energy Procedia*, vol. 105, pp. 2772–2777, 2017.

[29] H. Kim, J. Lee, S. Lee, I. B. Lee, J. Park, and J. Han, "Economic process design for separation of CO2 from the off-gas in ironmaking and steelmaking plants," *Energy*, vol. 88, pp. 756–764, 2015.

[30] R. M. Jones, L. L. Thomas, S. V Koshevets, R. T. Thatcher, and J. M. Hall, "Development and field validation of a large-frame gas turbine power train for steel mill gases," in *Proceedings of ASME Turbo Expo 2011: Power for Land, Sea and Air*, 2011.

[31] G. Di Marco and R. Manuzzi, "The recovery of waste and off-gas in large combustion plants subject to IPPC national permit in Italy," *Waste Manag.*, 2017.

TOWARDS A HEALTHCARE PERFORMANCE ASSESSMENT FRAMEWORK: A REVIEW

B. Joubert[1]* & W. Bam[2]

[12]Department of Industrial Engineering
Stellenbosch University, South Africa
[1]18175996@sun.ac.za, [2]wouterb@sun.ac.za

ABSTRACT

Improving health is the fundamental goal of all health systems, and with an evident gap between the potential of the public health system and its performance, bettering the services that provide care is critical to achieving this goal. This gap creates a need to identify and understand what drives performance so as to be able to assess how the potential can be reached. Thus, this research investigates the key inputs into a performance assessment framework and develops a core set of performance dimensions that is inclusive of all the factors that affect performance.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

Due to concerns regarding accountability and quality improvement strategies of health systems, as well as an increasing awareness of patient satisfaction, there has been a growing interest in the assessment of healthcare performance [1], [2]. Performance assessment helps in gaining an understanding of what drives performance, as well as which areas are performing above or below expectations, offering guidance on where to look for potential solutions and possible actions for improvement [3], [4]. It can therefore be used for the evaluation, and subsequent improvement of healthcare, thus initialising quality improvement activities [2], [5]. In the performance assessment process, measurement is a vital and necessary step, as it provides the information required for evaluation and decision making. Measurement allows for the comparison of how health systems or facilities are doing in contrast to the original targets so that opportunities for improvement can be identified [6].

Many frameworks aimed at contributing to the understanding of health systems, as well as the assessment of their performance, exist. In these frameworks, in order to successfully assess performance, indicators to track and measure are put in place to gain information about certain performance dimensions and their level of attainment. These indicators will provide information on how the performance dimensions and different goals are being met.

When constructing a performance assessment framework and deciding on which performance dimensions to include, it is clear that there are various differing views on how performance should be measured, and how a measurement system should be structured. Additionally, each framework defines a different set of dimensions along which to measure performance, adding to confusion on what exactly should be measured. The selection of which performance dimensions to include in a framework is very important, as these will determine the areas in which performance is measured, and thus the information available for assessing performance.

Simply including all the performance dimensions encountered in the literature will only over-burden the data collection process and complicate the assessment of the information gathered. There is already so much unnecessary data being collected, obstructing the proper flow and use of information [7]. The process of collecting all this data required large amounts of costly time and resources [7]. This creates a need to develop a list of only the relevant and necessary performance dimensions that can be used to garner a complete view of performance, while still being concise in order to reduce the burden of data collection. The aim of this paper is therefore to determine what is needed in a performance assessment framework, as well as to determine a core set of performance dimensions to be used.

The rest of this paper reviews various existing health system frameworks. These are then used to garner an understanding of what the ‘health system’ and ‘healthcare system’ is assumed to include. It then goes on to explore the goals of the health system, and what the improvement of their performance aims to achieve and investigates those factors that determine health. Lastly, the different performance dimensions are explored, with a core set being outlined.

2. METHODOLOGY

Documents published on performance frameworks, indicators, and dimensions were examined, along with documents published by or regarding the UK, Canada, Australia, the Commonwealth Fund, the World Health Organisation (WHO), and the Organisation for Economic Cooperation and Development (OECD) regarding their health system frameworks. These documents were gathered using basic strategies such as searching the official websites of the departments, ministries, or agencies of health of each country or organisation; checking the reference lists of selected documents and articles for any relevant articles; and lastly, electronic platforms such as Scopus and Google Scholar were used to identify performance frameworks, as well as doing a generic Internet search using the Google search engine.

Only recent sources regarding performance assessment were explored, especially those that were published after the World Health Report 2000. This is due to the increased interest in and attention given to developing a means to assess health system performance following the publication of this report. Only materials that dealt with the subject of study were included and seeing as the focus of this article is performance, only frameworks dealing with health system performance, and those health system frameworks that had a performance aspect to them were considered for this review.

3. REVIEW OF EXISTING CONCEPTUAL FRAMEWORKS

Many frameworks aimed at contributing to the understanding of health systems, as well as the assessment of their performance exist. There are a variety of different types of frameworks, but most important to this review are the health system and health system performance assessment (HSPA) frameworks. A health system framework aims to offer an overview of the healthcare system and therefore aims to act as a tool to describe it [8]. A HSPA framework assumes a certain health system structure but aims to specifically evaluate performance.

Although health system frameworks may also be used for performance assessment, they generally will not clearly have specified how this should be done. In both types of framework, how they are used to conduct evaluation activities depends on the framework itself and how clearly it defines what “good” performance is [3]. As the focus of this paper is on performance, all frameworks with a performance aspect were considered, regardless of the type of framework. An overview of the commonly discussed frameworks is given in this section.

3.1 UK

In the UK, the *Performance Assessment Framework* (PAF) was created for the National Health Service (NHS) as a performance measurement system [9]. This framework was conceptually based on the balanced scorecard approach (BSC), and is aimed at providing a broader view of performance in the NHS [9], [10]. Basing the framework on the BSC approach means moving away from traditional financial performance measures and undertaking a wider range of indicators to represent a balanced view of NHS performance [11].

The PAF identified six dimensions of performance to achieve this balanced view, namely: (1) health improvement; (2) fair access; (3) effective delivery of appropriate healthcare; (4) efficiency; (5) patient/carer experience; and (6) health outcomes of NHS care [11]. These dimensions are shown in Figure 1, where the circular representation of the areas of performance is aimed at showing their inter-dependence [11].

The two main purposes of the PAF is to: (1) assist the NHS in both improving the health of the population, as well as providing improved care and outcomes for the people who use the health services; and (2) to assess how well the NHS is achieving this objective [11]. The framework was created to support the aspiration toward a high-quality health system. This aim is supported through focusing the framework on the delivery of clinically cost effective, appropriate, and timely health services to meet the needs of the population [11].

In order to provide support and provide ways to monitor the six dimensions of the framework, a set of performance indicators was introduced, called high level performance indicators (HLPI) [9]. The indicator sets, which were chosen based on the aim of each of the dimensions, are there to highlight any issues that may need to be explored or action that needs to be taken, they are not a direct measure of quality [9], [11]. Both the PAF and HLPIs were first published in 1998 and are used to benchmark the performance of local health authorities [11]. As new information and data becomes available the HLPIs are constantly being developed and improved [10].

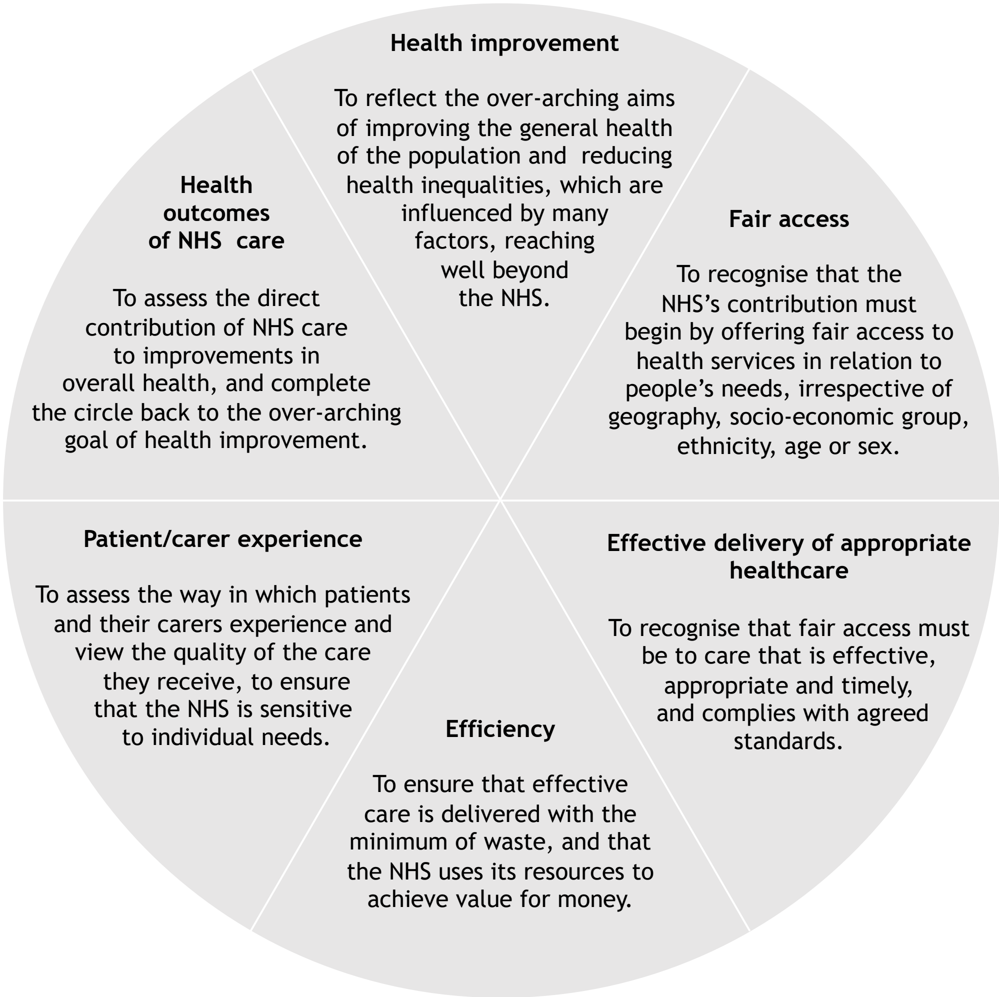


Figure 1 NHS Performance Assessment Framework [11]

3.2 Canada

The Canadian *Health Indicators* project was launched in 1999 and was based off the *Health Information Roadmap Initiative* [12], [13]. While the framework identifies important dimensions of health and health system performance for indicator development, it is important to note that it is not a health systems framework, but rather a framework used for classifying indicators for performance measurement. The goal of the framework is therefore to "identify and report on a set of indicators reflecting the health of Canadians and the health system" [12].

The *Health Indicator Framework* has four interconnected tiers representing the factors which have an influence on health, it is important to note that the tiers do not represent a hierarchy, but rather shows the relationship between the tiers and the impact each has on each other [12]. Dimensions along which indicators should be developed is suggested for each of the tiers. The tiers are as follows:

1) *Health status:* this tier represents the health and well-being of the Canadian population. The 4 ways in which health status can be measured, specified by this framework, include: well-being, health conditions, human function, and deaths [12];
2) *Non-medical determinants of health*: this reflects those factors that have an influence on the health of the population that isn't a result of the medical field, as well as, in some cases, when and how healthcare is used. These factors include: health behaviours, living and working conditions, personal resources, as well as environmental factors [12];
3) *Health system performance*: indicators in this tier measure various aspects in how well a health system is performing, and eight performance dimensions are defined: acceptability, accessibility, appropriateness, competence, continuity, effectiveness, efficiency, and safety [12];

4) *Community and health system characteristics*: this tier acknowledges the measures that provide useful contextual information but are not direct measures of health status or the quality of care. This includes the community, health system, and resources [12].

The final performance dimension identified is equity. Equity is not the listed in the health system performance tier and this is due to the fact that it is considered to span all dimensions of the framework, and is therefore included as an aspect the spans across all 4 of the tiers [14].

3.3 Australia

In February 2000, the National Health Performance Committee (NHPC) of Australia began to develop a new Australian health performance framework. The committee reviewed various existing frameworks for national reporting including Australian Institute of Health and Welfare conceptual framework for health, the United States' *Healthy People 2010*, the United Kingdom quality framework, and the Canadian Institute of Health Information framework. Following this review, it was decided that the framework developed by the Canadian Institute for Health Information (CIHI) as part of the Canadian Roadmap Initiative established in 1999 was to be used as a basis for the development of the Australian health performance framework [15].

It is important to note that the resulting framework is not a framework to model the health system, but rather one to provide structure as to how assess performance of the health system [15]. The framework was developed so it could be used to benchmark health system improvement, as well as provide information on the performance of the national health system [15].

As with the Canadian framework, the Australian framework consists of tiers, but only three tiers are represented in this framework:

1) *Health status and outcomes*: this tier proposed the following dimensions: health conditions, human function, life expectancy and wellbeing, and deaths. These dimensions aim to garner an understanding on the health of the population, and more specifically answer the following questions: "How healthy are Australians? Is it the same for everyone? Where is the most opportunity for improvement?" [15].
2) *Determinants of health*: the following factors are proposed as those that have an impact on the health status of the population: environmental factors, socioeconomic factors, community capacity, health behaviours, and person-related factors [15].
3) *Health system performance*: this tier encompasses the nine dimensions aimed at providing a better understanding on the performance of the health system. These dimensions are whether the health system is: effective, appropriate, efficient, responsive, accessible, safe, continuous, capable, and sustainable [15].

These three tiers represent the factors which have an influence on health, showing the relationship between the tiers and the impact each has on each other, and therefore the tiers do not represent a hierarchy. For each of the tiers, a number of dimensions are suggested along which indicators should be developed in order assess the performance of that tier to gain an understanding on the performance of a given tier. For each tier a question is posed, and the aim of the indicators developed for the dimensions should be to answer those questions, so a better understanding of the health system's performance can be gathered [15]. In this framework, one indicator may provide information for more than one of these dimensions [15].

Equity is also considered a measure of health system performance. In this framework it is integral to all three of the tiers, and therefore included as a dimension that spans across them using the question "is it the same for everyone?". [15]

3.4 The Commonwealth Fund

The Commonwealth Fund is a private foundation based in the United States, and was established in 1918 with the aim to "enhance the common good" [3], [16]. In trying to improve the performance of the healthcare system in the US, the Commission on a High Performance Health System was established [17]. The goal of this commission is to "promote a high performing healthcare system that achieves better access, improved quality, and greater efficiency, particularly for society's most vulnerable, including low-income people, the uninsured, minority Americans, young children, and elderly adults" [16].

In realising that the level of performance of the US healthcare system is not where it should be, the commission created the Framework for a High Performance Health System for the United States. This framework states that the goal of a healthcare system is "to help everyone, to the extent possible, lead long, healthy, and productive lives" [16].

In this framework, shown in Figure 2, four core goals are outlined:

1) High quality, safe care;
2) access to care for all people;
3) efficiency, high value care; and
4) system capacity to improve [16].

Each of these goals is made up of criteria that can be used to map indicators [3]. The commission defines what a high performance health system is, and sets targets to be tracked over time, as well as identifying and analysing any public policy or practice changes that need to be made on order to meet the targets that have been set [17].

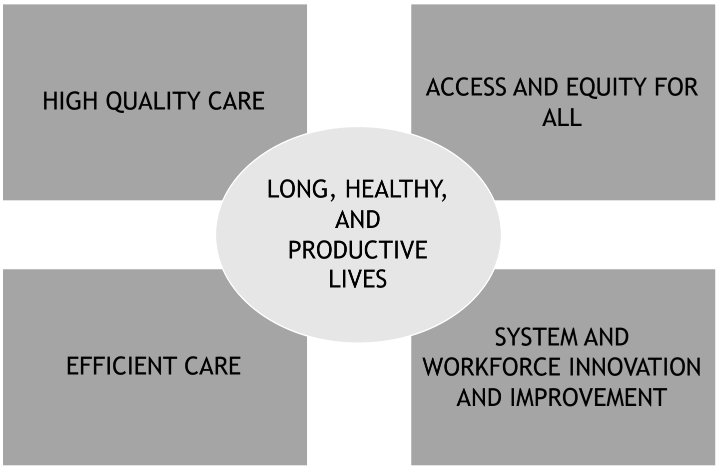


Figure 2 Framework for a High Performance Health System for the United States [16]

3.5 World Health Organisation

The WHO's framework assesses health system performance by relating the health system goals to its functions. By conceptualising performance in this way, the framework aims to aid in assessing performance, understanding the factors that contribute to and improve it, and as a result, allow better response to the needs and expectations of the people the health system serves [3], [7]. The framework for assessing health system performance was developed by Murray and Frenk, and subsequently used in the World Health Report 2000 to structure the statistical annexes [3], [18].

The framework begins by defining a health system and its boundaries. A 'health system' is defined as "comprising all the organisations, institutions and resources that are devoted to producing health actions", where 'health action' is defined as "any effort, whether personal healthcare, public health services or through intersectoral initiatives, whose primary purpose is to improve health" [7]. Therefore, any organisation with the primary objective being to contribute to health is included in this definition, while actions that may have an impact on health but does not have health-related primary objective is excluded [3].

The framework then explores the goals of a health system. While improving health is identified as the main objective of a health system, the WHO goes further by defining three intrinsic goals. The framework states that in addition to the defining goal, the other two goals common to all systems is *responsiveness* of the system to the expectations of the population, and *fairness* in the financial contribution required to make the system work [18]. Explicitly stated, the three main goals for the health system are: health improvement, responsiveness, and fairness in financial contribution. These goals form the basis for assessing the performance of a health system, and the progress towards achieving them depends on how the functions of the health system are being carried out [7], [18].

For both health attainment and responsiveness, the average level of achievement as well as the inequalities in the distribution across different groups are considered, whereas for the third intrinsic goal, only distribution is considered [3], [18]. It is worth taking note that efficiency and equity are not explicitly mentioned in the list of intrinsic goals, however they are considered to be present amongst the goals. Efficiency is how well the goals are being achieved, and is therefore taken into account in the average attainment of each goal, while their distributions represents total equity of the health system[3], [18].

Once the goals had been identified, the functions of a health systems are defined in order to relate them to the goals, so that performance can be measured. The framework identifies four basic functions: financing, service provision, resource generation, and stewardship [7].

3.6 Organisation for Economic Cooperation and Development

The OECD's Health Care Quality Indicator (HCQI) Project was undertaken with the goal of developing a common set of indicators that can be used to assess the quality of healthcare being provided by any of the OECD member countries [10]. The authors of this framework use same definition of a health system as defined by the WHO, the framework therefore not only considers healthcare, but also includes all activities that "have a primary purpose of promoting, restoring or maintaining health" [3].

Although the framework (Figure 3) captures a broad perspective on all the factors influencing health, the specific focus of the HCQI project is on the quality of care, this focus is shown by the shaded area [10]. The framework is heavily based on the Institute of Medicine's national healthcare quality indicator framework developed in the USA, as well as the Canadian Health Indicator Framework [10]. The HCQI Framework has the 4 interconnected tiers, each influencing each other, with the causal pathways being represented in the figure by the arrows between them.

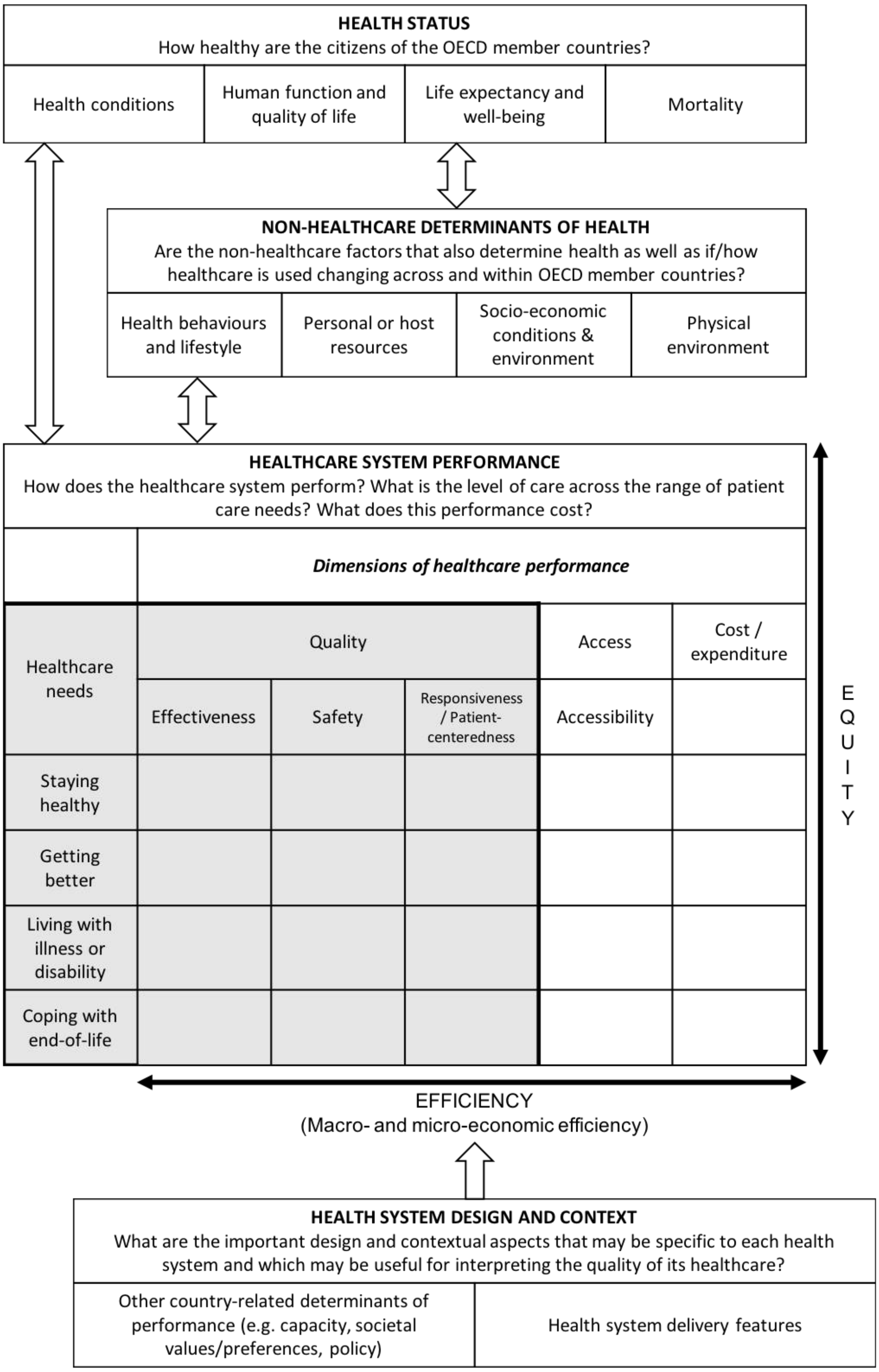


Figure 3 Conceptual framework for OECD HCQI Project [10]

The tiers are as follows:

1) *Health:* this tier represents the health of the population and is influenced by the two succeeding tiers, both non-medical determinants of health and health system performance [10];
2) *Non-medical determinants of health:* this tier reflects those factors that have an effect on the health of the population but aren't as a result of the medical field [10];

3) *Health system performance*: it is recognised that this tier may sometimes not only influence the health status of the population, but also the non-healthcare determinants of health; and is aimed at capturing the inputs, processes, and outcomes of the healthcare system as well as its efficiency and equity [10];
4) *Health system design and context*: this tier recognises the policy, responsibility, and structural differences between healthcare systems in managing the healthcare and non-healthcare determinants of health [10].

The third tier, health system performance, is illustrated using a matrix, where the columns represent the performance dimensions, and the rows represent the healthcare needs [10]. The OECD therefore defines the following performance dimensions: effectiveness, safety, responsiveness/patient-centeredness, access, and cost/expenditure. Seeing as the focus of this framework is on quality, the areas of effectiveness, safety, and responsiveness/patient-centeredness are grouped into the quality dimension. As can be seen in Figure 3, efficiency is considered across all dimensions, and equity across all tiers and healthcare needs. Macroeconomic efficiency is about finding the right level of health expenditure, and microeconomic efficiency is concerned with maximising the value for money [10]. Equity focuses on both health being distributed fairly across populations, as well as fairness in the payment for healthcare [10].

4. DISCUSSION

This section aims to determine the necessary elements for the development of a performance assessment framework by offering a discussion on the frameworks outlined in Section 3. Along with a brief look at the determinants of health, specific aspects from the frameworks are highlighted, namely: how they define health or healthcare systems and their boundaries; the overall and instrumental goals trying to be achieved; as well as which performance dimensions are defined in each framework. The performance dimensions are then critically analysed in order to define a core set of performance dimensions that will address all the areas considered to play a part in performance of a health system, while making the list as concise as possible to reduce the burden of data collection.

4.1 Health systems

Crucial to developing a framework for performance measurement is to define the boundaries of the system for which it is created. To gain an understanding of the context in which the frameworks are placed, how they define a health system is first explored. Multiple definitions exist for health systems, and where the boundaries should be drawn is an often-debated topic. In its World Health Report 2000, the WHO [7] defines health systems as: “comprising all the organisations, institutions and resources that are devoted to producing health actions”, where ‘health action’ is defined as “any effort, whether personal healthcare, public health services or through intersectoral initiatives, whose primary purpose is to improve health”. The OECD [10] states that “a health system includes all activities and structures whose primary purpose is to influence health in its broadest sense”, thus adopting the same idea as the WHO. Similarly, the Canadian framework adopts the same definition, therefore including services provided to both individuals and groups, as well as public health services and policies [12]. The Australian national framework does not explicitly state which boundaries of the health system are used, but after their previous framework that focused specifically on the care provided by the hospital sector, it was decided to take on a broader view of health systems for the current framework [15].

Many boundaries to separate the health system from elements outside of it have been proposed, showing that all boundary definitions are arbitrary, but to assess health system performance an operational boundary must be proposed [18]. It is also important to note the difference between frameworks developed to assess the performance of the health system as a whole, and those specifically for the healthcare system. The healthcare system is part of the overall health system, but refers specifically to the provision of, and investment in, health services [7]. The health system encompasses a much broader scope than the healthcare system, and the healthcare system is limited to personal healthcare services, including preventive, curative and palliative interventions [7]. Although most frameworks focus on the healthcare system, they often refer to the health system and even include elements from the wider set of factors [3].

The Framework for a High Performance Health System for the United States is an example of a framework developed for measuring the performance of the healthcare system. In this framework the Commonwealth Fund [16] defines a healthcare system as: “the ways in which healthcare services are financed, organised, and delivered to meet societal goals for health. It includes the people, institutions, and organisations that interact to meet the goals, as well as the processes and structures that guide these interactions”.

From different definitions, it is often unclear as to where public health and health promotion activities lie. But nevertheless, a framework’s function as an HSPA tool is influenced by how narrowly or broadly the boundaries are set. If the boundaries are too wide it becomes difficult to identify the influence each factor has on performance, however, narrowing the boundaries can exclude actions that have a great impact on health [3].

It is important in practice to align the definition of health system a closely as possible to the actors responsible for improving health. While additionally acknowledging the greater setting in which the system operates improves the understanding of how the system interacts with the wider economic, political and social surroundings [3]. Regardless of how narrow or broad the boundaries are defined, it should be made clear which activities are included in these boundaries [3].

4.2 Goals

Performance is the achievement of desired goals and objectives according to specific standards and guidelines [1], [19]. It must therefore be defined in relation to specific goals that reflect the values of all the stakeholders involved, such as patients, professions, insurers, etc. [6]. Consequently, it is of vital importance to realise the goals of the system being improved, as the shared goal unites the interests and activities of all the stakeholders [20]. In something as complex as a health system there will inevitably be a countless number of goals, and they will often be conflicting [20]. Through reviewing both the frameworks discussed in this paper and various literature it is clear that the overarching goal of a health system is to improve the health of the population it serves [7]. While the main objective is to better health, it is not the only one. Additionally there are many intermediate measures of both processes and outcomes [6].

Murray and Frenk [18] discussed the difference between intrinsic and instrumental goals. Intrinsic goals are those valued in themselves, whereas instrumental goals are there for the purpose of achieving something else. According to Murray and Frenk [18], intrinsic goals fulfil two criteria, any goals that do not satisfy both the criteria are most likely instrumental. For a goal to be intrinsic:

1) It is partially independent of the other intrinsic goals, in that the level of attainment of the goal can be raised while the other ones are kept constant; and
2) It is desired to increasing the level of achievement of the goal [18].

According to the World Health Report 2000 the objective of good health itself is twofold: goodness and fairness. Goodness refers to achieving the best attainable average level of health, meaning that the health system responds well to the expectations of the people it serves. Fairness aims to achieve the smallest feasible differences among individuals and groups, meaning it responds equally well to everyone, without discrimination [7]. The concept of performance in the WHO framework is therefore focused on three fundamental goals: "improving health, enhancing responsiveness to the expectations of the population, and assuring fairness of financial contribution. Improving health means both increasing the average health status and reducing health inequalities." [18].

The OECD framework identifies the following three goals of health policy: (1) improving health; (2) efficiency; and (3) equity [10]. For the goal of efficiency, the OECD subdivided it into (i) macroeconomic efficiency or sustainability; and (ii) microeconomic efficiency or value for money. Macroeconomic efficiency is especially important for public health expenditure, as macroeconomic efficiency entails setting the right level for health expenditure [21]. For the goal of equity, this applied to both fair financing and fair access of health across the population [10].

The Commonwealth Fund [16] states the goal of the healthcare system is "to help everyone, to the extent possible, lead long, healthy, and productive lives", with four core goals of a high performance health system being identified as (1) high quality, safe care; (2) access to care for all people; (3) efficient, high value care; and (4) system capacity to improve.

The PAF for the NHS was created to aid in assessing the attainment of the goal of improving the health of the population, as well as providing improved care and outcomes for the people who use the health services. In order to improve health the PAF states that it must be ensured that "everyone with healthcare needs (fair access) receives appropriate and effective care (effective delivery) offering good value for money for services (efficiency) as sensitive and convenient as possible (user/carer experience) so that good clinical outcomes are achieved (health outcome of NHS care), to maximise the contribution to improved health (back to health improvement)" [11].

Deciding on the goals of the system is an important initial step creating a performance assessment framework, defining the goals of a system allows for them to be measured, thus allowing the concept of performance and the key factors that influence it to be explored [18].

4.3 Determinants of health

In recognising that the overall goal of health systems is to improve health, and performance measurement aids in determining the extent to which goals are being met, it is crucial to recognise those elements that play a part in the health of a population. Health determinants, as defined by the National Health Performance Committee [22] are: "those factors that either have a positive or negative influence on health at the individual or population level". It would be a gross oversimplification to purely attribute improved health to better performing healthcare system. Health is determined by a large number of factors, and while healthcare is one way to maintain and improve health, it is important to realise the other factors that also determine the health of a population, many of which lie outside the health system [18], [23], [24].

McKeown [25] identifies three main fields that have an impact on health: environment, personal behaviour, and healthcare. Personal behaviour includes activities that the individual has a large influence on, such as smoking, the misuse of drugs or alcohol, diet, exercise, and other habits [25], [26]. The environment, conversely, refers to the physical and social environment that the individual can do little to control, these can include things such as contamination of drinking water, air pollution, and the effect of rapid social change on mental and physical health [26].

While all three of these categories are recognised as having an impact on the health of people, McKeown believes that a lot of emphasis is put on developing and improving healthcare, when personal behaviour and the environment are actually the primary determinants of health. There is no argument, however, that healthcare does indeed have an impact on health [25], [26].

4.4 Performance dimensions

In order to select the indicators required to measure performance, the dimensions of healthcare performance along which it will be measured needs to be established. The Joint Commission on Accreditation of Healthcare Organisations [27] defines a dimension of healthcare performance as “an attribute that is definable and preferably measurable, and related to the system’s functioning to maintain, restore, or improve health”.

There are many existing frameworks for performance measurement, each taking a different set of dimensions into account. For this paper, when deciding on the dimensions that will be used, the frameworks discussed in Section 3 were used. By considering existing conceptual frameworks for their performance dimensions, it allows a framework to build on the experience of previous studies. An initial list of all existing performance dimensions was gathered and tabulated in Table 1. As a consequence of the sheer number of dimensions presented in the various frameworks, some overlap and redundancy between the dimensions are inevitable [10].

Table 1 gives an overview of the most commonly used performance dimensions in the considered frameworks.

Table 1: Dimensions of healthcare performance

Dimensions	*UK*	*Canada*	*Australia*	*Commonwealth Fund*	*WHO*	*OECD*	*Count*
Acceptability		X		X			2
Accessibility	X	X	X	X		X	5
Appropriateness	X	X	X	X			4
Capability			X				1
Competence		X					1
Continuity		X	X	X			3
Effectiveness	X	X	X	X	X	X	6
Efficiency	X	X			X	X	4
Equity	X	X	X		X	X	5
Expenditure or cost					X	X	2
Patient-centeredness/ Responsiveness	X		X		X	X	3
Safety		X	X				2
Sustainability			X				1
Timeliness	X						1

To reduce this list, the definitions of all these concepts must first be explored. Donabedian [28] defines acceptability as the “conformity to the realistic wishes, desires and expectation of healthcare users and their families”, accessibility as “the ease with which health services are reached. Access can be physical, financial, or psychological and requires that health services are a priori available” and defines appropriateness as “the degree to which provided healthcare is relevant to the clinical needs, given the current best evidence”. Competence and capability deal with the same concept, which is having the knowledge and ability to appropriately provide care [15], [29]. And according to Donabedian [28]: “continuity addresses the extent to which healthcare for specified uses, over time, is coordinated across providers and institutions”.

Further, effectiveness is the degree to which attainable and desirable outcomes or objectives are achieved [23], seeing as improved health is often the objective of a health system, any dimension with clinical focus or aiming to improve health was included in this category. Whereas efficiency involves the optimal use of available resources to achieve the maximum benefits or results (i.e. maximising the outputs to inputs ratio) [4], [5], [10], [23].

Equity defines the degree to which a system fairly benefits those who use it [10], [23]. In the health context, equity deals with how fairly healthcare is distributed among people, as well as the distribution of the burden of paying for it [7], [10], [23]. While safety encapsulates the degree to which the structures, as well as processes of the healthcare system prevent harm or any adverse outcomes to the users of the system and the environment in which healthcare is delivered [4], [5], [15].

Responsiveness refers to how the needs and expectations of people are met [7]. With the WHO [4] stating that responsive governance "embraces the extent to which the hospital relates to community health needs, ensuring the continuity of care and the provision of health services irrespective of ethnical group, physical, cultural, social, demographic or economic characteristics". Similarly, a patient-centered system delivers care that is responsive and respective to the preferences, needs, and values of the people being served, and one whose clinical decisions are guided by these values [30]. It is clear from the definition of patient-centeredness that it is often considered to be synonymous with responsiveness, and therefore why they are included in the same category in Table 1.

The National Health Performance Committee [15] defines sustainability as the "system or organisation's capacity to provide infrastructure such as workforce, facilities and equipment, and be innovative and respond to emerging needs". And timeliness involves providing care in a time window that is most beneficial to both those who receive care and those who deliver care, reducing waiting time and therefore harmful delays [10], [30].

From the Table 1 it can be seen that the most commonly occurring dimensions were: accessibility, appropriateness, effectiveness, efficiency, equity, and patient-centeredness/responsiveness. Not including some of the dimensions does not mean they are disregarded or ignored, but rather most of them can be absorbed into the scope of the other dimensions. Cost, for example, can be included as a part of effectiveness. In the UK PAF effectiveness includes both clinical and cost effectiveness, and the OECD includes it as a part of efficiency [10]. Further, timeliness can be reflected in both accessibility or responsiveness by including the promptness of care [10]. Similarly, accessibility and equity are closely related and both deal with how fair and available the access is to health services [10]. Continuity, as well as acceptability can be subsumed in the dimension of responsiveness due to the fact that it can be measured from the perspective of the patient. In the Canadian framework acceptability is presented as a part of patient-centeredness [10].

Although responsiveness was not necessarily included in a lot of the frameworks, it will be included in this core list due to the other dimensions it can represent. The only dimension that was included in all 6 of the frameworks considered is effectiveness, highlighting the importance of including this dimension. Following this discussion, the core dimensions for measuring and assessing performance identified in this paper include: appropriateness, effectiveness, efficiency, equity, and responsiveness.

5. CONCLUSION

In the synthesis of the various frameworks reviewed in this chapter, the necessary elements to be included in a performance assessment framework has been determined and illustrated in Figure 4 below. As per the figure, when developing such a framework one should begin by first defining the boundaries of the system it aims to assess. The act of separating the health system from the elements outside of it acknowledges the greater setting in which it operates. By not just separating, but also classifying where different activities lie within the health and healthcare system, as well as outside of it, it improves the understanding of how the system interacts with its wider surroundings.

Secondly, both the intrinsic and instrumental goals must be clearly articulated in order for the framework to measure their level of attainment. Thirdly, it is clear that a broader view of health that acknowledges all the factors that may play a part in its improvement must be adopted. By stating both the healthcare and non-healthcare determinants of health the framework recognises that any changes in health status cannot solely be attributed to the changes in performance of the health system.

Lastly, the specific dimensions that are understood to impact the performance of the system being assessed must be determined. Following this review, a core set of performance indicators including: appropriateness, effectiveness, efficiency, equity, and responsiveness were realised.

The arrows in the figure represent the relationship between an improving health system and the healthcare it provides, as well as the goals it aims to achieve. Improving health system performance in the specified dimensions should be bringing the health system closer to achieving its stated goals, as well as improving the healthcare it provides. Additionally, the goals of, and healthcare provided by, the health system impact the performance of the health system. It is clear that identifying these dimensions and linking them in a comprehensive and structured way is what is needed for developing a performance assessment framework.

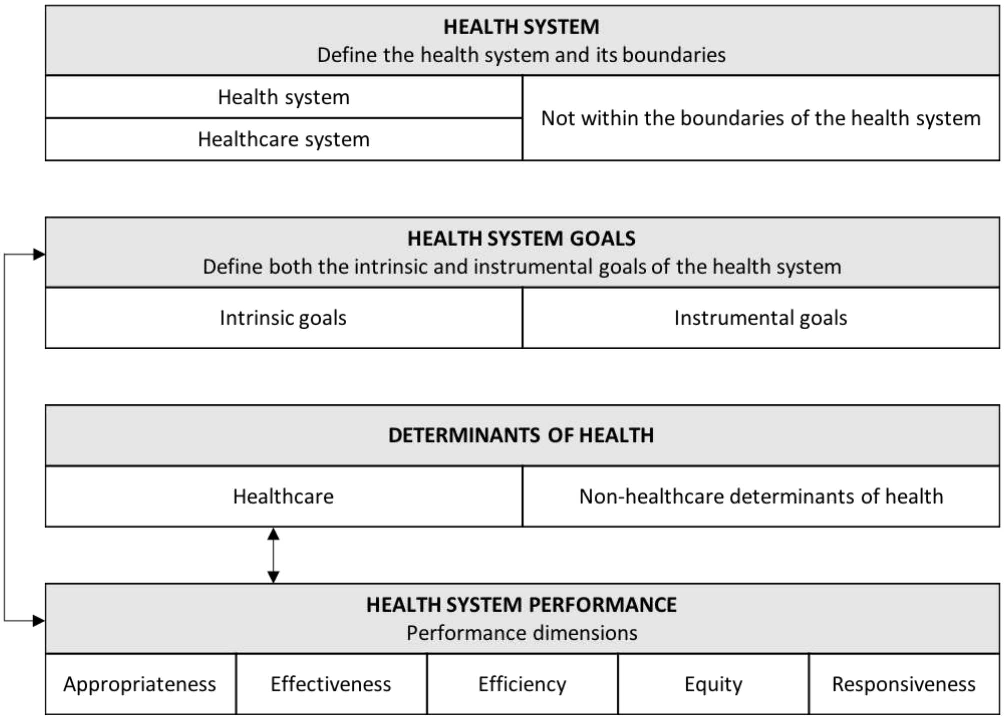


Figure 4 Outline for a health system performance assessment framework

6. REFERENCES

[1] World Health Organization, "Measuring hospital performance to improve the quality of care in Europe: A need for clarifying the concepts and defining the main dimensions," 2003.

[2] O. Groene, N. Klazinga, V. Kazandjian, P. Lombrail, and P. Bartels, "The World Health Organization Performance Assessment Tool for Quality Improvement in Hospitals (PATH): An Analysis of the Pilot Implementation in 37 Hospitals," *Int. J. Qual. Heal. Care*, vol. 20, no. 3, pp. 155–161, Mar. 2008.

[3] I. Papanicolas and P. C. Smith, *Health System Performance Comparison*, 1st ed. Berkshire: Open University Press, 2013.

[4] World Health Organization, "Performance assessment tool for quality improvement in hospitals." WHO Regional Office for Europe, 2007.

[5] J. Veillard, F. Champagne, N. Klazinga, V. Kazandjian, O. A. Arah, and A. L. Guisset, "A performance assessment framework for hospitals: The WHO regional office for Europe PATH project," *Int. J. Qual. Heal. Care*, vol. 17, no. 6, pp. 487–496, 2005.

[6] C. Shaw, "How can hospital performance be measured and monitored?," Copenhagen, 2003.

[7] World Health Organisation, "The World Health Report 2000: Health Systems: Improving Perfomance," Geneva, 2000.

[8] G. Shakarishvili, R. Atun, P. Berman, W. Hsiao, C. Burgess, and M. A. Lansang, "Converging Health Systems Frameworks: Towards A Concepts-to-Actions Roadmap for Health Systems Strengthening in Low and Middle Income Countries," *Glob. Heal. Gov.*, vol. 3, no. 2, 2010.

[9] L. Chang, S. W. Lin, and D. N. Northcott, "The NHS Performance Assessment Framework: A 'balanced scorecard' approach?," *J. Manag. Med.*, vol. 16, no. 5, pp. 345–358, 2002.

[10] O. A. Arah, G. P. Westert, J. Hurst, and N. S. Klazinga, "A conceptual framework for the OECD Health Care Quality Indicators Project.," *Int. J. Qual. Health Care*, vol. 18 Suppl 1, no. Supplement 1, pp. 5-13, 2006.

[11] NHS, "The NHS Performance Assessment Framework," 1999.

[12] Canadian Institute for Health Information, *Health Indicators 2013*. Ottowa, ON: CIHI, 2013.

[13] O. A. Arah, N. S. Klazinga, D. M. J. Delnoij, A. H. A. Ten Asbroek, and A. T. Custers, "Conceptual frameworks for health systems performance: a quest for effectiveness, quality, and improvement," *Int. J. Qual. Heal. Care*, vol. 15, no. 5, pp. 377–398, 2003.

[14] M. Wolfson and R. Alvarez, "Towards Integrated and coherent health information systems for performance monitoring: the Canadian experience," in *Measuring Up: Improving Health System Performance in {OECD} Countries*, Paris: Organization for Economic Cooperation and Development, 2002, pp. 133–155.

[15] National Health Performance Committee, *National Health Performance Framework Report*. Brisbane: Queensland Health, 2001.
[16] Commonwealth Fund, *Framework for a high performance health system for the United States*. New York: The Commonwealth Fund, 2006.
[17] K. Davis, "Toward a high performance health system: The Commonwealth Fund's new commission," *Health Aff.*, vol. 24, no. 5, pp. 1356–1360, 2005.
[18] C. J. L. Murray and J. Frenk, "A framework for assessing the performance of health systems," *Bull. World Heal. Organ.*, vol. 78, no. 6, pp. 717–731, 2000.
[19] UNICEF and World Health Organisation, *Monitor evaluate improve: TDR performance assessment framework*. 2013.
[20] M. E. Porter, "What is value in health care?," *N. Engl. J. Med.*, vol. 363, no. 23, pp. 2477–2581, 2010.
[21] Organization for Economic Cooperation and Development, *Towards High-Performing Health Systems*. Paris: Organization for Economic Cooperation and Development, 2004.
[22] National Health Performance Committee, "National report on health sector performance indicators. AIHW Cat. No. HWI 78," Canberra, 2004.
[23] E. Kelley and J. Hurst, "Health Care Quality Indicators Project: Conceptual Framework Paper," Paris, 23, 2006.
[24] WHO, "The world health report 2013: Research for universal health coverage," *World Heal. Organ. Press*, p. 146, 2013.
[25] T. McKeown, "Determinants of Health," in *Understanding and Applying Medical Anthropology*, Third., P. J. Brown and S. Closser, Eds. New York: Taylor & Francis, 2016, pp. 99–104.
[26] M. Lalonde, *A new perspective on the health of Canadians*. Ottowa, ON: Minister of National Health and Welfare, 1974.
[27] Joint Commission on Accrediation of Healthcare Organisations, *National Library of Healthcare Indicators: Health Plan and Network Edition*. Oakbrook Terrace, IL: Joint Commission on Accreditation of Healthcare Organisations, 1997.
[28] A. Donabedian, *An Introduction to Quality Assurance in Health Care*. Oxford: Oxford University Press, 2003.
[29] Canadian Institute for Health Information and Statistics Canada, *Roadmap Initiative... Launching the Progress*. Ottowa: Canadian Institute for Health Information, 1999.
[30] Institute of Medicine & Committee on Quality of Health Care in America, *Crossing the Quality Chasm: A New Health System for the 21st Century*, vol. 323, no. 7322. 2001.

ENERGY COST REDUCTION IN SECONDARY STEEL MAKING THROUGH IMPROVED PRODUCTION PLANNING

W.A. Pelser[1], W.J.J. Breytenbach[2] & J.H. Marais[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
wapelser@researchtoolbox.com

[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbreytenbach@researchtoolbox.com

[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
jhmarais@researchtoolbox.com

ABSTRACT

The international steel manufacturing industry is under financial pressure due to a surplus in production flooding the market. In a South African context, steel producers deal with additional challenges such as the increasing cost of raw materials, electricity tariffs, wages, and transportation. Steel manufacturing facilities consume approximately 18% of industrial energy worldwide. Further research indicates that between 20% and 40% of steel production costs originate from energy expenses.

In this paper, production scheduling is used in a novel way to reduce energy cost. The developed solution uses various energy awareness implementation techniques and supporting systems to assist production planners with scheduling tasks. Opportunities for energy cost reduction by improving production planning methods are evaluated, and suggestions are made for production planners to adapt schedules to reduce energy costs.

The solution is applied at a secondary steel making facility, and is used to manage electricity costs by utilising time-of-use tariffs. Electrical energy intensity of different steel qualities are evaluated, and production is scheduled accordingly. The practical implementation indicates an annual cost reduction of R 1.5-million. This solution makes provision for the addition and integration of other potential production scheduling solutions.

This work was sponsored by ETA Operations (Pty) Ltd

[1] Wiehan Pelser holds a Master's degree in Electric and Electronic Engineering from the North-West University. He is currently a PhD student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr Wynand Breytenbach holds a PhD in Development and Management Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

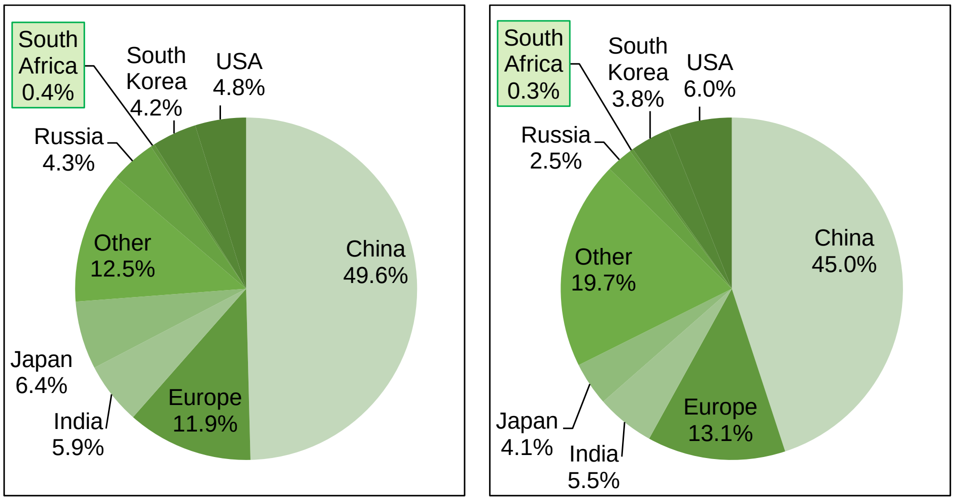
[3] Dr Johan Marais is a registered professional engineer and holds a PhD in Electrical Engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. BACKGROUND

1.1 The steel manufacturing industry

A surplus in international steel production has been reported to be flooding the market, placing the international industry under pressure [1, 2, 3, 4, 5]. Further reports indicate that the surplus most likely originates from China, where more steel is produced than consumed, with the remaining steel being exported to the rest of the world [6]. Of the 1 630 million tonnes of steel produced in 2016, China produced 49.6%, but only consumed 45.0% [7]. South Africa produced 0.4%, but only consumed 0.3% [7], indicating that they are a minority player in the international market, and easily affected by the actions of majority players.

Figure 1 (compiled from data obtained from the *World steel association* [7]) provides a summary of steel production versus consumption for major steel producing countries, and how they compare to South Africa. The comparison between production in Figure 1.a and consumption in Figure 1.b indicates the oversupply of 4.6% by China (approximately 75 million tonnes – more than 12 times the steel produced by South Africa). Due to subsidies provided to the steel manufacturing industry by the Chinese government, steel imported from China is often cheaper than countries can produce for themselves [8, 9].



a.) World steel production **b.) World steel consumption**

Figure 1: World steel production vs consumption for major countries and South Africa in 2016 [7]

Additional factors constraining the South African steel manufacturing industry include the increasing cost of raw materials, electricity tariffs, wages, and transportation [9]. These irregular increases are reported to be a result of a weakening economy and exchange rate volatility [10]. Such challenges reportedly resulted in a steel production decrease in South Africa from 9.7 million tonnes in 2006 [11] to 6.1 million tonnes in 2016 [7].

1.2 Energy consumption in steel manufacturing

The iron and steel industry is one of the most energy intensive energy consumers in the industrial sector, consuming about 18% of the sector's energy [12]. Research indicated that energy costs contributes 20% to 40% of steel manufacturers' operational costs [13, 14]. It was also indicated that the energy consumption of an older plant can be reduced by up to 60% by using technological improvements [13]. It is thus expected that plants with resistance toward technological solutions would have the potential to reduce energy costs, highlighting it as a critical area of focus.

1.3 Steel manufacturing processes

Different approaches exist for steel making. The most popular methods are the electric arc furnace (EAF) – basic oxygen furnace (BOF) and blast furnace (BF) – BOF process [12, 13], as depicted in Figure 2 [13]. The EAF-BOF method is used for 25% of applications, and the BF-BOF process for the remaining 75% [13]. An open hearth furnace is a rarely used method for steel making (about 0.5% of production) [12].

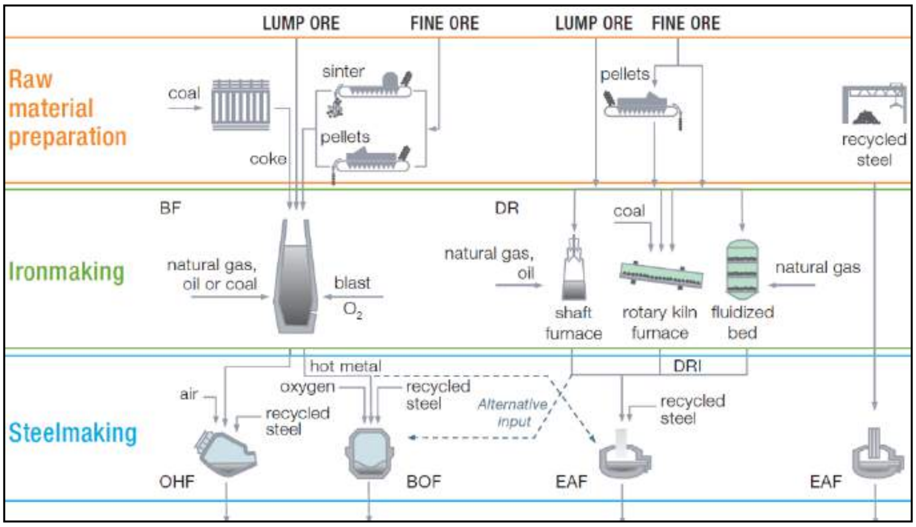


Figure 2: Different routes for steel production [13]

This paper focusses on an application in the BF-BOF approach, which will be discussed in more detail. It is, however, of importance to note that the EAF-BOF process makes use of a batch production process, making it possible to induce buffers into the production process. Such buffers are useful for production planning initiatives.

The BF-BOF route typically consists of preparatory processes, such as a coke plant, sinter plant, and hot blast stove. These processes provide some of the elements used in the BF to produce liquid iron. A BF is charged with iron ore, fluxes, and fuel, which works together to establish combustion, and the melting of iron ore to produce liquid iron. This iron production process is continuous, and delays can therefore not be induced if required [15, 16, 17].

Liquid iron is then moved to the BOF where it is refined to liquid steel by blowing oxygen into it. The oxygen removes impurities, and adjusts the carbon content as required for the steel quality [16]. Liquid steel is further refined to the steel quality requirements at the secondary metallurgy (SecMet) section [16]. During the refining at SecMet, the liquid steel temperature decreases due to the addition of material at ambient temperature.

Ladle furnaces are used to increase the temperature of steel to a specific set point before it can be casted [18, 19]. A ladle furnace uses a three-electrode system, and is thus a large consumer of electrical energy[4]. The main focus of the case study discussed in this paper is on the energy cost reduction of the ladle furnaces by means of improved production planning.

Liquid steel is sent to the continuous casters (ConCast) after it has reached its required specifications at SecMet. It is then casted through a tundish, continuously solidifying it into slabs at the bottom of the caster. More than one consecutive cast of the same steel quality is called a sequence. The interruption of such a sequence has critical cost and production delay implications [19].

This is a very important constraint to be considered when performing production planning on a steel plant. Other factors to be considered during production planning at such a facility have been reported to be production delivery dates, limitations of the casters, the requirements of the rolling mills, etc. [19].

1.4 Problem statement

Research by Dondofema *et al*. indicates that limited research on improved production processes have been published in South Africa (five publications by the South African Institute of Industrial Engineering) [11]. Dondofema *et al*. also referred to the review of South African industrial engineering by Van Dyk, indicating that few industrial engineers are employed by the iron and steel industry in South Africa [11, 20].

A problem exists in the South African steel making industry to remain profitable amid challenging international and local conditions. The high contribution of energy towards the operational cost of steel manufacturing highlights it as a key area of focus. Production planning has been identified as a possible tool that can be utilised to reduce energy costs without negatively affecting production.

[4] https://www.ingeteam.com/au/en-au/sectors/steel-metals/p15_29_528_525/secondary-metallurgy-ladle-furnaces.aspx

2. EXISTING RESEARCH FOR PRODUCTION PLANNING TO REDUCE ENERGY COST

Production planning is most commonly performed manually by experienced production planners [21]. The problem is, the complexity of production planning is continuously increasing due to higher production requirements, a wider variety of products, unstable orders from customers, and increased pressure to reduce production and energy cost [21]. Short-term production planning is gaining attention as a method to reduce energy cost. This has been described as an "enabler" for improved energy consumption [21].

The concept consists of taking energy consumption into consideration when performing general production planning – something that has not been done at the case study plant in the past. This makes it possible to predict and improve consumption trends, thereby managing energy consumption with the result of reducing energy costs [21].

As part of the review of previous work relevant to this study, the following fields were critically evaluated:

- Production scheduling with the focus on energy cost efficiency [22, 23, 24]
- Production scheduling with the focus on production efficiency [25, 26, 27]
- General iron and steel making energy initiatives [1, 11, 12, 28, 29, 30]
- Steel plant production scheduling research [31, 32, 33, 34, 35]

The most relevant of this research is the concept of electrical load shifting by utilising varying electricity prices. This is a well-known technique in various industries to reduce energy costs without affecting production [21]. It is typically done by making use of available buffers in a process to shift electrical energy intensive processes to less expensive time periods.

A study by Merkert *et al.* discusses the correlation between energy and scheduling, and highlights the value of electrical load shifting [21]. The study discusses the possibility of using the energy intensiveness of different batches in a process to shift electrical load, rather than inducing buffers in the process [21]. Existing research on performing electrical load shifts on ladle furnaces made use of the EAF-BOF production process [19, 36].

As discussed in section 1.3, buffers can be induced in the EAF-BOF process, making it possible to schedule delays during more expensive electricity time periods. This paper, however, uses a new approach of combining these concepts to perform an electrical load shift on the ladle furnaces in a BF-BOF production process, where the continuous production of liquid iron restricts the induction of delays in the process.

The concept of using the energy intensity of batches, as suggested by Merkert *et al.*, will thus be used to reduce peak time electricity consumption by evaluating the energy intensities of different steel qualities [21]. Even though the BF-BOF process is not considered as a batch process, this novel approach suggests viewing a sequence of steel qualities as a "batch", and using the energy intensity of different batches to reduce energy costs.

3. A COST MODEL FOR PRODUCTION PLANNING IN SECONDARY STEEL MAKING

3.1 Overview of the cost model for production planning

The cost model developed for this study forms part of a larger integrated production planning model, as presented in Figure 3. This paper only focusses on the development of a single initiative, and the implementation thereof. The integration between initiatives in the methodology of Figure 3 will not form part of the discussion or the implementation on the case study plant. The steps in the methodology that will be discussed are the identification, evaluation, implementation, and revision of a production planning initiative.

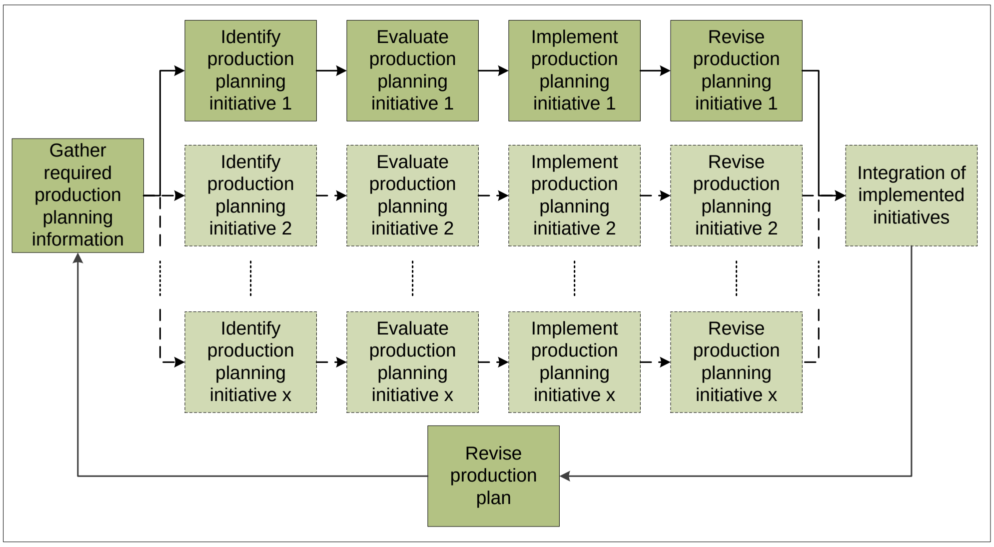


Figure 3: Cost model for production planning in secondary steel making

The output of the model is a set of "tools" for production planning to adapt the latest production schedule to reduce energy costs. It is important that other implicated costs be considered when implementing a solution. The steps in the methodology are discussed in more detail in this section. Before the steps are configured, a proper understanding of the production planning process of the specific facility is required. This includes continuously gathering the latest production planning information.

3.2 Identify production planning initiative

After obtaining the required production planning information, the first step in realising energy cost reduction is to identify an energy cost saving initiative. This can be done by means of various methods, such as investigating literature for existing solutions in other industries, or at other steel plants. In this case, the literature reviewed in section 2 led to the identification of an electrical load shift on the ladle furnaces by using the variation in electrical energy intensity of steel qualities. After an initiative has been identified, it has to be determined what the status of the initiative is on the facility.

It could be possible that the implementation of such an initiative has been attempted in the past, but is no longer in effect. This will serve as an indication of what the major risks are for implementing the solution. Alternatively, it has to be established why such an initiative has not been attempted before. This will include the identification of the risks and limitations of the initiative in a facility's unique conditions. Once the limiting factors have been identified, historic data should be collected to confirm the information obtained from discussions with the facility personnel.

Previous implementations of similar initiatives, and the risks and limitations that were identified, will serve as an indication of what data is required. The historic data that is collected should be representative of different conditions on the facility in order to be representative of the potential that exists for the initiative.

3.3 Evaluate production planning initiative

Once the initiative and its risks and limitations have been identified, it is important to evaluate the energy cost saving potential at the specific facility. The first part of this step is to evaluate the historic data that was collected. Different data analysis techniques can be used, depending on the conditions of the facility and the requirements for the initiative. From the evaluation of the data it should be possible to determine to what extent the initiative is already implemented on the plant. It can also be determined which conditions are favourable for optimal performance. This will serve as an indication of the theoretical potential benefit.

In this case, the electrical power profile will serve as an indication of the variation in electrical energy consumption throughout a typical day. Correlating this data with the steel qualities that were being processed at that instance will serve as an indication of the energy intensity of the steel quality. It has to be evaluated as to whether the energy intensity of a steel quality is the same on a recurring basis by identifying and evaluating several such instances.

After determining the theoretical potential benefit, the practical constraints that could prevent the implementation of the initiative need to be evaluated. In this case, energy intensity of steel qualities have never been considered as an input factor when performing production planning. Personnel was thus resistant towards such a solution, and the concept first had to be proven to them theoretically. This also led to the use of an awareness-based approach, rather than the development of an automated system.

The next factor that has to be considered is to determine the theoretical potential benefit. It is, however, important to make use of methods that can be used for real-time benefit quantification, as this will be required when the integration between initiatives is done. This quantification method also includes compiling a baseline for the evaluation of the initiative after implementation.

3.4 Implement production planning initiative

After it was determined whether scope exists for the implementation of the initiative on the facility, and the risks and limitations are known, the implementation step can be initiated. This consists of developing a solution and implementing it. The limitations and practical constraints that were identified for the initiative and the specific facility, will serve as a guide for the type of implementation approach that should be followed.

For most production planning initiatives, it is preferred by plant personnel to not make use of automated solutions. This is due to the complexity of scheduling and the human decision-making that is required. As part of the awareness-based approach production planners should be provided with assistive tools to use when performing production planning. The tools can be used to compare different scenarios, and flag opportunities for cost savings.

Additionally, the use of energy awareness is a useful tool to ensure that all parties are on-board with the implementation of the initiative. The awareness-based approach was also used to inform different parties of the benefits of the initiative, and to provide feedback on the progress and performance. Once all of the tools and assistive systems have been developed, it has to be implemented.

3.5 Revise production planning initiative

It is critical to regularly evaluate the performance of the initiative and to provide feedback to plant personnel. Discussions with the involved parties should also be used as inputs to adapt the system and the approach. The evaluation will make it possible to identify and adapt to practical limitations. Feedback on the progress and performance will also serve as a further awareness tool. The same evaluation techniques used in section 3.3 should be applicable when monitoring the performance of initiatives in this step.

4. CASE STUDY RESULTS

4.1 Facility description

The methodology was applied to a steel making facility. A simplified layout of the facility is presented in Figure 4. The facility makes use of the BF-BOF production process. The steel plant consists of three BOFs, a SecMet section with a vacuum degasser and two ladle furnaces, and ConCast with two casters. Depending on the steel quality requirements, liquid steel is either sent to the vacuum degasser from the BOF (referred to as “route 3”), or directly to the ladle furnaces (“route 2”).

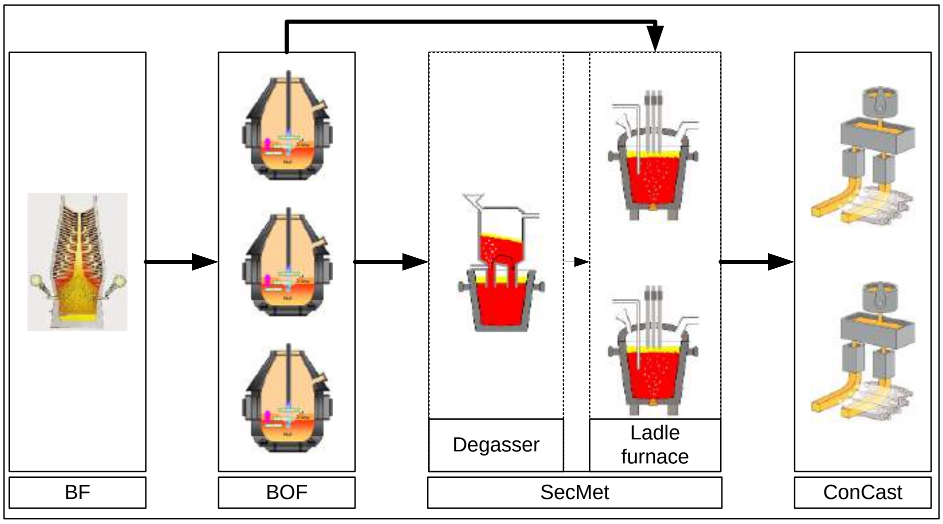


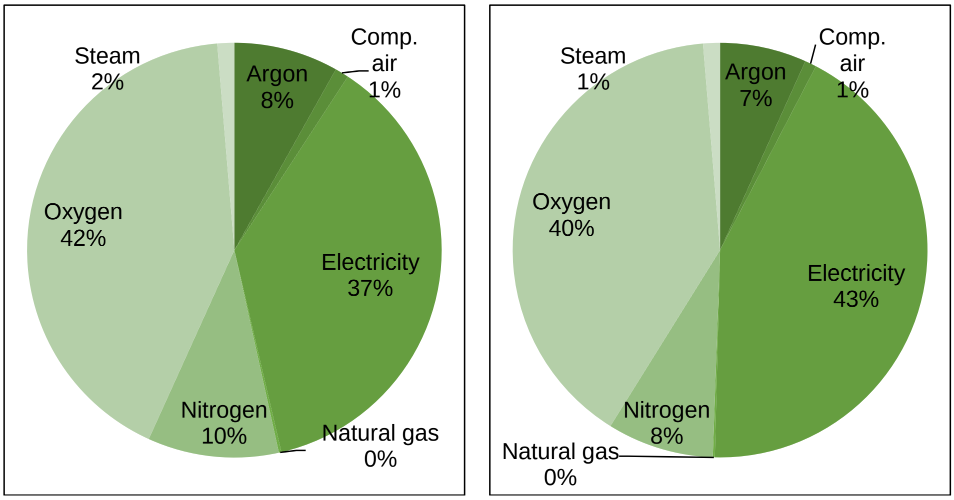
Figure 4: Basic layout for the case study plant

The facility makes use of time-of-use electricity tariffs, which vary depending on the time of day and the season. This makes it possible to perform an electrical load shift by reducing electrical energy consumption during more expensive peak times, and increasing it during the less expensive off-peak and standard times. The facility also consumes various other energy sources and utilities during this process such as steam, natural gas, compressed air, nitrogen, argon, and oxygen.

On this facility, a production planner is responsible for compiling a priority list three days in advance for the required steel qualities and its quantities. This is then provided to the scheduler, who is situated at ConCast. The scheduler is responsible for scheduling steel production throughout the day, ensuring that the steel is provided to ConCast on time, and that it is of the correct quality and temperature. The scheduler is in regular contact with the production planner and the operators in the different sections of the plant.

4.2 Practically identify production planning initiative

The first step of the methodology is the identification of an energy cost saving initiative. This was done from the literature review in section 2 of this paper. An electrical load shift on the ladle furnaces using the energy intensity of different steel qualities was identified. This focus is further justified by the analysis of the cost distribution of energy sources on the plant, as presented in Figure 5.



a.) Energy cost distribution for 2016 **b.) Energy cost distribution for winter 2016**

Figure 5: Steel plant energy cost distribution for case study facility

Figure 5.a presents the energy / utility cost distribution for 2016, from which it is seen that the largest cost contributor is oxygen (42%), followed by electricity (37%). In Figure 5.b, however, the winter months in 2016 are isolated, and it is seen that the higher electricity cost during these months increased the contribution of electricity to 43%, making it the largest cost consumer on the plant.

As per the methodology, the possibility of implementing the identified initiative was discussed with plant personnel. It was found that there has been some analysis done on the differences in electrical energy consumption for steel qualities, but it has not led to the implementation of any initiatives. There has also been several evaluations and initiatives to improve the efficiency of the ladle furnace electrodes. No attempts have, however, been made to reduce peak time electricity consumption.

Since no restricting factors could be identified from past projects, it was necessary to obtain sufficient data to properly assess the potential of the initiative. The data listed in Table 1 was collected for a period of 1 year. It is important to obtain a wide variety of available data to properly assess the potential of the initiative.

Table 1: Data collected for historic analysis of ladle furnace electrical load shift

Data collected	Description
Power consumption	Half-hourly power consumption of each ladle furnace
Steel quality description	Coded description of the steel qualities
Time spent at ladle furnace	Duration of heats at the ladle furnace
Ladle furnace used	Which ladle furnace was used
Date and time of heats	When the heat took place
Production route followed	Whether route 2 or route 3 production was used
Start temperature	Temperature at which steel was received at SecMet
End temperature	Temperature at which steel was required at ConCast

As part of the investigation of the process, it was found that a unique code is used for the identification of different steel qualities. The digits in the code represents specific requirements of the steel. The exact meaning of the different codes are withheld due to confidentiality reasons. Differences in steel quality requirements are expected to result in the variation of energy intensity for different qualities.

Thus, the idea is to use this to shift electrical load into off-peak periods by scheduling lower electrical energy intensive steel qualities in peak times. This had to be verified by an analysis of the available data in the next step of the methodology.

4.3 Practically evaluate production planning initiative

The collected historic data was analysed to determine whether enough variance existed between the electrical energy intensity of different steel qualities to verify the concept. An example of such an analysis is presented in Figure 6 in the form of frequency curves. The energy consumption of two random steel qualities is compared, from which it is seen that there is a correlation between steel quality and electrical energy consumption. This is typically due to differences in the additives (aluminium, silicon, etc.) and casting temperature requirements of different steel qualities. It is further seen that the difference in energy consumption for the two steel qualities is about 1.3 MWh per heat, providing sufficient variation to implement a load shift initiative.

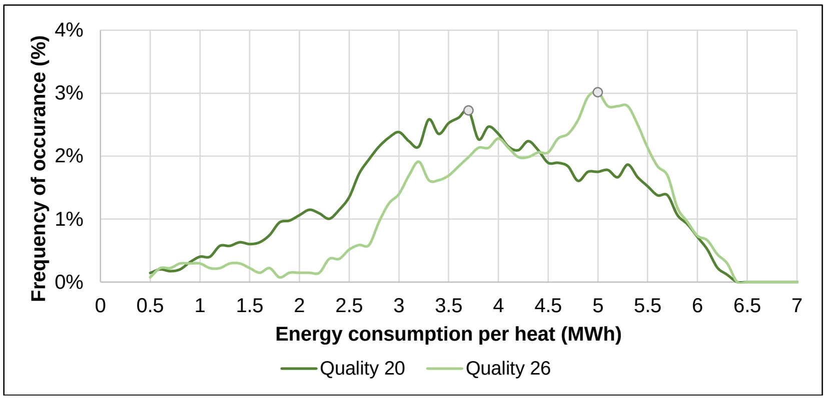


Figure 6: Frequency graphs indicating the difference in electrical energy intensity for two steel qualities at the ladle furnaces

The average electrical energy consumption at the ladle furnaces was determined for each steel quality that was produced in the past year, from which a database of electrical energy intensities for steel qualities was compiled. This database was used to plot the expected power profile and related energy cost for a planned schedule on a given day. An example is presented in Figure 7 for a randomly selected day.

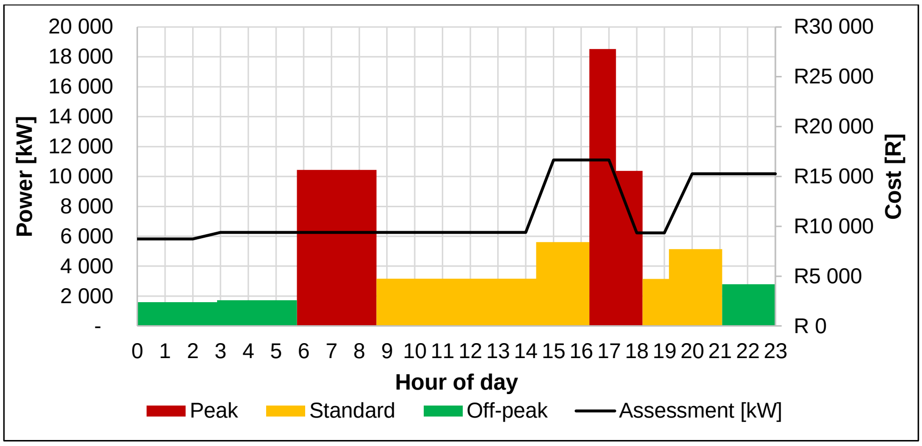


Figure 7: Original schedule compiled by schedulers

By considering the energy consumption of steel qualities, the planned schedule was adapted manually. The purpose was to attempt to reduce peak time energy consumption by switching around steel qualities throughout the day without affecting the day's production outputs. The main practical constraints that were identified and had to be taken into consideration were the level of liquid iron available, and that a sequence of a steel quality could not be interrupted at ConCast.

The power profile and related cost for the manually adapted schedule is presented in Figure 8. By using the winter electricity tariffs for the specific plant, the maximum theoretical potential cost reduction for the specific day was calculated as about R 35 000. This verified that the reduction of energy cost by scheduling steel qualities based on their energy intensities, is theoretically possible.

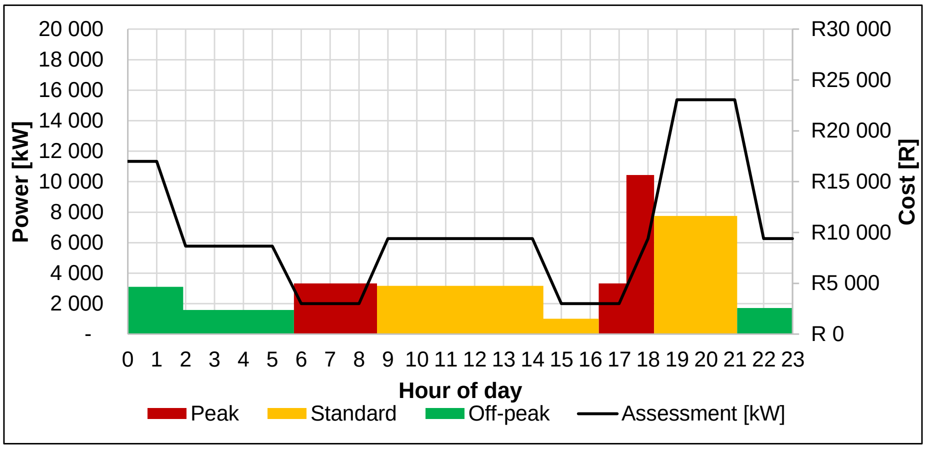


Figure 8: Suggested adapted schedule based on the same priority list

In order to quantifying the benefit of the initiative practically, it was decided to make use of average daily power profiles for weekdays, Saturdays, and Sundays, as presented in Figure 9. Average power profiles for a 6 month period were used due to the lack of a repetitive power profile in the past. This is due to the energy intensity of steel qualities historically not forming part of production planning inputs.

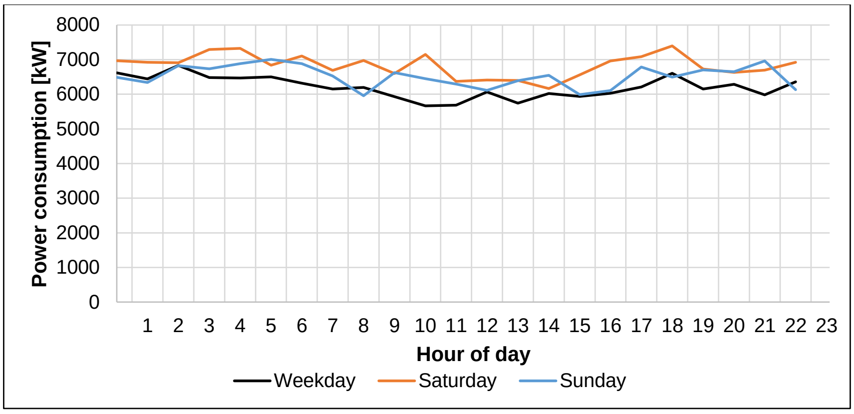


Figure 9: Average daily power profiles used for baseline

It is recommended that the baseline methodology make use of an adjustment method to account for differences in production during the baseline and assessment periods. This is, however, out of the field of study for this paper, and is not discussed in detail.

4.4 Practically implement production planning initiative

As previously discussed, the initiative was implemented using an awareness-based approach. The focus was on providing schedulers with the required tools to adapt production schedules to take the electrical energy intensity of steel qualities into account. The first step was thus to inform all of the involved parties of the initiative, and to provide production planners and schedulers with the necessary information to make informed decisions.

The database of calculated energy intensities for different steel qualities was used to classify the qualities into preferred time-of-use periods. Colour coding was used as a quick reference to determine which steel qualities to schedule in which period. A steel quality indicated in red was considered to be safe to produce in peak ("red") time. The sheet provides an alphabetical list to easily find steel qualities, as well as categorised lists to indicate the order of preference.

Regularly produced steel qualities were split from less-regularly produced qualities to simplify the reference process. This dataset sheet enable schedulers to easily determine what steel qualities are preferred during specific times of the day. An example of the summary for route 2 steel qualities is presented in Figure 10. A similar summary page was provided for route 3 steel qualities.

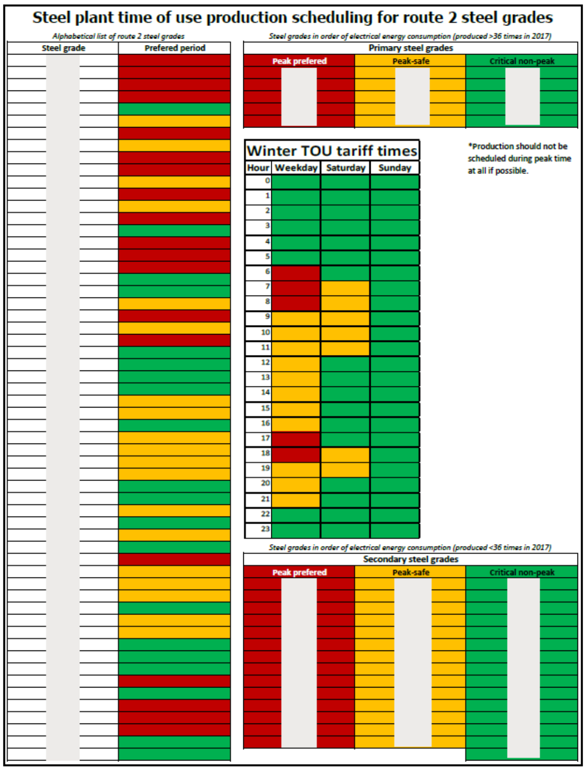


Figure 10: Example database of steel quality energy intensity at ladle furnaces

Additionally, an online interface was developed for schedulers to use as a tool to estimate the effect of changes that can be made in advance. The interface used the blast furnace production rate, current liquid iron level, and current production schedule as inputs. A second production schedule could then be compared to the original schedule to evaluate what the effect of changes would be. Figure 11 provides an example of the inputs required by the system.

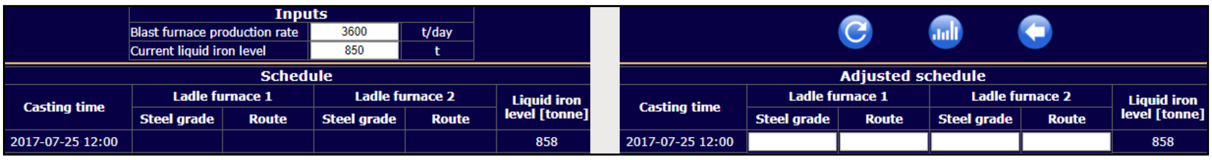


Figure 11: Example of inputs for online interface used for comparing scenarios

An example of the output provided by the system is presented in Figure 12. The output compares aspects of the original and revised schedules. These aspects include the predicted liquid iron level, expected energy cost distribution over the time-of-use periods for the next three days, and the total estimated energy cost. The scheduler can thus estimate the effect of changes in the schedule in advance, and pro-actively adapt the schedule to reduce the cost of steel making.

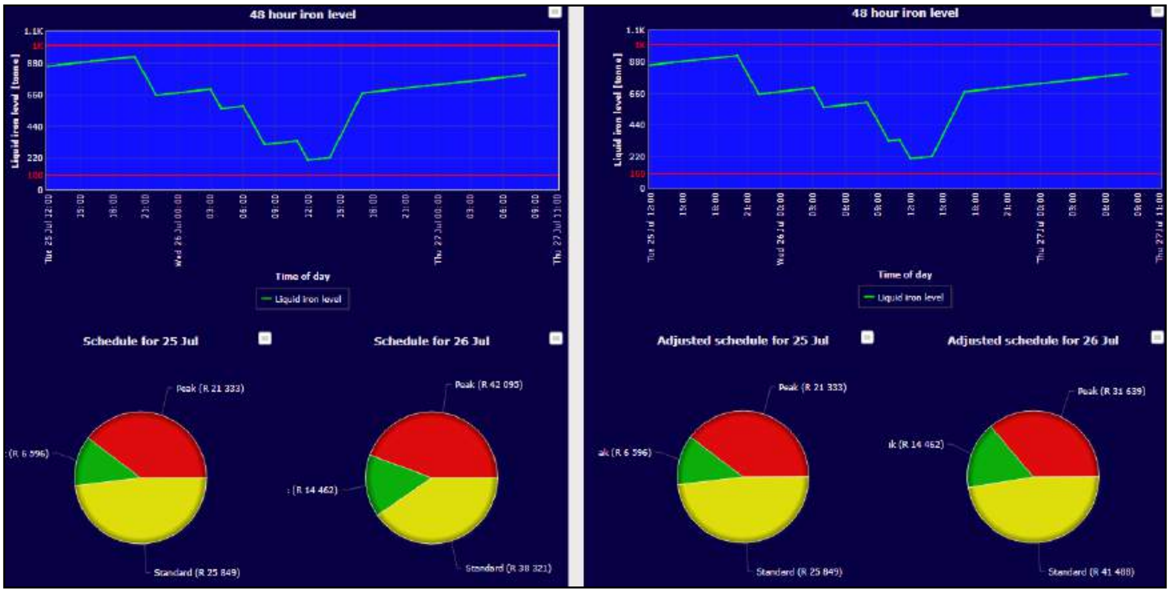


Figure 12: Example of online interface output

The project was implemented for a three month period, from 1 July 2017 to 30 September 2017 (two winter months and one summer month). The effect was evaluated on a weekly basis by using the discussed baseline approach. An example of the weekday profile for one week of the implementation is presented in Figure 13. This figure provides the weekday baseline profile, the actual profile, and the adjusted (scaled) baseline profile.

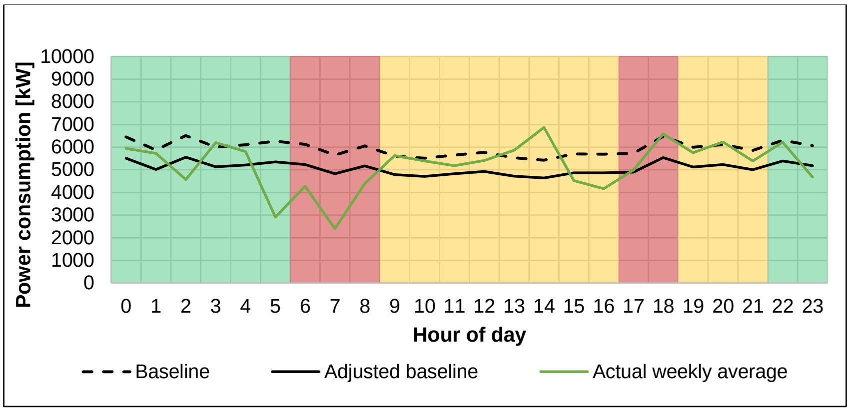


Figure 13: Example of a successful load shift and adjusted baseline

It is seen from this result that the actual power consumption reduced during the morning peak time, leading to a successful energy cost reduction by means of a load shift. This method was applied to every week in the assessment period, calculating the weekly energy cost savings. During this time a total cost saving of about R 500 000 was achieved, as indicated in Table 2 for each month.

Table 2: Achieved energy cost savings

Month	Energy cost saving
July 2017	R 240 000
August 2017	R 155 000
September 2017	R 105 000
Total	R 500 000

From the monthly energy cost savings in Table 2, it is seen that the achieved benefit during September 2017 was significantly less than during the two preceding months. This is due to the reduced summer billing tariffs for this month having a smaller difference between off-peak and peak time cost. It is estimated that the annual effect of this initiative is about R 1.5-million, by considering the results for the different seasonal tariffs in the implementation.

4.5 Practically revise production planning initiative

The final step of the methodology is to revise the implemented system and assistive tools. During this step, the performance of the initiative was evaluated and discussions with plant personnel were used to adapt the system accordingly. Revision of the implementation strategy took place continually throughout the process in order to take the experiences and feedback of plant personnel into account.

Feedback to plant personnel was provided in the form of weekly performance reports, indicating the effect that the initiative had had over the past week. The main indication in this report was the resulting power profile presented in Figure 13. Production schedulers also received daily feedback reports to evaluate the performance for the previous day and week-to-date.

From the revisions, limitations were identified and the approach was adapted accordingly. The most notable adaptions to the approach included:

- the separation of route 2 and route 3 steel qualities in the database sheets;
- the inclusion of the blast furnace production rate and liquid iron level in the online interface; and
- the addition of daily feedback reports for schedulers.

These changes were made at various times during the implementation, and were made possible by properly monitoring performance on a continual basis.

5. DISCUSSION OF RESULTS

The results indicate that it is possible to make use of production planning to reduce the cost of energy on a steel plant. A sequence of steel qualities was considered as a "batch", and the suggested approach by Merkert *et al.* in section 2 was used in a novel way to schedule batches according to their energy intensity [21]. This resulted in an electrical energy cost reduction of about R 500 000 (over three months). This unique approach makes it possible to perform a load shift on a ladle furnace in a steel plant using the BF-BOF production process.

The methodology is designed to integrate various production scheduling initiatives, while this paper only focusses on the implementation of one such an initiative. The result is used to validate the use of the individual solution per initiative before integrating multiple initiatives. It was found that production scheduling takes various factors into account, such as production rates, the available liquid iron levels, steel orders, etc., which have to be considered as part of the solution. The focus of this study was to include energy consumption as one of these factors.

6. CONCLUSION

This paper presents the development and implementation of an initiative for reducing energy cost in steel making by means of production planning. The discussed methodology forms part of an integrated cost model. The model is practically implemented on the ladle furnaces of the SecMet section of a steel plant in the BF-BOF production process, with the purpose of managing the power profile around time-of-use electricity tariffs. This makes it possible to reduce the energy costs without affecting production.

An awareness-based approach is used for the implementation, resulting in a cost reduction of R 500 000 over a three month period, and an estimated annual effect of R 1.5-million. The main focus is on scheduling steel qualities based on their electrical energy intensity. This unique implementation validates the novel method used for implementing a cost saving initiative by adapting production planning.

REFERENCES

[1] **W. J. Breytenbach, H. G. Brand and J. M. Marais** 2017. "Framework for operational changes to reduce electricity costs in the South African steel industry," in 2017 International Conference on the Industrial and Commercial Use of Energy (ICUE), Cape Town.

[2] **S. G. van Niekerk, W. J. Breytenbach and J. H. Marais** 2017. " Developing an optimisation model for industrial furnace gaseous fuel distribution for energy cost savings," in 2017 International Conference on the Industrial and Commercial Use of Energy (ICUE), Cape Town.

[3] **G. H. Popescu, E. Nica, R. O. Stefamescu-Mihaila and G. Lazaroiu** 2016. "The United States (U.S.) steel import crises and the global production overcapacity till 2016," Metalurgija, vol. 3, no. 55, pp. 538-540.

[4] **S. Njobeni** 2017. "Duty to protect SA steel industry," BusinessReport, 9 May 2017.

[5] **International Trade Administration Commission of SA** 2016. "Monitoring and Analysis of South African Steel Imports," Trade Monitor.

[6] **L. Prinsloo and K. Crowley** 2017. "ArcelorMittal SA plans to raise R3.5bn in debt," Fin24, 6 April 2017.

[7] **World steel association** 2017. "World steel in figures 2017," worldsteel.org, Beijing.

[8] **S. Nkabinde** 2015. "Dumping, China, and ArcelorMittal," Moneyweb, 20 July 2015.

[9] **R. Wilkinson** 2016 "Import prices on steel not the best solution," Creamer media's Engineering News, 1 July 2016.

[10] **T. Heiberg** 2017. "ArcelorMittal South Africa loss deepens," Moneyweb, 27 July 2017.

[11] **R. A. Dondofema, S. Matope and G. Akdogan** 2017. "South African iron and steel industrial evolution: An industrial engineering perspective," South African Journal of Industrial Engineering, vol. 28, no. 4, pp. 1-13.

[12] **K. He and L. Wang** 2017. "A review of energy use and energy-efficient technologies for the iron and steel industry," Renewable and Sustainable Energy Reviews, vol. 70, pp. 1022-1039.

[13] **World Steel Association** 2018. "Fact Sheet: Energy use in the steel industry," worldsteel.org.

[14] **Asia Pacific Partnership for Clean Development and Climate** 2010. The State-of-the-Art Clean Technologies (SOACT) for Steelmaking Handbook, 2nd Edition, Asia Pacific Partnership for Clean Development and Climate.

[15] **J. Bazan and J. Kret** 2015. "Iron and Steelmaking," Technical University of Ostrava, Ostrava.

[16] **M. Makela** 2012. "Assessing potential forest and steel inter-industry residue utilisation by sequential chemical extraction," Aalto University, Doctor of Science in Technology thesis, Espoo.

[17] **L. E. Zeelie** 2016. "Compressed air energy savings on an iron production plant," North-West University, Magister in Mechanical Engineering thesis, Potchefstroom.

[18] **M. Yi and J. Mao-fa** 2010. "Exergy analysis and optimisation of ladle furnace refining process," Journal or iron and steel research, International, vol. 17, no. 11, pp. 24-28.

[19] **T. Yuan-yuan, H. Ying-lei and L. Shi-xin** 2013. "Two-stage mathematical programming approach for steelmaking process scheduling under variable electricity price," Journal of iron and steel research, International, vol. 20, no. 7, pp. 1-8.

[20] **L. Van Dyk** 2014. "A census of South African industrial engineers, based on data extracted from Linkedin," in SAIIE26 Proceedings, Muldersdrift.

[21] **L. Merkert, I. Harjunkoski, A. Isaksson, S. Saynevirta, A. Saarela and G. Sand** 2015. "Scheduling and energy – Industrial challenges and opportunities," Computers and Chemical Engineering, vol. 72, pp. 183-198.

[22] **C. Gahm, F. Denz, M. Dirr and A. Tuma** 2015. "Energy-efficient scheduling in manufacturing companies: A review and research framework," European Journal of Operational Research, vol. 248, pp. 744-757.

[23] **M. Rager, C. Gahm and F. Denz** 2015. "Energy-oriented scheduling based on Evolutionary Algorithms," Computers and Operations Research, vol. 54, pp. 218-231.

[24] **J. A. Swanepoel** 2012. "Modelling for integrated energy optimisation in cement production plants," North-West University, Master of Engineering in Mechanical Engineering dissertation, Potchefstroom.

[25] **M. Lochmuller and G. Schembecker** 2016. "Simultaneous optimization of scheduling, equipment dimensions and operating conditions of sequential multi-purpose batch plants," Computers and Chemical Engineering, vol. 94, pp. 157-179.

[26] **M. Biondi, G. Sand and I. Harjunkoski** 2017. "Optimization of multipurpose process plant operations: A multi-time-scale maintenance and production scheduling approach," Computers and Chemical Engineering, vol. 99, pp. 325-339.

[27] **M. E. Moshidi** 2014. "An analysis of effective maintenance planning at a steel manufacturer," North-West University, Masters in Business Administration mini-dissertation, Potchefstroom.

[28] **E. Worrel, P. Blinde, M. Neelis, E. Blomen and E. Masanet** 2010. "Energy efficiency improvement and cost saving opportunities for the US iron and steel industry: An ENERGY STAR guide for energy and plant managers," Energy Analysis Department, Environmental Energy Technologies Division, University of California, Berkeley, CA.

[29] **Department of Trade and Industry** 2011. "Introduction and implementation of an energy management system and energy systems optimization," Department of Trade and Industry, Pretoria.

[30] **R. Remus, M. A. Monsonet, S. Roudier and L. D. Sancho** 2013. "Best available techniques (BAT) reference document for iron and steel production," European Commission JRC Reference report, Seville, Spain,.

[31] **Q.-K. Pan, Q.-d. Chen, T. Meng, B. Wang and L. Gao** 2017. "A mathematical model and two-stage heuristic for hot rolling scheduling in compact strip production," Applied Mathematical Modelling, vol. 48, pp. 516-533.

[32] **S.-l. Jiang, M. Liu, J.-h. Lin and H.-x. Zhong** 2016. "A prediction-based online soft scheduling algorithm for the real-world steelmaking-continuous casting production," Knowledge-Based Systems, vol. 111, pp. 159-172.

[33] **M. H. Fazel Zarandi and P. Ahmadpour** 2009. "Fuzzy agent-based expert system for steel making process," Expert Systems with Applications, vol. 36, pp. 9539-9547.

[34] **B. Karwat** 2012. "Optimization of production schedules in the steel production system," Archives of civil and mechanical engineering, vol. 12, pp. 240-252.

[35] C. Xu, G. Sand, I. Harjunkoski and S. Engell 2012. "A new heuristic for plant-wide schedule coordination problems: The intersection coordination heuristic," Computers & Chemical Engineering, vol. 42, pp. 152-167.

[36] H. Hadera, I. Harjunkoski, G. Sand, I. E. Grossmann and S. Engell 2015. "Optimization of steel production scheduling with complex time-sensitive electricity cost," Computers and Chemical Engineering, vol. 76, pp. 117-136.

CONTINUOUS EVALUATION OF OPERATIONAL RISKS ON DEEP-LEVEL MINE EQUIPMENT

S. van Jaarsveld[1], W van Blerk[2], J.H. Marais[3] & P. Goosen[4]

North-West University, CRCED-Pretoria, Pretoria, South Africa
[1] svanjaarsveld@researchtoolbox.com
[2] wvanblerk@researchtoolbox.com
[3] 10728023@nwu.ac.za
[4] pgoosen@researchtoolbox.com

ABSTRACT

Underground mining operations comprise complex systems that provide the service areas with cold water, compressed air and ventilation. Equipment is located on the surface and underground, which makes the monitoring thereof a challenging task.

A condition monitoring process on deep mines involves several types of parameters and equipment. Software tools and applications are used to analyse the raw data and identify operational risks. This automated analysis results in a substantial amount of risk information being generated on a regular basis. It is therefore necessary to examine the information for knowledge discovery to take place.

A software-based application was developed to categorise the risk information according to system class, parameter type and risk severity over a selected date range. The solution enables site managers to determine where critical risks occur repeatedly and what the maintenance impact is.

The newly developed application was used to evaluate the risk information of six mining sites over a period ranging from six to 12 months. Up to 96 daily risk identifiers per site were evaluated resulting in more than 17 000 values over a period of six months. The application and associated reports facilitated the identification of problem areas within the respective operation.

This work was sponsored by ETA Operations Pty (Ltd).

[1] Dr. Stephan van Jaarsveld holds a PhD in Computer and Electronic engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Mr. Wihan van Blerk completed his Bachelor's degree in Mechanical engineering at the University of Pretoria. He is currently enrolled for a Master's degree at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr. Johan Marais is a registered professional engineer and holds a PhD in Electrical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[4] Mr. Pieter Goosen is a registered professional engineer and is currently enrolled for a PhD at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

Too much data and too little information is a common occurrence in the 21st century. Enhanced hardware and software have given users access to greater processing power, immense storage capacity and advanced analysis methods [1]. Many industries are finding it difficult to gain insights from the data they now have at their disposal [2-4]. One example is the challenge that the mining industry is facing in the field of preventative maintenance. Effective maintenance strategies play an integral part in system availability. On deep-level mines, equipment up-time is vital to ensure uninterrupted production and adherence to safety regulations [5]. A condition monitoring system was therefore developed to assess the operational condition of mine equipment on a daily basis [5]. The risk assessment is based on the *SCRF* methodology developed by *Van Jaarsveld* [5, 6]. The methodology will be briefly discussed to provide some context.

Four regions of operation namely, safe, caution, risk and failure (SCRF) are used to characterise each input signal. The boundary for each region depends on equipment attributes such as type and size. A safe vibration amplitude for a large dewatering pump will differ from that of a small booster pump or a ventilation fan. Parameter specific boundaries are therefore used to assess each individual input.

The system performs a daily assessment to determine a risk score, and corresponding risk category, for each parameter. The risk category is translated to a health status level (1, 2 or 3) which the online platform converts to a colour marker (or format). Figure 1 shows the relationship between the risk category, health status and online formatting.

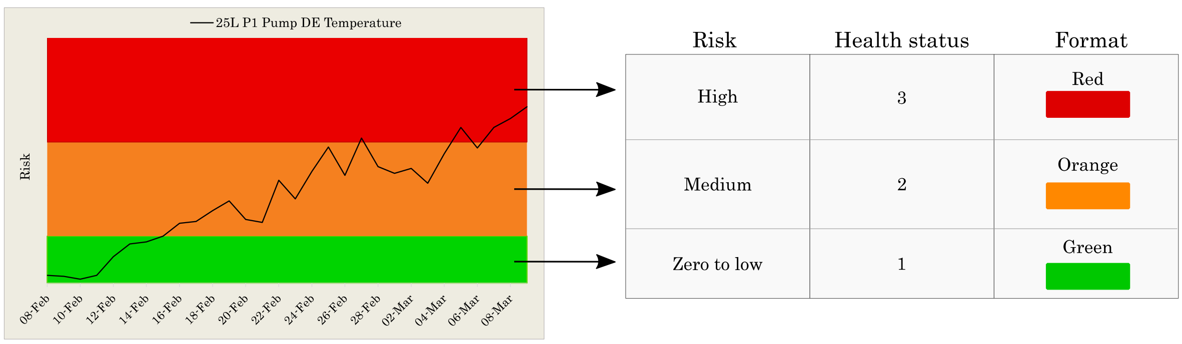


Figure 1: Risk category and health status relationship [5]

Parameters are grouped together, based on the type of measurement. As seen in Figure 2 single formatting value can thus be assigned to each parameter type according to the health status of the respective parameters. This is achieved by using the highest health status value from the input list.

Parameter	Health status	Format
Main Fan 1 - Fan DE Temp	1	Temp format: Main Fan 1 Orange
Main Fan 1 - Fan NDE Temp	2	
Main Fan 1 - Motor DE Temp	1	
Main Fan 1 - Motor NDE Temp	1	

(A)

Parameter	Health status	Format
Main Fan 1 - Fan DE Vibration	1	Vibration format: Main Fan 1 Red
Main Fan 1 - Fan NDE Vibration	1	
Main Fan 1 - Motor DE Vibration	3	
Main Fan 1 - Motor NDE Vibration	2	

(B)

Figure 2: Process to determine format label [5]

The same procedure is followed to determine a format label for each system, for example, the maximum health status of all the fans is used to format the collective fan system. The risk information is compiled in automated reports and made available on an online platform. A screenshot of the online platform's overview page can be seen in Figure 3.

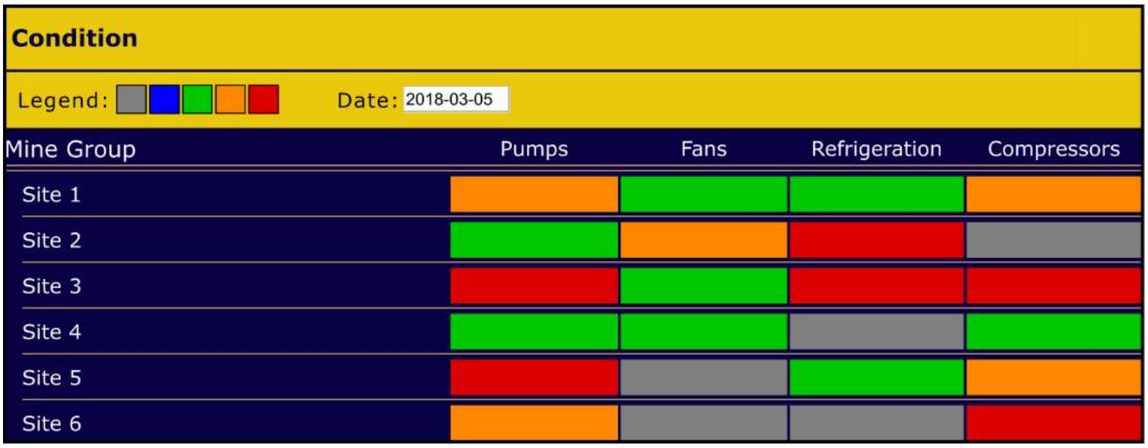


Figure 3: Online platform overview page

The risk information shown here are daily risk indicators, and therefore, over longer periods of time, it becomes overwhelming and difficult to evaluate manually. It was therefore necessary to develop a software application capable of interpreting and categorising the risk information.

2. BACKGROUND

Deep-level gold and platinum mines can reach depths of up to 4 km below the surface [7]. Successful ore extraction at these depths require complex systems operating together. These systems include compressed air, cooling, dewatering, hoisting and ventilation systems [8]. These operation-critical systems are typically distributed across the mine, from the surface to the deepest levels [9]. The physical conditions in a typical mining environment, particularly at deep levels, can be very harsh [5].

2.1 Need for maintenance

The harsh conditions in deep-level mines holds multiple health and safety risks for mining personnel. Health and safety risks include, hot air, polluted air and flooding. Multiple systems are relied upon for mitigation of these risks.

The high virgin rock temperatures at these depths cause high air temperature, while mining equipment's exhaust fumes pollutes the air [10–12]. Cooling systems are typically situated on the surface, but can be installed underground if deemed necessary [9, 11]. The cooling systems dehumidify and cool ambient air, which is then fed down the mine to deeper levels via the ventilation systems [13]. Cooling and ventilation systems are therefore integrated to achieve safe air temperatures and safe atmospheric vapour compositions in underground mining environments [14].

In underground mining operations, water is constantly released from subsurface fractures (fissure water). Water is also sent down the mining levels for cooling and mining purposes [9]. These large volumes of water create the risk of flooding in underground mining operations, especially at deeper levels. Multistage pumping systems mitigate this risk by continuously pumping water to the surface [9, 15–17].

Operational systems are not only responsible for safe working conditions. Compressed air systems power mining equipment such as pneumatic drills, mechanical ore loaders and refuge bays [18]. Equipment availability is a major contributor to a mine's productivity [19, 20].

Equipment up-time and reliability in deep-level mines are thus critical for workplace safety and production and can be largely affected by the maintenance strategy applied. Ill-maintained equipment tends to develop malfunctions and will eventually break down [21]. Maintenance after equipment failure (reactive maintenance) typically results in extended equipment downtime and increased maintenance compared to maintenance before equipment failure (proactive maintenance) [19, 22, 23].

Maintenance costs of mining equipment vary depending on the type of the equipment and the maintenance strategy used. These costs can amount to 25-40% of overall equipment costs (including procurement and operational costs) [24, 25]. Since mines are on a tight budget, avoidable operational and capital expenditures should be kept to a minimum [10, 26]. These expenditures include component replacement cost, secondary damage to additional components in vicinity and loss of production due to equipment downtime. A cost optimised maintenance strategy is therefore crucial in mines.

Maintenance strategies can be divided into 5 types [19]:

- **Breakdown maintenance**: Repairs to failed or broken-down equipment. This typically entails replacement of components.
- **Corrective maintenance**: Upgrades to components with a flawed design to improve reliability.
- **Preventative maintenance**: Aim is to prevent equipment failure. Two sub-categories are described below:
    - *Time-based maintenance (TBM)*: Maintenance done on a predetermined schedule.
    - *Condition-based maintenance (CBM)*: Maintenance done based on the current condition of equipment.
- **Reliability-centred maintenance**: Only perform preventative maintenance on parts which directly affect overall system reliability.
- **Total productive maintenance**: In addition to production responsibilities, operators are responsible for reporting on maintenance needs. The aim is to maximise production, while maintaining equipment reliability. This strategy requires highly trained and knowledgeable operators.

Preventative maintenance strategies are desirable as they aim to avoid expensive repair and replacement costs and unplanned downtime [24, 27]. CBM's statistical analysis relies more on actual data, while TBM's statistical analysis relies more on theoretical equipment life cycles. CBM is therefore preferred over TBM as it is more practical and often more accurate than TBM [28].

2.2 Condition monitoring and CBM

Condition monitoring is a tool to assist with CBM. Condition monitoring involves data acquisition and data processing. Data acquisition consists of constant measurement and logging of equipment's operational parameters. Data processing involves analysis of the logged data, with the purpose of detecting changes in the operational parameters that may be indicative of a potential fault. CBM can utilise this condition monitoring information to make maintenance decisions [22, 29].

Various specialised condition monitoring methods can be used in deep-level mines. These methods include vibration frequency monitoring (spectral analysis), acoustic emission monitoring, visual monitoring, oil particle analysis and infrared thermography [30–32]. These specialised condition monitoring methods are expensive to implement and tend to require highly skilled personnel to utilise and maintain. Basic protection systems are used to prevent machines from operating outside of their design specification. Input values are continuously compared to a relevant trip limit. These systems will therefore initiate an automatic trip sequence when a violation occurs. The measurements are typically only used on a real-time basis to ensure safe operation instead of being used for monitoring purposes. However, the available data from these systems can be used to develop simpler and cheaper methods for condition monitoring [33].

Basic protection system's data can be analysed to determine mining equipment health and to identify operational risks. This simplified method of condition monitoring was used to successfully apply CBM in the mining industry [5, 6, 33]. This condition monitoring method is useful for daily CBM decision making, but lacks long-term tracking of operational risks.

Long-term tracking of operational risks can provide valuable insight into the maintenance strategy performance. Literature on performance tracking of maintenance strategies is limited [34]. A SWOT (strength, weakness, opportunity and threat) analysis was done on maintenance strategies in the manufacturing industry. The analysis indicated multiple benefits of evaluating an organisation's maintenance strategy. The efficiency of the maintenance strategy was improved. The analysis assisted with improving the adaptability of the maintenance strategy to the changing business environment. The analysis also assisted with the development of contingency plans [34].

3. SOLUTION DEVELOPMENT

An application was developed to categorise the risk information according to system class, parameter type and risk severity. Considering the system class information, four mining systems were available. These systems include Pumps (P), Compressors (C), Refrigeration plants (R) and Ventilation fans (F). Two parameter types were available via the basic protection system, namely temperature (T) and vibration amplitude (V). Depending on instrumentation availability and equipment type, Table 1 lists the parameters that were considered for analysis.

Table 1: Parameters used for analysis

Pump drive end (DE) temperature	Pump drive end (DE) vibration
Pump non-drive end (NDE) temperature	Pump non-drive end (NDE) vibration
Motor drive end (DE) temperature	Motor drive end (DE) vibration
Motor non-drive end (NDE) temperature	Motor non-drive end (NDE) vibration
Fan drive end (DE) temperature	Fan drive end (DE) vibration
Fan non-drive end (NDE) temperature	Fan non-drive end (NDE) vibration
Compressor drive end (DE) temperature	Compressor drive end (DE) vibration
Compressor non-drive end (NDE) temperature	Compressor non-drive end (NDE) vibration
Gearbox drive end (DE) temperature	Gearbox drive end (DE) vibration
Gearbox non-drive end (NDE) temperature	Gearbox non-drive end (NDE) vibration

The notation that will be used consists of the relevant system identifier, followed by a subscript of the relevant parameter type identifier. The pumping system's vibration risk information can therefore be denoted as P_V. Figure 4 displays the various system classes and parameter types that were evaluated and how they can be combined for different types of analyses. For example, a user can choose to view the combined compressor risk information for all the parameter types (C_T and C_V).

The risk analysis is performed daily and the corresponding risk information is added to a database. Three levels of severity are used, namely low (1), medium (2) and high (3). The medium-risk occurrences are given a warning label, while the high-risk occurrences are given a critical label. It is therefore possible to only view the critical count, or view both the warning and critical counts when selecting the system and parameter types.

A few simple steps are followed to configure and generate a new report. A script-based application provides the user with several options to compile the required information. The setup procedure can be summarised as follows:

- The user exports the relevant risk information for the required date range.
- The user selects whether only critical counts are included.
- The user selects daily or monthly graphing intervals.
- The user adds charts by defining a system and parameter type for each chart.

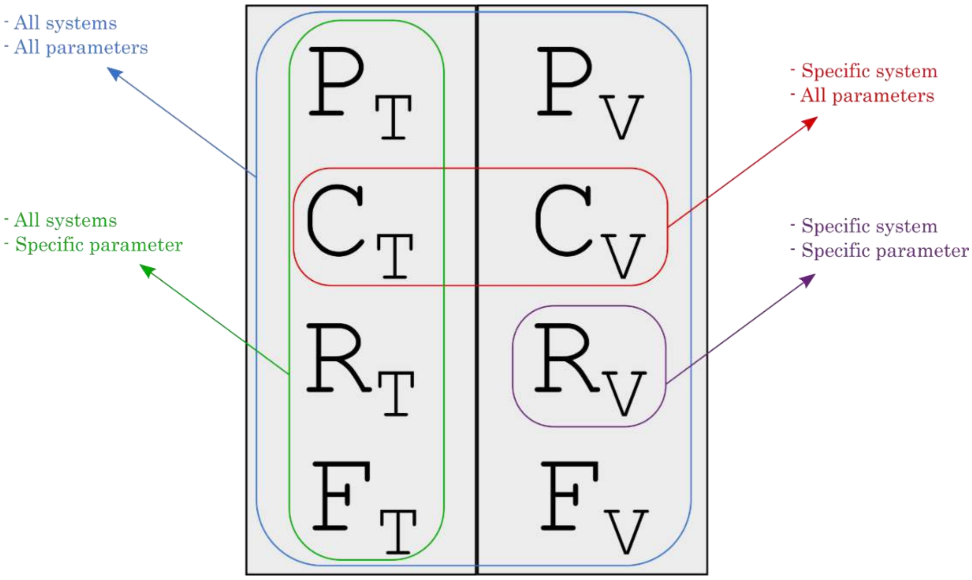


Figure 4: System classes and parameter types

Figure 5 is an example of risk information that was exported for an arbitrary number of input tags. The different types of evaluation options are illustrated in the figure. The evaluation options determine which of the input values are included in the analysis. For example, if a parameter type evaluation is performed, all the input values corresponding to the selected parameter type (highlighted in blue) will be considered irrespective of the system class. Each of the analyses are performed over the entire date range to determine a warning and critical count for each day. The results can then be compiled in monthly totals.

Tag name	01-Jan	02-Jan	03-Jan	04-Jan	05-Jan		26-Aug	27-Aug	28-Aug	29-Aug	30-Aug	31-Aug
Pump_1_Risk_Temp	1	2	1	2	1		3	1	1	1	1	3
Pump_1_Risk_Vib	2	1	2	1	2		1	1	2	2	2	3
Pump_2_Risk_Temp	3	3	1	2	2		1	3	2	3	2	1
Pump_2_Risk_Vib	1	2	2	1	3		2	1	2	3	2	2
Pump_3_Risk_Temp	3	1	3	1	3		1	3	3	3	3	1
Pump_3_Risk_Vib	3	1	3	2	2		3	3	2	3	2	3
Pump_4_Risk_Temp	3	2	2	2	2		3	1	3	1	1	1
Pump_4_Risk_Vib	3	1	1	3	3		2	3	2	1	1	3
Pump_5_Risk_Temp	2	1	2	1	2		3	2	1	2	2	2
Pump_5_Risk_Vib	2	1	3	2	3		2	2	1	2	1	2
Comp_1_Risk_Temp	2	2	3	3	3		3	3	3	1	1	2
Comp_1_Risk_Vib	1	1	1	1	3		3	3	1	2	1	2
Comp_2_Risk_Temp	1	2	1	2	2		3	1	2	1	3	3
Comp_2_Risk_Vib	3	3	1	1	2		3	3	1	3	2	1
Comp_3_Risk_Temp	2	3	3	3	1		3	3	3	1	1	1
Comp_3_Risk_Vib	1	3	1	1	2		3	1	1	1	2	1
Comp_4_Risk_Temp	2	2	1	3	2		3	3	2	2	2	1
Comp_4_Risk_Vib	3	1	2	2	1		1	1	3	2	1	3
Fan_1_Risk_Temp	3	3	2	1	2		2	1	3	2	2	1
Fan_1_Risk_Vib	2	2	2	2	2		2	3	2	3	2	3
Fan_2_Risk_Temp	1	1	3	2	1		2	2	3	3	2	1
Fan_2_Risk_Vib	1	2	2	2	1		2	3	2	1	1	3

Parameter type evaluation

System evaluation

Parameter and system evaluation

Figure 5: Risk information with evaluation options

Three charts are added to the report by default. For illustration purposes, both the critical and warning counts will be included in all the examples given. The first chart (Figure 6) displays the totals of all the systems with both parameter types. The second chart (Figure 7) provides a distribution between the relevant systems, while the third chart (Figure 8) provides a distribution between the parameter types. The user can add additional charts as needed e.g. a system evaluation of the compressed air system (Figure 9).

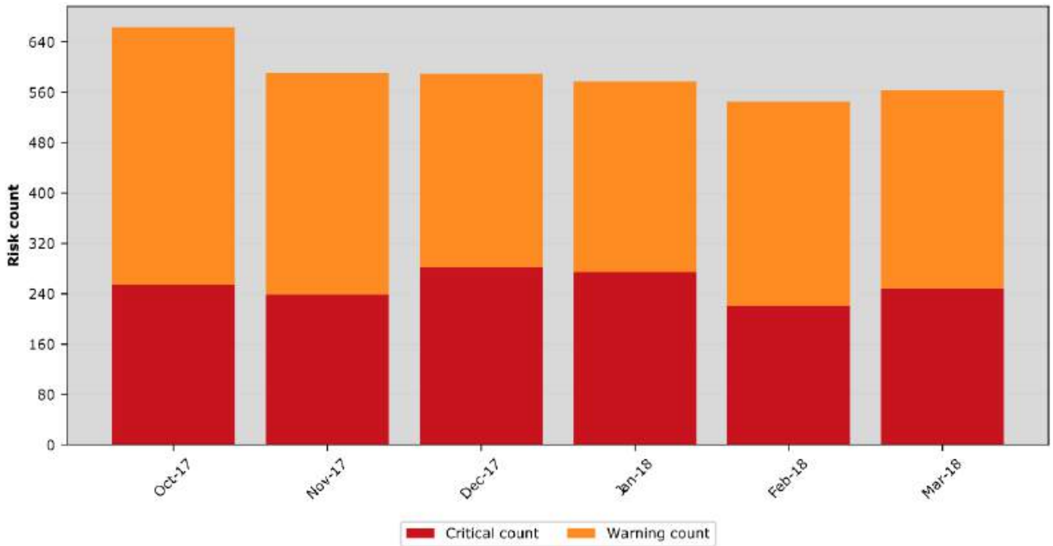


Figure 6: Risk information overview

Figure 7 displays the risk distribution between the four major systems that were evaluated.

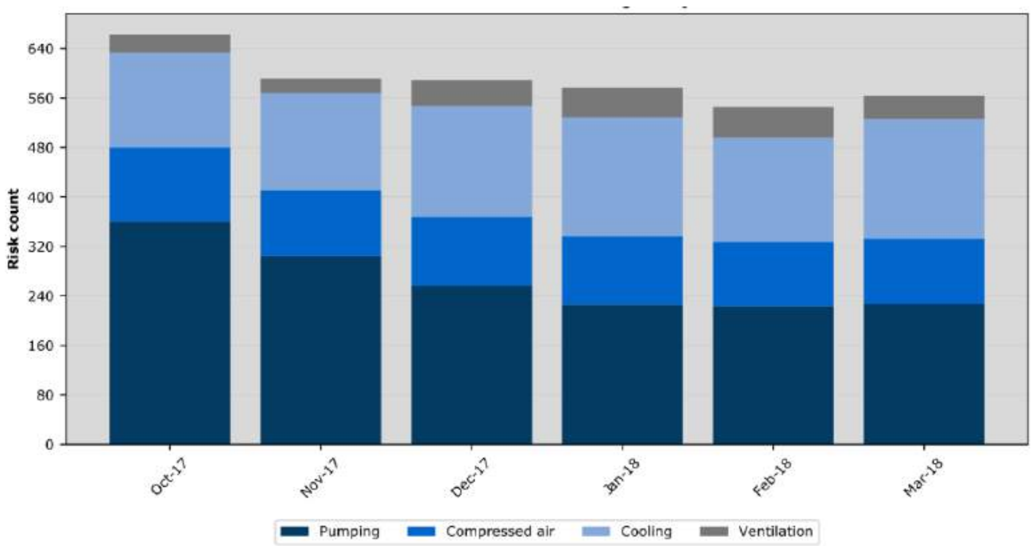


Figure 7: System class risk distribution

Figure 8 displays the risk distribution between the two parameter types that were evaluated.

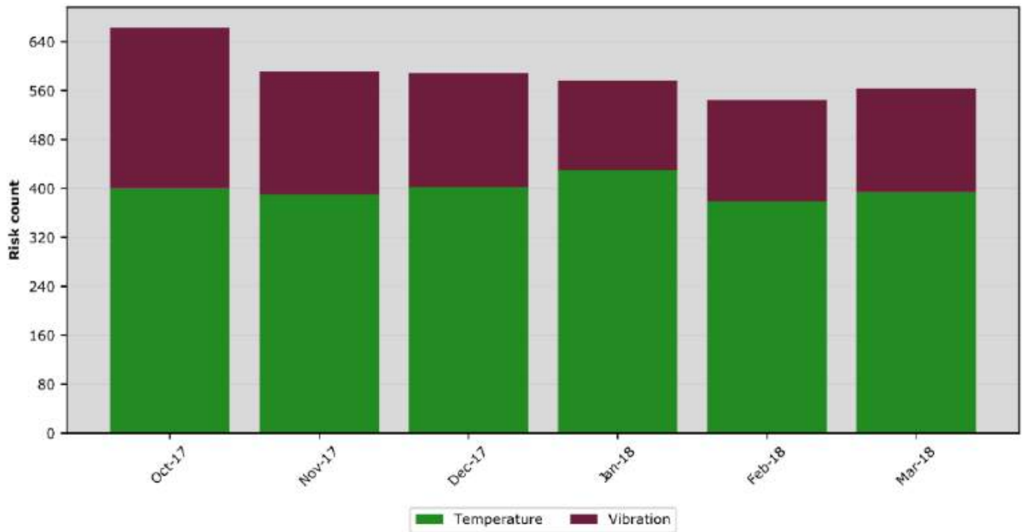


Figure 8: Parameter type risk distribution

Figure 9 displays an overview of the risk information related to the compressed air system.

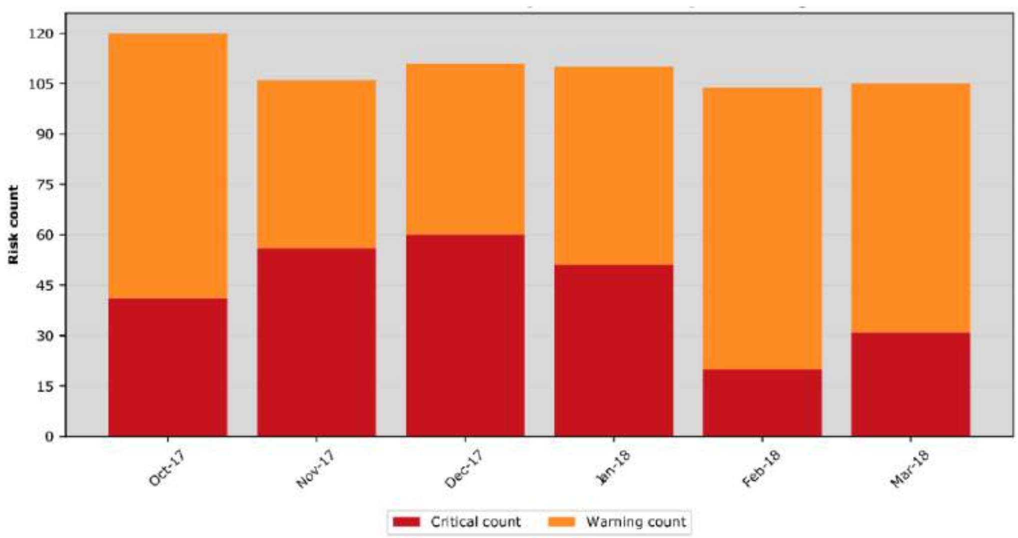


Figure 9: Compressed air system evaluation

4. IMPLEMENTATION AND RESULTS

The risk information of six deep-level mines was evaluated. The availability of the risk information depends on the respective site implementations. It was therefore possible to evaluate some sites' information over a period of 12 months. The number of individual risk identifiers depends on the number of mining systems and subsystems on each site. 96 daily risk identifiers were used in one of the site evaluations, which resulted in more than 17 000 input values for a period of six months. Two site evaluations will be discussed in more detail.

4.1 Case study 1: Mine A

The systems on Mine A that were evaluated included pumping, cooling and ventilation. 48 risk identifiers were evaluated over a period of eight months. The evaluation therefore consisted of 11 520 input values. Figure 10 displays the system distribution of Mine A's risk information. This risk distribution clearly indicated that the majority of the risks occurred on the cooling and ventilation systems.

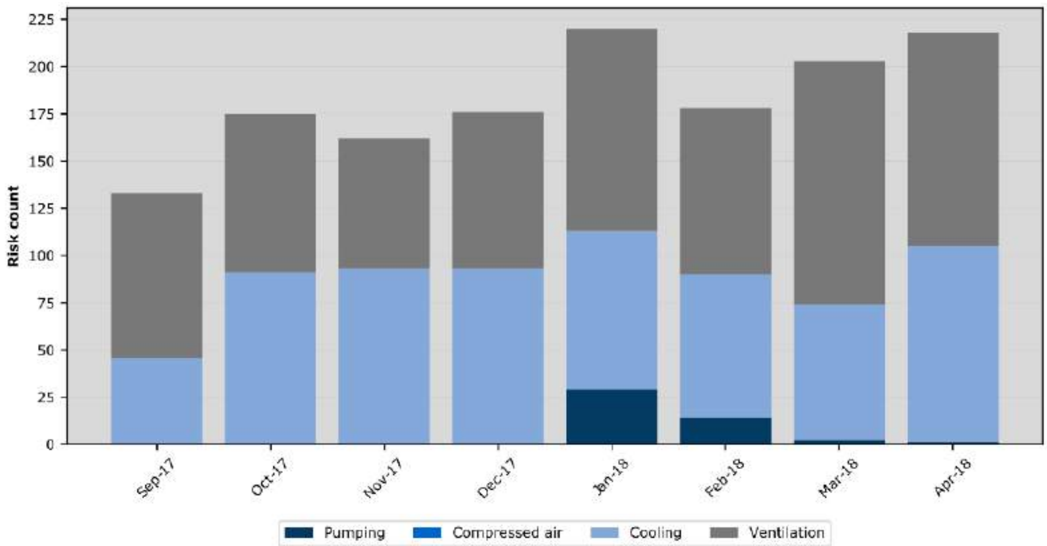


Figure 10: Mine A – System distribution

The parameter distribution is shown in Figure 11. It is evident that most of the risks are related to vibration.

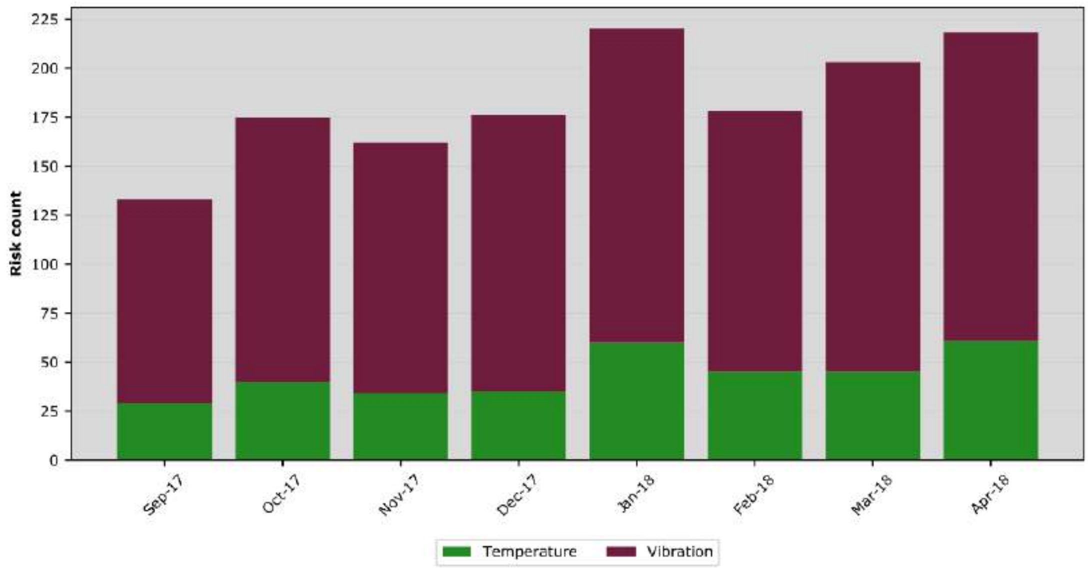


Figure 11: Mine A – Parameter distribution

Another interesting revelation was the increase in critical risks (high risk severity) on the cooling system (Figure 12). These risks were related to high vibration measurements on three refrigeration plants. This was seen as a high priority due to the fact that the refrigeration plants provide the underground operations with cold water and a cold-water supply interruption could force the mine to halt production.

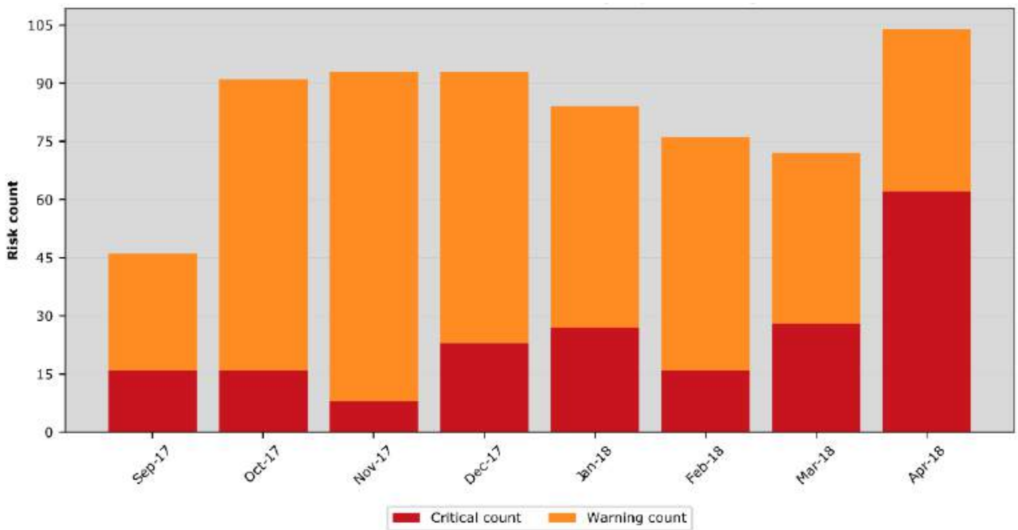


Figure 12: Mine A – Cooling system risks

4.2 Case study 2: Mine B

Four mining systems were evaluated on Mine B. 40 risk identifiers were assessed for a period of 12 months. The assessment therefore consisted of 14 400 input values. A risk overview, for all the systems, is shown in Figure 13. It is clear that the number of operational risks decreased during the first nine months. During the next three months, however, there was an increase in risks. These risks mainly occurred on the pumping and compressor systems. This assessment highlights the importance of a continuous evaluation: to immediately be made aware of new or re-emerging system risks.

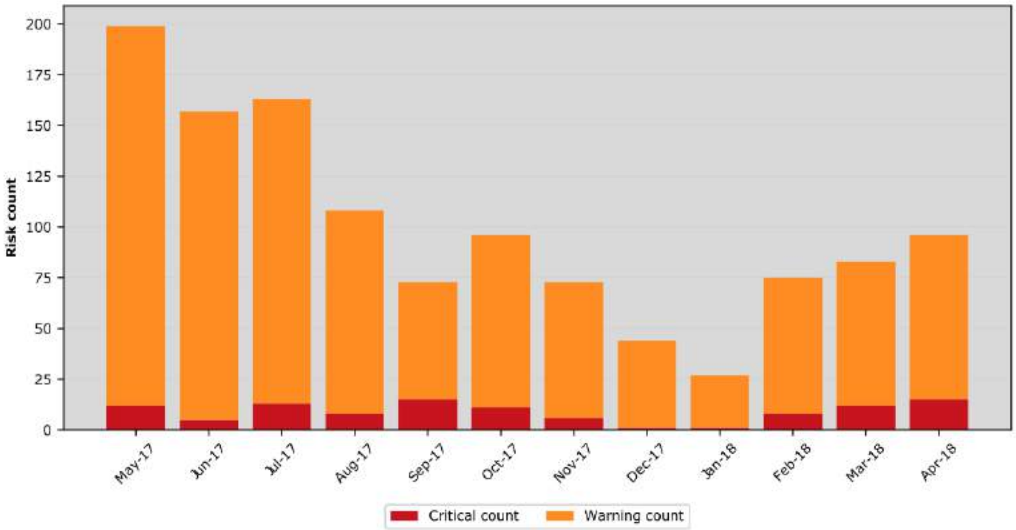


Figure 13: Mine B – Risks overview

The system-specific evaluation for Mine B's compressor is shown in Figure 14. The risks that were initially identified were resolved. During the last month, there were a number of critical risks that were detected. Maintenance personnel can therefore schedule the required maintenance. Site supervisors can now also verify, on a monthly basis, that the issues have been resolved.

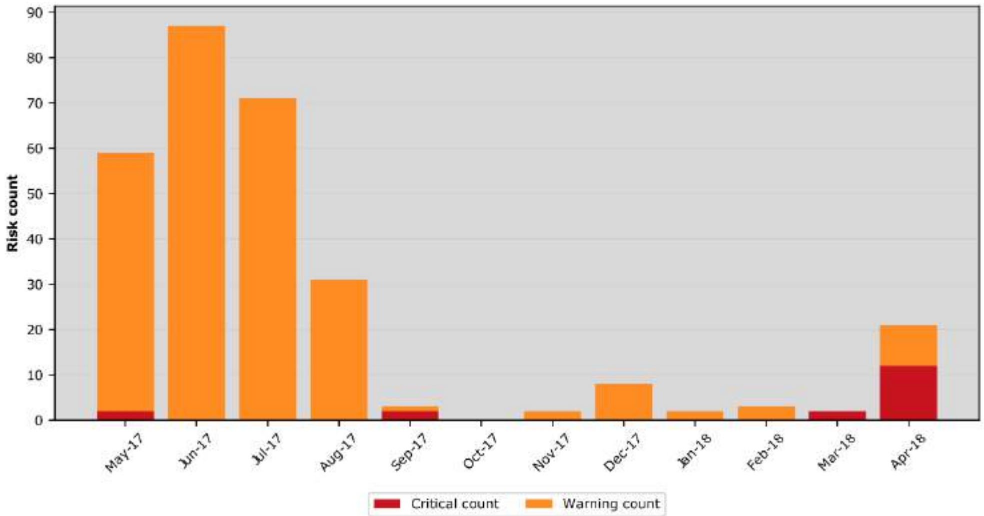


Figure 14: Mine B – Compressed air system risks

5. CONCLUSION

Due to the operating depths of deep-level mines, equipment maintenance is a challenging task. It is not feasible for mine personnel to perform manual inspections regularly. An automated system was therefore developed to perform a condition monitoring analysis to identify operational risks. However, over longer periods of time, the risk information becomes too substantial for supervisors to gain insights from.

A new software application was subsequently developed to categorise the risk information according to the system class, parameter type and risk severity. The application is used on a monthly basis to evaluate the change in equipment condition.

The results from the system have already enabled mine personnel to identify problem areas within their operations. Maintenance efforts can therefore be focussed on critical and recurring issues. Site supervisors are also able to follow up on equipment overhauls that were performed. Future work includes an assessment of the long-term use of the application to quantify the benefit thereof.

REFERENCES

[1] **Avram, M. G.** 2014. Advantages and challenges of adopting cloud computing from an enterprise perspective, *Procedia Technol.*, vol. 12, pp. 529–534.

[2] **Woods, D. D.**, **Patterson, E. S.**, **Roth, E. M.**, and **Christoffersen, K.** Apr. 1999. Can we ever escape from data overload? A cognitive systems diagnosis, *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 43, no. 3, pp. 174–178.

[3] **Jess, T.** et al. 2015. An industrial data recommender system to solve the problem of data overload, *23rd Eur. Conf. Inf. Syst. (ECIS 2015).*

[4] **Elkinson, K.** and **McGrail, A.** 2012. The risk of too much data, in *IEEE Power and Energy Society General Meeting*.

[5] **van Jaarsveld, S.** 2017. Developing an integrated information system to assess the operational condition of deep level mine equipment, PhD Thesis, North-West University.

[6] **Van Jaarsveld, S.**, **Van Heerden, S. W.**, and **Van Rensburg, J. F.** 2017. Development of a condition monitoring information system for deep level mines, in *SAIIE28 Proceedings*.

[7] **Cilliers, C.**, **Brand, H. G.**, and **Kleingeld, M.** 2016. Benchmarking the electricity use of deep-level mine compressors, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[8] **Schutte, A. J.**, **Kleingeld, M.**, and **Van Der Zee, L.** 2014. An integrated energy efficiency strategy for deep mine ventilation and refrigeration, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[9] **Cilliers, C.** 2015. Benchmarking electricity use of deep-level mines, PhD Thesis, North-West University.

[10] **Buys, J. L.**, **Kleingeld, M.**, and **Cilliers, C.** 2015. Optimising the refrigeration and cooling system of a platinum mine, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[11] **Schutte, A. J.**, **Kleingeld, M.**, and **Groenewald, H. J.** 2016. A holistic energy-cost evaluation of ice vs chilled water in mining, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[12] **Zhu, S.**, **Wu, S.**, **Cheng, J.**, **Li, S.**, and **Li, M.** 2015. An underground air-route temperature prediction model for ultra-deep coal mines, *Minerals*, vol. 5, no. 3, pp. 527–545.

[13] **Maré, P.**, **Marais, J. H.**, and **Van Rensburg, J.** 2015. Improved implementation strategies to sustain energy saving measures on mine cooling systems, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[14] **Van Jaarsveld, S.**, **Du Plessis, J. N.**, and **Pelzer, R.** 2015. A control system for the efficient operation of Bulk Air Coolers on a mine, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*, pp. 133–137.

[15] **Zhang, H.**, **Zhang, W.**, **Lv, L.**, and **Feng, Y.** 2010. Effect of fissure water on mechanical characteristics of rock mass, *Min. Sci. Technol.*, vol. 20, no. 6, pp. 846–849.

[16] **Bridgwood, E. W.**, **Singh, R. N.**, and **Atkins, A. S.** 1983. Selection and optimization of mine pumping systems, *Int. J. Mine Water*, vol. 2, no. 2, pp. 1–19.

[17] **Zhang, H.**, **Xia, X.**, and **Zhang, J.** 2012. Optimal sizing and operation of pumping systems to achieve energy efficiency and load shifting, *Electr. Power Syst. Res.*, vol. 86, pp. 41–50.

[18] **Mare, P.**, **Bredenkamp, J. I. G.**, and **Marais, J. H.** 2017. Evaluating compressed air operational improvements on a deep-level mine through simulations, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[19] **Groenewald, H. J.**, **Kleingeld, M.**, and **Vosloo, J. C.** 2015. A performance-centred maintenance strategy for industrial DSM projects, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*, pp. 50–53.

[20] **de Jager, J. P.**, **Groenewald, H. J.**, and **Cilliers, C.** 2016. Investigating the effect of pump availability on load shift performance, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[21] **Al-Najjar, B.** and **Alsyouf, I.** 2004. Enhancing a company's profitability and competitiveness using integrated vibration-based maintenance: A case study, *Eur. J. Oper. Res.*, vol. 157, no. 3, pp. 643–657.

[22] **Tsang, A. H.** 1995. Condition-based maintenance: Tools and decision making, *J. Qual. Maint. Eng.*, vol. 1, no. 3, pp. 3–17.

[23] **Yam, R. C. M.**, **Tse, P. W.**, **Li, L.**, and **Tu, P.** 2001. Intelligent predictive decision support system for condition-based maintenance, *Int. J. Adv. Manuf. Technol.*, vol. 17, no. 5, pp. 383–391.

[24] **Ballal, M. S.**, **Jaiswal, G. C.**, **Tutkane, D. R.**, **Venikar, P. A.**, **Mishra, M. K.**, and **Suryawanshi, H. M.** 2017. Online condition monitoring system for substation and service transformers, *IET Electr. Power Appl.*, vol. 11, no. 7, pp. 1187–1195.

[25] **Oberholzer, P. J.** 2014. Best practices for automation and control of mine dewatering systems, North-West University.

[26] **Hamer, W.**, **Vosloo, J. C.**, and **Pelzer, R.** 2015. Analysing electricity cost saving opportunities on South African gold processing plants, in *Proceedings of the Conference on the Industrial and Commercial Use of Energy, ICUE*.

[27] **Alsyouf, I.** 2009. Maintenance practices in Swedish industries: Survey results, *Int. J. Prod. Econ.*, vol. 121, pp. 212–223.

[28] **Ahmad, R.** and **Kamaruddin, S.** 2012. An overview of time-based and condition-based maintenance in

industrial application, *Computers and Industrial Engineering*, vol. 63, no. 1. pp. 135–149.

[29] **Jardine, A. K. S.**, **Lin, D.**, and **Banjevic, D.** 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance, *Mechanical Systems and Signal Processing*, vol. 20, no. 7. pp. 1483–1510.

[30] **Thorsen, O. V** and **Dalva, M.** 1997. Condition monitoring methods, failure identification and analysis for high voltage motors in petrochemical industry, in *Eighth International Conference on Electrical Machines and Drives*, pp. 109–113.

[31] **Nair, A.** and **Cai, C. S.** 2010. Acoustic emission monitoring of bridges: Review and case studies, *Eng. Struct.*, vol. 32, pp. 1704–1714.

[32] **Ruvo, P. De**, **Ruvo, G. De**, **Distante, A.**, **Nitti, M.**, **Stella, E.**, and **Marino, F.** 2008. A visual inspection system for rail detection and tracking in real time railway maintenance, *Open Cybern. Syst. J.*, vol. 2, no. 1, pp. 57–67.

[33] **Stols, J. A.**, **Brand, H. G.**, and **Schoeman, W.** 2017. A condition monitoring system using basic equipment protection measurements, in *SAIIE28 Proceedings*.

[34] **Jasiulewicz-Kaczmarek, M.** 2016. SWOT analysis for planned maintenance strategy - A case study, *IFAC-PapersOnLine*, vol. 49, no. 12, pp. 674–679.

INCORPORATING PRODUCT CATEGORIZATION TO IMPROVE THE PERFORMANCE OF SA'S PUBLIC HEALTHCARE SUPPLY CHAIN: A RESEARCH AGENDA.

Newton Mapowo[1], Louzanne Bam[2*], Imke H. de Kock[3] & Joubert van Eeeden[4]

[123]Health Systems Engineering and Innovation Hub, Department of Industrial Engineering
Stellenbosch University, South Africa
[1]21664056@sun.ac.za, [2]louzanne@sun.ac.za, [3]imkedk@sun.ac.za

[4]Department of Industrial Engineering
Stellenbosch University, South Africa
[4]jveeden@sun.ac.za

ABSTRACT

In an influential publication from the late 1990's, Marshall Fisher argued that many of the challenges in supply chains could be traced back to a lack of alignment between the type of product and the type of supply chain. Subsequently, the idea of tailoring supply chain management practices and policies to the characteristics of the products being supplied has received significant research attention, and various researchers have worked on modifying the premise as well as on its application to diverse sectors, with promising findings.

The South African National Department of Health is in the process of rolling out the Visibility and Analytics Network (VAN) reference framework, with the aim of ensuring sustained availability of and access to commodities. At present, the VAN strategy does not incorporate a product categorization element. This paper proposes a research agenda for determining how product categorization could be incorporated into the VAN strategy to enable supply chain practices and policies to be tailored to the characteristics of products.

*Corresponding author

1. INTRODUCTION: BACKGROUND

Supply chains incorporate the end-to-end streaming of information, products, and money [1]. Consequently, how supply chains are managed has significant implications for an organization's competitiveness in the context of product cost, working capital requirements, lead time to market and service delivery, among others [1]. Supply chain management is therefore defined as:

> *"the management of upstream and downstream relationships with suppliers and customers in order to deliver superior customer value at less cost to the supply chain as a whole"* [2].

In this setting, the appropriate alignment of the business strategy with the supply chain is critical to achieving a high level of business performance. Harris, Componation and Farrington [3] added by stating that a product with steady demand and dependable supply cannot be managed similarly as a product with unpredictable demand and inconsistent supply. Fisher [4] proposed that the reason why supply chains do not perform as expected despite increased investments in effort and resources is improper alignment of product attributes and supply chain strategies.

The South African National Department of Health (NDoH) is in the process of rolling out the Visibility and Analytics Network (VAN) reference framework, with the aim of ensuring sustained availability of, and access to, medicines [5]. One of the objectives of the VAN strategy is to transform South Africa's public healthcare pharmaceutical supply chain from an 'uninformed pull' system, to an 'informed push' system; the distinction between these respective systems being that specialized supply chain management professionals will be utilized in each province. These respective professionals will analyse and optimize complex links in the public healthcare pharmaceutical supply chain and make inventory planning and management recommendations to primary healthcare facilities (PHCF), rather than the PHCFs doing it on their own [5]. An informed push model will relieve the PHCFs staff of sophisticated and time-consuming supply chain planning work and enable them to focus more on delivering healthcare.

At present, the VAN strategy does not incorporate product categorization—which entails the organization of products into categories according to shared attributes; more specifically, when supply chains are considered, attributes that are related or could influence the best supply chain management strategy for said attributes [6]. Simchi-Levi, Clayton and Raven [7] reasoned that "one size does not fit all" in the formulation of a supply chain strategy highlighting the fact that different supply chain management strategies are most likely needed for products with varying attributes. Various products (e.g. anti-hypertensive, anti-viral, thrombolytic and anti-cancer medicines etc.) [8] require different supply chain management strategies (e.g. continuous-flow, efficient, fast, agile supply chains etc.) [1] depending on the product attribute context (e.g. demand volume, lead-time, uncertainty, life cycle etc.) [9]. Findings from a survey conducted by the Stop Stock-outs Projects (SSP) stated that in 2013, 21% of South African public healthcare facilities experienced a stock-out or shortage of a Tuberculosis (TB) or Human Immunodeficiency Virus (HIV) medicine within three months prior to SSP conducting the survey [10]. In 2014, a second national SSP survey exposed that the abovementioned percentage further increased to 25% meaning that a quarter of public healthcare facilities in South Africa experienced HIV and TB medicine stock-outs within the three months prior to the SSP survey. In 2015, the status quo remained unchanged [11]. The product categorization concept therefore proposes that such essential medicines should have separately managed supply chain strategies due to possible severe risks associated with non-availability of medicines such as the interruption of treatment of chronic diseases patients which poses a risk of developing drug resistance [10].

Alignment of product attributes with the appropriate supply chain strategy has benefits for the organization that go well beyond the positive effects on cost, efficiency, productivity and competitiveness [1]. Payne and Peters [12] added to Fisher's framework of product classification ascertaining that "no matter how good the supply chain characteristics are, if the product fundamentally does not fit with the dominant supply chain design, optimum service and cost cannot be achieved". Research has quantitatively analyzed and explored the validity of Fisher's framework for improving efficiency, applying the concept in a wide range of sectors like the electronic commerce, fashion industry, manufacturing industry and wine industry, which are elaborated in section 3.1.

In terms of the benefits of product categorization for the healthcare sector specifically, Prinja, Bahuguna, Tripathy and Kumar [8] proposed that the concept of product categorization poses benefits in sustainable availability of medicines and reduction in healthcare expenditure, while others have highlighted that it is vital to business attributes such as spend analysis, financial analysis, strategic sourcing, tendering, enterprise resource planning, and merchandising in healthcare [13]. Product categorization has come as a concept to be employed in conjunction with robust Information Technology (IT) systems for better use in scientific warehousing and inventory management, transparent integrated procurement, real-time stock monitoring, and decentralised distribution methodologies [8].

Effective application of the concept of product categorization into the South African pharmaceutical supply chain has the potential of reaping the same or more benefits as established in other industries. This paper aims to propose a research agenda for determining the viability and value of incorporation of product categorization in the South African public healthcare pharmaceutical supply chain. The focus will be on the feasibility of tailoring the supply chain practices and policies to the attributes of products within the said context. To establish the said aim, the first objective of providing an overview and trends in research of product categorization dimensions as applied in other industries and then more particularly the healthcare sector was established. Furthermore, product categorization dimensions were defined through determination of product attributes in relation to supply chain strategies.

2. METHODOLOGY

To understand the trends in research, the current applications of product categorization, and how product categorization relates to supply chain strategies, a systematic literature review was conducted. The systematic literature review consisted of firstly gathering and then analyzing research studies and publications concerning product categorization and supply chain strategies that incorporated the practice of product categorization. An overview of the application of product categorization dimensions across industries and more specifically healthcare sector are subsequently provided.

2.1 Systematic literature review methodology

The literature review was conducted using the Scopus[1] database to identify the application trends of product categorization. The literature review thus sought to address the following research questions (RQ):

- RQ 1: For the past decade (2007 to 2017), what have been the trends and concepts employed in applying the product categorization dimensions?
- RQ 2: What are the impacts of incorporating product categorization in the various industries?

Both research questions are considered within a supply chain management context. Keywords from the research questions were derived and a Scopus database search was conducted, and the search line (which also catered for variations between American English and British English where 'z' can be used for 's'; as well as word variations for instance: 'distribute', distributary' and 'distribution' etc.) was developed. The search line also took into consideration the understanding that product categorization, segmentation or classification are used interchangeably in literature as synonyms. Since the aim is to consider product categorization within the context of supply chain management, the terms 'supply chain' and 'value chain' are also included as search terms as well as 'demand', 'supply' and 'distribution'. Supply chain and value chain do not necessarily mean the same in literature, but they have various aspects that intersect hence the two in context were incorporated in the search line.

The aim is to understand the contemporary concepts and impacts of the application of the concept of product categorization, and therefore the literature available over the past decade (2007-2017) was considered in the Scopus database search. The search line that was used to identify relevant literature is as follows:
(((categori* OR segment* OR classif*) W/5 (demand* OR supply* OR distrib*)) AND ("supply chain" OR "value chain")).

The search field included the article title, abstract and keywords. The initial search, that was not limited by publication year, yielded 792 documents. The search line used was then limited to the publication years 2007 to 2017 (consistent with the research question of finding the focus of research on product categorization for the past 10 years) and yielded 636 documents. In screening the 636 documents for relevance; two criteria were used:

i. Firstly, documents that addressed the development of frameworks and models of product categorization dimensions and/or supply chain strategies were accepted (documents that referred to product categorization once were part of the search results; however, they were not related to developing frameworks and models for product categorization, so they were screened out),
ii. Secondly, documents that at least detailed how and why the aforementioned dimensions and strategies should be applied in any setting were included too.

Furthermore, a review was done into whether the frameworks and models in the different sectors were developed from parameters related or applicable to supply chain strategies and those that satisfied this question were accepted. After the titles together with the abstracts of the 636 documents were reviewed for relevance using the above stated criteria, the screening yielded 61 documents that were concerned with detailing and/or development of frameworks, models and impacts of incorporation of product categorization and/or supply chain strategies. Care was taken not to perform extensive filtering of the documents lest important documents might be left out. This meant a set of documents totalling 61 publications were scrutinized. Four documents were

[1] Scopus is a large abstract and citation database of peer-reviewed literature: books, scientific journals and conference proceedings. https://www.scopus.com/search/form.uri?display=basic

found to be inaccessible and this gave a remaining total of 57 documents. Seven documents from references of the 57 documents that were significantly referenced by authors and satisfied the above stated criteria were incorporated in the systematic literature review and gave a total set of 64 relevant documents. A summary of the approach followed is provided in Figure 1.

To fully pursue the aim of proposing a research agenda for the incorporation of the product categorization concept in the South African pharmaceutical supply chain, an overview of the application of the concept in other industries as well as the healthcare sector is considered. This will facilitate the understanding of the impacts of incorporating product categorization in the various industries found in literature and these insights can be used to develop suggestions on how these impacts can possibly be applicable to the South African pharmaceutical supply chain. A framework established in 1997 by Fisher [4], an acknowledged pioneer in the linking of supply chain strategies with product attributes will then be used to define the dimensions of the product categorization concept, consistent with the second objective of this paper. Existing product categorization methods (product characteristics attributes and the categorization selection methods) which premised on Fisher's framework are used as basis for this proposition.

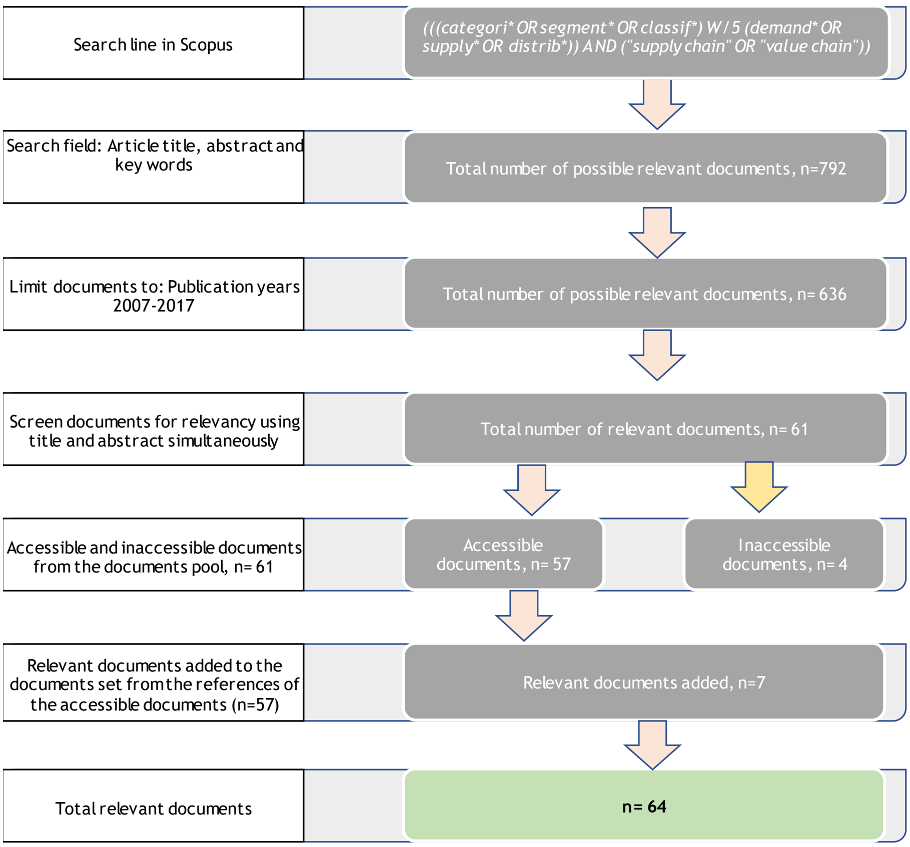


Figure 1: Scopus database search methodology

3. OVERVIEW OF PRODUCT CATEGORIZATION APPLICATION

This section provides an overview from the systematic literature review of the application of the product categorization concept, firstly in other industries which are not healthcare (where there were 18 industry-application specific documents from literature), then secondly in the healthcare industry itself (where there were three industry-application specific documents from literature). This section is consistent with the two aforementioned research questions as well as the first objective of this paper.

3.1 Application in other industries (not healthcare)

The concept of product categorization has been applied in a wide range of industries, but for this paper the manufacturing industry, fashion and retail industry, E-commerce industry and wine industry have been chosen for analysis. This is because it has been understood that the nature of the industries' supply chains shares certain characteristics with the healthcare supply chain, including:

- ✓ Agility to seasonal and demographic variations, in the case of the fashion industry;
- ✓ The visibility, variability, analytics and categorization of a large innovative product portfolio and how they are supplied that underscores the need for continuous improvement, in the case of the e-commerce industry;
- ✓ Responsiveness to high demand, especially in the case of the wine industry; and
- ✓ Combination of lean and agile principles, in the manufacturing industry.

Table 1: Impacts of application of product categorization in selected industries

Industry	Literature	Concepts in application	Impacts
Wine industry	Wen et.al [14]	Quality Function Deployment (QFD)	Product management, transport and package management and marketing.
Manufactu-ring	Nagashima et.al [15]	Ordered feature evaluation analysis (OFEA). Collaboration intensity.	Demand forecasting accuracy
	Naim et.al [16]	Lean and agile paradigms giving 'Leagile' manufacturing	Optimal determination of decoupling point
	Ramkumar et.al [17]	Vendor Managed Inventory (VMI)	Reduction in lead time and demand variability. Optimal stocking quantities at warehouses
	Ervolina et.al [18]	Available-To-Sell (ATS)	Optimal demand forecasting. Profitable demand response. Profitable product portfolio.
	Hiremath et.al [19]	Outbound logistics network (OLN). Multi objective genetic algorithm (MOGA)	Minimize the total network cost. Maximize resource utilization
	Taylor et.al [20]	Concurrent product and supply chain design (CP-SCD)	Provide managerial insights in design trade-off analysis
	Li [21], Fichtinger et.al [22]	Two-Echelon Dual-Channel Supply Chain	Development of profitable market segments
	Micieta et.al [23]	Group technology	Improvement in quality, productivity and inventory management. Increase costing accuracy. Increase customer service. Reduction in overall production times. Gain in sustainable competitive advantage
	Rajan and Wang [24]	Hierarchical clustering	Obsolescence Reduction
Fashion industry	Sen [25]	Efficient supply chain management practices	Production of a variety of sizes, styles and colors in shorter lead-time. Agility to changing demand requirements. Ability to better forecast
	Roscoe and Baker [26]	Supply chain segmentation	Aligning of demand planning, marketing, sales and supply chain functions.
	Martinez et.al [27]	Pronto Moda or Rapid-Fire Fulfilment	Reduction of variety and wide range of garments per collection. Standardizing fabrics and product platforms. Introduction of mini-collections
	Sardar et.al [28]	Outsourcing strategies. Goal programming	Cost saving. Capacity flexibility

E-commerce	Rofin and Mahanty [29]	Cournot model	Determination of optimal product categories' dual-channel supply chain configuration for different customer online channels preference
	Hofbauer et.al [30]	Ontology	Derivation of web ontologies. Efficient and reliable electronic product data exchange across organizations.
	Liao et.al [31]	Two-step data mining approach	Mining customer knowledge on online channels and product segments preferred. Better market segmentation

From the systematic literature review conducted, these areas have received significant attention. Product categorization has had various impacts to these industries which are in the context of demand planning, distribution planning, forecasting, supply planning; and more specifically in product value, product shelf space, lead time, demand variability etc. Table 1 above details the concepts of product categorization application in the industries and the impacts of these.

Deducing from the selected wine industry, it can be understood that responsiveness to high demand peaks is potentially feasible through the incorporation of product categorization with positive impacts. As information about product categories (volume, volatility of ordering or value) becomes available, informed decisions about sourcing and outsourcing (local or international) of medicines can possibly be made. With delivery frequencies and minimum lot sizes determined, the cost on inbound and outbound transportation can potentially be reduced [14].

Examples from the manufacturing industry show that optimal determination of decoupling point, reduction in lead time and demand variability, optimal stocking quantities at depots and demand forecasting accuracy are dimensions that can possibly be achieved by the incorporation of product categorization. As seen in the fashion industry, a reduction of variety and wide range of garments per delivery as well as standardizing product platforms is potentially possible through the incorporation of product categorization. Mini-collections and supply flexibility are likely to be achieved too. The E-commerce industry shows that visibility, analytics and categorization of a large product portfolio is likely feasible through the incorporation of product categorization. Determination of optimal product categories' multi-channel supply chain configuration for products is potentially possible as well.

3.2 Application of product categorization in healthcare

Unlike in the commercial sector where the focus is largely on profit and competitive advantage, the healthcare sector poses a different configuration in that it mainly focuses on effective and sustainable availability and administration of healthcare [32]. The application of product categorization in literature in healthcare is summarized in Table 2.

Table 2: Impacts of application of product categorization in healthcare

Industry	Literature	Concepts in application	Impacts
Healthcare	Yadav et.al [33]	Vaccines supply chain generic map	Efficient, effective and sustainable delivery of vaccines to end recipients.
	Hua et.al [34]	Integrated reverse supply chain model	Increase in percentage of unexpired medications grew, the rate of expired medications collection remained the same but the profit of the whole reverse supply chain increased. Cost reduction.
	Kritchanchai and Meesamut [35]	Inventory management model	Minimized total inventory costs while maintaining drug administration safety levels. Reduction in demand variability. Increase in demand forecasting accuracy.

The concept of product categorization has been used in the healthcare sector by Yadav, Lydon, Oswald, Dicko, and Zaffran [33] in developing a framework for decision making in the integration of vaccine supply chains with other health commodities supply chains. The framework enabled the optimization of immunization supply chains as vaccines were delivered to the end recipients efficiently, effectively and sustainably. Optimal product categories which aided in aligning of demand planning and supply chain functions were determined. Hua, Tang, and Wu [34] employed the product categorization concept in developing an integrated reverse supply chain model. The model was meant to investigate the impact of various unwanted medications categories and government's publicity and penalty investment on the reverse supply chain profit as well as collection rate of expired medications. The findings showed that as the percentage of unexpired medications grew, the rate of expired medications collection remained the same, but the profit of the whole reverse supply chain increased [34]. There was a significant reduction on cost as well. Kritchanchai and Meesamut [35] employed the concept in developing an inventory management model for a hospital in Thailand. They argued that a single inventory management system could not be used on all medicines effectively. The inventory management system they developed, which incorporated the product categorization concept, enabled the hospital to minimize the total inventory costs while maintaining patient drug administration safety levels. Provision for demand forecasting accuracy was made as well as reduction in demand variability [35].

The application of product categorization, as seen in other industries (section 3.1) and the healthcare industry (section 3.2), can be used to suggest possible impacts of application of product categorization in the South African public healthcare pharmaceutical supply chain (section 3.3).

3.3 Possible impact of product categorization in the South African public healthcare pharmaceutical supply chain

As seen in the industries discussed in section 3.1 and 3.2, the incorporation of product categorization in the South African public healthcare pharmaceutical supply chain can potentially reap positive results. Product categorization will potentially enable determination of optimal lot sizes for various pharmaceutical products, replenishment frequency and safety stock levels for different product categories. The lower the replenishment lead time from the provincial depots, the lower the amount of safety stock that needs to be held by the PHCFs. Product categorization has the potential to aid in ensuring sustainable availability of pharmaceuticals. Some products are highly sensitive to shortages and stockouts and these pose an excessive cost to both the healthcare system and its patients, thus on-time performance, which also affects the variability of lead time, will possibly be optimized by the incorporation of product categorization. Product categorization in the South African public healthcare pharmaceutical supply chain has the capability of increasing supply flexibility of various pharmaceuticals. The more flexible the supply is, the less lead time variability will be potentially displayed as order quantities. Product categorization based on product attributes (demand, value volatility etc.) will possibly enable a better match between demand and supply chain strategy (continuous-flow, efficient, fast, agile supply chains) to ensure sustainable medicine availability. The incorporation of product categorization in the South African public healthcare pharmaceutical supply chain will potentially enhance the design collaboration capability for demand planning, supply planning and distribution planning and this will possibly decrease required inventories and transportation costs.
Sections 3.1, 3.2 and 3.3 have detailed the impacts of the concept of product categorization when applied in various industries including the potential benefits product categorization can have in the South African public healthcare pharmaceutical supply chain. Section 4 below will detail how product categorization has been generically conceptualized.

4. CONCEPTUALIZATION OF PRODUCTION CATEGORIZATION

This section seeks to define product categorization dimensions through the analysis of product attributes in relation to supply chain strategies consistent with the second objective. It endeavors to bring to the topical discussion of the concept of product categorization in general (i.e. non-industry specific). Supply chains in various industries suffer from an excess of some products and a shortage of others due to an inability to effectively predict demand [3]. According to Fisher [4], the root cause of the problems such supply chains face is a mismatch between the product-type and the supply chain-type. As shown in Figure 2, Fisher [4] proposed that if products are classified based on their demand configurations, they fall into one of two categories: they are either primarily functional or primarily innovative. Furthermore, Fisher [4] proposed that supply chains can either be physically efficient or market responsive. These concepts are explored in more depth in the remainder of this section.

	Functional Products	Innovative Products
Efficient Supply Chain	Match	Mismatch
Responsive Supply Chain	Mismatch	Match

Figure 2: Fisher's framework, matching supply chains with products. (Reproduced from [4].)

4.1 Functional and Innovative Products

Functional products are the widely available products which satisfy basic needs. They are characterized by a relatively extensive life-cycle, with little change over time and little variance in their offerings [3]. Demand for functional products is typically predictable and stable, and they tend to possess low profit margins because of the many competitors in the market [2]. Inventory is used to buffer demand because the cost of obsolescence is low.

Sullivan et.al [5] highlighted that innovative products are typically distinguished as trendy and have highly volatile demand that is difficult to predict. They are associated with significantly more uncertainty than functional products [4]. Innovative products have larger product variety and short life-cycles, but the profit margin is high, therefore lost sales (opportunity cost) exert a significant effect on company performance [3]. Table 3 shows the demand aspects for classification of products as either functional or innovative.

Table 3: Product demand aspects versus functional or innovative (adapted from [4].)

Demand Aspects	**Functional** (Predictable Demand)	**Innovative** (Unpredictable Demand)
Average stock-out rate	1% to 2%	10% to 40%
Product Variety	Low (10 to 20 variants per category)	High (often millions of variants per category)
Contribution to Margin	5% to 20%	20% to 60%
Product Life Cycle	more than 2 years	3 months to 1 year

Various products from the pharmaceutical product portfolio can probably be determined and categorized according to either being functional or innovative. This, as established by Fisher [4], would be dependent on the pharmaceuticals demand aspects (e.g. average stock-out rates, product life cycle and product variety etc.) of each product. It can then be reasoned that chronic medication which has a stock-out rate of 25% as established in section 1.0 above, will not be considered the same as other medication with little or no stock-outs. Chronic medication stock-outs can probably be argued to be in the innovative products category since their stock-out rate is between 10% and 40% as identified in Table 3. However, it can contrarily be argued that since normally chronic patients are registered and known, then chronic medication should have predictable demand and thus should be categorized as functional. This uncertainty in categorization brings the need to consider all chronic products attributes, and this is the aim of this research agenda.

4.2 Physically Efficient and Market Responsive Supply Chains

Innovative products with their high profit margins and volatile demand, require an essentially different supply chain than stable and low-margin functional products [3]. According to Fisher [4], two supply chain types—physically efficient and market responsive—exist [4].

Physically efficient supply chains' fundamental focus is cost reduction and the efficient use of resources. Sullivan et.al [5] emphasised that this type of supply chain pursues creation of the lowest possible cost of operation through the removal of all non-value adding activities, chasing economies of scale and optimizing resource utilization. As shown in Figure 2, Fisher [4] proposes that companies that offer functional products should employ an efficient supply chain.

In contrast, the market responsive supply chain is fundamentally focused on meeting the customer delivery expectations irrespective of demand variability. As shown in Table 4, Fisher [4] proposes that market responsive supply chains are most appropriate for innovative products. The possibility of a stock-out increases when product demand is uncertain and volatile [4]. Lee [36] also asserted that supply disruption risks are mitigated in market responsive supply chains by the strategic placement of inventories and thus such a supply chain can adapt to customer, market, and supply uncertainty. Table 4 details the physically efficient and market responsive supply chains attributes.

Table 4: Physically Efficient and Market Responsive supply chains (adapted from [4].)

	Physically Efficient	**Market-Responsive**
Primary Purpose	Supply predictable demand efficiently at the lowest possible cost	Respond quickly to unpredictable demand to minimize stock outs, forced markdowns, and obsolete inventory
Inventory Strategy	Generate high turns and minimize inventory throughout the chain	Deploy significant buffer stocks of parts or finished goods
Lead-Time focus	Shorten lead time if it does not increase cost	Invest aggressively in ways to reduce lead time
Product-design Strategy	Maximize performance and minimize cost	Use modular design to postpone product differentiation for as long as possible

Given the above discussion, it is proposed that the South African public healthcare pharmaceutical supply chain could potentially benefit from aligning various pharmaceutical products to various supply chain strategies. Therefore, the concept of fitting supply chain strategies to the demand attributes of pharmaceuticals to ensure sustainable availability of medicines should be investigated.

4.3 Generic supply chain models

Perez [1] defined six generic supply chain models, grouped into two clusters that align to the types of supply chains defined by Fisher [4]. Three of Perez's [1] generic supply chains are oriented to achieve physical efficiency and three are oriented to achieve market responsiveness.

The three physically efficient supply chain models defined by Perez [1] are:

- *Continuous-flow supply chain*: This supply chain employs a 'make to stock' decoupling point where production is scheduled to replenish predefined stock levels which are based on a specific reorder point for inventory in the production cycle. It pursues high service levels and low inventory levels. It is mainly proposed for businesses with short shelf-life products, for example bread and dairy products.
- *Efficient supply chain*: This supply chain has production scheduled based on sales expectations for the duration of the production cycle, using a 'make to forecast' model as a decoupling point. It has been proposed for businesses with commoditized products, for example cement and steel.
- *Fast supply chain:* This supply chain has production scheduled in a single batch per stock keeping unit (SKU), with the size being defined by the season's sales expectations, and utilizing a 'make to forecast' decoupling point. The fast supply chain has been proposed for companies that engage in catalogue sales and trendy apparel for example fashion clothing.

The three market responsive supply chain models defined by Perez [1] are:

- *Custom-configured supply chain:* This supply chain is characterized by multiple configurations of the finished product on a unique platform, using a 'configurable to order' decoupling point. The custom-configured supply chain has been recommended for assembly of personalized products, for example computers and vehicles.
- *Agile supply chain:* This supply chain employs a 'make to order' decoupling point, where items are produced after a purchase order has been placed by the customer. It has been proposed for businesses that are characterized with unpredictable demand and essential for companies that use unique specifications for each customer to manufacture products for example chemical specialties and packaging.
- *Flexible supply chain:* This supply chain is characterized by adaptability, which entails the capability to reconfigure internal processes to meet a specific need (or solve a problem) of a customer. It is mainly proposed for service companies that encounter unexpected situations and emergencies faced with long periods of low workload and high demand peaks for example the medical emergency response sector.

Different products in the healthcare pharmaceutical industry can probably fall into each of these six generic supply chain models and thus need to be treated differently. Aligning each pharmaceutical product category with each supply chain model has the potential to optimize performance of the pharmaceutical supply chain and possibly enhance sustainable medication availability.

4.4 Product Categorization Methods

Over the years, numerous categorization methods using various product attributes have been developed for various sectors, whether manufacturing or e-commerce or distribution logistics. For several of them, the difference in the product attributes used is basically on semantics but in concept being the same, for example one would use 'flexibility' and the other would use 'agility' basically addressing the same concept. Perhaps the most salient product attributes definition is the one proposed by Sullivan et al. [5] which also details the measurable characteristics of the products attributes. Examples of these product characteristics include demand, cost, quality, lead-time, life-cycle, and the level of certainty with regards to customer demand and the market environment. The complete set of product characteristics defined by Sullivan et al. [5], together with the measurable attributes that have been defined for each, are presented in Table 5. These product attributes have been used to derive the three well documented methods that address product characterization selection as shown in Table 5 below, which are: DWV^3 [37], the three-dimensional global classification system [38] and the Product Supply Characterization (PSC) model [7].

4.5 The DWV^3; The three-dimensional global classification system; & The PSC model product characterization selection methods

The DWV^3 classification system utilizes five of the product attributes defined in Table 5, namely the <u>D</u>uration of the product life cycle, the time <u>W</u>indow for delivery, the <u>V</u>olume, the <u>V</u>ariety and the <u>V</u>ariability, which build the acronym DWV^3. This classification method is mainly used to develop focused demand chains where processes are prioritised as a sequence of events with the end view of serving the ultimate consumer [38].

The three-dimensional global classification system utilizes three product characteristics from those listed in Table 5 namely: product, demand and lead-times [5]. Each characteristic is classified in one of two gradations [13]:

- Product (standard or special);
- Demand (stable or volatile); and

- Lead-time (short or long).

The three-dimensional global classification system focuses on linking the supply chain strategy with the product life-cycle management signifying that the most suitable supply chain strategy of a product differs depending on its stage in the product life cycle [13].

The Product Supply Characterization (PSC) model utilizes seven (7) product characteristics which are volume, volatility, order line value, frequency of order lines, order line weight, substitutability of a product and number of customers buying each product. The PSC's focus is on addressing total supply chain costs and service performance to the customer [5].

The DWV^3 and the PSC are perhaps the best suitable methods to be applied in a public healthcare pharmaceutical supply chain. This is largely because DWV^3 is employed in developing focused demand supply chain, which in this case the end goal will be to serve the patient. Furthermore, the PSC focuses on service performance to the customer, which in the case of the healthcare pharmaceutical supply chain would be the patients.

Sullivan et.al detailed the product attributes and the measurable characteristics as in Table 5 below, and the author interpreted the product attributes used in the DWV^3, the three-dimensional global classification system and the PSC, and provided the last three columns.

Table 5: Product attributes and their measurable characteristics (adapted and developed from [5].)

Product attributes	*Examples of measurable characteristics*	*DWV^3*	*The three-dimensional global classification system*	*PSC*
Cost	***Supply chain, inventory and manufacturing***			
Demand	***Variability, predictability, volatility and volume***	X	X	X
Quality	***Defects and yield percentage***			
Financial	***Profit margin per part***			
Product	***Product characteristics***		X	X
Life cycle	***Phase and length of time in phase***	X		
Design	***Manufacturability of the product***			
Standardization	***Few customized features of the product***			
Customer	***Responsiveness in service***			X
Uncertainty	***Customer demand and market environment***			
Delivery	***On-time or on-schedule***	X		
Flexibility	***Handling of change in demand, design and delivery***			X
Inventory	***Product held in kanban/JIT inventory***			
Lead time	***Response time to deliver product***		X	
Production	***Capability and capacity to produce in lean environment***			

It can be seen in Table 5 that out of the 15 product attributes, demand and product have been the mostly used with demand being used by all three methods. This can be attributed to the fact that whether the focus point of product categorization is cost based, value based or needs based (as detailed in section 4.5), information about demand (variability, predictability, volatility and volume) seems essential. Moreover, product characteristics seem vital in the application of product categorization hence the product has been considered in two of the three methods in Table 5. It can be reasoned that depending on the focus point (as shown in section 4.5) of the South African public healthcare pharmaceutical supply chain, demand and product attributes will potentially be critical in the incorporation of product categorization too.

4.6 Which focus of the supply chain should product categorization target?

Dawe, Pittman and von Koeller [39] suggested that there are various focus points for supply chain product categorization and the methodologies employed depend upon the purpose. These include:

- **Cost-based.** Costs (and profits) cannot be disregarded in the development of the product categorization concept, however, cost-based analyses only (estimating, allocating and assigning costs) leave much unanswered particularly the shortcoming in tracking of costs directly to vital business entities [40]. The cost-based approach revolves around the connotation of resolving problems rather than seeking opportunities [39].
- **Value-based**. This categorizes products by economic value, for instance total revenue generated, which then disregards the cost-assigning-only approach to categories and actually segments them to determine profitability [13]. It is not categorizing just for the sake of categorizing therefore the categories should be sizeable enough to complement the supply chain strategy [36].
- **Needs-based**. Categorization is done on differentiated product drivers that clients have for a distinct supply chain service. Products are categorized based on a common set of clients' needs and internal resources, such as sales, and can provide insight on determining and validating the needs (met or unmet) [39]. The purpose is to match sector needs with the correct supply chain service with the aim of gaining competitive advantage [4].

In the healthcare pharmaceutical supply chain, the product categorization concept would probably focus on the needs-based and cost-based requirements of the supply chain. This is primarily due to healthcare supply chains' focus on the availability of medication in a cost-effective manner.

5. CONCLUSIONS, RECOMMENDATIONS AND FURTHER STUDY

Trends in research and current applications of product categorization dimensions have been analyzed using a systematic literature review. It is evident that a significant focus in research and application of the product categorization concept in the last ten years is shifting towards e-commerce probably owing to the growing widespread use of big data over the years. An acknowledgement of the concept of product categorization as applied in other industries as well as in healthcare supply chains has been established in this paper. The methods employed when applying product categorization as well as focus of the supply chain which it should target is established too. Furthermore, the concepts of supply chain management and generic supply chain models are discussed. This all laid a research agenda platform for determining the viability and value of incorporation of product categorization in the South African public healthcare pharmaceutical supply chain.

It is therefore recommended that, in harmony with the implementation of VAN, where it is stated [5] that the first three years of implementation will focus only on 'getting the pharmaceuticals to the point of need', the next focus after three years being 'selecting the right pharmaceuticals', research should thus be conducted on categorizing the pharmaceuticals according to their demand aspects and product attributes so as to optimize the selection and handling of the right pharmaceuticals in the healthcare pharmaceutical supply chain. This can be facilitated by the effective incorporation of product categorization in VAN in order to match the right product attributes with the right supply chain strategy and develop the right supply chain practices and policies. This will most probably optimize the performance of the South African healthcare supply chain and ensure sustainable medicine availability in a cost-effective manner, with probable benefits in scoring factors like on-time performance, supply flexibility, delivery frequency, minimum lot sizes, supply quality, inbound and outbound transportation cost, information coordination capability and design collaboration capability etc. A further study should be undertaken on how to incorporate product categorization in VAN (and NDOH) for demand planning; supply planning; and distribution planning, matching product categories with the right supply chain strategies and enhance the supply chain performance.

6. REFERENCES

[1] Perez, D. (2013). Supply chain strategies: Which one hits the mark. *CSCMP's Supply Chain Quarterly*, *2013*(1), 30. https://doi.org/10.1111/j.1461-9563.2008.00379.x

[2] Christopher, M. (2011). *Logistics & Supply Chain Management*. *Communications of the ACM* (Vol. 48). https://doi.org/10.1007/s12146-007-0019-8

[3] Harris, G. A., Componation, P. J., & Farrington, P. A. (2010). An exploration of fisher's framework for the alignment of supply chain strategy with product characteristics. *EMJ - Engineering Management Journal*, *22*(4), 31–43. https://doi.org/10.1080/10429247.2010.11431877

[4] Fisher, M. (1997). What is the Right Supply Chain for Your Product? *Harvard Business Review*, *75*(2), 105–116. https://doi.org/Article

[5] Llewellyn, E. (2017). National Strategy for Improved Availability and Access: Visibility and Analytics Network VAN High Level Operating Model CONFIDENTIAL DRAFT v1.0 Not for Distribution Hidden Slides are NOT included in other versions Executive Summary: Slide 1 of 5

[6] Qi, Y., Boyer, K. K., & Zhao, X. (2009). Supply chain strategy, product characteristics, and performance impact: Evidence from Chinese manufacturers. *Decision Sciences*, *40*(4), 667–695. https://doi.org/10.1111/j.1540-5915.2009.00246.x

[7] Simchi-Levi, D., Clayton, A., & Raven, B. (2013). When One Size Does Not Fit All. MITSloan Management Review, 54(2), 15–17

[8] Prinja, S., Bahuguna, P., Tripathy, J. P., & Kumar, R. (2015). Availability of medicines in public sector health facilities of two North Indian States. *BMC Pharmacology and Toxicology*, *16*(1), 1–11. https://doi.org/10.1186/s40360-015-0043-8
[9] Sullivan, K. W., Harris, G. A., Farrington, P. A., & Componation, P. J. (2007). Classification Of Products for Alignment with Supply Chain Strategy, pp.1–6.
[10] Stock Outs in South Africa – A National Crisis. (2013). Retrieved from http://stockouts.org/Download/stop_stockouts_report_2013pdf_1.pdf
[11] 2015 Stock Outs National Survey Third Annual Report – South Africa. (2016). Retrieved from http://stockouts.org/Download/2015_stock_outs_national_survey.pdf
[12] Payne, T., & Peters, M. J., "What Is the Right Supply chain for Your Products?" The International Journal of Logistics Management, Vol. 15 No. 2 (2004), pp. 77-92.
[13] GS1 (2015). Executive Summary. Product Classification in Healthcare, (April), 1–18.
[14] Wen, H., Tong, J., & Yao, S. (2010). The Demand Chain Management for New World Winery : Key Features of Chinese Wine Market, 4385–4391.
[15] Nagashima, M., Wehrle, F. T., Kerbache, L., & Lassagne, M. (2015). Impacts of adaptive collaboration on demand forecasting accuracy of different product categories throughout the product life cycle. *Supply Chain Management: An International Journal*, *20*(4), 415–433. https://doi.org/10.1108/SCM-03-2014-0088
[16] Naim, M. M., & Gosling, J. (2011). On leanness, agility and leagile supply chains. *International Journal of Production Economics*, *131*(1), 342–354. https://doi.org/10.1016/j.ijpe.2010.04.045
[17] Ramkumar, N., Subramanian, P., Narendran, T. T., & Ganesh, K. (2013). An Approach for Setting Inventory Norms for a Two-Echelon Supply Chain. *International Journal of Information Systems and Supply Chain Management*, *6*(4), 102–119. https://doi.org/10.4018/ijisscm.2013100106
[18] Ervolina, T. R., Ettl, M., Lee, Y. M., & Peters, D. J. (2009). Managing product availability in an assemble-to-order supply chain with multiple customer segments. *Supply Chain Planning: Quantitative Decision Support and Advanced Planning Solutions*, 145–168. https://doi.org/10.1007/978-3-540-93775-3_6
[19] Hiremath, N. C., Sahu, S., & Tiwari, M. K. (2013). Multi objective outbound logistics network design for a manufacturing supply chain. *Journal of Intelligent Manufacturing*, *24*(6), 1071–1084. https://doi.org/10.1007/s10845-012-0635-8
[20] Taylor, P., Gan, T., Grunow, M., Gan, T., & Grunow, M. (2015). International Journal of Computer Integrated Manufacturing Concurrent product and supply chain design : a literature review , an exploratory research framework and a process for modularity design, (July). https://doi.org/10.1080/0951192X.2015.1067908
[21] Li, F., & Huang, B. (2012). The impact of product category on pricing decision in a two-echelon dual-channel supply chain. *Proceedings of 2012 IEEE International Conference on Service Operations and Logistics, and Informatics, SOLI 2012*, (71171085), 269–272. https://doi.org/10.1109/SOLI.2012.6273544
[22] Fichtinger, J., Chan, C., & Yates, N. (2017). A joint network design and multi-echelon inventory optimisation approach for supply chain segmentation. *International Journal of Production Economics*, (August), 1–9. https://doi.org/10.1016/j.ijpe.2017.09.003
[23] Micieta, B., Lieskovsky, R., & Binasova, V. (2016). Design of Product segmentation in assembly line. *MM Science Journal*, *2016*(December), 1608–1612. https://doi.org/10.17973/MMSJ.2016
[24] Rajan, R., & Wang, Y. (2016). *Obsolescence Reduction Through Product Segmentation*. *Proceedings - 2011 4th IEEE International Conference on Utility and Cloud Computing, UCC 2011*. https://doi.org/10.1109/UCC.2011.36
[25] Şen, A. (2008). The US fashion industry: A supply chain review. *International Journal of Production Economics*, *114*(2), 571–593. https://doi.org/10.1016/j.ijpe.2007.05.022
[26] Roscoe, S., & Baker, P. (2014). Supply chain segmentation in the sporting goods industry. *International Journal of Logistics Research and Applications*, *17*(2), 136–155. https://doi.org/10.1080/13675567.2013.837869
[27] Martínez, S., Errasti, A., & Rudberg, M. (2015). Adapting Zara's "Pronto Moda" to a value brand retailer. *Production Planning and Control*, *26*(9), 723–737. https://doi.org/10.1080/09537287.2014.971526
[28] Sardar, S., Lee, Y. H., & Memon, M. S. (2016). Multi-objective outsourcing strategies for functional and fast fashion products in textile supply chain. *International Journal of Engineering and Technology*, *8*(2), 870–886
[29] Rofin, T. M., & Mahanty, B. (2017). Optimal dual-channel supply chain configuration for product categories with different customer preference of online channel. *Electronic Commerce Research*, 1–30. https://doi.org/10.1007/s10660-017-9269-4
[30] Hofbauer, P., Withalm, J., & Wölfel, W. (2007). Superposition of Supply Chains classification by ontological properties. *IFAC Proceedings Volumes (IFAC-PapersOnline)*, *1*(PART 1), 220–224. https://doi.org/http://dx.doi.org/10.3182/20071002-MX-4-3906.00036
[31] Liao, S. H., Chen, Y. J., & Yang, H. W. (2013). Mining customer knowledge for channel and product segmentation. *Applied Artificial Intelligence*, *27*(7), 635–655. https://doi.org/10.1080/08839514.2013.813195
[32] Abdulsalam, Y., Gopalakrishnan, M., Maltz, A., & Schneller, E. (2015). Health Care Matters: Supply Chains In and Of the Health Sector. *Journal of Business Logistics*, *36*(4), 335–339. https://doi.org/10.1111/jbl.12111

[33] Yadav, P., Lydon, P., Oswald, J., Dicko, M., & Zaffran, M. (2014). Integration of vaccine supply chains with other health commodity supply chains: A framework for decision making. *Vaccine*, *32*(50), 6725–6732. https://doi.org/10.1016/j.vaccine.2014.10.001

[34] Hua, M. N., Tang, H. J., & Wu, Z. L. (2016). Analysis of a pharmaceutical reverse supply chain based on unwanted medications categories in household. *IEEE International Conference on Industrial Engineering and Engineering Management*, *2016–Decem*, 1493–1497. https://doi.org/10.1109/IEEM.2016.7798126

[35] Kritchanchai, D., & Meesamut, W. (2015). Developing Inventory Management in Hospital. *Int. J Sup. Chain. Mgt*, *4*(2), 11–19. https://doi.org/10.1108/IMDS-12-2015-0500

[36] Lee, H. L. (2002). Aligning Supply Chain Strategies with Product Uncertainties California Management Review Reprint Series Aligning Supply Chain Strategies with Product Uncertainties. *Cmr*, *44*(3), 104–119. https://doi.org/10.2307/41166135

[37] Christopher, M, & Towill, d "Marrying lean and agile paradigms," *Proceedings of EUROMA Conference*, Ghent, Belgium (June 3-6, 2000).

[38] Christopher, M., & Towill, D. R. "Developing Market Specific Supply Chain Strategies," *International Journal of Logistics Management*, Vol. 13 No. 1 (2002), pp. 1-14.

[39] Dawe, P., Pittman, A., & von Koeller, E. (2015). Segmentation in the Consumer Supply Chain: One Size Does Not Fit All. Retrieved April 17, 2018, from https://www.bcg.com/publications/2015/segmentation-in-the-consumer-supply-chain-one-size-does-not-fit-all.aspx

[40] Ketchen, D. J., & Hult, G. T. M. (2007). Bridging organization theory and supply chain management: The case of best value supply chains. *Journal of Operations Management*, *25*(2), 573–580. https://doi.org/10.1016/j.jom.2006.05.010

USING DATA-DRIVEN ANALYTICS TO DEVELOP A MATERIAL BALANCE OVER A FERROCHROME FURNACE

K. Campbell[1], W. Booysen[2] & M. Kleingeld[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
kcampbell@rems2.com
[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
wbooysen@rems2.com
[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
mkleingeld@rems2.com

ABSTRACT

Furnaces used for ferrochrome production are complex systems. There are a significant number of inlet and outlet streams with various parameters. However, some of these parameters are not always measured, which can limit decision-making abilities. Linking available data and additional information together with data analytics can possibly produce estimates of the unknown streams. It is, therefore, necessary to perform a material balance on a typical ferrochrome furnace to evaluate the underlying fundamentals of this concept.

This paper provides a brief overview of the furnace parameters measured in practice, before presenting an approach to perform the material balance. The available measurements of input and output streams are used together with literature-based compositions. This analytical approach links the known composition together with the known mass to estimate the unknown streams. The analytics are structured in such a way that it can later be automated.

The approach is applied to several industrial case studies and the results are presented in order to provide a proof of concept. The analysis manages to balance all elements (in and out) within an accuracy margin of 1.25%. The results are further discussed to illustrate the potential benefit in various application areas.

This work was sponsored by ETA operations

[1] Ms. K. Campbell holds a Master's of Engineering in Mechanical Engineering from the North-West University. She is currently enrolled for a PhD degree at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr. W. Booysen is a registered professional engineer and holds a PhD in Electrical Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Prof. M. Kleingeld is a registered professional engineer and holds a PhD in Mechanical Engineering. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

Furnaces used for ferrochrome (FeCr) production are complex systems [1]. There are a significant number of inlet and outlet streams with various parameters [2]. However, due to difficulty (and in some cases financial restraints), some of these parameters are not always measured [3] which can limit decision-making abilities. Linking available data and additional information together with data analytics can possibly produce accurate estimates of the unknown streams. It is, therefore, necessary to perform a material balance on a typical ferrochrome furnace to evaluate the underlying fundamentals of this concept.

Section 2 of this paper will focus on research done on the FeCr production process as well as mass and composition measurements typically conducted at a furnace. Data quality analysis was also researched and summarised. The method to be followed for conducting a material balance over a typical FeCr furnace is described in Section 3. Even though every furnace is unique, generic steps and examples are provided. The methodology from Section 3 was applied to a real-life FeCr furnace and used as case study for this investigation. The results are provided and discussed in Section 4, while the paper is concluded in Section 5.

2. RESEARCH BACKGROUND

2.1 Ferrochrome production process

The process of FeCr production is an energy intensive one [4, 5]. The total electricity consumption of such a process typically ranges between 3.3 and 4.2 kWh/t of FeCr produced [6, 7, 8]. FeCr is mainly used for stainless steel production, where approximately 1 tonne is needed to produce 3 – 3.5 tonnes of stainless steel [9].

Production is accomplished by feeding raw materials in the form of chromite ores, carbon-rich materials (reductant such as anthracite, char, and coke), and additives (fluxes, in the form of quartz, limestone, dolomite, etc.) to an arc furnace [10, 11, 12]. In some cases, the raw materials are prepared before being fed to the furnace: pelletising, sintering, and drying techniques are often performed on raw materials to produce a dry and uniform feed to the furnace, which would increase furnace stability [13, 14].

Electricity is used to heat up the furnace and melt the raw materials by means of an electric arc [15, 16]. Due to the heat provided, various chemical reactions take place, causing reduction of the metal oxides within the chromite ore to a final metal product, FeCr. Together with the main metal product, waste material (slag) as well as off-gas also exit the furnace as by-products [17]. Figure 2-1 shows an illustration of this process [18].

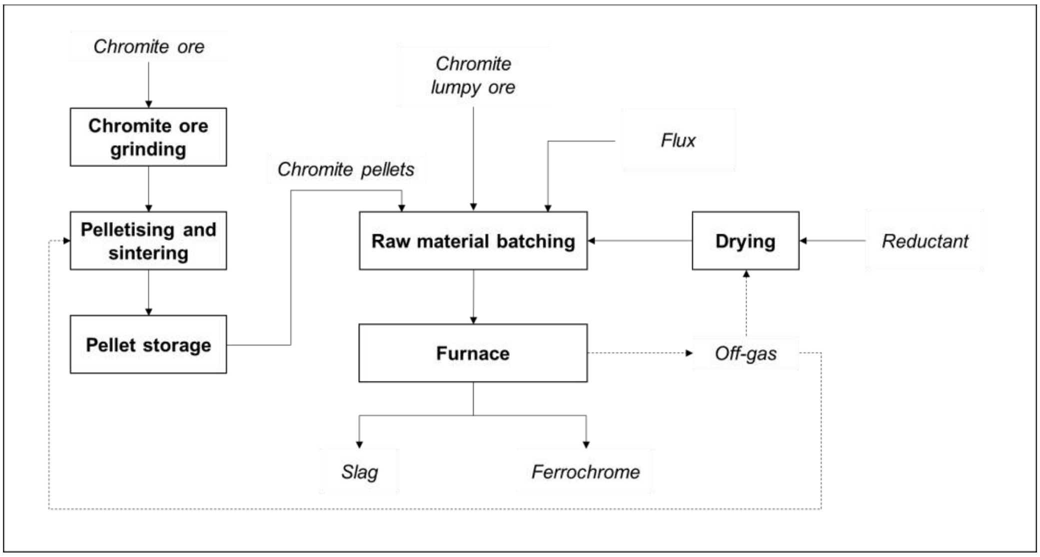


Figure 2-1: Ferrochrome production process

For the purpose of simplification, this diagram has been summarised to establish the focus area of this study: FeCr furnace, input, and output streams. A simplified illustration of this process is thus shown in the diagram below (Figure 2-2).

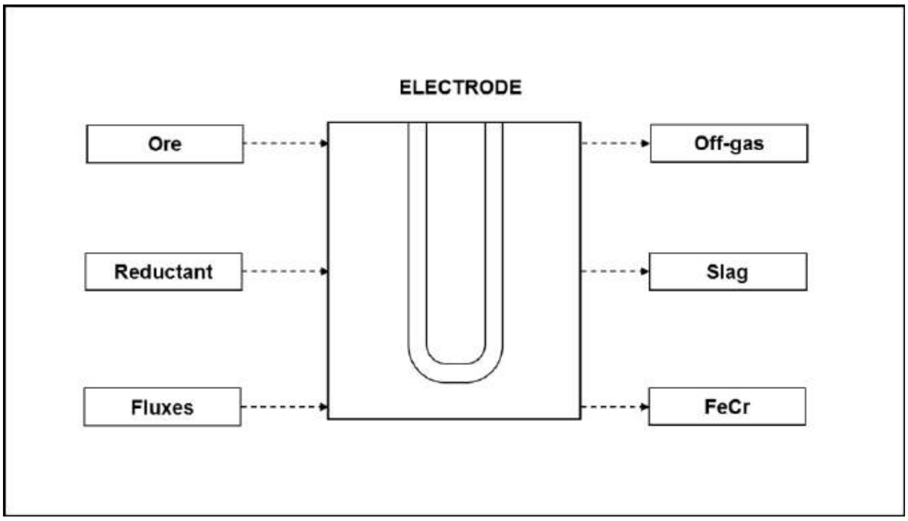


Figure 2-2: Simplified illustration of the FeCr production process

As mentioned, various reactions take place within the furnace which cause the FeCr product to be formed. The most significant reactions included the following [1, 18, 17, 19]:

$Cr_2O_3 + 3C \rightarrow 2Cr + 3CO$ Equation 2-1

$FeO + C \rightarrow Fe + CO$ Equation 2-2

$SiO_2 + 2C \rightarrow Si + 2CO$ Equation 2-3

$H_2O + C \rightarrow H_2 + CO$ Equation 2-4

$CO_2 + C \rightarrow 2CO$ Equation 2-5

$CaCO_3 \rightarrow CaO + CO_2$ Equation 2-6

From these reactions it can be assumed that the prominent elements present within the FeCr furnace are the following: Cr, Fe, C, Si, O, H, and Ca.

2.2 Typical measurements

Not all parameters are always measured at ferroalloy furnaces [20]. Mass balances can be used to predict mass and compositions of certain unmeasured streams [14]. The measurements typically available for the general furnace will be discussed briefly.

FeCr metal product

The main objective of the process referred to within this study is the production of FeCr metal, with a certain alloy grade. The mass and composition of the metal produced is therefore monitored closely and measured continuously [20]. The metal is tapped a few times per day, solidified, crushed, and then weighed on weigh-bridges[i]. Metal samples are also sent for regular (usually daily) composition analysis. A typical composition of FeCr product is as follows: 56.7% Cr, 33.8% Fe, 7.2% C, and 2.3% Si [21].

Raw materials

The mass of raw materials (chromite ore, reductant, and fluxes) are usually measured at weigh bins before being batched to the furnace [22]. The composition, however, is rarely known on site. Occasionally sampling of the reductants takes place on site. However, this is generally done long before batching (before materials are stored on stockpiles) creating a significant buffer capacity. From literature, the composition of each of the raw materials normally used in FeCr production is summarised below:

Cr ore: 50% Cr_2O_3, 25% FeO, 9% MgO, 10% Al_2O_3, 5% SiO_2, 1% CaO [21]

Anthracite: 88.94% C, 3.4% H, 2.32% O, 1.55% N, 0.8% S [23]

Char: 77.84% C, 0.34% H, 21.11% O, 0.71% N [24]

Coke: 89% C, 3.6% H, 1.56% H, 4.95% S [25]

Dolomite: 100% $CaMg(CO_3)_2$ [26]

Limestone: 100% $CaCO_3$ [27]

Quartz: 100% SiO_2 [28]

By-products

The two by-products (slag and off-gas) have often been considered waste streams and are rarely measured accurately[ii]. The slag composition is, however, estimated to ensure the required slag composition [29]. From time to time, the slag mass is calculated by using a slag to metal ratio (usually between 1.1 and 1.8 tonnes of slag produced per tonnes of metal [30]), which is estimated by random sampling. The amount of slag produced can also be estimated based on an aluminium (Al_2O_3) balance. This is done with the assumption that the slag analysis is done accurately and representatively [31].

Then slag and off-gas compositions are given below:

Slag: 23.2% SiO_2, 24.7% Al_2O_3, 19.8% MgO, 3.0% CaO, 10.7% FeO, 18.6% Cr_2O_3 [31]

Off-gas: 75-90% CO, 2-10% CO_2, 2-15% H_2, 2-7% N_2 [2, 19]

The typical compositions are converted to element-based compositions by using the molecular weight of each formula and element. The results are summarised in Table 2-1:

Table 2-1: Typical compositions for streams entering and exiting a FeCr furnace (element-based)

	Fe	Cr	Si	S	C	Al	O	Ca	Mg	H	N
A	17.5	34.2	2.3	-	-	5.2	34.6	0.7	5.4	-	-
B	-	-	-	0.8	88.9	-	2.3	-	-	3.4	1.6
C	-	-	-	-	77.8	-	21.1	-	-	0.3	0.7
D	-	-	-	5.0	89.0	-	0	-	-	3.6	1.6
E	-	-	-	-	13.0	-	52.2	21.7	13.0	-	-
F	-	-	-	-	12.0	-	48.0	40.0	-	-	-
G	-	-	46.7	-	-	-	53	-	-	-	-
H	33.8	56.7	2.3	-	7.2	-	-	-	-	-	-
I	8.3	12.7	10.8	-	-	12.8	41.3	2.1	11.9	-	-
J	-	-	-	-	33-41	-	45-59	-	-	2-15	2-7

** Chrome ore = A, anthracite = B, char = C, coke = D, dolomite = E, limestone = F, quartz = G, FeCr metal = H, slag = I, off-gas = J.*

The typical mass ratio of input and output streams are as follows [17, 18]:

- 2.1 – 2.4 tonnes Cr ore fed per tonne FeCr metal produced
- 0.5 – 0.55 tonnes reductants fed per tonne FeCr metal produced
- 0.1 – 0.45 tonnes fluxes per tonne FeCr metal produced
- 1.1 – 1.3 tonnes slag per tonne FeCr metal produced
- 0.9 – 1.1 tonnes off-gas per tonne FeCr metal produced

The chromium recovery of this process is usually around 80% - 85% [18].

2.3 Data quality

Data quality is an important factor for a number of reasons and needs to be monitored pro-actively [32]. When evaluating a FeCr furnace, a significant amount of data needs to be collected, processed, and analysed. Therefore, thorough data evaluation is necessary.

A study done by Booysen [33] provided a data quality evaluation method that was developed to identify any potential errors and abnormalities. This method consists of four steps: the first three steps aim to identify abnormal measurements, whereas the final step identifies abnormal operation. A schematic flow of this method is shown in Figure 2-3.

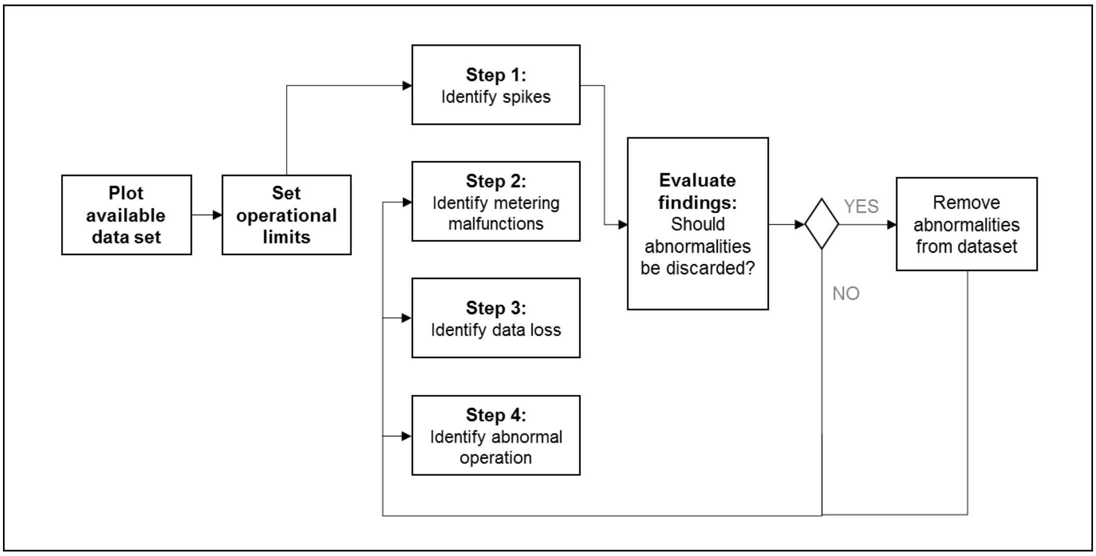


Figure 2-3: Data quality evaluation [33]

Figure 2-4 provides a simplified visualisation of a dataset containing typical measurement abnormalities. The minimum and maximum limits are selected based on the variable being assessed. Steps 1 – 3 are indicated on this figure as follows: data spikes (step 1), faulty data (step 2), and data loss (step 3).

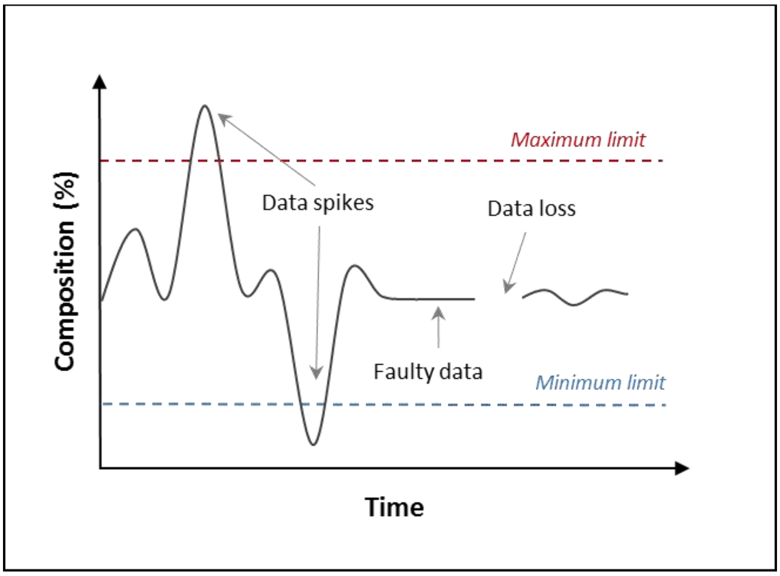


Figure 2-4: Data evaluation – Identifying abnormal measurements

The first step of identifying abnormal measurements within the data evaluation method, aims to detect data spikes. Data spikes happen when there is a failure on measurement equipment or when communication is briefly lost. Even though these tend to happen over short periods of time, their amplitude (very high or very low) can still significantly influence the accuracy of calculations.

The next step is to identify metering malfunctions, which could lead to faulty data being logged. This occurrence is illustrated as a constant value in Figure 2-4, where the last data reading is typically repeated for a number of resolutions, until the malfunction has been resolved. Even though this data will still fall within the operational limits, the results will be influenced by this incorrect, constant value.

Step 3 aims to identify data loss. This can generally be detected where no data has been recorded, as indicated by the blank space in Figure 2-4. The final step of this method aims to identify abnormal system operation. This is illustrated by Figure 2-5.

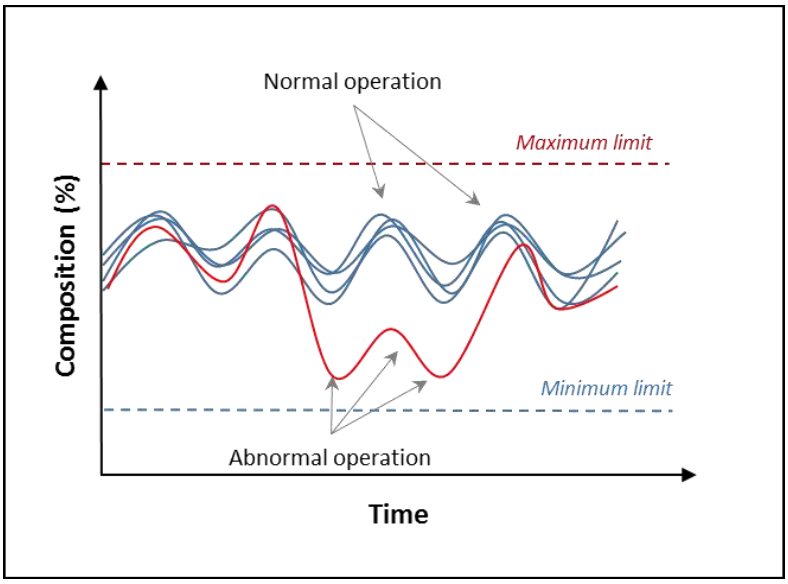


Figure 2-5: Data evaluation – Identifying abnormal operation

In Figure 2-5 there are numerous profiles that follow the same trend, indicating normal operation. The red profile however is deemed abnormal, as it differs from the trend. The utmost care should be taken when investigating these results as there is, unfortunately, no fixed rule as to what is defined as "abnormal operation". This step therefore requires a thorough understanding of the process being evaluated.

The aim of this data quality evaluation method is to evaluate a dataset, remove any measurement abnormalities, and identify operational abnormalities. An accurate, high quality dataset is the final outcome of this method.

2.4 Research background conclusion

The research background section focused on the following:

- The basic process of FeCr production, so that all the theoretical components and elements are known to the reader.
- The typical setup and measurements, so that the reader can understand the practical implications and challenges faced when conducting a mass balance.
- Data quality, so that the practical data can be assessed to be usable before being included in the mass balance.

This was applied to develop a functional method on constructing a mass balance for a typical FeCr furnace.

3. METHODOLOGY

The methodology is divided into three main steps: *Collect information, layouts, and data*; *evaluate data quality*; and *construct a material balance*. An overview of this strategy is presented in Figure 3-1.

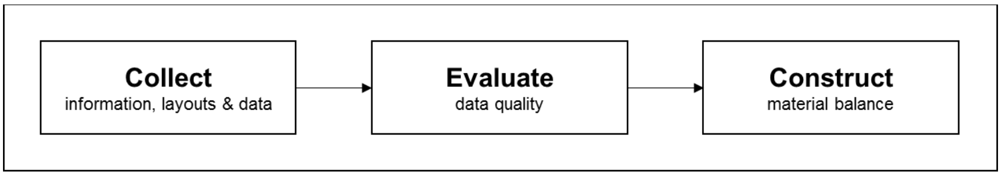
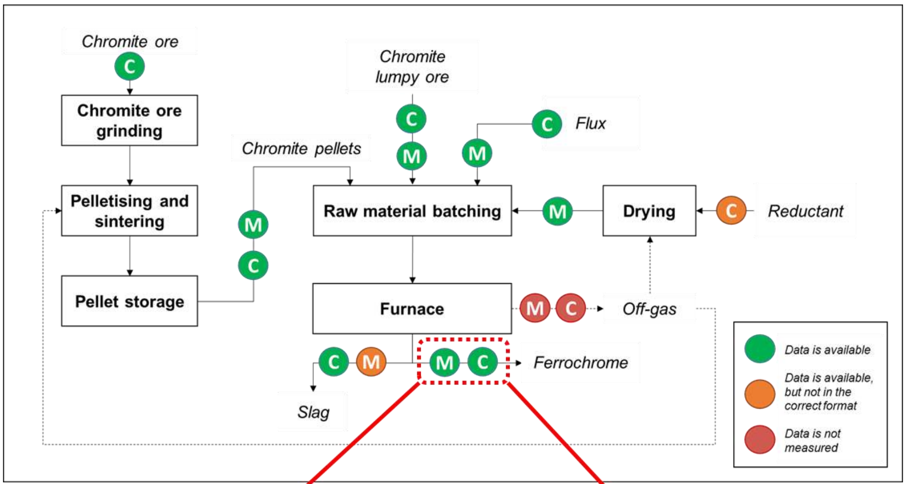


Figure 3-1: Basic three-step methodology for developing a FeCr material balance

This strategy provides the basic knowledge to developing a material balance, based on data-driven analytics. Each step is discussed in more detail in Sections 3.1 through 3.3.

3.1 Collect information, layouts, and data

The first step is to collect all relevant information from the site being evaluated. This includes layouts of the furnace, the relevant points of measurement indicated on layouts, as well as the corresponding mass and composition data for the evaluation period. Figure 3-2 indicates the outcome of this step. Note that for illustration purposes, only data collected for ferrochrome metal produced is shown on the layout (mass and composition data). However, data from all available streams must be collected.



Date	Ferrochrome Metal produced (MASS, tonnes)
01-Jan-18	x
02-Jan-18	x
03-Jan-18	x
04-Jan-18	x
05-Jan-18	x
06-Jan-18	x
07-Jan-18	x
08-Jan-18	x
09-Jan-18	x
10-Jan-18	x
11-Jan-18	x
12-Jan-18	x
...	...
31-Dec-18	x

Date	Ferrochrome Analysis data (COMPOSITION, mass%)			
	Cr	Fe	C	Si
01-Jan-18	x%	x%	x%	x%
02-Jan-18	x%	x%	x%	x%
03-Jan-18	x%	x%	x%	x%
04-Jan-18	x%	x%	x%	x%
05-Jan-18	x%	x%	x%	x%
06-Jan-18	x%	x%	x%	x%
07-Jan-18	x%	x%	x%	x%
08-Jan-18	x%	x%	x%	x%
09-Jan-18	x%	x%	x%	x%
10-Jan-18	x%	x%	x%	x%
11-Jan-18	x%	x%	x%	x%
12-Jan-18	x%	x%	x%	x%
...	...	...	...	...
31-Dec-18	x%	x%	x%	x%

Figure 3-2: Collect information, layouts, and data

The circles on the layout represent points of (required) measure. Green circles indicate that data is available, the orange represent data available, however not in the correct format, while the red circles indicate that data is not available, or not even measured. The "M" symbols refer to mass measurements (by means of weighbridges or weigh bins), whereas the "C" denotes composition analysis sampling taking place.

If data is to be required for an orange-indicated measurement point (data that is available, but not in the correct format), data needs to be processed to the correct analysis by making use of various assumptions or methods. An example of such a case is when the reductant composition is based on proximate analyses instead of ultimate analyses. The slag mass may also not be available, however, slag to metal ratio data can be used to calculate a theoretical slag mass.

3.2 Evaluate data quality

Having an accurate and "good" quality dataset is essential. It is important that all the data received in step 1 is a reflection of the truth. Step 2 of the methodology is thus to evaluate the data quality.
The method provided in section 2.3 (Figure 2-3) needs to be followed for all mass and composition measurements in order to clean the dataset. This will ensure a representative, "good" quality dataset. Only then can the data be processed and used in further calculations.

3.3 Constructing a mass balance

A basic mass, or material balance is based on the principle of "mass in equals mass out" [34]. A material balance can be performed on the total mass entering and exiting the FeCr furnace, but also based on the individual chemical elements. These two different approaches are shown in Equation 3-1 and Equation 3-2, respectively:

$$M_{ore} + M_{reductant} + M_{flux} = M_{FeCr} + M_{slag} + M_{off-gas} \quad \text{Equation 3-1}$$

$$\sum(M_{X_{in}}) = \sum(M_{X_{out}}) \quad \text{Equation 3-2}$$

Where "M" represents the total mass of a certain stream or element, and "X" refers to a certain chemical element. Constructing a mass balance is the third and final step of the method, as referred to by Figure *3-1*. This step will however take place over four different phases:

3.3.1 Calculate total mass of unmeasured streams

From the research section in 2.2, it is gathered that the mass of all batching streams (ore, reductant, and flux) are generally measured, as well as the mass of FeCr produced. Slag mass data is usually in the form of a ratio; however, it can be derived to estimate the total mass of slag. Thus, the only unknown stream mass is that of off-gas. By using Equation 3-1, the mass of the total off-gas stream can be calculated.

3.3.2 Assume the composition of all streams

Section 2.2 was summarised into Table 2-1 which provided the typical compositions for streams entering and exiting a FeCr furnace (element-based). This table can be used to assign compositions to all material streams. Since the total mass of each input and output stream is known, the theoretical mass of each element entering and exiting the furnace can be calculated. This is illustrated in Table 3-1, where A_1 (lit) refers to the mass of element "1" present in stream "A" (based on literature composition).

Table 3-1: Mass per element based on literature compositions

Stream	**Mass per element** *(tonnes)*				
	1	2	3	…	x
A	A_1 (lit)	A_2 (lit)	A_3 (lit)	…	A_x (lit)
B	B_1 (lit)	B_2 (lit)	B_3 (lit)	…	B_x (lit)
C	C_1 (lit)	C_2 (lit)	C_3 (lit)	…	C_x (lit)
…	…	…	…	…	…
n	n_1 (lit)	n_2 (lit)	n_3 (lit)	…	n_x (lit)

3.3.3 Update the assumed compositions with measured plant data

After the literature compositions were used to assume the mass of each element, actual plant data can be used to replace the data from literature. For the purpose of this example, note that composition data for streams A and C was available. This is illustrated in Table 3-2, where "A_1 (lit)" was replaced by "A_1 (actual)".

Table 3-2: Mass per element updated with measured compositions (where applicable)

Stream	**Mass per element** *(tonnes)*				
	1	2	3	…	x
A	A_1 (actual)	A_2 (actual)	A_3 (actual)	…	A_x (actual)
B	B_1 (lit)	B_2 (lit)	B_3 (lit)	…	B_x (lit)
C	C_1 (actual)	C_2 (actual)	C_3 (actual)	…	C_x (actual)
…	…	…	…	…	…
n	n_1 (lit)	n_2 (lit)	n_3 (lit)	…	n_x (lit)

3.3.4 Visualising results

Once all streams are updated with actual plant data (where available), the results of the mass balance can be visualised in various different ways. An example of a desired plot is the total mass in vs. total mass out (based on Equation 3-2). This plot is shown in Figure 3-3:

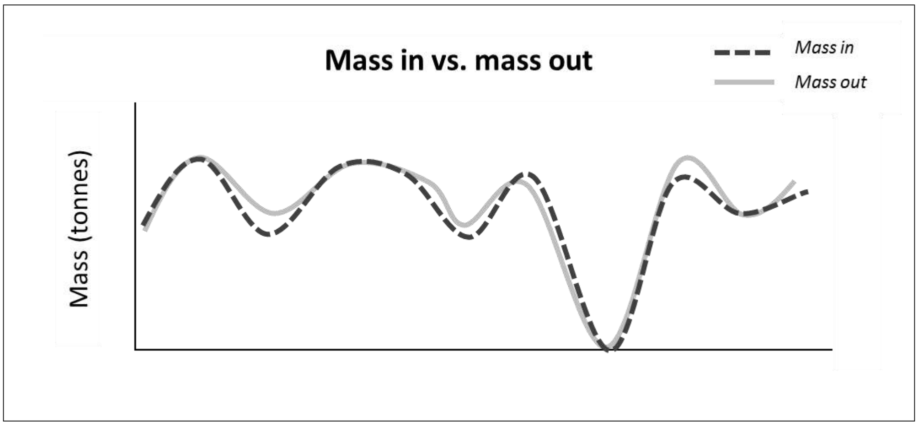


Figure 3-3: Visualising results

A final mass balance error can also be calculated by using the following equation:

$$\textbf{Mass balance error}\ (\%) = \frac{\sum(M_{X_{in}}) - \sum(M_{X_{out}})}{\sum(M_{X_{in}})} \qquad \text{Equation 3-3}$$

For the purpose of this study, an error margin of 3% has been chosen as an acceptable error.

3.4 Final methodology

The three steps (from section 3.1, 3.2 and 3.3) can be combined to present the final methodology to be followed for developing a material balance over a FeCr furnace (presented in Figure 3-4).

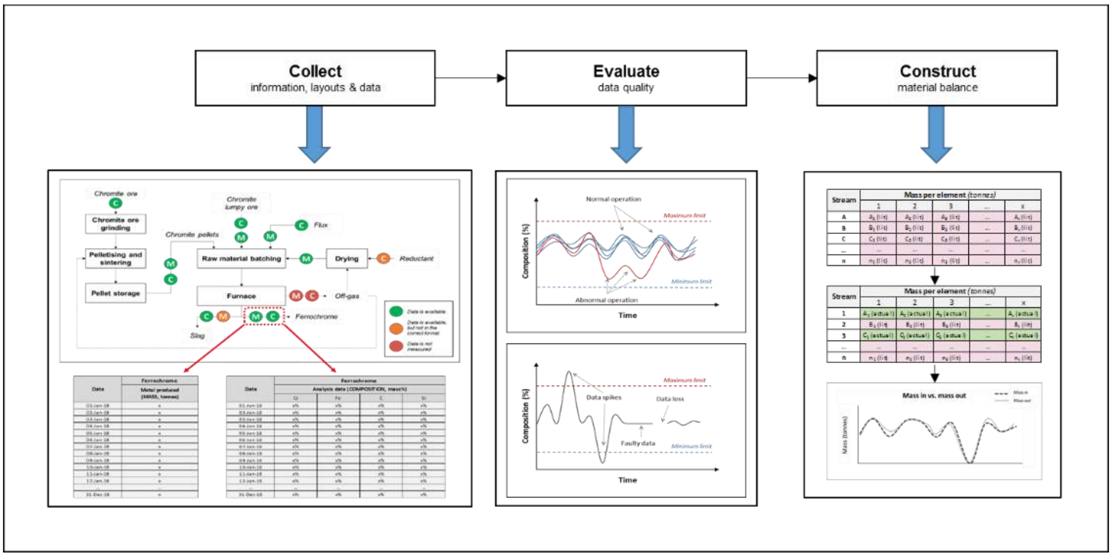


Figure 3-4: Final three-step methodology for developing a FeCr material balance

4. CASE STUDY: RESULTS AND DISCUSSION

The final methodology is applied to an industrial case study and is used to highlight the specific outcomes of the method.

4.1 Collect information, layouts, and data

Furnace X is evaluated in terms of the methodology. A layout of the furnace, the relevant points of measure as well as the corresponding mass and composition data over a three-year evaluation period has been collected. This is shown in Figure 4-1.

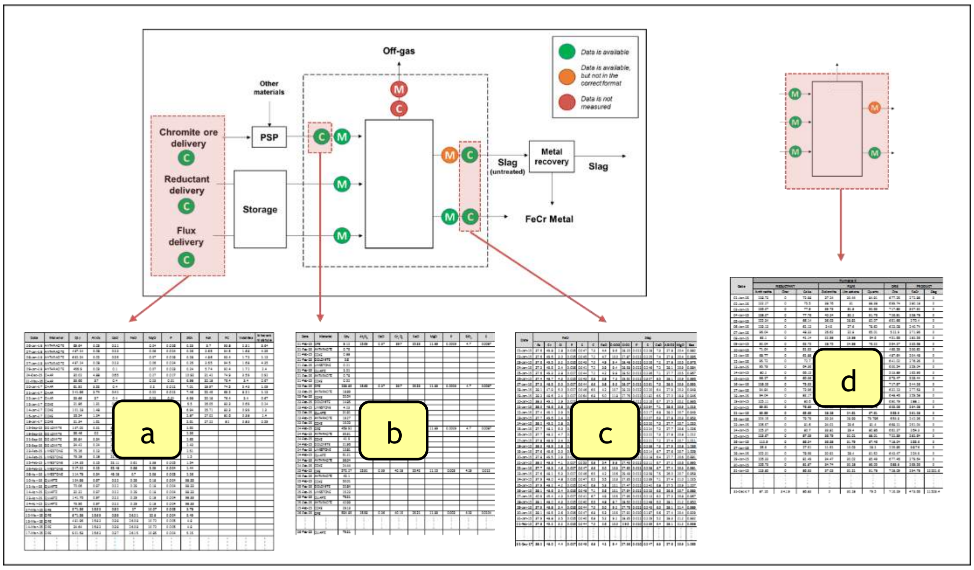


Figure 4-1: Collect information, layouts, and data

Chrome ore, reductant (anthracite, char, and coke), and fluxes (dolomite, limestone, and quartz) are sampled on delivery (dataset “a”). Chromite ore is pelletised (PSP), and sampled again before batching (dataset “b”). The composition of the FeCr metal and slag is also determined when exiting the furnace (dataset “c”). The mass of all raw materials is determined in weigh bins right before batching, whereas the FeCr metal and slag mass are measured at weigh bridges (dataset “d”).

These four datasets have been collected so as to be evaluated throughout the next step (4.2). Note that the slag mass meter is indicated in orange. This is due to the slag mass only being logged per month, when the rest of the data is in daily resolution. Thus, all data will be converted to monthly resolution when processing and calculations commence.

The time delay that may occur between mass and composition measurements is uncertain. All compositions will therefore be averaged to a constant annual value, in order to compensate for any possible storage capacities.

4.2 Evaluate data quality

The raw datasets as received in 4.1 (datasets “a”, “b”, “c”, and “d”) were evaluated based on the method discussed in Section 2.3. Figure 4-2 illustrates one of the datasets before and after the dataset has been cleaned.

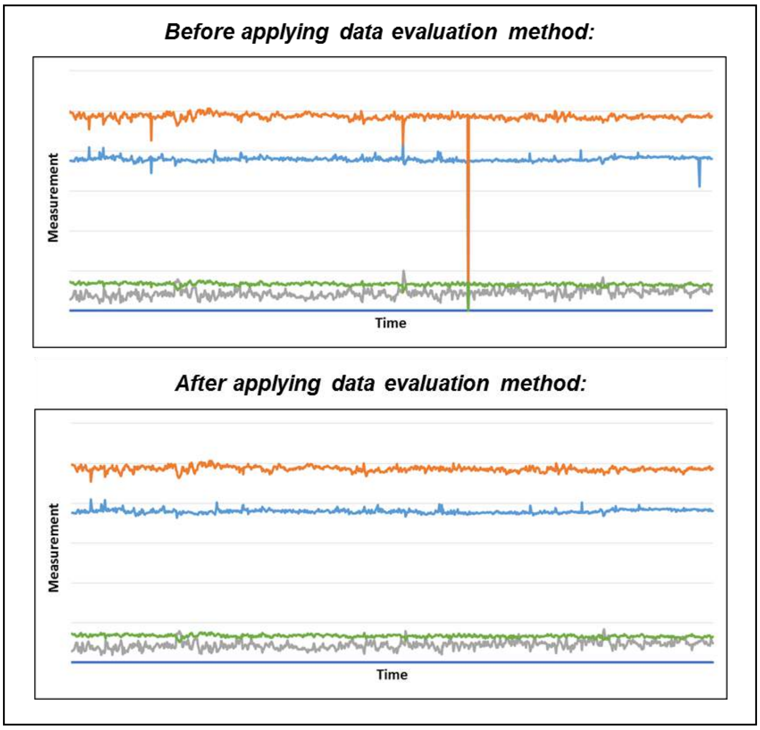


Figure 4-2: Data quality evaluation

Any abnormal measurements within the data have been identified, investigated, and removed if necessary. Annual shutdowns were detected (where furnace was shut down for a month or two), however data was not removed since this is not an abnormal operational occurrence. Consequently, after applying the method of data quality evaluation, all datasets are of high quality and can be used in the final step to construct a mass balance.

4.3 Constructing a mass balance

4.3.1 Calculate total mass of unmeasured streams

From Section 4.1 it was noted that the only unknown mass is that of the off-gas stream. By using Equation 3-1, the mass of the total off-gas stream can be calculated:

$$M_{ore} + M_{reductant} + M_{flux} = M_{FeCr} + M_{slag} + M_{off-gas}$$
$$M_{off-gas} = M_{ore} + M_{anth} + M_{char} + M_{coke} + M_{dolomite} + M_{limestone} + M_{quartz} - M_{FeCr} - M_{slag}$$

Table 4-1: Calculating the total mass of unmeasured (off-gas) streams

Date	Consumption							Production		
	REDUCTANT			FLUX			ORE	FeCr	Slag	Off-gas
	Anth.	Char	Coke	Dolomite	Limestone	Quartz	Ore			
Jan-15	x xxx	0	x xxx	xxx	xxx	x xxx	xx xxx	x xxx	xx xxx	x xxx
Feb-15	x xxx	0	x xxx	xxx	xxx	x xxx	xx xxx	x xxx	x xxx	x xxx
Mar-15	x xxx	0	x xxx	xxx	xxx	x xxx	xx xxx	x xxx	xx xxx	x xxx
Apr-15	x xxx	0	x xxx	xxx	xxx	x xxx	xx xxx	x xxx	xx xxx	x xxx
May-15	x xxx	0	x xxx	xxx	xxx	x xxx	xx xxx	x xxx	xx xxx	x xxx
Jun-15	0	0	0	0	0	0	0	0	0	0
Jul-15	0	0	0	0	0	0	0	0	0	0
Aug-15	xxx	0	xxx	xxx	0	xxx	x xxx	x xxx	x xxx	x xxx
Sep-15	x xxx	xxx	x xxx	xxx	0	x xxx	xx xxx	x xxx	xx xxx	x xxx
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Dec-17	x xxx	xxx	x xxx	0	xxx	x xxx	xx xxx	x xxx	xx xxx	x xxx

Since all mass streams are now known, the next phase of constructing a mass balance can be investigated.

4.3.2 Assume the composition of all streams

Table 2-1 is used to assign compositions to all material streams. Since the total mass of each input and output stream is known, the theoretical mass of each element entering and exiting the furnace is calculated. This is illustrated in Table 4-2, where the pink blocks represent mass per element (calculated by using literature compositions).

Table 4-2: Mass per element based on literature compositions

Date	Anthracite S	Anthracite C	Anthracite O	Anthracite H	Anthracite N	Char C	Char O	Char H	Char N	Coke S	Coke C	Coke H	Coke N	Dolomite Fe	Dolomite Si	Dolomite C	Dolomite Al	Dolomite O	Dolomite Ca	Dolomite Mg	Limestone Fe	Limestone Si	Limestone C	Limestone Al	Limestone O	Limestone Ca	Limestone Mg
Jan-15	xx	xxxx	xx	xxx	xx	0	0	0	0	xxx	xxxx	xx	xx	0	0	xxx	0	xxx	xxx	xxx	0	0	xxx	0	xxx	xxx	0
Feb-15	xx	xxxx	xx	xx	xx	0	0	0	0	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Mar-15	xx	xxxx	xx	xxx	xx	0	0	0	0	xxx	xxxx	xx	xx	0	0	xxx	0	xxx	xxx	xxx	0	0	xxx	0	xxx	xxx	0
Apr-15	xx	xxxx	xx	xxx	xx	0	0	0	0	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xxx	0	xxx	xxx	0
May-15	xx	xxxx	xx	xxx	xx	0	0	0	0	xxx	xxxx	xx	xx	0	0	xx	0	xxx	xx	xx	0	0	xx	0	xxx	xxx	0
Jun-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-15	xx	xxx	xx	xx	xx	0	0	0	0	xx	xxx	xx	x	0	0	xx	0	xxx	xx	xx	0	0	0	0	0	0	0
Sep-15	xx	xxxx	xx	xxx	xx	xxx	xx	x	x	xx	xxxx	xx	xx	0	0	xxx	0	xxx	xxx	xxx	0	0	0	0	0	0	0
Oct-15	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xxx	0	xxx	xxx	xxx	0	0	x	0	xx	xx	0
Nov-15	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Dec-15	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xx	xx	0	0	xx	0	xxx	xxx	0
Jan-16	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xx	xx	0
Feb-16	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	0	0	0	0	0
Mar-16	xx	xxxx	xx	xxx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	xx	0	xxx	xx	xx	0	0	xxx	0	xxx	xxx	0
Apr-16	xx	xxxx	xx	xx	xx	xxx	xx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xx	xx	0	0	xx	0	xxx	xxx	0
May-16	xx	xxxx	xx	xx	xx	xxx	xx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Jun-16	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Jul-16	x	xxxx	xx	xx	xx	xxx	xx	x	x	xx	xxx	xx	xx	0	0	xx	0	xx	xx	xx	0	0	xx	0	xxx	xxx	0
Aug-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sep-16	xx	xxxx	xx	xx	xx	xxx	xx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Oct-16	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xx	xxxx	xx	xx	0	0	xx	0	xxx	xxx	xx	0	0	xx	0	xxx	xxx	0
Nov-16	xx	xxxx	xx	xxx	xx	xx	x	0	0	xxx	xxxx	xx	xx	0	0	x	0	xx	x	x	0	0	xxx	0	xxx	xxx	0
Dec-16	xx	xxxx	xx	xxx	xx	xxx	xx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Jan-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Feb-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Mar-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Apr-17	xx	xxxx	xx	xx	xx	xxxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
May-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Jun-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Sep-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Oct-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Nov-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0
Dec-17	xx	xxxx	xx	xx	xx	xxx	xxx	x	x	xxx	xxxx	xx	xx	0	0	0	0	0	0	0	0	0	xx	0	xxx	xxx	0

Date	Quartz Fe	Quartz Si	Quartz Al	Quartz O	Quartz Ca	Quartz Mg	Ore Fe	Ore Cr	Ore Si	Ore P	Ore S	Ore C	Ore Al	Ore O	Ore Ca	Ore Mg	FeCr Fe	FeCr Cr	FeCr Si	FeCr C	Slag Fe	Slag Cr	Slag Si	Slag P	Slag S	Slag Al	Slag O	Slag Ca	Slag Mg	OFF-GAS C	OFF-GAS O	OFF-GAS H	OFF-GAS N
Jan-15	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-15	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
Mar-15	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-15	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
May-15	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-15	0	xxx	0	xxx	0	0	xxx	xxxx	xxx	0	0	0	xxx	xxxx	xx	xxx	xxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxx	xxx	xx	xxx	xxx	xxx	xxx	xxx
Sep-15	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Oct-15	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-15	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-15	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jan-16	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-16	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Mar-16	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-16	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
May-16	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-16	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jul-16	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxx	xxxx	xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xx
Aug-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sep-16	0	xxx	0	xxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
Oct-16	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-16	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-16	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jan-17	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Mar-17	0	xxxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
May-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Sep-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Oct-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-17	0	xxx	0	xxxx	0	0	xxxx	xxxx	xxx	0	0	0	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxxx	xxxx	0	0	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx

4.3.3 Update the assumed compositions with measured plant data

Actual plant data can be used to replace the data used from literature. As indicated in Figure *4-1*, all streams' compositions are available, except for the off-gas. This is illustrated in Table *4-3*, where the pink blocks represent mass per element calculated by using literature compositions, and the green blocks represent mass per element calculated by using actual measured plant data.

Table 4-3: Mass per element updated with measured compositions (where applicable)

Date	ANTHRACITE S	ANTHRACITE C	ANTHRACITE O	ANTHRACITE H	ANTHRACITE N	CHAR C	CHAR O	CHAR H	CHAR N	COKE S	COKE C	COKE H	COKE N	DOLOMITE Fe	DOLOMITE Si	DOLOMITE C	DOLOMITE Al	DOLOMITE O	DOLOMITE Ca	DOLOMITE Mg	LIMESTONE Fe	LIMESTONE Si	LIMESTONE C	LIMESTONE Al	LIMESTONE O	LIMESTONE Ca	LIMESTONE Mg
Jan-15	0	xxxx	xxx	xxx	xx	0	0	0	0	0	xxxx	xx	xx	x	xx	xx	x	xxx	xxx	xxx	x	x	xx	x	xxx	xxx	xx
Feb-15	0	xxxx	xxx	xxx	xx	0	0	0	0	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	x	x	xx	x	xxx	xxx	xx
Mar-15	0	xxxx	xxx	xxx	xx	0	0	0	0	0	xxxx	xx	xx	x	xx	xx	x	xxx	xxx	xxx	x	x	xx	x	xxx	xxx	xx
Apr-15	0	xxxx	xxx	xxx	xx	0	0	0	0	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
May-15	0	xxxx	xxx	xxx	xx	0	0	0	0	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Jun-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-15	0	xxx	xxx	xx	xx	0	0	0	0	0	xxx	xx	xx	x	x	xx	0	xxx	xx	xx	0	0	0	0	0	0	0
Sep-15	0	xxxx	xxx	xx	xx	xxx	xx	xx	x	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	0	0	0	0	0	0	0
Oct-15	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	0	x	x	0	xx	xx	x
Nov-15	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	x	x	xx	x	xxx	xxx	xx
Dec-15	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Jan-16	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	x	x	x	x	xxx	xxx	x
Feb-16	0	xxxx	xxx	xx	xx	xxx	xxx	xx	x	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xxx	0	0	0	0	0	0	0
Mar-16	0	xxxx	xxx	xxx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Apr-16	0	xxxx	xxx	xx	xx	xxx	xx	x	x	0	xxxx	xx	xx	x	x	xx	x	xxx	xx	xx	x	x	xx	x	xxx	xxx	xx
May-16	0	xxxx	xxx	xx	xx	xxx	xx	x	x	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Jun-16	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Jul-16	0	xxx	xxx	xx	xx	xxx	xx	x	xx	0	xxx	xx	xx	x	x	xx	x	xx	xx	xx	x	x	xx	x	xxx	xxx	xx
Aug-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sep-16	0	xxxx	xxx	xx	xx	xxx	xx	xx	x	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Oct-16	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	x	x	xx	x	xxx	xxx	xx	x	x	xx	x	xxx	xxx	xx
Nov-16	0	xxxx	xxx	xxx	xx	xx	x	x	0	0	xxxx	xx	xx	0	0	xx	0	xx	x	x	x	x	xx	x	xxx	xxx	xx
Dec-16	0	xxxx	xxx	xxx	xx	xxx	xx	xx	x	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Jan-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Feb-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Mar-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Apr-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
May-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Jun-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Sep-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Oct-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Nov-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx
Dec-17	0	xxxx	xxx	xx	xx	xxx	xxx	xx	xx	0	xxxx	xx	xx	0	0	0	0	0	0	0	x	x	xx	x	xxx	xxx	xx

Date	QUARTZ Fe	QUARTZ Si	QUARTZ Al	QUARTZ O	QUARTZ Ca	QUARTZ Mg	ORE Fe	ORE Cr	ORE Si	ORE P	ORE S	ORE C	ORE Al	ORE O	ORE Ca	ORE Mg	FECR Fe	FECR Cr	FECR Si	FECR C	UNTREATED SLAG Fe	UNTREATED SLAG Cr	UNTREATED SLAG Si	UNTREATED SLAG P	UNTREATED SLAG S	UNTREATED SLAG Al	UNTREATED SLAG O	UNTREATED SLAG Ca	UNTREATED SLAG Mg	OFF-GAS C	OFF-GAS O	OFF-GAS H	OFF-GAS N
Jan-15	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-15	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
Mar-15	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-15	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
May-15	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-15	x	xxx	xx	xxx	0	x	xxx	xxxx	xxx	0	x	0	xxx	xxxx	xx	xxx	xxxx	xxxx	xxx	xxx	xx	xxx	xxx	0	xx	xxx	xxx	xxx	xxx	xxx	xxx	xxx	xx
Sep-15	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Oct-15	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-15	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-15	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
Jan-16	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-16	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Mar-16	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-16	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	x	xx	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
May-16	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-16	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jul-16	x	xxx	x	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	x	xx	xxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xx
Aug-16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Sep-16	x	xxx	xx	xxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxx	x	xx	xxxx	xxxx	xxx	xxx	xxxx	xxxx	xxx	xxx
Oct-16	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-16	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-16	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jan-17	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Feb-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Mar-17	x	xxxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Apr-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
May-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Jun-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Jul-17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Aug-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Sep-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Oct-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Nov-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx
Dec-17	x	xxx	xx	xxxx	x	x	xxxx	xxxx	xxx	x	x	0	xxxx	xxxx	xx	xxxx	xxxx	xxxx	xxx	xxx	xxx	xxx	xxxx	x	xx	xxxx	xxxx	xxx	xxxx	xxxx	xxxx	xxx	xxx

4.3.4 Visualising results

The data of the mass balance is visualised in various ways in order to see the results of certain key parameter indicators. The following visualisations will be provided:

- Cr from ore vs. Cr in FeCr product
- Fe from ore vs. Fe in FeCr product
- Total C in vs. C in product
- Total mass in vs. total mass out

Each of these visualisations will be provided and discussed briefly.

Cr from ore vs. Cr in FeCr product:

The amount of Cr in the ore, trended together with the amount of Cr present in the final metal product, ultimately gives an idea of the recovery of Cr from the furnace. This trend is shown in Figure 4-3.

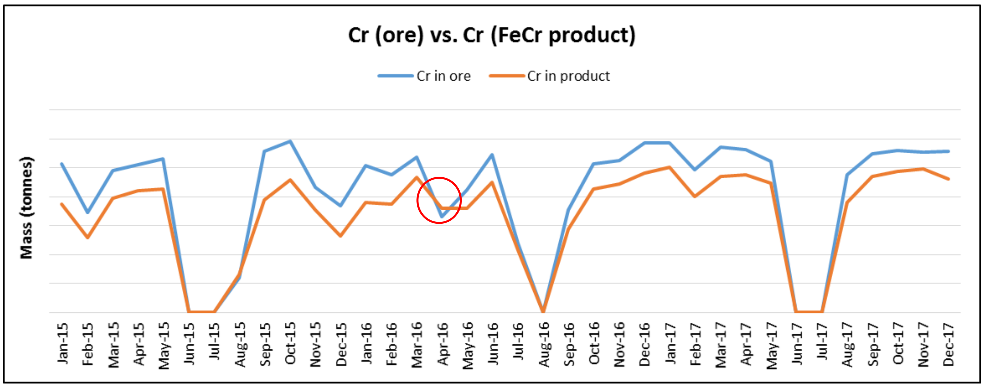


Figure 4-3: Cr from ore vs. Cr in FeCr product

These two lines must be as close together as possible, which would indicate that most of the Cr fed to the furnace is recovered in the final FeCr metal product. In this case, the relationship between the two streams seems to be typical, with the exception of April 2016. Here, the amount of Cr in the product exceeds the amount of Cr fed by the ore, which is not possible. Further data investigation can be done when such an incident occurs.

The variance between the two lines is calculated to be 18%, which means that there was a 100% - 18% = 82% recovery of chrome. This correlates well with the value found in literature (as stated in section 2.2: between 80 and 85%).

Fe from ore vs. Fe in FeCr product:

The visualisation of Fe is similar to that of the Cr, discussed in the previous point. The Fe trend is shown in Figure 4-4.

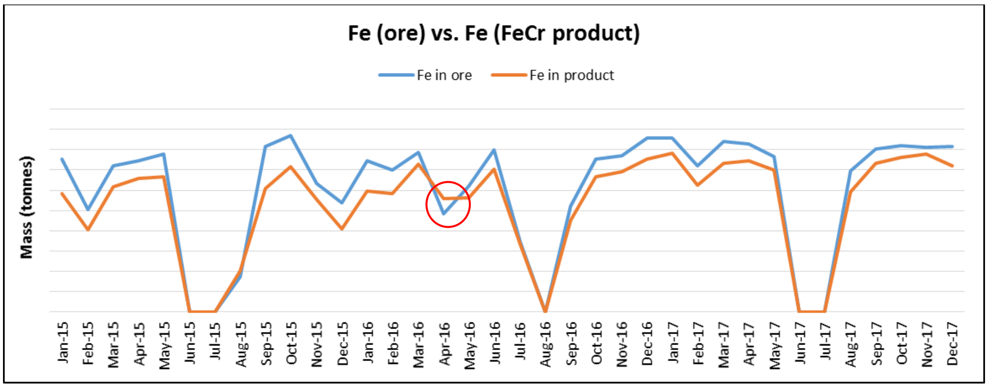


Figure 4-4: Fe from ore vs. Fe in FeCr product

Once again, the relationship between the two streams seems to be typical, with the exception of April 2016, as discussed previously. The difference between the variables is calculated to be 12%.

Total C in vs. C in product:

The total amount of carbon entering the furnace, trended together with the amount of carbon in the metal product, ultimately shows how much carbon is captured in the solid phase. The difference between the two lines would typically represent the amount of carbon emitted as part of the off-gas. The trend is shown in Figure 4-5.

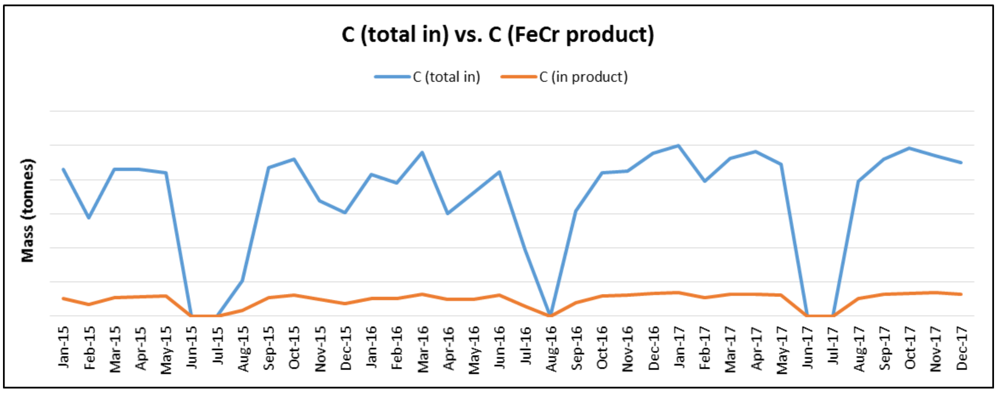


Figure 4-5: Total C in vs. C in product

It can be seen that the difference between the two lines is quite significant. The calculated difference is 86%, which means that 86% of the carbon used in FeCr production is emitted to the atmosphere (usually in the form of CO and CO_2). Further studies are in progress, focusing on off-gas emissions specifically [35].

Total mass in vs. total mass out:

The aim of this study is to perform a material balance on a typical ferrochrome furnace, which will be the final visualisation. The total mass of elements entering the furnace will be plotted against the total mass of elements exiting. This is shown in Figure 4-6.

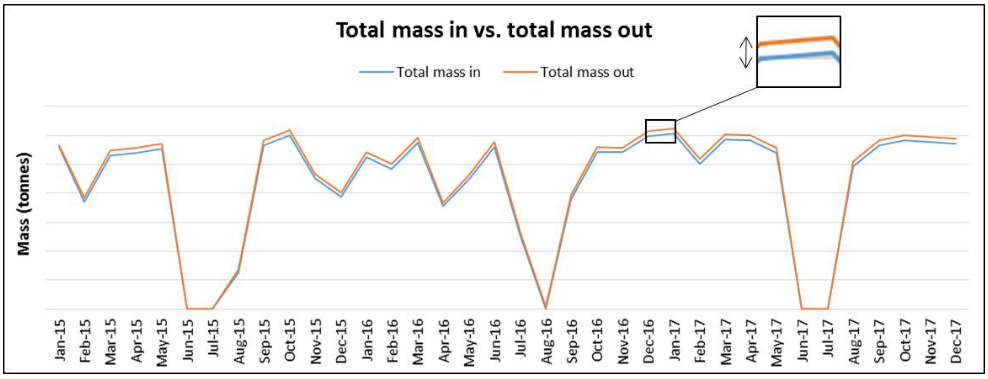


Figure 4-6: Total mass in vs. total mass out

Ideally, one would expect the two lines to follow the exact same trend, which would indicate a 'good' mass balance. This is almost the case, where a 1.25% discrepancy is found between the mass in and the mass out. This remains within the 3% error margin of an acceptable error, as chosen for this study.

Generally, the remaining results (not shown) also correlated well within the limits of typical mass ratios of input and output streams found in literature (as provided in section 2.2, p. 4).

5. CONCLUSION

This paper provided a brief overview of the furnace parameters measured in practice, before presenting an approach to perform the material balance. The available measurements of input and output streams were used together with compositions from literature. The analytical approach linked the known composition together with the known mass, to estimate the unknown streams.

The approach was applied to an industrial case study. The analysis managed to balance all elements (in and out) within an accuracy margin of 1.25%. Completing the mass balance up to this point can significantly decrease limitations on decision-making abilities.

6. BIBLIOGRAPHY

[1] **Gasik, M.M.** 2013. *Handbook of ferroalloys: Theory and technology*, Espoo: Elsevier.

[2] **Beukes, J., Dawson, N. & Van Zyl, P.** 2010. Theoretical and practical aspects of Cr(VI) in the South African ferrochrome industry, Infacon XII - Proceedings of 12th International Ferroalloys Congress, Helsinki.

[3] **Kühn, R., Geck, H.G., & Schwerdtfeger, K.** 2005. Continuous off-gas measurement and energy balance in electric arc steelmaking, *ISIJ International*, vol. 45, nr. 11, pp. 1587-1596.

[4] **Roos, J. & Hearn, A.** 2004. Optimising the effective use of energy in the ferroalloy industry through innovative technology, *Infacon X - Proceedings of 10th International Ferroalloys Congress*, Cape Town.

[5] **Xiao, Y., Yang, Y. & Holappa, L.** 2006. Tracking chromium behaviour in submerged arc furnace for ferrochrome production, *2006 TMS Fall Extraction & Processing Meeting: Sohn Int'l Symposium*, pp. 583-594.

[6] **Otani, Y. & Ichikawa, K.** 1974. Manufacture and Use of Prereduced Chromium-ore, *Infacon I - Proceedings of 1st International Ferroalloys Congress*, Johannesburg,

[7] **Pan, X.** 2015. Effect of South African reductants on ferrochrome production, *Infacon XIV - Proceedings of Fourteenth International Ferroalloys Congress*, Kiev.

[8] **Pan, X.** 2015. Learning and teaching using process modelling and simulation, *International Conference on Mining, Minerals and Metallurgical Engineering*, Harare.

[9] **Ideas 1st Research.** 2010. *Ferrochrome*, Ideas 1st Research, Mumbai,.

[10] **Hattingh, J. & Friend, J.** 2003. Environmental and economic implications of slag disposal practices by the ferrochromium industry: A case study, *South African water research commission*, vol. 29.

[11] **Du Preez, S.** 2014. *Cr(VI) generation during flaring of off-gas from closed ferrochromium submerged arc furnaces*, North West University, Potchefstroom.

[12] **Basson, J., Curr T., & Gericke W.** 2006. *South Africa's ferroalloys industry - present status and future outlook*.

[13] **Sager, D., Grant, D., Stadle, R., & Schreiter, T.** 2010. Low cost ferroalloy extraction in DC-arc furnace at Middleburg Ferrochrome, *The Journal of The Southern African Institute of Mining and Metallurgy,* vol. 110, pp. 717-724.

[14] **Biermann, W., Cromarty, R. & Dawson, N.** 2012. Economic modelling of a ferrochrome furnace, *The Journal of The Southern African Institute of Mining and Metallurgy,* vol. 112, pp. 301-308.

[15] **Anon.** 2004. *Reductant characteisation in ferroalloys*, University of the Witwatersrand, Johannesburg.

[16] **Holappa, L.** 2010. Toward sustainability in ferroalloys, *Infacon XII - Proceedings of 12th International Ferroalloys Congress*, Helsinki.

[17] **Kapure, G., Kari, C., Rao, S.M.en Raju, K.S.** 2007. Use of chemical energy in submerged arc furnace to produce ferrochrome: prospects and limitations, *Infacon XI - Proceedings of 11th International Ferroalloys Congress*, New Delhi.

[18] **Riekkola-Vanhanen, M.** 1999. *Finnish expert report on best available techniques in ferrochromium production*, Finnish environment insitute, Helsinki.

[19] **Niemelä, P., Krogerus, H. & Oikarinen, P.** 2004. Formation, characteristics and utilisation of cogas formed in ferrochromium smelting, *Infacon X - Proceedings of 10th International Ferroalloys Congress*, Cape Town,

[20] **Broekman, B. & Ford, K.** 2004. The development and application of a HCFeMn furnace simulation model for Assmang Ltd., *Infacon X - Proceedings of 10th International Ferroalloys Congress*, Cape Town.

[21] **Bergman, K. & Kjellberg, B.** 1997. *DC Arc Furnace Technology Applied to Smelting Applications*, Helsinki.

[22] **Harman, C. & Rama Rao, N.** 2007. Use of sintered pellets in production of high carbon ferrochrome, *Infacon XI - Proceedings of 11th International Ferroalloys Congress*, New Delhi.

[23] **Van Doorn, J., Bruyn, P. & Vermeij, P.** 1996. Combined combustion of biomass, municipal sewage sludge and coal in an atmospheric fluidized bed installation, *9th European bioenergy conference*.

[24] **Sun, Y., Sridhar, S., Seetharaman, S., Wang, H., Liu, L., Wang, X. & Zhang, Z.** 2016. A Fe-C-Ca big cycle in modern carbon-intensive industries: toward emission reduction and resource utilization, *Energy science and technology*.

[25] **Van der Sloot, H. & Cnubben, P.** 2000. *Verkennende evaluatie kwaliteitsbeinvloeding poederkoolvliegas*, ECN-report.

[26] **Klein, C. & Philpotts, A.** 2017. *Earth materials - Introduction to mineralogy and petrology*, Cambridge: Cambridge University Press.

[27] **O'Brien, M. & Wood, W.** 1998. *The prehistory of Missouri, Columbia*: University of Missouri Press.

[28] **Götze, J. & Möckel, R.** 2012. *Quartz: Deposits, mineralogy and analytics*, Heidelberg: Springer.

[29] **Hockaday, S.A.C. & Bisaka, K.** 2010. Some aspects of the production of ferrochrome alloys in pilot DC arc furnaces at Mintek, *Infacon XII - Proceedings of 12th International Ferroalloys Congress*, Helsinki.

[30] **Sprinzl, F.R.** 2016. *Valorization of ferrochrome slag: Towards increasing the beneficial utilisation of ferrochrome slag in South Africa*, Stellenbosch University, Stellenbosch.

[31] **Nkohla, M.A.** 2006. *Characterization of ferrochrome smelter slag and its implications in metal accounting*, Cape Peninsula University of Technology, Cape Town.

[32] **Herzog, T.N., Scheuren, F.J. & Winkler, W.E.** 2007. *Data quality and record linkage techniques*, Springer.

[33] **Booysen, W.** 2014. *Measurement and verification of industrial DSM projects*, North-West University, Potchefstroom.

[34] **Felder, R.M. & Rousseau, R.W.** 2005. *Elementary principles of chemical processes*, Hoboken (NJ): John Wiley & Sons, Inc.

[35] **Campbell, K.** 2018. *A critical analysis of emission quantification methods in the ferrochrome industry*, North-West University, Potchefstroom.

[i] Knowledge gained from site experience and interviews with site personnel

[ii] Knowledge gained from site experience and interviews with site personnel

OPTIMISING PRODUCTION THROUGH IMPROVING THE EFFICIENCY OF MINE COMPRESSED AIR NETWORKS WITH LIMITED INFRASTRUCTURE

D Nell[1], JIG Bredenkamp[2] & JF v Rensburg[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
dnell@rems2.com
[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
jbredenkamp@rems2.com
[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
jvanrensburg@rems2.com

ABSTRACT

The gold mining industry of South Africa has been experiencing a decline in gold production. This decline is due to a unique set of challenges faced by this industry; encouraging deep-level mines to use existing infrastructure much more efficiently. One area offering large potential for optimisation is addressing deep-level gold mine compressed air network inefficiencies. Through addressing these inefficiencies, an increase in rock penetration rate can be achieved on the pneumatically hand operated rock drills, leading to reduced drilling times and potentially improved production figures.

An investigation performed on Mine A indicated that a specific compressed air network inefficiency contributed to a pressure drop of approximately 19% during peak drilling periods. Simulations indicated that replacing specific undersized pipe sections with the correct sized pipes, would reduce the pressure drop to only 5% during peak drilling periods. The simulated solution was implemented and resulted in a peak drilling pressure increase of approximately 62 kPa. This supply pressure increase translated into a potential production increase of approximately R 11-million per annum.

This work was sponsored by ETA operations

[1] Diaan Nell holds a master in mechanical engineering from the North-West University. He is currently a doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr Johan Bredenkamp is a registered professional engineer and holds a PhD in mechanical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr Johan van Rensburg is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

1.1 South African gold mining

For the 2016 financial year, the gold mining industry in SA contributed 4.4% to the total global gold production, resulting in R28 billion in employee earnings [1]. Unfortunately, for more than the last four decades, the South African gold mining sector faced numerous challenges influencing production. Figure 1-1 illustrates the declining gold mining production trend for the past 13 years.

Figure 1-1: Gold production trends in South Africa [2]

The South African gold mining industry, which was once the leading global gold producer from as early as 1970 (1000 tonnes a year) to 2007, has dropped significantly to sixth place in 2016.On the contrary, the global gold mining sector grew between 2004 to 2016, except durin for the global financial crisis in 2008 [3]. However, South Africa still holds the number three spot regarding gold reserves [2].

1.2 Challenges faced in the South African gold mining industry

It is evident that South African gold mining companies face a clear decline in production as a result of a unique set of challenges, these challenges are summarised as follow [4]:

- Electricity cost increases [5-8];
- Labour cost increases [9, 10] and
- Infrastructure limitations [11, 12]

These challenges have a large impact on the production trends of the South African gold mining sector. To assist in addressing this production decrease, the focus should be on optimising current infrastructure to ensure larger profit margins are secured. This infrastructure specifically entails the equipment contributing to large operating cost. This optimisation will be especially true for mining operations which will not necessarily benefit from mine mechanisation due to inherent limitations[13].

1.3 The efficiency of a compressed air network infrastructure and it's correlation to production

In literature, gold mine compressed air network are strongly related to production as it plays a crucial part in drilling processes [14-16]. On the other hand, it is also regarded as one of the most inefficient and energy intensive systems on deep-level mines[16]. These inefficiencies can mainly be ascribed to the complexity of these systems, with piping networks extending over several kilometres[17]. In some instances, these systems may even have overall efficiencies, of close to 2% when summing the various system inefficiencies, such as compressor efficiency, network losses, drill efficiency etc. together [18].

The reticulation network of these systems consists of all the pipes connecting the supply with the demand side. It is common to find pipe sizes decreasing from the main shaft to the various levels, haulages, cross-cuts and stoping areas. It is therefore essential to remain within design specifications as mining progresses deeper to ensure unwarranted line restrictions are not created [19].

Addressing these inefficiencies is therefore crucial in optimising the mining process which in turn will lower operational costs. Unfortunately, identifying these inefficiencies are difficult and tedious, due to the limited availability of monitoring infrastructure on most South African mines [16]. A simplified and feasible solution strategy is therefore required.

2. LITERATURE REVIEW

2.1 Preamble

Compressed air networks form the backbone of several mining processes. Their main purpose is, however, to supply adequate compressed air pressure and airflow required for underground drilling [18].

"Drilling" in the gold mining industry refers to the process of drilling holes in the rock face with the use of rock drills [20]. These holes are used for explosives charging during the blasting process.

The performance of a pneumatic drill is highly subject to the supplied pressure and directly influences the drill's efficiency [21]. The supply pressure can be drastically influenced by reticulation network losses, an inefficiency plaguing mines in South Africa [18].

2.2 Compressed air network inefficiencies

Compressed air network inefficiencies can primarily be divided into two categories, namely major (linear) head losses and minor head losses, [22]. These losses are accounted for in the Bernoulli equation:

Equation 2-1: Bernoulli equation with hydraulic loss [23, 24]

$$\frac{v_1^2}{2g}+\frac{p_1}{\rho g}+z_1=\frac{v_2^2}{2g}+\frac{p_2}{\rho g}+z_2+\Delta h_{ls}$$

Where

v_1	=	Initial fluid velocity	m/s
p_1	=	Initial fluid pressure	Pa
v_2	=	New fluid velocity	m/s
p_2	=	New fluid pressure	Pa
g	=	Gravitational acceleration	m/s^2
ρ	=	Fluid density	kg/m^3
z_1	=	Initial fluid elevation	m
z_2	=	New fluid elevation	m
Δh_{ls}	=	Hydraulic losses	m (sum of the major and minor head losses)

These inefficiencies greatly affect the supplied pressure to the hand-held rock drills which in turn affects the rock penetration rate. The next section will elaborate more on this.

2.3 Rock penetration rate and correlation with production

Rock penetration rate refers to the penetration of the drill bit into the rock within a given amount of time. The parameter is normally measured in millimetres per second mm/s and serves as a measure to evaluate drill performance [25]. Various studies have been conducted to estimate how penetration rate varies when parameters such as pressure, thrust, drill bit size, rock composition etc. changes [18, 21, 26, 27]. These studies are further scrutinised in the following section.

Comparing the rock penetration rate at different drilling conditions, such as at varying pressures, gives an indication of how changes to the compressed air reticulation network may affect drilling and in turn influence production. The influence on production is explained as follows:

- If drilling targets are met, a decrease in drilling time will result in more blasting holes being drilled, which leads to more panels being blasted. This ultimately leads to an increase in ore volume that can be excavated [18, 21].
- Where drilling targets are met, decreasing the drilling time will lead to drill operators finishing earlier. This will especially be advantageous when travelling time to workplaces increases [4].
- During drilling periods, more electrical energy is consumed on the compressors due to the larger compressed air demand. Decreasing the drilling times will therefore automatically lead to energy savings [28].

There is also a health and safety benefit coupled with the decrease in drilling times. Pneumatic drills operate at high noise levels, and continuous exposure could be hazardous to drill rig operators [34]. Reducing the exposure times to high noise levels and the inherently dangerous mining environment, specifically encountered closest to the currently mined area, increases overall health and safety of the workforce [29].

2.4 Previous research relevant to the study

Several studies have been evaluated to gain better insight into compressed air network inefficiencies and their influence on production These studies specifically focussed on the effects on the compressed air network when addressing these inefficiencies and the impact on drilling rates at various supply conditions respectively. Table 2-1 shows the focus of each study.

Table 2-1: Analysis of previous studies

	Study 1 [17]	Study 2 [30]	Study 3 [14]	Study 4 [21]	Study 5 [18]	Study 6 [26]	Study 7 [27]
Research performed on addressing compressed air line inefficiencies	✓	✓	✓				
Quantifying parameters influencing drill penetration rate				✓	✓	✓	✓
Correlating compressed air network improvements to production	✗	✗	✗	✗	✗	✗	✗

From this summary table it is evident that little to no focus has been placed on the correlation between changes in a compressed air network and production improvements. Limited literature was available regarding this exact topic, and therefore this study aims to address this shortcoming.

3. METHODOLOGY

3.1 Evaluating the compressed air network

The first step is to evaluate the compressed air network thoroughly. Evaluating the compressed air network entails characterising the network and developing a solution strategy which will address the identified problems. A key step in this evaluation process is to ensure the correct measuring points are identified in the network to ensure the information can be used to scrutinise the problem. These points should be selected prior to main level intakes, splits or cross-cut tap-offs. It is also important to account for any supply changes which might occur during the study. Using a well recorded reference point is imperative, should baseline scaling be required at a later stage. This is especially common at mines with limited infrastructure which require manual measurements to be taken.

The baselines for each measurement point should be evaluated to determine the pressure difference present between any two measuring points. Figure 3-1 illustrates a typical baseline profile.

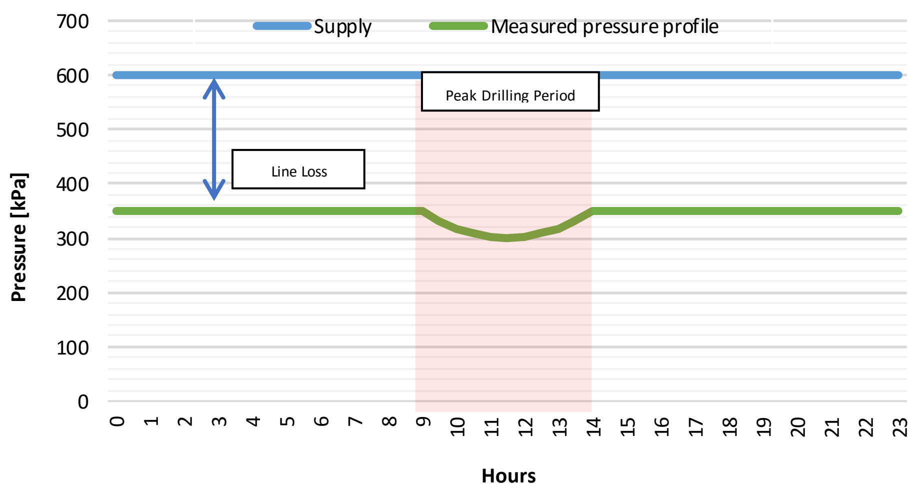


Figure 3-1: Compressed air reticulation problem

The measured baselines can be used to prioritise the identified network inefficiencies through the use of a scoring system with reference to specific criteria as shown in Table 3-1. The total score will indicate how great the optimisation potential is and will be obtained through summing all the sub criteria's scores together. The score for each criterion will be assigned on a scale of 0 (minimum) to 10 (maximum) using Equation 3-1[31].
Equation 3-1: Optimisation potential score calculation

$$Score = \left[\frac{(Measured\ Value - Minimum\ Measured\ Value)}{(Maximum\ Measured\ Value - Minimum\ Measured\ Value)}\right] * 10$$

Table 3-1 will be populated for each measured pipe section within the reticulation network. The criteria for the scoring system are presented as follow:

Table 3-1: Criteria description of prioritising inefficiencies

Criteria	Description
Pressure difference (Low – 0; High – 10)	Pressure difference refers to the difference in pressure between the measured starting point of the pipe section and the measured end.
Pipe length (Long – 0; Short – 10)	Pipe length refers to the total pipe length between the two measured points. The shorter the pipe, the larger the concern for a given pressure difference.
Nominal pipe size (Large – 0; Small – 10)	Nominal pipe size refers to the nominal pipe size of the measured pipe section. The smaller the pipe diameter, the larger the concern for a given pressure difference.
Ease of alteration (Difficult – 0; Easy – 10)	Ease of alteration refers to how easily the pipe can be replaced. This accounts for accessibility, production influence and general repair.
Repair time frame (Time-consuming – 0; Quick – 10)	Repair time frame refers to how long it will take to repair an identified concern. The faster the repair, the more appeal in addressing the problem.
Implementation/ Repair Cost (Expensive – 0; Cheap – 10)	Implementation refers to the costs involved with repairing the identified pipe section/sections. The cheaper the repair, the more appeal in addressing the problem.
Total Score	Summation of the assigned scores.

The next step entails evaluating the expected impact when addressing the highest optimisation potential.

3.2 Verification, Simulation and Validation

The effect of addressing the identified inefficiency will be simulated before the recommendation of any implementation strategies. This will ensure the most optimal solution is selected. The accuracy of th e simulation model should be verified and calibrated by using a controlled data set, which can be used to compare the measured pressure drop across a pipeline.

The simulated results will then be validated once the all recommended changes have been impleme nted and measurements have been taken at the previous measuring points

3.3 Impact on Production

The model developed in Study 4 (see Table 2-1) to approximate drill penetration rates, was selected due to its flexibility at various pressures with the smallest number of required input parameters. [21]. This model (represented by Equation 3-2) is used to calculate the theoretical penetration rate which can be achieved through improved pressures (simulated).

Equation 2-2: Penetration rate approximation

$$RPR = 0.0879242 + 0.0111569 * A - 0.246978 * B + 0.0070986 * C - 0.0000100938 * A^2 + 0.003057 * B^2 - 0.00000760976 * C^2 + 0.0000103687 * A * C - 0.0000546415 * B * C$$

Where

A	=	Pressure	kPa
B	=	Drill bit diameter	mm
C	=	Thrust	N

Table 3-2 summarise the inputs required to calculate the potential financial benefit as shown in Equation 3-3. The production improvement should be calculated for each improved working area to determine the total impact on the mine's production.

Table 3-2: Production impact analysis inputs

Workplace Parameter	Index	Unit
Panel height	*A*	m
Panel width	*B*	m
Number of holes per panel	*C*	-
Hole depth/ drill length	*D*	m
Number of drills per panel/manifold	*E*	-
Forward advancement during blasting	*F*	m
Previous rock penetration rate	*G*	m/s
New rock penetration rate	*H*	m/s
Travelling times	*I*	hr
Gram gold per tonne	*J*	g/tonne
Rock density	*K*	kg/m^3
Gold price	*L*	R/g

From Equation 3-3 the impact on production output and financial benefits through the implementation of the improvements can be determined. These figures will ultimately highlight the feasibility of the study.

Equation 3-3: Financial benefit of additional gold per panel [R]

$$Financial\ benefit\ of\ additional\ gold\ per\ panel\ [R] = A * B * F * H * J * L * \frac{K}{1000 * E} * \left[\frac{1}{G} - \frac{1}{H}\right]$$

4. RESULTS

4.1 Background

The developed methodology was implemented on the compressed air system of a South African gold mine, further referred to as Mine A. The compressed air reticulation network of Mine A is intricate and supplies compressed air from either a compressor house at 1# or 2# to underground mining sections and a gold plant located on surface. Figure 4-1 shows a simplified surface layout of the compressed air network at Mine A.

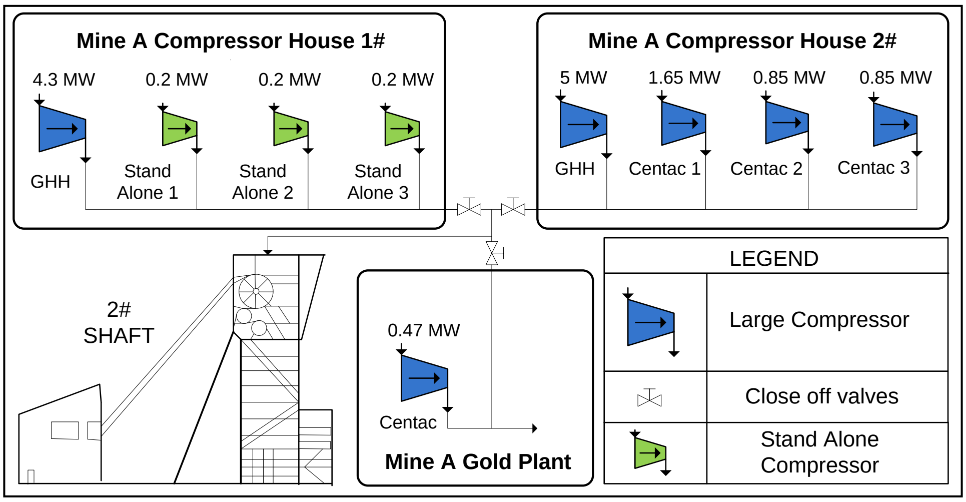


Figure 4-1: Surface compressed air network layout at Mine A

Figure 4-2 illustrates a simplified layout of the various underground mining levels at Mine A, which form part of the compressed air reticulation network. Compressed air is only supplied via 2# shaft.

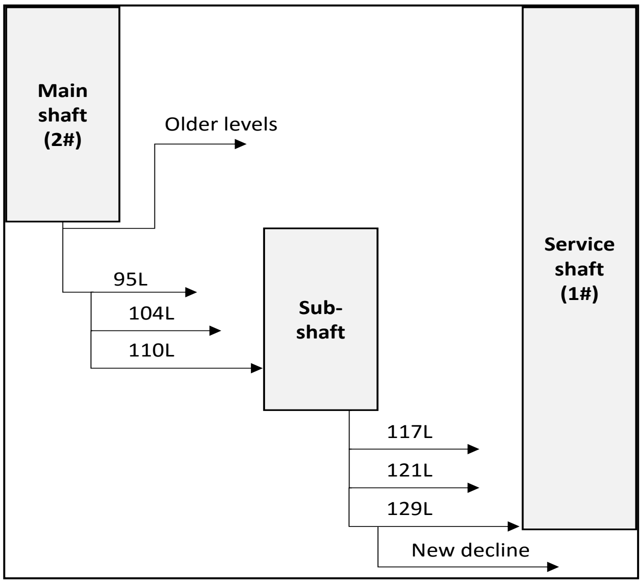


Figure 4-2: Underground level layout at Mine A

Most mining activity commences on 129L, which contains the newly developed decline. The levels from 95L to 117L are known as the older levels and have already been mined out. Compressed air is supplied down the shaft through 95L and 110L to the sub-shaft. From the sub-shaft, compressed air is supplied via 117L, to121L and 129L, respectively.

4.2 Evaluating the compressed air network

One of the largest challenges at Mine A is limited infrastructure in terms of instrumentation and communication to equipment. No active Supervisory Control and Data Acquisition SCADA system is in place, which results in no availability of historical data. Alternative means of data collection had to be implemented to identify the compressed air network inefficiencies.

- ***Measuring Points Identification***

Figure 4-3 illustrates the measurement points (A-G) that were identified to establish the pressure losses from the supply to the point of use. These points served as an initial indication of where the largest pressure losses occurred. Points were selected prior to each level's intake or at the splits on active levels.

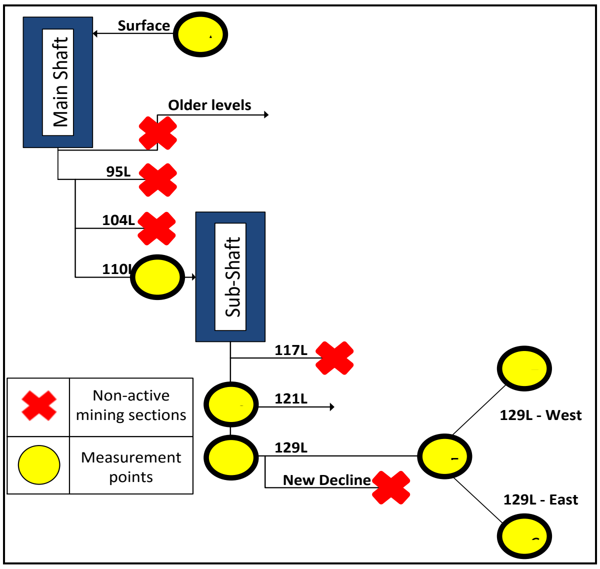


Figure 4-3: Simplified layout of measurement points

The absence of measurement points greatly affected the time-frame of the investigation and meant all measured data had to be translated to a common reference point (due to measurements on different periods) to ensure accurate comparisons. The reference point chosen was the surface conditions due to the ease of obtaining the data.

4.3 Baseline construction and evaluation

A maximum of 1.2% surface pressure deviation was observed during the different measurement periods, whereas the power consumption deviated to a maximum of 4%. This enables the comparison between levels without the need of scaling measured values. These measurements were taken using pressure transmitted logging for a no less than 7 consecutive days. Figure 4-4 shows a summary of the average pressure profiles (baselines) for all the measured data points.

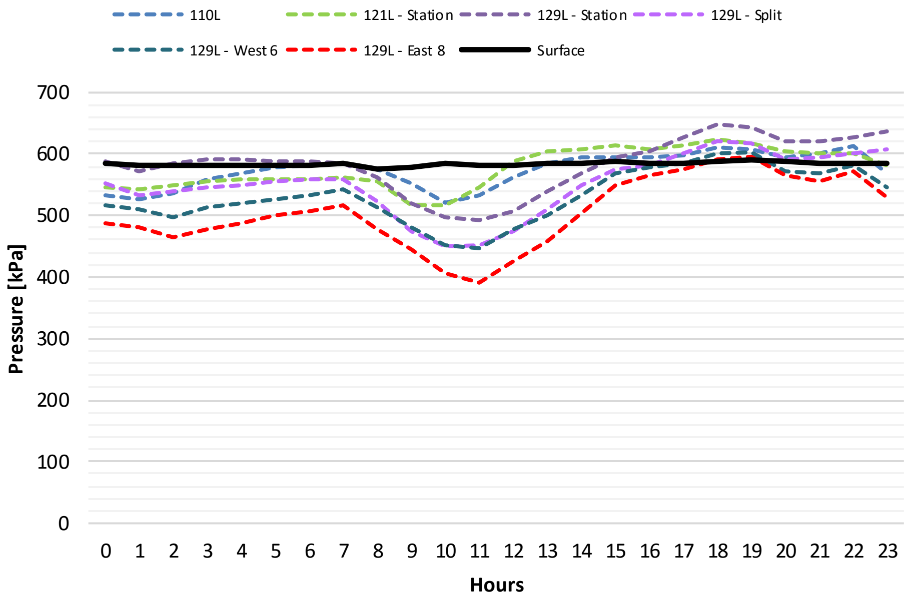


Figure 4-4: Combined average pressure profiles

The large pressure drops during peak drilling periods (09:00 to 14:00) are clearly visible in Figure 4-4 and were much more prominent on the active mining levels due to the increase in compressed air demand. One will also notice that after 16:00 pm certain measuring points recorded a higher pressure than that of supply pressure. This can be attributed to the effect of auto compression combined with low compressed air demand. Table 4-1 prioritises the obtained scoring value for the baselines constructed.

Table 4-1: Case study A - Baseline scope prioritisation

Criteria	Surface to 110L	110L to 121L Station	121L to 129L Station	129L Station to Split	129L Split to West	129L Split to East
Pressure difference	1	1	0	4	7	10
Pipe length	0	10	10	10	9	9
Nominal pipe size	0	8	8	8	8	10
Ease of alteration	1	1	1	10	9	9
Repair time frame	1	2	3	4	9	10
Implementation/ Repair Cost	1	1	1	8	10	10
Total Score	4	22	22	44	51	**58**

From Table 4-1, the baseline constructed for the far end of the east haulage on 129L proved to have the highest potential for improvement. It is also interesting to note that larger potential exists closer to the mining operations due to the convention of pipe diameter reductions.

- ***Strategy Selection***

The East haulage was found to have compressed air pipes installed with various dimensions. These pipe size variations contributed to the large pressure losses over this section. Figure 4-5 illustrates the pressure profiles which were measured from start to finish over this pipe section.

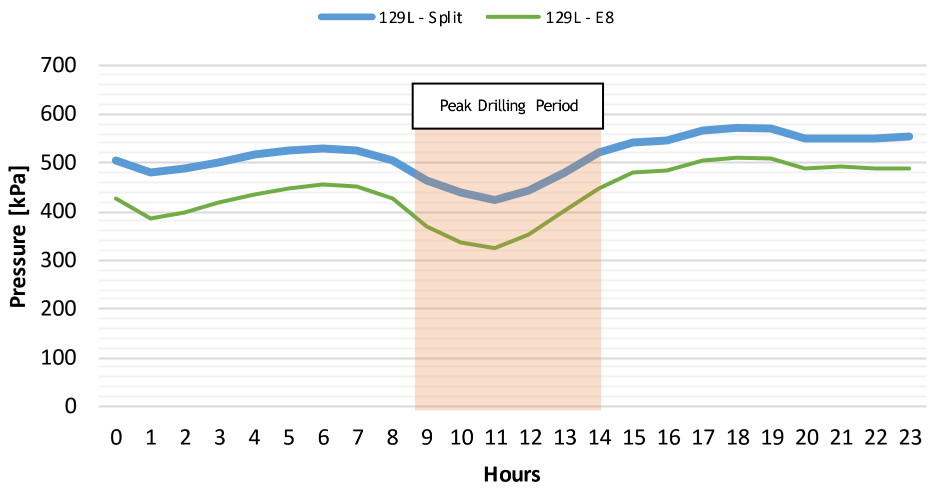


Figure 4-5: 129L - East haulage average pressure profile

The average pressure at the far end of the East haulage, East 8, was recorded to be 438 kPa. Table 4-2 summarises the pressure drops measured from the start of the East haulage split to the end of the east haulage. These values also served as the baselines prior to any improvements were implemented.

Table 4-2: Baseline pressure drop measurements

Baseline	Pressure Drop
Daily average	76 kPa
Average during peak drilling	87 kPa
Daily max	99 kPa
Daily min	60 kPa

4.4 Simulation, Verification and Strategy Implementation

The replacement of the 6" pipe sections with 8" pipe sections was further investigated. Figure 4-6 illustrates a simplified layout of the east haulage compressed air pipe section that was investigated.

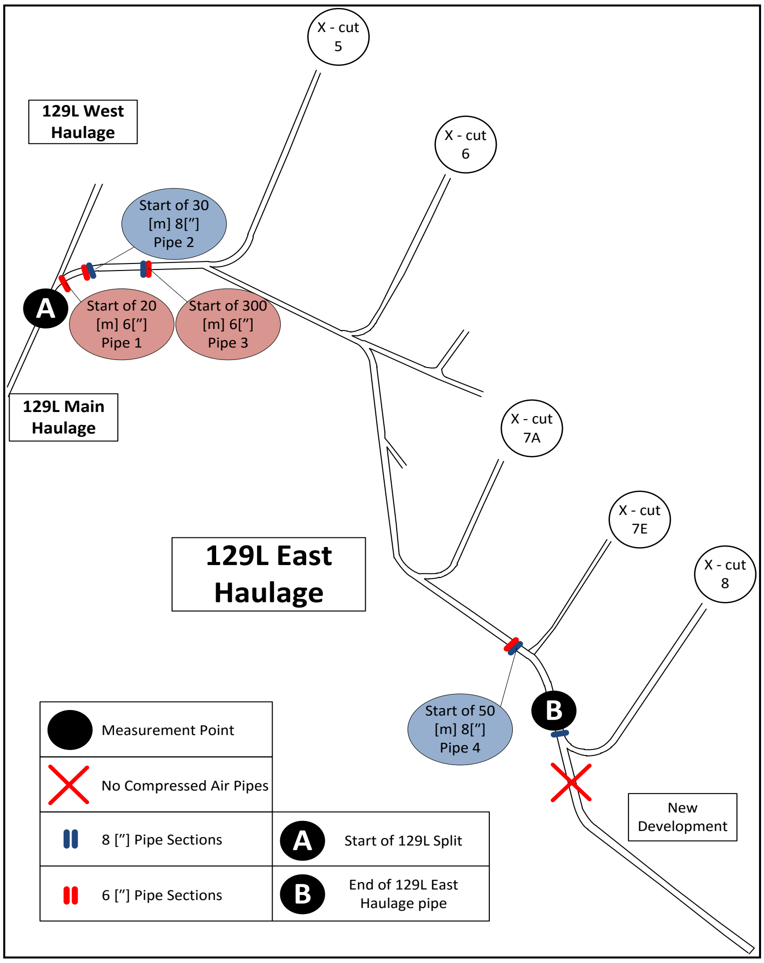


Figure 4-6: 129L East haulage layout

The proposed solution strategy, of increasing the 6" pipe sections to 8" pipe sections, was simulated using Process Toolbox (PTB). The purpose of the simulation was to investigate the potential service delivery improvement that can be expected with the pipe replacement before recommending the solution to Mine A's personnel. This was done through simulating the various connected pipe sizes and calibrating for the required to match the measured pressure drop. The flow was kept constant when simulating the new pressure drop with the increase pipe sizes. The developed PTB simulation model also had to be verified to ensure the model was accurate. Verification entailed using existing pressure and flow data to calibrate the model accordingly and ensure the simulated values correlate to actual measured values. Data from another site, which consisted of accurate flow and pressure measurements, were taken and used to verify the model's accuracy in predicting the final, measured pressure. The correlation yielded an accuracy of 92% [31].

The simulation indicated a potential pressure improvement of 57 kPa which yielded the largest potential impact, and the solution strategy was implemented accordingly. Figure 4-7 illustrates the pressure profile measured after implementation. The same measuring points are used prior to implementation.

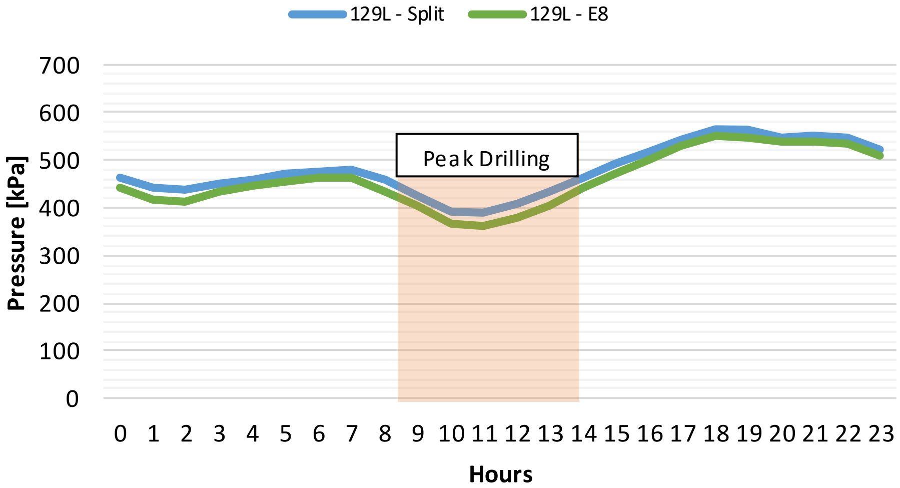


Figure 4-7: 129L - East haulage average pressure profile after implementation

Figure 4-7 shows a significant reduction in pressure losses compared to the original baseline. Table 4-3 summarises the pressure drops comparison after implementation.

Table 4-3: Measurements after implementation of solution strategy

	Pre-implementation	Post-implementation	Improvement
Average	76 kPa	19 kPa	57 kPa
Average during peak drilling	87 kPa	25 kPa	62 kPa
Max	99 kPa	28 kPa	71 kPa
Min	60 kPa	13 kPa	47 kPa

Table 4-3 indicates a large pressure drop improvement with the implementation of the proposed solution.

- ***Result Validation***

The implementation of the optimisation strategy stretched over an 8-month period. During this time the conditions and requirements of the 129L east haulage changed since the baseline period. The changes are mostly related to the dynamic nature of mining activity on the production levels. Figure 4-8 illustrates all the newly developed areas (indicated with yellow lines) as well as the additional stope faces (indicated by coloured blocks) which were mined during the implementation period of the solution strategy.

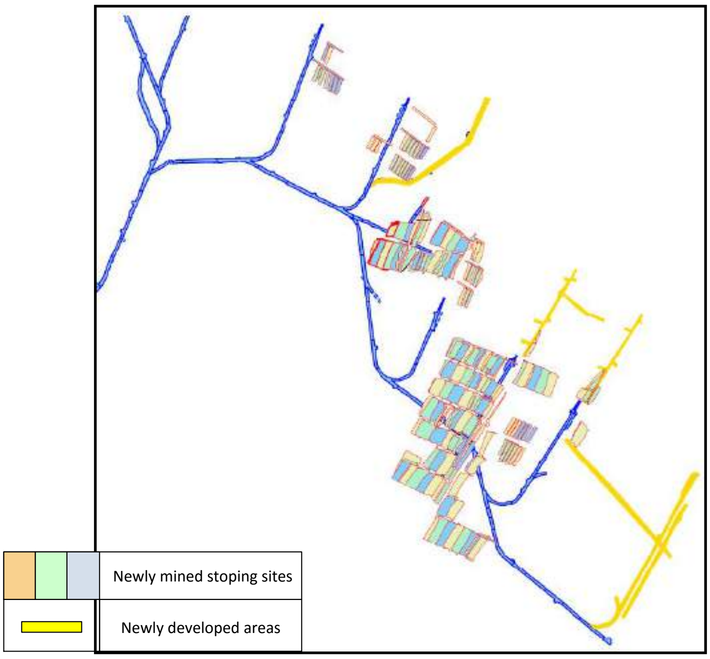


Figure 4-8: Newly developed and mined areas after implementation

These system changes led to an increase in compressed air demand, which directly means higher volumes of compressed air and consequently lower pressures. The compressed air reticulation network also expanded over the implementation period, with new pipe sections installed to reach newly developed mining areas. It was therefore required to compare apples with apples to quantify the true impact of the solution.

The measured data was evaluated for instances where the inlet pressure for the pipe section (point A on Figure 4-6) was identical before and after the new pipe sections were installed. The pressure most frequently achieved in both instances was determined to be 570 kPa. The predicted pressure drop before implementation was therefore also simulated for the given inlet pressure of 570 kPa, with the same demand requirements. Table 4-4 shows the predicted pressure drop simulated in comparison with the measured pressure drop.

Table 4-4: Validated results

	Predicted Pressure Drop	Measured Pressure Drop
Pressure Drop	14.18	13.91 kPa
Simulation error	2%	
Average Pressure drop improvement	**45 kPa**	

From Table 4-4 it is evident that, when simulating the predicted pressure drop for the same inlet pressure conditions, the simulation yields a low error percentage. The average measured pressure drop improvement of 45 kPa is lower than the simulated pressure drop improvement of 57 kPa.

This is due to most of the inlet pressure values of 570 kPa occurs outside peak drilling periods. During these periods lower flow demand is present and results in lower pressure drops, as can be seen in Figure 4-7. The newly simulated pressure improvement therefore reflects the pressure improvement which can be expected outside of peak drilling periods and should be compared with the minimum measured pressure drops. Doing this results in a 4% error.

4.5 Predicted production impact analysis

A theoretical production impact analysis was performed, as discussed in Chapter 3, to determine what the impact would be on production due to the improved compressed air pressures in the network.

The rock penetration rate increase was calculated for a drill bit size of 34 mm at a design thrust pressure of 700 N. The pre- and post-implementation delivery pressure to East 8 crosscut were used as inputs to calculate the improved rock penetration rate. These pressures were calculated as an average across the 24-hour pressure profile.
Table 4-5 shows the improved rock penetration rate resulting from the service delivery improvement.

Table 4-5: Rock penetration rate improvement

	Baseline	Post-implementation
The pressure at East 8	438 kPa	483 kPa (@45 kPa improvement)
Rock penetration rate	1.6520 mm/s	2.0623 mm/s
Rock penetration rate improvement	0.4103 mm/s	

The service delivery pressure improvement on 129L – East haulage directly affects three active mining areas. Therefore, the theoretical annual financial improvement of this study amounts to an annual saving of approximately R11-million. Production data for 129L East level could only be collected for the period of 9 September 2016 to 28 April 2017, due to the poor record keeping of Mine A.

Production data will be required to validate the calculated results but will only serve as a guideline and should not be evaluated in isolation. The correlation between production impact and compressed air network alterations will be less direct as one increases the distance from the exact area of production. This is because numerous other factors start to play a role, factors such as labour issues, decrease in ore deposits, starting/stopping of other work place areas, number of leaks and increasing travel distances.

5. CONCLUSION

This study highlighted the fact that optimisation of current infrastructure is imperative for the struggling South African gold mining sector. It was further revealed that large potential exists for optimisation on the compressed air networks of these mines due to the large inefficiencies.

A case study was performed where an inefficiency was identified causing large pressure losses to active mining areas. A solution strategy was proposed and implemented yielding large pressure improvements. These pressure improvements were theoretically translated into a production correlation and indicated a potential profit increase of approximately R11–million per annum.

6. REFERENCES

[1] **Chamber of Mines** 2017. Mine SA 2017 Facts & Figures Pocketbook, *Chamb. Mines*.

[2] **Chamber of Mines of South Africa** 2017. Facts and Figures 2016, no. June, pp. 16–18.

[3] World gold production 2005-2017 | Statista, [Online]. Available: https://www.statista.com/statistics/238414/global-gold-production-since-2005/. [Accessed: 13-May-2018].

[4] **Neingo, P. N.** and **Tholana, T.** 2016. Trends in productivity in the South African gold mining industry, *J. South. African Inst. Min. Metall.*, vol. 116, no. 2, pp. 283–290.

[5] The energy efficiency series.

[6] The world's deepest gold mine is in South Africa - Business Insider, [Online]. Available: http://www.businessinsider.com/the-worlds-deepest-gold-mine-is-in-south-africa-2016-12?IR=T. [Accessed: 13-May-2018].

[7] Eskom tariff increase to cost SA 5 000 jobs - Moneyweb, [Online]. Available: https://www.moneyweb.co.za/news/south-africa/eskom-tariff-increase-to-cost-sa-5-000-jobs/. [Accessed: 13-May-2018].

[8] 2015. Deep level mining in South Africa - Where we are, where we're going, p. 7.

[9] Fact sheets - This is GOLD, [Online]. Available: http://www.thisisgold.co.za/facts-and-figures/fact-sheets. [Accessed: 13-May-2018].

[10] **Strydom, E.** and **Russel, C.** 2012. Wages And Benefits In The Gold Industry, *Chamb. Mines*, pp. 1–2.

[11] Modern mining: Re-thinking the way we mine – a challenge in contrasts - This is GOLD, [Online]. Available: http://www.thisisgold.co.za/news/conversations-about-gold/2014/126-modern-mining-re-thinking-the-way-we-mine-a-challenge-in-contrasts. [Accessed: 13-May-2018].

[12] Modernisation - Chamber of Mines South Africa, [Online]. Available: http://www.chamberofmines.org.za/work/modernisation. [Accessed: 13-May-2018].

[13] **Chamber of Mines of South Africa** 2015. Modernisation: Towards the mine of tomorrow, pp. 1–8.

[14] **Bredenkamp, J. I. G.** and **Rensburg, D. J. van** 2014. Reconfiguring mining compressed air networks for cost savings, Potchefstroom Campus of the North-West University.

[15] **Neser, H.** 2008. Energy savings through the automatic control of underground compressed air demand, Potchefstroom.

[16] **Vergne, J. N. de la** 2003. Hard Rock Miner's Handbook, 1438 West Broadway Road, Suite 101 • Tempe • Arizona • USA: McIntosh Engineering.

[17] **Marais, J. H.** 2012. An integrated approach to optimise energy consumption of mine compressed air systems.

[18] **Fraser, P.** 2010. The energy and water required to drill a hole, no. 1, pp. 211–222.

[19] **Dudić, S.**, **Ignjatović, I.**, **Šešlija, D.**, **Blagojević, V.**, and **Stojiljković, M.** 2012. Leakage quantification of compressed air using ultrasound and infrared thermography, *Meas. J. Int. Meas. Confed.*, vol. 45, no. 7, pp. 1689–1694.

[20] Glossary of Mining Terms, [Online]. Available: http://www.rocksandminerals.com/glossary.htm. [Accessed: 13-May-2018].

[21] **Kivade, S. B.**, **Murthy, C. S. N.**, and **Vardhan, H.** 2015. Experimental Investigations on Penetration Rate of Percussive Drill, *Procedia Earth Planet. Sci.*, vol. 11, pp. 89–99.

[22] **Lahiouel, Y.** and **Lahiouel, R.** 2015. Evaluation of Energy Losses in Pipes, *Am. J. Mech. Eng.*, vol. 3, no. 3A, pp. 32–37.

[23] **Rieutord, M.** 2015. Fluid Dynamics, .

[24] **Rodgers, T.** 2013. Fluid Flow 2013.

[25] **Petit, P. J.** 2006. Electric rock drilling system for in-stope mining in platinum operations, pp. 209–216.

[26] **Kahraman, S.** 2003. Performance analysis of drilling machines using rock modulus ratio, *J. South African Inst. Min. …*, no. October, pp. 515–522.

[27] **Kelessidis, V. C.** 2011. Rock drillability prediction from in situ determined unconfined compressive strength of rock, *J. South. African Inst. Min. Metall.*, vol. 111, no. 6, pp. 429–436.

[28] **Kriel, C. J. R.** 2014. Modernising underground compressed air DSM projects to reduce operating costs, North-West University.

[29] **Eisler, R.** 2003. Health risk of gold miners: A synoptic review, *Env. Geochem Heal.*, vol. 25, no. December 2002, pp. 325–345.

[30] **Lodewyckx, J. W.** 2007. Investigating the effects of different DSM strategies on a compressed air ring, p. 136.

[31] **Nell, D.** 2017. Optimising production through improving the efficiency of mine compressed air networks with limited infrastructure Dissertation submitted in fulfilment of the requirements for the degree Master of Engineering in Mechanical Engineering at the Potchefstroom, no. November.

INVESTIGATING THE SUSTAINABILITY AND FEASIBILITY OF DIFFERENT DISPOSABLE CUPS: A COFFEE SHOP PERSPECTIVE

C. Vermeulen[1*] & W.G. Bam[2]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
carlivermeulen.home@gmail.com

[2]Department of Industrial Engineering
Stellenbosch University, South Africa
wouterb@sun.ac.za

ABSTRACT

With convenience being a sought-after factor in today's society, disposable coffee cups play a significant role in many people's daily routines. There still, however, appears to be a general confusion in many coffee shop environments regarding optimal waste treatment of disposable coffee cups to minimise their environmental impact. This study thus compares the different disposable coffee cup options that are available to South African coffee shops to determine which options (in terms of different materials) are more environmentally friendly under different conditions, whilst also satisfactorily meeting the requirements of coffee shops and customers.

[1] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

The purchase of disposable coffee cups is embedded within the contemporary human lifestyle due to the convenience it offers. However, consumers rarely consider the environmental costs that are linked with disposable cups. Contrary to what many might think, the majority of disposable cups (including the majority of the “paper cups” used for hot take-away beverages) are not widely recyclable due to the polyethylene (PE) coating found inside of many of these cups. To cater for this problem, various coffee shops have started implementing biodegradable coffee cups with the aim of decreasing their environmental impact. The main differentiation of biodegradable cups is that they are lined with a bioplastic called Polylactic Acid (PLA) instead of PE [1]. Even though biodegradable cups have grown in popularity in recent years, there are still some concerns regarding their implementation. Firstly, biodegradable products are meant to be composted and not recycled [2]. It is therefore necessary to separate them from conventional plastics such as milk and soft drink bottles prior to the recycling process to prevent contamination [2]. Secondly, various manufacturers are claiming that their bioplastic products are “biodegradable”, “compostable” and “environmentally friendly” when this is only the case when certain requirements are fulfilled [3]. In general, the three main requirements for biodegradable cups to decompose as intended is a temperature higher than 60 degrees Celsius, a humid environment and they must be mixed with organic material [4].

Given the complexity regarding the environmental impact of biodegradable coffee cups, guidance for coffee shops as to which cup is optimal when considering economic, social and environmental factors, remains ambiguous. The current study forms part of a growing literature, that seeks to address this gap, with a specific focus on evaluating the optimal waste management system configuration within the South African context.

1.1 PROBLEM STATEMENT

The environmental concern regarding coffee cup packaging, as well as its associated waste, has grown significantly in recent years. Biodegradable packaging has been touted as a solution to this problem due to its ability to biodegrade under certain conditions. However, it remains unclear whether biodegradable cups are better for the environment on balance, given the end-of-life (EOL) treatment of these cups in practice. There is thus a need for research that investigates the environmental impact of different disposable coffee cup alternatives. This research should consider the various disposal systems that these coffee cups end up in and make recommendations towards how these systems could be improved to reduce the possible negative effects of coffee cup disposal.

1.2 RESEARCH AIM

The aim of this project was to evaluate the different disposable coffee cup options available to South African coffee shops to derive recommendations regarding how disposable coffee cup disposal practices can be improved. These recommendations aim to reduce the environmental impact of disposable coffee cups, whilst still ensuring that the utility and affordability of disposable coffee cups use are not compromised.

1.3 PROJECT OBJECTIVES

In support of the above-mentioned aim the following research objectives were defined:

1. Define the key terms for the study and identify the different types of disposable cups that are currently on the market.
2. Identify the coffee shop owner and coffee shop customer requirements for disposable coffee cups, as well as their environmental friendliness.
3. Consider the combination of coffee shop requirements, customer requirements and environmental analyses to identify the most environmentally friendly yet feasible disposable coffee cup options and evaluate them by means of two case studies.
4. Make recommendations for the case studies evaluated and reflect on the possible implications for South African coffee shops while highlighting the possible shortcomings of the research and implications for future research.

1.4 SCOPE OF THE STUDY

This study focuses on disposable cups that are suitable for hot beverages. To simplify the study, only coffee cups and not the lids or sleeves are considered. Only existing technologies are considered and their feasibility is determined for use in South African coffee shops.

2. STRUCTURE OF THE METHODOLOGY

This section of the article covers the methodology that was followed towards achieving the objectives stated in Section 1.3. Sections 2.1, 2.2 and 2.3 discuss the first, second and third phase of the methodology, respectively. Furthermore, it is mentioned which objective is addressed.

The methodology is divided into three phases:

i. literature study phase (Section 2.1);
ii. comparative phase (Section 2.2); and
iii. case study and recommendations phase (Section 2.3).

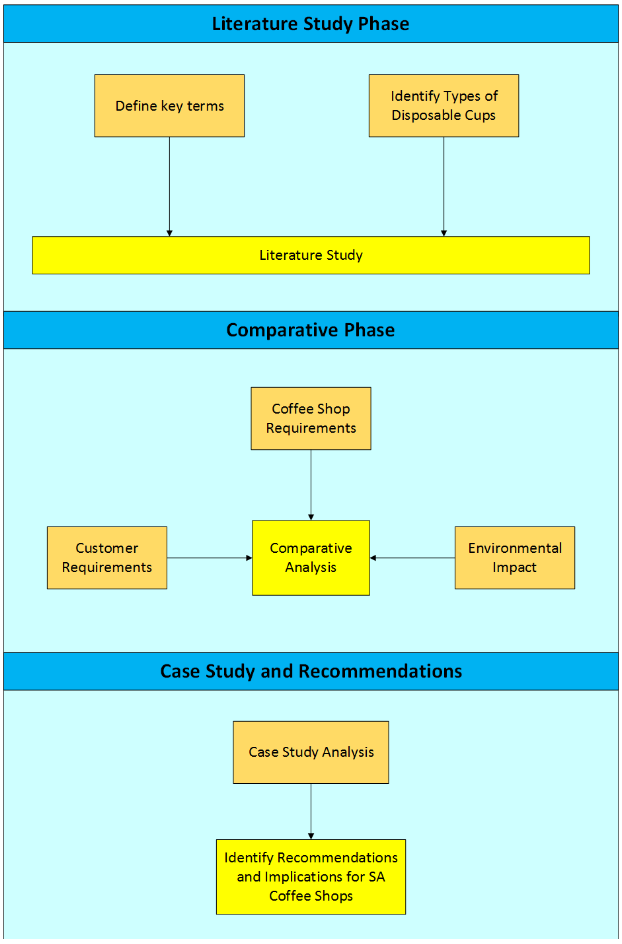


Figure 1: Structure of the Methodology.

2.1 LITERATURE STUDY PHASE

The results from the execution of the literature study phase are presented in Section 3 and addresses first objective from Section 1.3.

In Section 3.1 key terms of the study are discussed. Furthermore, the most common disposable cups are listed in Section 3.2. Thereafter, it is mentioned which disposable cups are suitable for hot beverages and therefore suitable for this study. Simplified production processes for the materials of these cups are also included.

2.2 COMPARATIVE PHASE

The results from the comparative phase are presented in Sections 4, 5 and 6 and addresses the second and third objective from Section 1.3.

This phase takes three aspects into account when comparing the disposable coffee cups that were identified in the first phase. The first aspect that the coffee cups are ranked on is how well they meet the coffee shop owner's requirements. Two coffee shop owners were interviewed to set up a list of coffee shop requirements that are listed in Section 4.1. The first coffee shop is Damascus Road that is located at the launchlab of Stellenbosch University. The second coffee shop is Deluxe coffees that is also located in Stellenbosch. The second aspect is how well the cup meets the coffee shop customer requirements. These customer requirements were identified by means of interviews with regular coffee drinkers and are discussed in Section 4.2. The final aspect against which the cups were compared is their CO_2 impact under different circumstances. This is discussed in depth in Section 5. Once all the cups were ranked according to the three aspects, their final rankings were compared. This comparison is presented in Section 6 and indicates the cup with the highest overall ranking.

2.3 CASE STUDY AND RECOMMENDATIONS

The results of this third phase of the methodology are presented in Sections 7, 8 and 9 and addresses the fourth and final objection from Section 1.3.

In Section 7 two case studies are performed to identify whether stakeholders are aware of the implications of implementing compostable coffee cups. The first case study involves a student cafeteria and is discussed in Section 7.2. Thereafter, a further case study involving a Western Cape municipality is presented in Section 7.3. In Section 8 the results presented in Section 1 to 7 are evaluated. Thereafter, further recommendations and practical considerations are presented in Sections 8.2 and 8.3, respectively. In Section 9 a summary of the report is provided and the most significant limitations, that were dealt with during the course of this study, are discussed.

3. LITERATURE STUDY

This section defines terms, relevant to the study. Furthermore, the different disposable cup options, as well as which ones are focused on in this research, are discussed in Section 3.2.

3.1 IMPORTANT DEFINITIONS

To gain a better understanding of this article, it is useful to define relevant terms such as "sustainability", as well as distinguishing between terms such as "recyclability", "degradability", "biodegradability", "compostability" and "digestion". The definitions are discussed in more detail in the following sections.

3.1.1 DEFINING SUSTAINABILITY AND THE TRIPLE BOTTOM LINE

According to [5] the overall goal of sustainable development is the "long-term stability of the economy and environment; this is only achievable through the integration and acknowledgment of economic, environmental, and social concerns throughout the decision making process". Moreover, in many businesses "sustainability" is considered as an important goal, however, quantifying the degree to which a company is sustainable is often complex. John Elkington therefore set up a framework in the 1990s to measure the sustainability of companies [6]. His framework, called the "Triple Bottom Line", combines three dimensions of performance: environmental, financial and social [6]. The "Triple Bottom Line" differed from traditional performance measures as it incorporated social and environmental means of measurement.

3.1.2 BIODEGRADABILITY, COMPOSTABILITY AND DIGESTION

Further terms that are beneficial to define include "biodegradability", "compostability" and "digestion". The meanings of the terms "biodegradability" and "compostability" are often confused. There are two significant differences between the two terms. The first differentiation is that compostable materials break down into humus and supply the soil with nutrients, whereas biodegradable materials are generally broken down into more basic molecules that integrate with the environment as basic organic building blocks [7]. The second difference between biodegradability and compostability is that biodegradable materials are generally able to disintegrate under the conditions prevalent in landfills. Compostable materials, on the other hand, generally require specific composting conditions in order to decompose to the desired level and through the desired pathways [8]. These conditions generally include an elevated temperature, high humidity and a mixture with other organic materials [4]. Compostable cups, for example, require a temperature of above 60 degrees Celsius before they decompose along the desired pathways and within a generally acceptable timeframe [4]. A further term that is beneficial to define is "digestion. Similarly to composting, digestion is a biological waste treatment. The major difference between the two is that during digestion organisms can feed off the organic waste itself without additional oxygen, whereas during composting organic matter is naturally oxidised by organisms [9].

3.2 DISPOSABLE CUP TYPES

Towards addressing the first objective in Section 1.3, it is necessary to review the types of disposable coffee cups that are currently on the market.

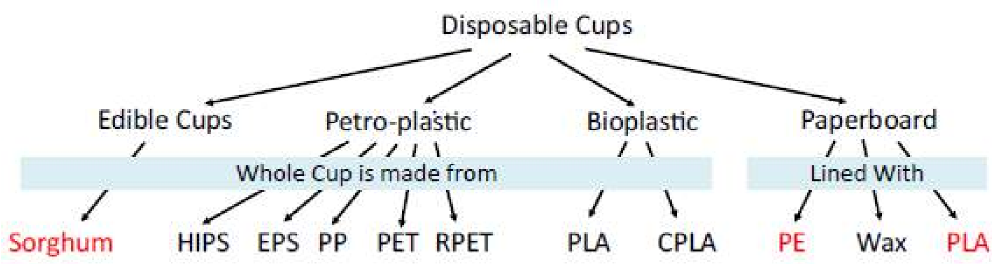


Figure 2: Types of Disposable Cups adapted from [4].

As can be seen in Figure 2, disposable cups can be categorised into four different material types: edible cups, petro-plastic, bioplastic and paperboard [4]. Figure 2 also shows the materials that the disposable cups consist of or are lined with. For example, Petro-plastic cups are either made from HIPS (high impact polystyrene), EPS (expanded polystyrene), PP (poly propylene), PET (poly ethylene terephthalate) or RPET (recycled PET). As mentioned in Section 1.4, this article focuses on cups that are suitable for hot beverages. The cups that are suitable for hot beverages are therefore edible cups and paperboard cups lined with PLA or LDPE as indicated in red in Figure 2. Even though EPS Petro-plastic cups can also be used for hot beverages, they are considered to pose a considerable health risk and are therefore excluded from further consideration in this article [10]. In Sections 3.2.1, 3.2.2 and 3.2.3 a brief overview is given of PE-coated, PLA-coated and edible cups, respectively.

3.2.1 PAPERBOARD CUPS LINED WITH PE

The manufacturing of heat-resistant paper cups involves a thin line of polyethylene (PE) on the inside to ensure that they are waterproof [11]. Moreover, a simplified diagram of the production of PE can be seen in Figure 3. The process starts with the coupling of a natural gas, namely methane or ethane, with an oxygen source. This process is referred to as Oxidative Coupling of Methane (OCM). The outputs of this process are ethylene and other associated products. These are then put under a pressure of between 10 and 80 bar and a temperature of between 250 and 300 degrees Celsius to create polyethylene. Polyethylene is then used to coat the paperboard cups, which increases the recycling process complexity of these cups and generally renders them infeasible to recycle [12].

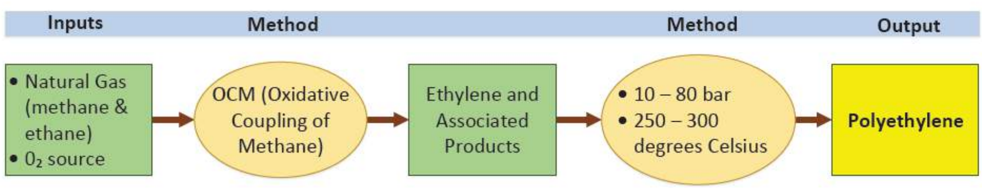


Figure 3: Manufacturing of Polyethylene collated from [13].

3.2.2 PAPERBOARD CUPS LINED WITH PLA

One of the most common compostable coffee cups is made from paperboard that is lined with PLA. A simplified manufacturing process of PLA can be seen in Figure 4. The key factor that differentiates PLA from thermoplastics, such as PE, is that it is manufactured from renewable resources such as corn starch or sugar cane [14]. These renewable resources are then fermented to produce lactic acid. The lactic acid is polymerised to achieve lactide and then condensed to finally achieve PLA. PLA-coated disposable cups can be composted, however, similarly to the PE-coated cups are infeasible to recycle. The reason for this is that the PLA must be separated from the paperboard before the two materials can be recycled separately. This process is generally too expensive to make it economically feasible to recycle the coffee cup.

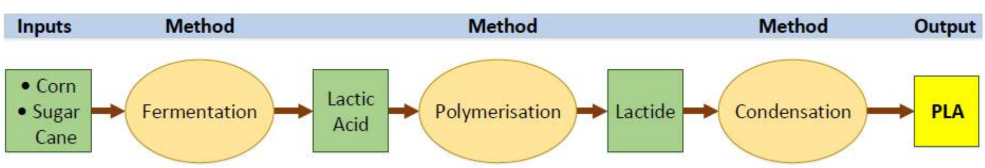


Figure 4: Production of PLA adapted from [15].

3.2.3 THE EDIBLE CUP

The third disposable cup that was investigated is the edible coffee cup. An application of this idea can be seen at an Indian edible cutlery manufacturer called "Bakeys". All their products consist of Sorghum as the primary ingredient [16]. Kentucky Fried Chicken (KFC) implemented their own edible disposable coffee cup in 2015 in the United Kingdom. However, this cup was not made from Sorghum, but rather a biscuit that is wrapped in sugar paper and lined with white chocolate [17]. This cup was discontinued in the same year that it was introduced. This was due to the fact that there is a trade-off between the cup's sturdiness and its taste. In order to make the cups durable and able to withstand hot liquids, they would be extremely hard and therefore not have the best taste. Another downside of these cups was that coffee shop customers did not always want to eat cookies while buying coffee. If the cup therefore remained uneaten it would most likely be thrown away and considered food waste. Even though the implementation of KFC's edible cups was unsuccessful, edible cups made from Sorghum are still compared with PE and PLA cups in this study as they could potentially still be successful if implemented using a different business model or design.

4. IDENTIFYING COFFEE SHOP AND CUSTOMER REQUIREMENTS

This section focuses on determining the various requirements, that stakeholders have, regarding the properties of disposable coffee cups. Two perspectives were considered: the coffee shop owner perspective is covered in Section 4.1 and the coffee shop customer perspective is covered in Section 4.2. Finally, the two perspectives were combined in Section 4.3. The interview questions can be found in Appendix A.

4.1 REQUIREMENTS AND PRIORITIES OF WESTERN CAPE COFFEE SHOPS FOR COFFEE CUPS

To identify the coffee cup requirements from a coffee shop perspective, two coffee shop owners who operate coffee shops in the Western Cape in South Africa were interviewed. Both shops have a large focus on coffee and currently sell coffee in take-away cups. Their first requirement was that the cup must be of low cost. Economically it would therefore only make sense for the company to purchase a cup that is price competitive. Their second requirement involved the strength of the cup. It was deemed to be of importance that the cup is sturdy as it deals with hot liquids that could lead to burn wounds. Thirdly, the outer appearance of the cup was also of importance. Fourth, the coffee shops mentioned that they would prefer using environmentally friendly cups over a non-environmentally friendly cups. Lastly, the coffee shop owners both agreed that the cup must have good insulation properties. If the cup has poor insulation, additional sleeves must be purchased at an additional cost, which might or might not be beneficial for the company depending on whether they use sleeves for marketing purposes.

4.2 REQUIREMENTS AND PRIORITIES OF CUSTOMERS FOR COFFEE CUPS

Three regular coffee drinkers identified at a South African university cafeteria were interviewed and asked what their disposable coffee cup requirements are. These coffee drinkers were selected as they on average buy at least one cup of take-away coffee per day from a local coffee shop. Of all their requirements, there were four that were common amongst all of the coffee drinkers. The first requirement is that the coffee cup must have good insulation properties. A further requirement, that all three coffee drinkers expressed, was that the coffee cup must be able to be sealed at the top. Further customer requirements included good grip and environmentally friendliness.

4.3 COMBINING COFFEE SHOP AND CUSTOMER REQUIREMENTS

It is evident that many of the coffee cup requirements are common amongst coffee shops and customers. To get an overall perspective of what is expected from disposable coffee, the coffee shop and customer requirements were combined and the final list can be seen in Figure 5.

Low cost
High strength
Good aesthetics
Good insulation properties
Must be able to be sealed
Good grip
Environmental friendliness

Figure 5: Combined Disposable Coffee Cup Requirements

5. EVALUATING THE CUP OPTIONS IN TERMS OF THEIR EINVIRONMENTAL IMPACT

This section aims to compare the environmental impacts of the three viable hot cup types that were identified in Section 3.2. To evaluate the environmental impact of cups under different usage scenarios, different EOL scenarios are investigated. The various EOL scenarios for PE-coated, PLA-coated and edible cups are shown in Figure 6.

Section 5.1 discusses the CO_2 equivalents of landfilling PE-coated and PLA-coated cups. Seeing as different assumptions about decomposition affect the CO_2 equivalents, three situations were investigated. The first assumes maximum decomposition, the second assumes 50 percent decomposition and the third assumes zero decomposition in the landfill. Section 5.2 focuses on the digestion and industrial composting of PLA-coated cups. Seeing as PE-coated cups cannot be digested or composted, only PLA-coated cups are investigated in this section. Moreover, Section 5.3 compares the different EOL options of PLA cups to determine which scenario has the lowest CO_2 equivalent footprint. In Section 5.4 the EOL scenario of edible cups is discussed. Thereafter, in Section 5.5 the degradation times of landfilling, composting and digestion are compared.

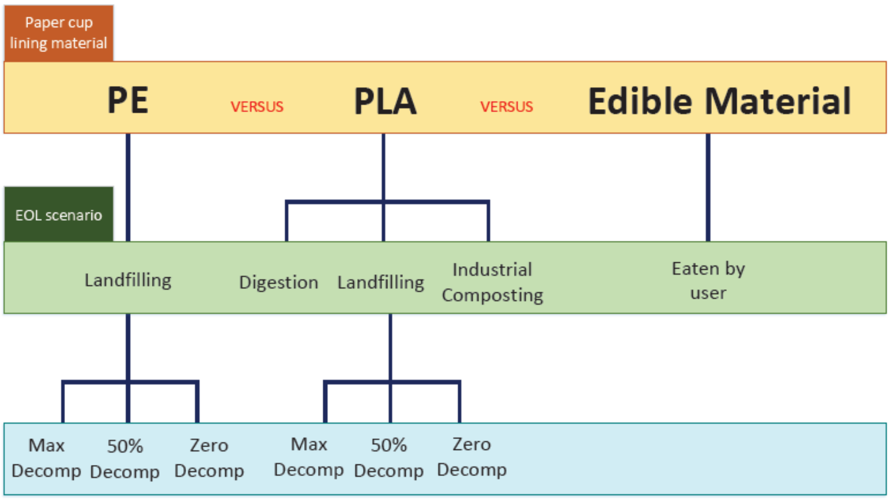


Figure 6: Paper Cup Lining Materials with their EOL Scenarios.

5.1 LANDFILLING OF DISPOSABLE COFFEE CUPS

The data used in this section regarding CO_2 impacts of landfilling cups was obtained from an existing LCA study [18]. In their LCA, 16oz "hot cups" were investigated with a functional unit of 10 000 cups. Considering that the LCA study that was worked from is based in America, various assumptions were based on American statistics. For example, in the United States, 80 percent of the municipal solid waste, that is not recycled or composted, is landfilled and 20 percent is incinerated. This assumption was therefore held throughout this analysis to ensure comparability. During this LCA three situations were considered. The first scenario assumes a maximum decomposition at the landfill, whereas the second scenario assumes 50 percent decomposition at the land fill and the third assumes zero percent decomposition. Furthermore, even though an additional corrugated sleeve could be added to improve the user insulation experience, this optional feature was disregarded as this study specifically focuses on the disposable coffee cup itself.

Figure 7 shows the comparison between PE and PLA cups for the three landfilling EOL scenarios. It can be seen that the less the cup decomposes the less the CO_2 equivalent is. This can be explained by considering that no atmospheric gases are produced and all the biomass carbon is sequestered in the paperboard, which results in a large CO_2 sequestration credit [18]. This is also the reason why there is a sequestration credit for the EOL carbon equivalents at zero percent decomposition. For all three decomposition levels the PLA cup has a lower CO_2 equivalent than the PE cup. The differences between the CO_2 equivalents of PLA and PE are, however, not large. At maximum decomposition, for example, the CO_2 equivalent of PE-cups is only 3 percent greater than that of PLA-cups. This therefore suggests that landfilling compostable (PLA) cups is not significantly better for the environment than landfilling PE cups in terms of its CO_2 equivalent.

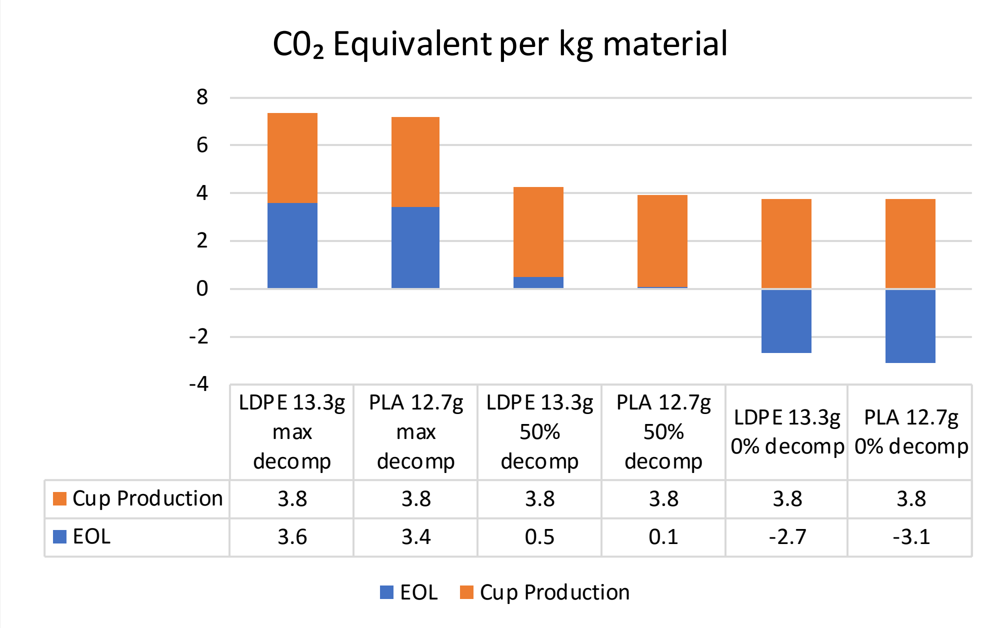


Figure 7: Landfilling of PE and PLA Disposable Coffee Cups sourced from [18].

5.2 COMPOSTING AND DIGESTION OF DISPOSABLE COFFEE CUPS

As mentioned in Section 3.1.2, there are four different biological waste treatments [19]. Seeing as no composting data for PLA-coated disposable coffee cups could be found, separate data for PLA and for paper (chemical pulp) was combined to achieve an approximation for the cup. Environmental data for chemical pulp was used as this comprises the "paper" part of the cup that is lined with PLA. Chemical pulping is a method of producing paper by chemically cooking wood chips under high pressure [20]. In Table 1 the four biological waste treatments are shown, as well as whether or not the treatment can be used to compost PLA and chemical pulp and therefore also compost coffee cups. The two waste treatments that can be used to compost both PLA and chemical pulp are highlighted in green in Table 1.

Table 1: Four Biological Waste treatments adapted from [19].

	Anaerobic Degradation (without oxygen)	Aerobic Degradation (with oxygen)
50-60°C	**Thermophilic digestion:** Suitable for PLA and chemical pulp	**Industrial composting:** Suitable for PLA and chemical pulp
≤35°C	**Mesophilic digestion:** Not suitable for PLA Suitable for chemical pulp	**Home composting:** Not suitable for PLA Suitable for chemical pulp

The data for the analysis was obtained from a LCA, in which different bio-based and biodegradable materials were compared based on their post-consumer waste treatment phase [19]. This LCA only investigated PLA and not the disposable coffee cups in particular. To estimate the CO_2 equivalent of the PLA-coated coffee cup, the individual CO_2 equivalents of chemical pulp and PLA were combined. Considering that the CO_2 equivalents of PLA and chemical pulp were found to be very similar, it was assumed that the average could be taken of the two to get an estimate of the overall cup. The individual CO_2 equivalents can be seen in Figure 8. The combined CO_2 equivalents for composting and digesting PLA-coated paper cups are shown in Figure 9. The data is expressed as CO_2 equivalents per kilogram of material.

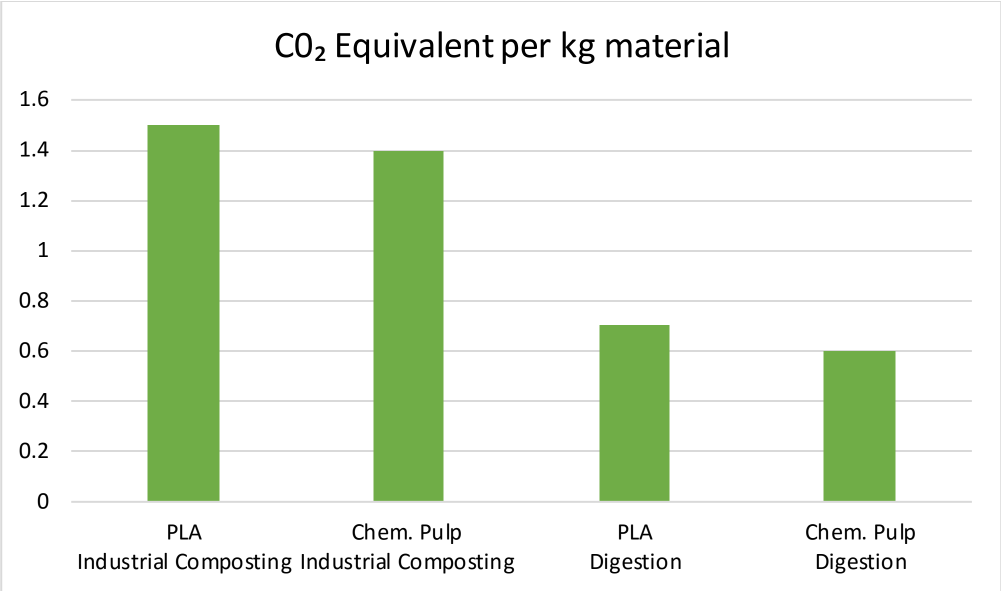


Figure 8: PLA and Chemical Pulp: Industrial Composting vs. Digestion adapted from [19].

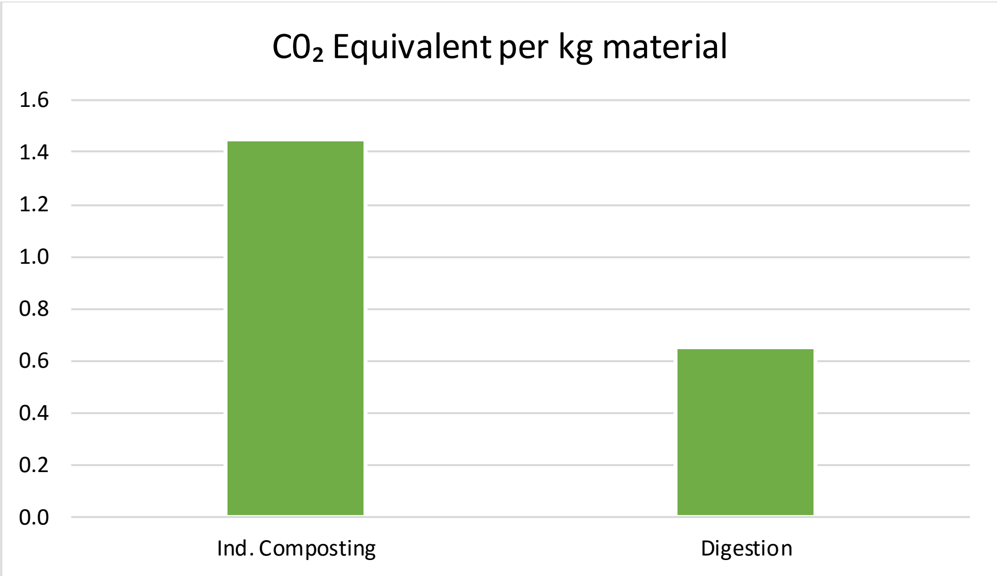


Figure 9: Combined Environmental Impact of Composting PLA Coated Cups adapted from [19].

As can be seen from Figure 9, digestion as an EOL waste treatment results in a significantly lower CO_2 equivalent value for PLA. This is due to the fact that digestion occurs below the surface, which traps the emissions and therefore prevents them from escaping. Composting, however, occurs above the surface, which results in more gases being released into the atmosphere.

5.3 COMPARISON OF ALL EOL SCENARIOS

In Section 5.1 it was established that compostable cups do not have a significantly lower carbon footprint in a landfilling EOL scenario when compared to the alternative non-compostable PE-coated cups. Furthermore, Section 5.2 showed that digesting PLA cups has a significantly lower carbon footprint than composting PLA cups. This section aims to compare the different EOL options of PLA cups in terms of their CO_2 equivalents to determine which one has the lowest CO_2 equivalent value. The LCA, that was used in Section 5.1, investigates the entire life cycle of the cup, whereas the LCA that was worked from in Section 5.2 only considers the EOL carbon footprints. It was therefore necessary to only use the EOL carbon impact from the LCA in Section 5.1 in order to make the two LCAs comparable.

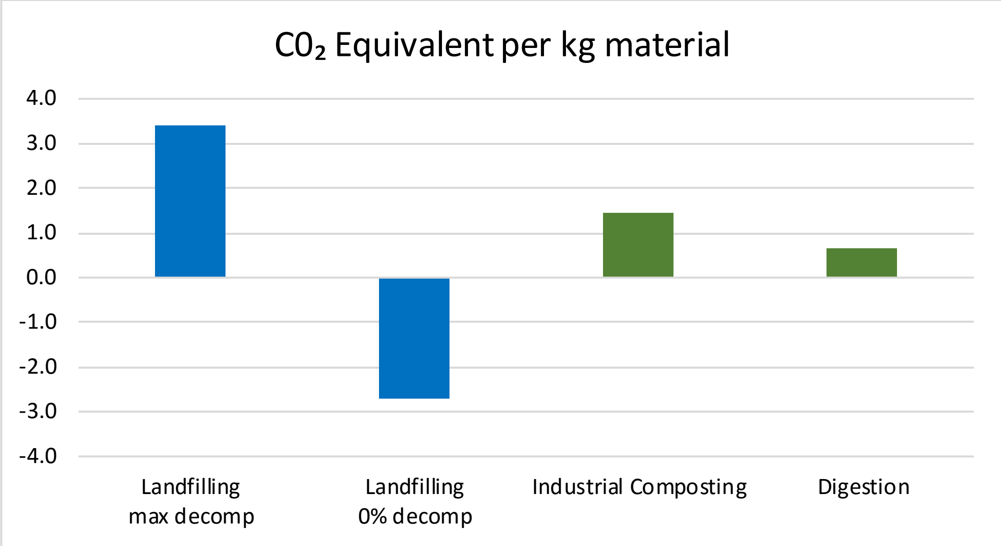


Figure 10: Waste Treatment Comparisons for PLA Cups adapted from [19]**.**

As can be seen in Figure 10 a maximum decomposition of PLA-coated cups in a landfill have by far the highest CO_2 equivalent. A zero percent decomposition in a landfill has a negative CO_2 equivalent, because gases are not being produced and are instead sequestered within the material. Considering that this is not a sustainable waste treatment, because of the cups taking up space in landfills, the negative CO_2 equivalent does not necessarily indicate environmental friendliness. When comparing the biodegradation options for PLA-coated cups, digestion has less than half the CO_2 equivalent of industrial composting. Digestion could thus be considered to be the most sustainable waste treatment option for PLA-coated coffee cups, as it does not lead to the impact of landfilling and produces less CO_2 than the other options.

5.4 EDIBLE CUP EOL SCENARIO

As can be seen in Figure 6, it is assumed that edible cups are eaten by the user. During the course of this article it is therefore assumed that the EOL CO_2 equivalent of edible cups is zero. The CO_2 emissions of manufacturing edible cups are thus disregarded in this study.

5.5 A FURTHER ENVIRONMENTAL CONSIDERATION

A further way to compare the environmental friendliness of PLA-coated and PE-coated disposable coffee cups is by how long it takes them to degrade in a landfilling, composting or digestion EOL scenario. Table 2 been set up to show the relevant degradation times.

Table 2: Degradation Times sourced from [21] **and** [22]**.**

Degradation Time		
EOL Scenario	PLA Cup	PE Cup
Landfilling Composting Digestion	50 years 3-6 months 6-8 months	50 years Not Applicable Not Applicable

Even though information regarding the time it takes for PLA cups to degrade when they are digested was not attained, it was assumed that this time is slightly longer than that of composting due to the fact that oxygen is not present to accelerate the process [19]. Even though the degradation time for digestion is slightly longer than that of composting, materials disintegrate significantly faster when they are composted or digested, as opposed to being landfilled as can be seen from Table 2. This therefore indicates that composting and digestion are significantly better EOL options for PLA than landfilling, because of the faster disintegration.

6. COMBINING COFFEE SHOP REQUIREMENTS, CUSTOMER REQUIREMENTS AND ENVIRONMENTAL ANALYSES

The aim of this section is to determine the best overall cup from a coffee shop owner, coffee shop customer and environmental perspective. This is done by ranking the three cups (PE-coated, PLA-coated and edible cups) according to three sets of requirements. The first criterion that the cups are ranked against is how well they meet the coffee shop owner requirements that were established in Section 4.1. Thereafter, the cups are ranked according to how well they meet the coffee shop customer requirements from Section 4.2. Lastly, the cups are ranked in terms of their environmental impacts with reference to Section 5. The cups are ranked on a scale of one to five. One being a poor score and five being an excellent score. Furthermore, two cases are compared during the ranking evaluation. The first case is covered in Section 6.1 and assumes that there is no special composting system being implemented for compostable coffee cups. The coffee cups therefore either end up in the recycling or the non-recycling waste stream and not the composting waste stream. As previously established, it is generally infeasible to recycle disposable coffee cups. Therefore, even when disposable cups are channelled to the recycling waste stream, they generally still end up in landfills. The second case is considered in Section 6.2 and assumes that there is a special composting system being implemented for compostable disposable coffee cups. This leads to the PLA (compostable) coffee cups landing in the composting waste stream to be composted by means of an industrial composter or digestion. These two specific cases were chosen to emphasise the different rankings that PLA cups have under different assumptions.

6.1 OVERALL CUP RANKING: CASE 1

As mentioned previously, the first case assumes no special composting system for compostable cups and therefore an EOL scenario of landfilling. The disposable coffee cup ratings for case 1 can be seen in Table 3.

Table 3: Case 1 Cup Rankings.

Case 1: No Special Composting System			
Requirement	PE Cups	PLA (Bio) Cups	Edible Cups
Low cost	5	3	1
High strength	5	5	1
Good aesthetics	5	5	3
Good insulation	5	5	3
Must be able to be sealed	5	5	3
Good grip	5	5	3
Environmental friendliness	1	1	5
Achieved ranking	31	29	19
Total possible ranking	35	35	35

As can be seen from the rankings, the PE cups achieved the highest score. However, this is despite the fact that they are not environmentally friendly as they have a large CO_2 footprint (significantly larger than the edible cups). The PLA cups achieved a ranking of 29 out of a possible 35, putting them 2 points behind the PE cups. PLA cups received a lower rating than PE cups in the cost area, because the cups are more expensive than PE cups (approximately 28 percent according to the interviewed coffee shop owners). Similar to the PE cups, the PLA cups achieved a poor rating for their environmentally friendliness, because of the assumption that there is no special composting system for them, which causes them to end up in landfills along with nonbiodegradable (PE) cups. This is also explained in Section 5.1, which indicates that the difference between the CO_2 equivalents of PE and PLA cups is smaller than 3 percent. It can therefore be concluded that in a landfilling EOL scenario, PLA and PE cups have equally large CO_2 footprints. The edible cups achieved the lowest score of 19 out of a possible 35, because based on an interview with an employee at the Bakeys company, they are flimsy and expensive and therefore achieved a rating of 1 for both low cost and high strength. Considering that Sorghum edible cups have not yet been manufactured, the majority of coffee shop owner and customer requirements were difficult to rank. A ranking of 3 (neutral) was therefore given to the requirements of good aesthetics, good insulation, must be able to be sealed and good grip. Considering that edible cups will be eaten, they were considered to have little to no EOL environmental impact. The edible cups were thus given a rating of 5 for environmentally friendliness.

In order to account for the uncertainty related to edible cups, a sensitivity analysis can be performed. Specifically, it is possible to (in addition to the baseline case already considered) also consider the best-case scenario for the elements regarding edible cups that were uncertain. In this case the uncertain edible cup elements, namely aesthetics, insulation, sealability and grip, were given a rating of 5. For the best-case scenario, the rankings can be seen in Table 4.

Table 4: Case 1 Cup Rankings (with Sensitivity Analysis).

Case 1: No Special Composting System			
Requirement	PE Cups	PLA (Bio) Cups	Edible Cups
Low cost	5	3	1
High strength	5	5	1
Good aesthetics	5	5	5
Good insulation	5	5	5

Must be able to be sealed	5	5	5
Good grip	5	5	5
Environmental friendliness	1	1	5
Achieved ranking	31	29	27
Total possible ranking	35	35	35

Even though the uncertain elements of edible cups were given a rating of 5, edible cups still achieved the lowest ranking of the three cups as can be seen in Table 4. It can therefore be concluded that the cup rankings are not sensitive to the uncertainty related to the information regarding edible cups, as the results remained unchanged.

6.2 OVERALL CUP RANKING: CASE 2

As previously mentioned, the second case evaluates the different coffee cups under the assumption that there is a special composting system for compostable cups. The EOL scenarios for PLA (compostable) cups are therefore either industrial composting or digestion. Table 5 shows the disposable coffee c up ratings for case 2.

Table 5: Case 2 Cup Rankings.

Case 2: Special Composting System			
Requirement	PE Cups	PLA (Bio) Cups	Edible Cups
Low cost	5	3	1
High strength	5	5	1
Good aesthetics	5	5	3
Good insulation	5	5	3
Must be able to be sealed	5	5	3
Good grip	5	5	3
Environmental friendliness	1	5	5
Achieved ranking	31	33	19
Total possible ranking	35	35	35

Contrary to the first case, the cups with the overall best score in the second case are the PLA cups. The higher ranking of PLA cups was caused by them having a rating of 5 for their environmentally friendliness in the second case. This is justified in Section 5.3 that illustrates that both industrial composting and digestion of PLA cups have a significantly lower CO_2 equivalent than landfilling of PLA cups and therefore also PE cups, assuming maximum decomposition. The ratings of PE cups and edible cups remain unchanged from case 1 seeing as they are not affected by the implementation of a composting system. Similar to the first case, the uncertainty of edible cups was accounted for by means of a sensitivity analysis. The uncertain elements of edible cups (aesthetics, insulation, sealability and grip) were again given a rating of 5. This, however, did not change the overall ranking of edible cups as they still achieved the lowest ranking. It can therefore be concluded that the cup ranking for the second case is not sensitive to the uncertainty of edible cups. The ranking results can be seen in Table 6.

Table 6: Case 2 Cup Rankings (with Sensitivity Analysis).

Case 2: Special Composting System			
Requirement	PE Cups	PLA (Bio) Cups	Edible Cups
Low cost	5	3	1
High strength	5	5	1
Good aesthetics	5	5	5
Good insulation	5	5	5
Must be able to be sealed	5	5	5
Good grip	5	5	5
Environmental friendliness	1	5	5
Achieved ranking	31	33	27
Total possible ranking	35	35	35

7. INVESTIGATING SHORT LISTED OPTIONS

In Section 6 it was identified which cups are best under which conditions. This section aims to investigate whether decisions in practice correspond with the previous section's findings i.e. are coffee shops implementing the optimal strategies. To achieve this, Deluxe Cofee and Damascus Road were considered in Stellenbosch in the Western Cape and their cup choice given the current EOL scenario of the cups was evaluated. If a mismatch was found, recommendations were made as to how waste management practices could be improved. Section 7.1 considers two additional coffee shops and one university cafeteria - Blue Crane, Hazz and Chalkboard Cafeteria, respectively. It is identified how many of the five interviewed coffee shops use compostable disposable coffee cups. Thereafter, Section 7.2 presents a case study at Chalkboard Cafeteria to investigate whether or not they are following an optimal waste plan with regards to disposable coffee cups. Lastly, Section 7.3 performs a further case study at the municipality of Stellenbosch to identify what the current waste treatment of compostable cups is.

7.1 CUP RANKING EVALUATION

As can be seen from the overall rankings of the two cases, compostable cups are only a better option if they are treated correctly and there is a correct composting method being implemented. By means of interviewing five coffee shops and restaurants within the Western Cape, it was found that all five use compostable coffee cups.

Table 7: Coffee Shop Cup Type.

Which coffee cup type do these coffee shops/restaurants use?	
Coffee Shop	Cup Type
Deluxe Coffee	Compostable
Chalkboard	Compostable
Blue Crane	Compostable
Hazz	Compostable
Damascus Road	Compostable

In order to evaluate whether or not these cups are being treated correctly two in-depth case studies are performed. The first is a coffee shop whose waste stream is handled as part of the university waste system. The second coffee shop's waste is handled as part of the local municipality's waste system. These case studies thus provide a more in-depth perspective of two parallel waste management systems.

7.2 CASE STUDY: STUDENT CAFETERIA

This case study discusses a student cafeteria. Section 7.2.1 analyses the waste treatment process that disposable coffee cups currently followed at the cafeteria. Thereafter, a suggested situation is described in Section 7.2.2 to potentially optimise the waste treatment process of disposable coffee cups at the cafeteria. To gain more information about the disposable coffee cup waste treatment at the cafeteria, an interview was held with a stakeholder involved with facilities management at the cafeteria.

7.2.1 CURRENT SITUATION

Whilst considering the cafeteria specifically, it was noticed that there is currently no special waste stream to ensure that disposable coffee cups are composted. At this stage there are either bins for all waste forms or the disposable coffee cups are thrown in the "Recyclable Bin". In both of these cases disposable coffee cups end up in the waste storage room with the aim of them being recycled. However, as previously established disposable coffee cups cannot be recycled. Thus, during waste segregation, the disposable coffee cups are separated from the recyclable waste and sent to a landfill. At this stage the compostable waste is sent to a processing facility, where it is composted, bagged and sold.

7.2.2 SUGGESTED SITUATION

In order to compost the disposable coffee cups, there either needs to be a separate dustbin for cups and compostable waste or a coffee cup icon must be added to the compostable bins. This therefore ensures that the disposable coffee cups are put in the compostable waste stream and sent to the processing facility, where it is composted. Instead of bagging and selling the generated compost, it is suggested to rather use this compost at the university gardens. This could potentially reduce the carbon footprint due to the decrease in transport and packaging.

7.3 CASE STUDY: WESTERN CAPE MUNICIPALITY

This case study involves a Western Cape municipality. In Section 7.3.1 the current waste treatment of disposable coffee cups in the municipality is described. Thereafter, in Section 7.3.2 a suggested situation to potentially optimise the waste treatment of disposable coffee cups is discussed.

7.3.1 CURRENT SITUATION

Currently the municipality implements a system to collect waste from restaurants/coffee shops three times a week. According to the manager of solid waste at the municipality, it costs the restaurant/coffee shops R560 to service a bin for a month. Coffee cups are currently put in black bags with the rest of the waste as can be seen in Figure 12. The recyclable waste is separated from the rest to go to waste processing. However, considering that PLA cups are not recyclable they are classified as "food packaging" and sent to landfills.

Figure 12: Black Bin Bags collected by the municipality.

7.3.2 SUGGESTED SITUATION

To potentially better the waste treatment of disposable coffee cups, the municipality is currently considering a new scenario. In this scenario it is recommended that restaurants/coffee shops should be given one white bag for every black bag they receive. The restaurants and coffee shops are asked to put food and compostable coffee cups into the white bags. As an incentive, the municipality will only charge a service fee of R460 per month and collect the bags five times a week. The restaurants/coffee shops will therefore also have more storage space. The aim of these white bags is to send their contents to an anaerobic digestion plant. Disposable coffee cups, along with the food waste and will therefore be anaerobically digested. The rest of the waste will be separated into recyclable and non-recyclable waste. Considering that food waste will no longer be in the same bag as recyclable waste, the chance of contaminating recyclable waste also decreases. This system can be implemented as a pilot test at a few restaurants/coffee shops throughout the municipality to test feasibility and iron out any issues.

8. CONCLUSION

This section discusses the results from the previous sections in Section 8.1 and identifies various recommendations and practical considerations in Sections 8.2 and 8.3, respectively.

8.1 EVALUATE RESULTS

Throughout the course of this study it has become evident that there is a growing awareness of the global disposable cup problem as all the interviewed coffee shops are opting for compostable disposable coffee cups. However, as can be seen by the environmental impact comparisons of PLA cups and PE cups in Section 5.3, compostable (PLA) cups only have a lower CO_2 equivalent if they are composted or digested and not landfilled. If PLA cups are landfilled along with PE cups, their CO_2 equivalent is only slightly lower than that of PE cups and their degradation time is significantly longer than that of composting or digestion. This therefore suggests that they have the same impact on the environment as PE cups in a landfilling EOL scenario. Coffee shops and consumers therefore think they are buying eco-friendly cups, but they are not aware that if these cups are not disposed of correctly, they have the same eco-impact as non-compostable cups. Moreover, disposable coffee cups are often not consumed at the coffee shop premise. This limits the coffee shop's influence on the waste management of the disposable cups. There is therefore a need to raise awareness regarding the EOL waste management scenarios of coffee cups and their impact on environmental outcomes.

8.2 RECOMMENDATIONS

Various recommendations were developed throughout this article. The most important ones are highlighted in the following list:

1. An important recommendation is to ensure that compostable coffee cups end up in composting waste streams and are not landfilled. The suggested situations in Sections 7.2.2 and 7.3.2 are

recommendations for two specific cases to ensure that compostable cups are sent to the correct waste stream. However, from a more general perspective, an improved waste management ow of disposable coffee cups can be achieved by coffee shops investigating different options and liaising with their municipalities.
2. As mentioned in Section 8.1, the consumption of disposable cups often does not take place at the coffee shop premise. This therefore limits the coffee shop's influence on the waste management of disposable coffee cups. An important recommendation is thus to raise the consumer's awareness of the correct disposal procedure for disposable coffee cups. Furthermore, the municipality and waste management company need to be aware and make the correct infrastructure available.
3. A further recommendation to help consumers understand the significance of compostable coffee cups and their disposal method is communication between the coffee shops and coffee shop customers. If coffee shops, for example, use compostable coffee cups they can have a sign that describes the benefits of using compostable coffee cups and into which bins they need to be thrown. Coffee shop customers will therefore be made aware of the importance of throwing the compostable cups into the correct bins.
4. To ensure that the compostable coffee cups are in fact compostable, coffee shops could implement various quality checks to validate this. For instance they could ensure that their supplier is trustworthy by sending sample coffee cups to an industrial composter to test whether they decompose.

8.3 PRACTICAL CONSIDERATIONS

Whilst working on this study various practical concerns, that need to be considered, were identified. These are summarised in the following list:

1. A large concern that is raised by an increase in composting efforts of PLA cups within the municipality specifically, is whether there is enough space for industrial composters or digestion heaps. Seeing as the municipality is already limited for space, the added pressure of finding locations to place large composting facilities might become a problem. However, over time, this should be offset by the reduction of waste sent to landfill.
2. Another concern that was identified during this study was whether disposable coffee cups, that are sold at coffee shops, are compostable. Often disposable cup manufacturers claim that their cups are compostable, however, as mentioned in Section 3.1.2 this does not necessarily indicate that they are compostable according to the composting requirements. It is therefore of high importance that the coffee shop owners find a trustworthy compostable disposable coffee cup supplier.
3. As mentioned in Section 7.3.2, the recommended situation at the Western Cape coffee shops is to send their food waste and coffee cups to an anaerobic digestion plant. However, according to stakeholders of the municipality, the digestion plant it would be sent to is used for mesophilic digestion and not thermophilic digestion. As mentioned in Section 5.2, PLA-coated coffee cups require thermophilic digestion to disintegrate. The anaerobic digestion plant would therefore not be suited for PLA-coated coffee cups and the municipality would have to consider other options to cater for these coffee cups. This practical consideration shows that there are several points of failure that could cause compostable coffee cups to end up in landfills. It is therefore clear from the study that the necessary systems are not yet in place and many improvements still need to be made to ensure that disposable coffee cups are indeed composted. Moreover, the results from the study highlight the important of using reusable coffee cups as opposed to disposable coffee cups.
4. Considering that PLA is primarily made food sources, such as corn, corn prices might rise if there is a significant increase in PLA-coated disposable coffee cups. This could also increase world hunger, especially in third world countries. This implication might cause coffee shops to reconsider implementing PLA-coated coffee cups.

8.4 LIMITATIONS

To practically complete this project, various assumptions were required. These introduce various limitations to the research that need to be considered when interpreting the results of the study. These include:

1. There was no South African data found for the landfilling of disposable coffee cups. It was therefore necessary to use data from a different country, namely the United States of America. Even though it was assumed that South African data would not differ significantly from American data, climate and waste management system differences between the two countries may cause differences that were not accounted for.
2. This study uses a very small sample size of coffee shop owners and coffee shop customers. This was mainly due to the similar answers the two coffee shop owners and the three coffee shop customers had. However, there is still further research that can be done to build on this study.
3. There is no data for the composting of PLA-coated paperboard cups. There were, however, LCAs that had composting data for PLA and chemical pulp (paper) individually. It was therefore necessary for the student to integrate the two separate CO_2 equivalents. Considering that the data for PLA and chemical pulp did not differ significantly, the average was taken to get the overall CO_2 equivalent. However, the way these materials are combined in coffee cups may lead to unanticipated differences in emissions during degradation.

REFERENCES

[1] Bioplastics, “PLA for Paper Coating,” vol. 6, pp. 34–37, 2011.

[2] R. Gade, M. Siva Tulasi, and V. Aruna Bhai, “Seaweeds: A novel biomaterial,” *Int. J. Pharm. Pharm. Sci.*, vol. 5, no. SUPPL. 2, pp. 40–44, 2013.

[3] E. B. Arikan and H. D. Ozsoy, “Bioplastics as a Green Material,” pp. 1–5, 2014.

[4] E. van der Harst and J. Potting, “A critical comparison of ten disposable cup LCAs,” *Environ. Impact Assess. Rev.*, vol. 43, pp. 86–96, Nov. 2013.

[5] R. Emas, “The Concept of Sustainable Development : Definition and Defining Principles,” *Br. GSDR*, pp. 1–3, 2015.

[6] T. J. H. Timothy F. Slaper, “The Triple Bottom Line: What is it and how does it work?,” 2011. [Online]. Available: http://www.ibrc.indiana.edu/ibr/2011/spring/article2.html.

[7] Bridge Gate, “What’s the Difference? - Compostable vs. Biodegradable vs. Recyclable,” 2013. [Online]. Available: http://bridge-gate.com/2013/07/whats-difference-compostable-vs-biodegradable-vs-recyclable/.

[8] HPCG, “The Difference Between Biodegradable and Compostable Packaging Materials,” 2017. [Online]. Available: http://www.hpcorporategroup.com/the-difference-between-biodegradable-and-compostable-packaging-materials.html.

[9] Ianfranco, “Composting vs. Anaerobic Digestion, and the potential of Biogas,” 2017. [Online]. Available: http://alanfranco.com/composting-vs-anaerobic-digestion-potential-biogas/.

[10] F. C. Trust, “The Dangers Of Polystyrene,” 2010. [Online]. Available: http://businessbarbados.com/trending/green-business/the-dangers-of-polystyrene/.

[11] T. Häkkinen and S. Vares, “Environmental impacts of disposable cups with special focus on the effect of material choices and end of life,” *J. Clean. Prod.*, vol. 18, no. 14, pp. 1458–1463, 2010.

[12] A. Potter, “Let’s grab a coffee - billions of them,” 2017. [Online]. Available: https://www.choice.com.au/food-and-drink/drinks/tea-and-coffee/articles/are-takeaway-coffee-cups-recyclable.

[13] Siluria Technologies, “Natural Gas to Ethylene,” pp. 1–13, 2014.

[14] J. G. Araujo, “Everything You Need To Know About Polylactic Acid (PLA),” 2016. [Online]. Available: https://www.creativemechanisms.com/blog/learn-about-polylactic-acid-pla-prototypes. [Accessed: 23-Mar-2017].

[15] Hitachi, “Production process for polylactic acid (PLA),” 2017. [Online]. Available: http://www.hitachi.com/businesses/infrastructure/product_site/ip/process/pla.html.

[16] S. Munir, “Edible Cutlery: The Future of Eco Friendly Utensils,” 2016.

[17] A. Lynch, “KFC unveils the UK’s first EDIBLE coffee cup,” 2015. [Online]. Available: http://metro.co.uk/2015/02/25/kfc-unveils-the-uks-first-edible-coffee-cup-the-scoff-ee-cup-geddit-5078253/.

[18] Franklin Associates, “Life cycle inventory of foam polystyrene, paper-based, and PLA foodservice products,” p. 149, 2011.

[19] B. G. Hermann, L. Debeer, B. De Wilde, K. Blok, and M. K. Patel, “To compost or not to compost : Carbon and energy footprints of biodegradable materials ’ waste treatment,” *Polym. Degrad. Stab.*, vol. 96, no. 6, pp. 1159–1171, 2011.

[20] European Paper Industries, “Types of Pulping Processes,” 2017. [Online]. Available: http://www.paperonline.org/paper-production/pulping/types-of-pulping-processes.

[21] World Centric, “Compostable Plastics,” 2017. [Online]. Available: http://worldcentric.org/biocompostables/bioplastics.

[22] J. Hanna, “Meet the Change Makers: Starbuck’s Quest for a Better Cup,” 2012. [Online]. Available: http://archive.onearth.org/article/meet-the-change-makers-starbucks-quest-for-a-better-cup. [Accessed: 22-Feb-2017].

APPENDIX A

Deluxe Coffee interview

List four requirements that you as a coffee shop owner have for your disposable coffee cups?

Low cost, high strength, good insulation properties, environmentally friendliness

Do you currently use biodegradable coffee cups?

- (•) Yes
- ( ) No

How are the disposable cups currently disposed of?

The Stellenbosch municipality collects our black bags three times a week. The coffee cups are in those black bags.

Figure A1: Interview with Deluxe Coffee

Luke Simmonds Interview

Damascus Road

List four requirements that you as a coffee shop owner have for your disposable coffee cups?

Low cost, good insulation, environmentally friendly, good aesthetics

Do you currently use biodegradable or compostable cups?

- (•) Yes
- ( ) No

How much do biodegradable/compostable cups cost when compared to non-biodegradable/non-compostable cups?

Biodegradable cost 80 cents/cup
Non-biodegradable cost 58 cents/cup

How are the disposable cups currently disposed of?

Facility management is in charge of that, we have little control over the disposal of the coffee cups.

Figure A2: Interview with Damascus Road

SHAPING THE FUTURE OF RESIDENTIAL HOUSING PROJECTS – BREAKTHROUGH IN MINDSET AND TECHNOLOGY

J.E. Stimie[1]

[1]Stellenbosch University
johann.stimie@gmail.com

ABSTRACT

The objective of this paper is to illustrate how the application of Fourth Industrial Revolution principles and technology during the construction and management of a residential housing development can add value and contribute to the well-being of property owners, managers and residents. In the pursuit of this objective the methodologies involved in the development of formal theory seemed most appropriate to achieve these objectives. Building theory from case studies is a research strategy that involves using one or more cases to create theoretical constructs, from case-based, empirical evidence. In the paper the nature and extent of the South African residential housing landscape is firstly described. Some of the challenges inherent to the development and management of large scale residential developments are then discussed. A solutions framework to address some of these challenges consisting of six core elements is then presented. These core elements are a Centrailised Hot Water System powered by a PV plant, smart metering of three utilities (electricity, cold water and hot water), an automated shut-off mechanism for hot water, an Integrated Utilities Management Platform, an optimized utilities payment protocol and a Smart Payment Application. The paper concludes with an evaluation of the impact of the framework measured against the three core design objectives namely increased control, optimized operational efficiency and sustainability.

*Corresponding author

1. INTRODUCTION AND BACKGROUND

"The world is on the brink of a technological revolution that will fundamentally alter the way people live, work, and relate to one another. In its scale, scope, and complexity, the transformation will be unlike anything humankind has experienced before. It is not yet clear just how it will unfold, but one thing is clear -- the response to it must be integrated and comprehensive, involving all stakeholders of the global polity, from the public and private sectors to academia and civil society." [1]

Klaus Schwab, the Founder and Executive Chairman, of the World Economic Forum made the above statement in a paper presented at for the world economic summit at Davos in 2016, profiling what he calls The Fourth Industrial Revolution.

The world is confronted with many challenges that could impact the well-being and happiness of society. Among the most severe are rapid urbanization, climate change and resource scarcity. According to the World Economic Forum [2] around the world, 200,000 people a day are migrating to urban areas, where they need healthy, affordable, sustainable housing and infrastructure. However, the buildings and infrastructure assets the industry constructs account for a substantial portion of global greenhouse gas emissions, which are a major cause of climate change. On top of that, the industry is the largest consumer of raw materials, further depleting scarce resources.

The Infrastructure and Urban Development (IU) industry is important in the shaping of a desirable future for our planet. The sector plays a vital role in creating value for society by designing innovative human- and social-centric solutions. The IU industry lies at the heart of the global economy, directly affecting the quality of our lives and the well-being of society. The industry must, therefore, react quickly and with appropriate action to changing conditions and opportunities for new business to provide society with sustainable, affordable assets that fulfil human needs.

Digital technologies have launched the Fourth Industrial Revolution (FIR), transforming entire industries. The FIR will have a significant impact on the world of commerce and work specifically. In the future fewer commercial, office and public assets will be required, because so much happens virtually and because more workplaces and other spaces will be shared. In an automated world and against the backdrop of large scale urbanization the demand for residential and recreational buildings will however increase significantly. In an economy that primarily runs on automation, people spend more time at home or pursuing leisure activities, increasing the demand for residential assets. More leisure time increases the need for recreational structures and facilities, including swimming pools, fitness centers, theatres and arenas, and large assets such as stadiums and amusement parks.

However, the Infrastructure and Urban Development industry (IU) has been slow to adopt and did not kept up with recent FIR developments. Most companies in the industry's many sectors still use primarily manual methods, offer traditional products and services and operate according to established practices and business models [2]. The IU industry can no longer afford to stand still. Global megatrends such as climate change, resource scarcity, demographic shifts, and automatization and digitalization are affecting economies, governments and society at large. Players along the IU value chain need to prepare strategically and make the right moves to thrive amid the disruptions these trends could cause [2].

1.1 Objective of this paper

The objective of this paper is to illustrate how the application of FIR technology and principles during the construction and management of a residential housing development can add value and contribute to the well-being of property owners, managers and residents.

1.2 Research design strategy and methodology

The methodologies involved in the development of "Formal theory" seemed most appropriate to achieve these objectives. During the development of the chosen methodology for this research the work done by Eisenhardt [3] was found to be extremely useful.

Building theory from case studies is a research strategy that according to Eisenhardt [3], involves using one or more cases to create theoretical constructs, propositions and/or midrange theory from case-based, empirical evidence. Theory building through case-studies is according to Eisenhardt [4] an increasingly popular and relevant research strategy that forms the basis of a disproportionately large number of influential studies.

Although the writing of the emergent theory or in this case the development of the solution framework is presented as the last step in linear process of steps, it should not be construed as such. Theory building through case studies is a highly iterative process and already starts during the justification of the need for new theory. During the research process and as observations lead to new insights it was anticipated the new literature would be explored and that the scope of research may potentially widen. With a full understanding of this risk great care was taken during both the literature review as well as the during the ultimate development of the solutions framework to not widen the scope of the study beyond the constraints of the problem statement.

1.3 Overview of the remaining sections

In this paper the following aspects will addressed:

- The FIR will be briefly defined and contextualized in Section 2;
- In Section 3 a brief overview of the South African residential housing landscape will be presented;
- In Section 4 some of the challenges inherent to the development and management of residential rental properties will be presented;
- In Section 5 the implementation context and research approach are briefly described;
- In Section 6 the solutions framework will be presented. In this section the various components as well as the interaction between these components will be discussed briefly;
- In Section 7 the impact of the framework is evaluated against three core design objectives; and
- In Section 8 a number of concluding remarks will be presented.

2. THE FOURTH INDUSTRIAL REVOLUTION – THEORETICAL CONSIDERATIONS

According Darth and Horch [5], to the first three industrial revolutions spanned almost 200 years. First, mechanical looms driven by steam engines in the 1780s started a significant change. Fabric production left private homes in favour of central factories, followed by an extreme increase in productivity. The second industrial revolution began about 100 years later in the slaughterhouses in Cincinnati, Ohio, and found its climax with the production of the Ford Model T in the United States. The development of continuous production lines based on both division of labour and the introduction of conveyor belts resulted in another productivity explosion. Third, in 1969, Modicon presented the first programmable logic controller that enabled digital programming of automation systems. The programming paradigm still governs today's modern automation system engineering and leads to highly flexible and efficient automation systems.

It is according to these authors remarkable that the Fourth Industrial Revolution (FIR) was announced a priori. It is characterized by a fusion of technologies that is blurring the lines between the physical, digital, and biological spheres, collectively referred to cyber physical systems (CPS) [5]. It is marked by emerging technology breakthroughs in a number of fields, including robotics, artificial intelligence, nanotechnology, quantum computing, biotechnology, the Internet of Things, the Industrial Internet of Things, fifth generation wireless technologies (5G), additive manufacturing/3D printing and fully autonomous vehicles. Digital technologies have launched the Fourth Industrial Revolution (FIR), transforming entire industries. The FIR will have a significant impact on the world of commerce and work specifically.

However, the Infrastructure and Urban Development industry (IU) has been slow to adopt and did not kept up with recent FIR developments. Most companies in the industry's many sectors still use primarily manual methods, offer traditional products and services and operate according to established practices and business models [2]. Players along the IU value chain need to prepare strategically and make the right moves to thrive amid the disruptions these trends could cause [2].

In the next section a brief overview of the some of the current and future South African housing realities will be presented.

3. THE SOUTH AFRICAN RESIDENTAIL HOUSING LANDSCAPE

The ANC, after coming to power in April 1994, promised that a democratic state would act to steer the mixed economy down a new economic growth path through various macro-economic policies. The first major attempt at creating a policy framework was a document called the Reconstruction and Development Programme (RDP). The RDP was the political manifesto of the ANC during its election campaign for the first democratic elections in South Africa in 1994 and would be used as the framework for the transition to a democratic state. Nation building and improving the living standards of all South Africans through a local government sphere is at the heart of the RDP [6].

Williams [7] is of the opinion that the RDP did achieve some of its social security objectives by establishing an extensive welfare system, helping the aged, disabled and others who are unable to meet their basic needs. However, social welfare was not the main target of the RDP but rather a focus on housing, with the goal of building more than one million houses in five years, and pro- viding water and electricity to households.

Subsequent to the RDP the ANC led government has developed no less that four additional macro economic frameworks. These include:

- Growth, Employment and Re-distribution (GEAR) -- 1996
- Accelerated and Shared Growth Initiative for South Africa (ASGISA) – 2005
- New Growth Path (NGP) -- 2010
- National Development Plan (NDP) -- 2012

Each one of these macro economic frameworks did make a contribution to creating a better life for all South Africans, but according to Ansari [8], the ANC government has largely failed to deliver on the two decades of promises of economic redistribution and industrialization. A consequence of expectations, created by these policies and unmet thus far, are the waves of service delivery protests across South Africa. Service delivery protests mostly stem from the lack of access to basic services, including access to sanitation, water, refuse removal, electricity and basic housing.

3.1 Important housing statistics

Housing is thus one of the major challenges the South African government has to overcome. According to Stats SA [6] there were 51.8 million people living in 14.5 million households in 2011 in South Africa. The average household thus consisted of 3,57 individuals in 2011. Research done by Stats SA [3] indicates that the South African population will grow to 70 million people by 2030. 70% of these people will be urbanized. In the period 2001 to 2011 the number of household grew by 2,6% per annum, while the population grew by 1,5%. The increase in the proportion of households that comprise of one person is driving the trend in declining household sizes. During the 2006 census there were only 16% of households that comprised on a single person, while the 2011 census found that 27% of households comprised on a single person. These demographic realities are summurised in Table 3.1

The department of human settlements [6], identifies four broad housing categories, Formal Dwelling; Traditional Dwelling; Shack not in backyard and Shack in backyard. Table 2.2 illustrates the changes in distribution of households per housing category in the period 2001 to 2018.

Table 3.1: Key South African demographic realities

	2001	**2011**	**2018**	**2030**
Total Population	44,9 mil	51,8 mil	55,9 mil	70 mil
Average size of household	3,57	3,57	3,57	3,57
Total number of households	12,5 mil	14,5 mil	15,6 mil	19,6 mil
Level of Urbanization	57%	62%	66%	70%
Need for Urban Houses	7,1 mil	8,9 mil	10,2 mil	13,7 mil

Table 3.2: Households per housing category

Housing category	**2001**	**2011**	**2018**
Formal Dwelling	64%	74%	81%
Traditional Dwelling	15%	8%	6%
Shack not in back yard	12%	9%	8%
Shack in back yard	4%	5%	5%
Total households living in informal dwellings	2 mil	2 mil	2,1 mil

Over that period 2001 to 2011 the private sector build around 660 000 new units. Although the number of households living in a formal dwelling thus increased by over 3,5 million between 2001 and 2011, there were still no less than 2 million household in 2011 living in some form of informal settlement or shack. In the period 2014 to 2018 an additional 410 000 low cost houses were build. The current Minister of Human Settlements, Nomaindia Mfeketo recently however confirmed that although some progress have been made regarding the delivery of more low cost houses the backlog of low-cost homes are still estimated to be at least 2.1 million.

Between 2001 to 2011 the proportion of South African households that rent their primary dwellings increased from 19% to 25%. In urban areas it increased from 26% in 2001 to 32% in 2011 [4]. In the next section the above statistics will be used to determine the urban housing requirements and thus the demand for urban housing over the next 12 years.

3.2 Urban housing requirements 2030

In an attempt to estimate the 2030 urban housing requirements a number of extrapolations from the statistics presented in Section 2.1 were made. For the purpose of these calculations the following assumptions were made:

- The trend regarding declining household sizes (discussed above) were not considered, and the average household size was considered to be 3,57 [9];
- The South African population will grow to 70 million by 2030 and 70% of the population or 49 000 000 people will live in urban areas;
- The implication is that there will be 13,7 million urban households in 2030;
- There are currently (2018) 10,2 mil urban households; and
- Of the 10,2 mil urbans households 2,1 million are however living in informal settlements.

An additional 3,5 million urban housing units need to be constructed just to cater for the growing urban population. An additional 2,1 million units however have to be constructed if the low-cost housing backlog are to be eradicated. The implication is thus that 5,6 million new urban homes need to be constructed over the next 12 years to cater for the ever increasing demand for urban houses. This equates to 460 000 housing units per year.

In Section 2.1 it was mentioned to no less than 32% of households rent their homes. If it is accepted that this proportion will not change over the next 12 years, a total of 4,4 million households will rent their homes by 2030. Of the 5,6 million new urban households at least 1,8 million will thus be rental homes. The implication is thus that at least 150 000 new housing units will have to be purposely build every year over the next 12 year for rental purposes.

4. INSTITUTIONAL INVESTMENT AND MANAGEMENT OF RESIDENTIAL RENTAL PROPERTIES

Investors typically has a large number of investment opportunities. Against the backdrop of the growing demand for residential rental properties a major South African Life Insurance company ABC[1] created a fund dedicated to the development and management of affordable residential rental stock. Over the past 10 year no less than 4000 rental units have been developed by the fund. These investment opportunities are however not risk free and require careful analysis and planning. The performance of the fund is measured through generally accepted financial indicators such as Yield, Net Operating Income and Asset Value Growth [10]. A detailed discussion of these measure fall outside the scope of this document.

The management of the units are outsourced to a number of dedicated property management companies. Company XYZ[2] was appointed by ABC to manage some of the rental assets on their behalf. In the next section some of the challenges faced by XYZ during the management of the institutionally owned rental stock will be presented.

4.1 Challenges inherent to the management of a residential rental portfolio

The long term management of an institutionally owned large scale residential rental portfolio has a number of inherent challenges. These challenges are grouped together under the following headings:

4.1.1 Asset strategy and management

The Portfolio Yield is often the most important indicator for success when a Residential Rental portfolio is evaluated. Portfolio Yield is a function of a number of factors including Total Operating Cost, Average Rental Income and Total Occupancy Rates. The Asset Strategy considers all these aspects. It is thus essential to develop a realistic and achievable strategy at the outset of every project. Market conditions are however extremely volatile and strategy development has thus become an ongoing activity. It requires constant market analysis and plan iteration. This include advise on unit design, roll-out plans and rental increases and decreases.

4.1.2 Management reporting and feedback

The key to making good Asset Management decisions is access to appropriate information. Asset intensive organizations rely on asset data, information and ultimately asset knowledge as key enablers in undertaking both strategic Asset Management activities and operational Asset Management activities. There are currently no integrated rental management system that ensures both the capturing of operational and performance data, and that enables integrated and portfolio wide reporting and analysis.

---

[1] Company ABC is a pseudo name and used to ensure confidentiality.

[2] Company XYZ is a pseudo name and used to ensure confidentiality

4.1.3 Property marketing and tenant management

Total Occupancy Rate is a function of both growth and retention. Effective property marketing and tenant management thus requires a detailed understanding of both these aspects. Within the context of existing property– and consumer protection legislating[3], the placement of high quality and reliable tenants are critical. Unreliable and non-paying tenants quickly become a huge liability and result in huge value destruction. According to the Rental Monitor report [11],

- 69% of renters pay their rent on time;
- 11% of renters pay late but within the month;
- 5% of renters have some form of payment arrangement in place; and
- 5% of renters have outstanding rent that are never paid.

Within the context of South African Legislative framework the eviction of non-paying tenants has become extremely cumbersome. The eviction process normally takes extremely long and is very costly. The tenant management process thus requires constant interaction and communication.

4.1.4 Physical Asset Management (Facilities Management and Maintenance)

The contemporary business environment has raised the strategic importance of the Physical Asset Management (PAM) function in organizations which have significant investment in Physical Assets such as residential properties. Organizations the world over are increasingly becoming more and more capital intensive and it has been found in various studies that spending on assets and asset maintenance has been increasing steadily over the past number of decades.

The Institute for Asset Management (IAM (2011)) [12], defines Asset Management Strategy as follows:

"Long term optimized approach to the management of assets, derived from, and consistent with the organizational strategic plan and the asset management policy"

Within this context PAM has become a critical support function. The definition of relevant and adaptive PAM strategies is thus critical to the long-term success of any asset owner.

4.1.5 Utilities Management

The vending and management of utilities within large residential estates are inherently complex. The ineffective management of this process not only cause huge tenant frustration, but can also result in costly under recovery.

5. THE CATALYST FOR CHANGE

Company XYZ has a very strong focus on Operational Excellence. The management team is constantly seeking ways to optimize and redefine ineffective processes. During the management of the property portfolio's under its control the challenges and inefficiencies inherent to the management of this type of asset (and briefly introduced above) become apparent.

In an attempt to address the inefficiencies the XYZ management team in collaboration with a number of stakeholders developed a solutions framework and implemented the framework during the development and management of a new residential estate – Sun Village[4]. The purpose of this solution framework was not address each one of these challenges (referred to in Section 4) but to apply Fourth Industrial Revolution technology and principles to add value and contribute to the well-being of property owners, managers and residents. It was agreed that the impact of the framework would be evaluated against the following three design objectives presented in Table 5.1:

---

[3] There are at least 20 statutes and Acts relating to property in South Africa. A detailed discussion of the current legislative framework governing the management of rental properties in South Africa falls outside the scope of this document. It is however important to note the suit of legislation and regulations is extremely cumbersome and complicates the management of rental portfolios significantly.

[4] Sun Village is a pseudo name to ensure confidentially.

Table 5.1: Solution framework design objectives

Design Objective	Description
Control	The extent to which the framework provide more control to all stakeholders.
Operational Efficiency	The extent to which the implementation of the framework lead to higher levels of operational efficiency.
Sustainability	The extent to which the implementation of the framework ensures higher levels of both environmental and financial sustainability.

Sun Village consist of 330 rental properties. The development consists of 18 blocks of 3 story apartments that range in size and typology. The size and unit typology is presented in Table 4.1.

Table 5.2: Size and typology of apartments at Sun Village

Typology	Number	Size
Studio	17	28m2
One Bedroom	21	36m2
Two Bedroom	257	48m2
Three Bedroom	35	56m2

This solutions framework is presented in this next section.

6. THE SOLUTION FRAMEWORK

During the design phase of the Sun Village the XYZ management team conceptualized a number of solutions to address some of the above mentioned inefficiencies. Although these solutions were initially considered to be stand-alone, close collaboration with various stakeholders enabled the management team to implement an integrated optimized operations management process. The Solutions Framework is presented in Figure 5.1.

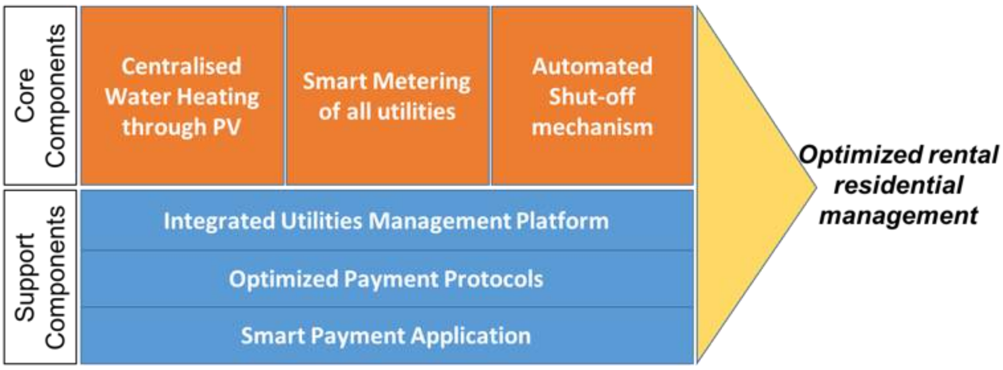


Figure 6.1: The Solutions framework

The solution consists of the following components:

- A Centralised Water Heating System (CWHS) and a Photovoltaic (PV) plant;
- Smart Metering of all utilities;
- An automated shut-off mechanism for hot water;
- An Integrated Utilities Management Platform (IUMP);
- An optimized utilities payment protocol; and
- A Smart Payment Application (SPA)

Each one of these components are briefly introduced in the next section.

6.1 A Centralised Water Heating System (CWHS) and the use of sustainable energy

In this section the focus will firstly fall on different types of water heating systems, and secondly on sustainable sources of energy. The section will conclude with a presentation of the water heating solution that was crafted for the Sun Village estate.

6.1.1 Water heating systems

Domestic hot water systems can be divided into centralised and localised. A *localised system* is one in which the water is heated locally to its needs. It may be chosen where a long distribution pipe would mean an unnecessarily long wait for hot water to be drawn off at the appliance. Traditionally most residential developments installed localised Hot Water Heating Systems (HWHS).

A *centralised system* is one in which the water is heated and possibly stored centrally within the building, supplying a system of pipework to the various draw-off points. Traditionally these systems also utilized direct electrical resistance heating elements to heat water.

6.1.2 Sustainable sources of energy

Concerns for the environment, combined with rising fuel costs, mean that more and more organizations are looking to alternative and more sustainable sources of energy to power and heat homes. The use of alternative energy source is often referred to as microgeneration. For the purpose of this document reference will only be made to three methods of alternative energy generation to power homes and more specifically to heat water.

Solar Thermal -- Solar thermal is according to Conran [13] a tried and tested technology that uses the sun's energy to heat water. Although solar thermal systems are most effective when they are installed in new houses, especially as integral building elements, they are also highly suitable for retro-fitting. Solar termal systems has become the gold standard in most affordable residential property developments within South Africa. It is however not possible to use Solar Thermal technology to heat water within a centralized water heating system. The use of Solar Thermal water heating technology thus requires the installation of one Solar Geyser per apartment or house.

Photovoltaics (PV) – Another type of solar generator is the photovoltaic cell or PV, as the name suggests converts the sun's energy into electricity. According to Chu et.al [14] PV cells, made from cilicon, are grouped to form modules, which are then arranged in panels or arrays of panels. Huge advance have taken place in this area of technology in recent years. One new development is a complete roofing system that uses PV panels in the same fashion as ordinary roof tiles. Sunlight falling on PV panels is converted electricity. The electricity is produced in Direct Current (DC), which then has to be converted to Alternating Current (AC) in an invertor. During the day (when the sun shines) the PV system generates electricity continuously, with the surplus or spare capacity flowing into the grid. Even in countries were households receive only a relatively small amount of money for the surplus energy they produce, it is still possible to make a net saving [15]. Despite the fact that South African electricity production is under severe pressure very few South African municipalities allow microgenerators of electricity to feed electricity back into the grid. A detailed discussion of the reasons for this unwillingness falls outside the scope of this document. Most PV systems in South Africa is thus run of the grid and surplus energy has to be stored in a battery that acts as the main power supply.

The use of PV in large residential developments within the South African context has been limited until recently. This is mainly due to two reason:

- The generation of electricity through PV and the demand for electricity within a residential complex is out of sync as illustrated in Figure 5.2. The yellow line illustrates the typical generation of electricity during the course of any given day. The orange area represents the demand for communal electricity, while the blue and grey areas represents the combined demand for general electricity and the combined generation of hot water respectively. Due to the fact that South African municipalities does not allow micorgenerators to feedback energy to the grid, the excess electricity either goes to waste or has to be store in batteries; and
- The acquisition and maintenance of batteries on this scale is at this juncture still not economically viable.

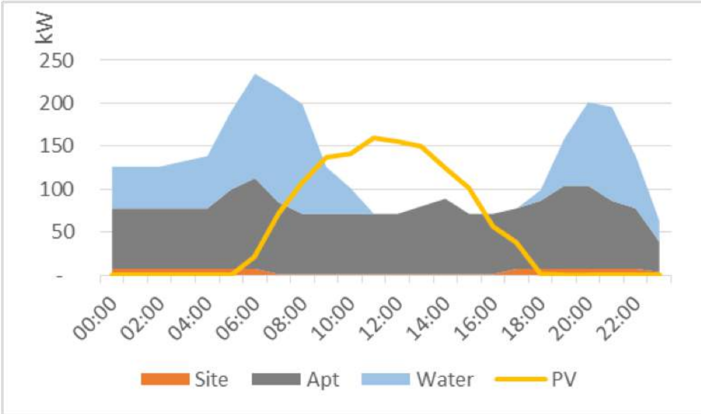


Figure 6.2: Electricity generation and demand cycles in a residential estate

Heat Pumps – Like solar thermal heat pumps have been around for some time. Heat Pump technology involves according to Bianco et.al. [16] the moving of heat from where it is plentiful, to where it can be used for space of water heating, with the assistance of an electric powered pump. The basic principle is similar to vapour compression used in refrigeration, but is makes use of the heat producing end of the thermodynamic cycle. Heat pumps are more efficient when they are powered by reneable energy.

Heat pump water heaters use electricity to move heat from one place to another instead of generating heat directly. Therefore, they can be two to three times more energy efficient than conventional electric resistance water heaters. To move the heat, heat pumps work like a refrigerator in reverse. While a refrigerator pulls heat from inside a box and dumps it into the surrounding room, a stand-alone *air-source heat pump* water heater pulls heat from the surrounding air and dumps it -- at a higher temperature -- into a tank to heat water.

6.1.3 The Sun Village Water Heating solution

The management team of XYZ in consultation with their energy partners realized that a combination of PV, Heat Pumps and the installation of large well isolated cisterns might enable them to make use of more sustainable energy sources and might enable them to reduce operational cost. Figure 5.3 illustrates the conventional warm water generation methodology within a residential estate.

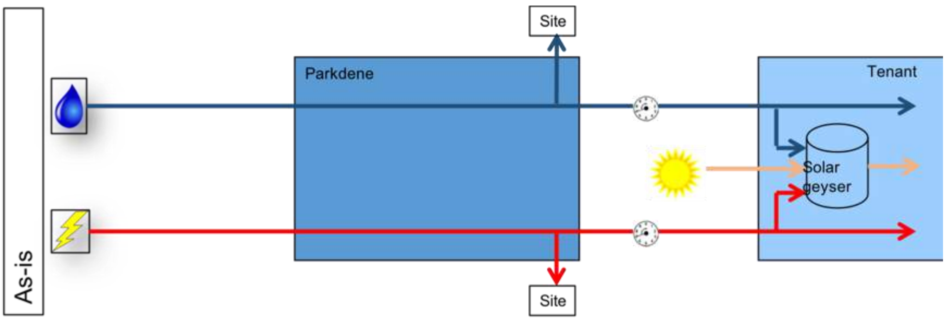


Figure 6.3: Conventional warm water generation

During the conventional process, cold water is acquired from the bulk service provider. Some of this water is used unheated and some water is heated through the solar geyser. The solar geyser is partly heated by the sun and partly heated by conventional electricity acquired from the electricity grid.

The optimized process is illustrated in Figure 5.4. During this process water is still acquired from the bulk service provider, water is however not heated through a solar thermal geyser but through a heat pump and stored centrally in well isolated cisterns. The heat pumps are further-more not powered by conventional electricity acquired from the grid but through PV panels strategically placed throughout the residential complex. All carports roofs for example are PV panels. The dilemma caused by the non-synchronized nature of PV electricity generation and demand within a residential complex as illustrated in Figure 5.1 is thus solved because excess electricity generated through PV are now used to heat water.

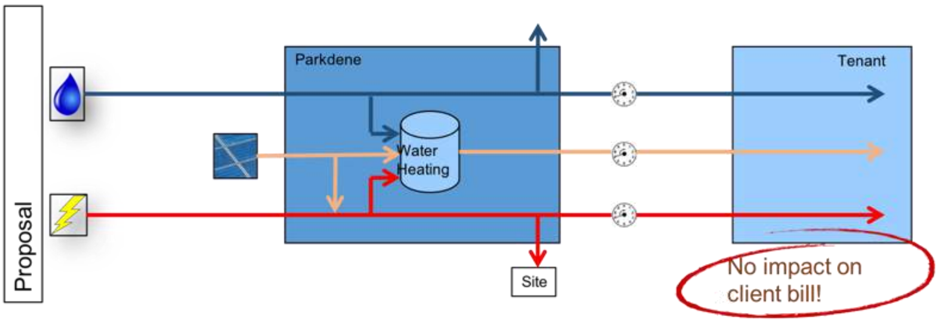


Figure 6.4: Warm water generation through PV and Heat Pump

The optimized supply and demand curves are illustrated in Figure 5.5. During the day when the sun shines the excess electricity generated through PV is used to power heat pump water heaters.

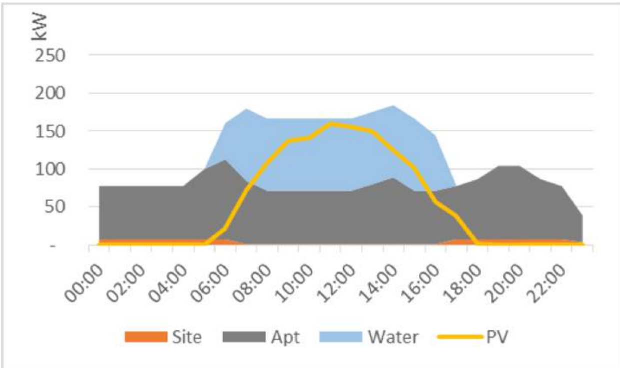


Figure 6.5: Optimized use of electricity generated by PV

The solution thus enabled company XYZ to generate warm water at a fraction of the cost of water generated through conventional methods. The management team of XYZ realized that access to relatively affordable and centrally stored hot water was the key to a number of solutions. A number of additional components were however required to ensure that the full potential of the solution could be unlocked. These components are presented in the following sections.

Although the installation of the Centralised Warm Water System (CWWS) as well as the three smart meters gave XYZ far more control, the absence of an effective utilities management system, an automated warm water shut-off mechanism and a well-defined payment protocol and payment platform however still defeated the efficiency and sustainability objectives of the solution. The development and functioning of these components are briefly discussed in the following sections.

6.2 Smart metering of all utilities

The management of utilities within a large scale residential complex is inherently complex and no standard operating industry mechanism are defined. In newer complexes electricity is normally sold on a pre-paid basis. Tenants acquire tokens at dedicated service providers. A number of relatively recent innovations has streamlined the token acquisition process to a large extent.

Access to potable water is a constitutional right. Due to the perceived high cost involved in the installation of smart meters to regulate the use of water, most residential complexes however make use of conventional water meters. These meters are read manually on a monthly basis and tenants are invoiced a month in arrears. Property managers thus has very little control over water cost recovery. Under recovery of water cost is a huge challenge within the industry and has a very negative impact on total operational cost.

Since access to electricity and warm water is not a constitutional right, the management team of XYZ realized that controlled access to these utilities would place a large amount of control in the hands of the management team. The most effective way to achieve the required levels of control was the installation of smart meters. Every apartment was equipped with three smart meters measuring and the consumption of electricity, cold water and hot water. The distribution and measurement of utilities is presented in Figure 5.6.

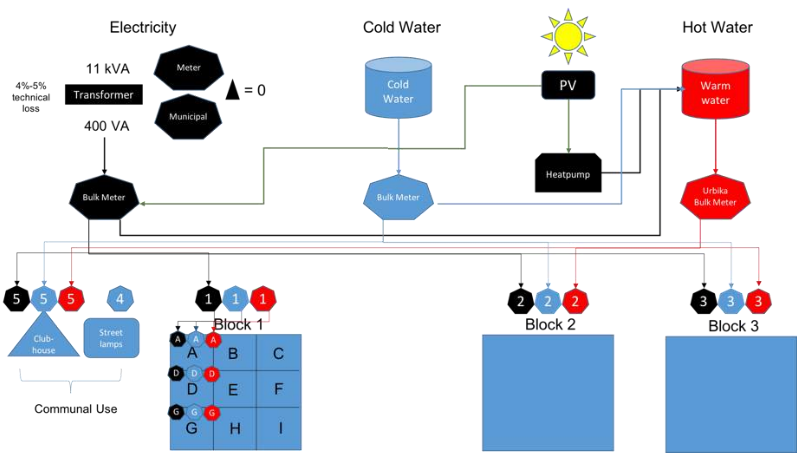


Figure 6.6: Electricity, Cold and Hot Water distribution and measurement

6.3 Shut-off mechanism for the hot water meters

The management team of XYZ realized that the installation of a Centralised Water Heating System can only yield real results if it was possible to efficiently control tenant access to warm water. If warm water control required any form of manual intervention the large scale implementation and maintenance of the solution would become unsustainable. An automatic warm water shut-off mechanism does had to be developed. The mechanism was developed in conjunction with company DEF[5] and consists of a smart meter and a simple ball valve that is remotely controlled via a GSM signal by the Integrated Utilities Management Platform (IUMP) (see discussion in Section 5.4). In the event that a tenant account is in arrears for more than 24 hours, a signal is simply send to the ball valve control mechanism, the ball valve close and warm water is shut down. As soon as the account is no longer in arrears the ball valve open and the tenant has access to the warm water.

6.4 An Integrated Utilities Management Platform

Company DEF had an existing Integrated Utilities Management Platform (IUMP). After minor configuration the platform provided a comprehensive micro-utility solution by incorporating smart meters for each of the three utilities, including a new automated shut-off mechanism for the hot water meters. The advanced tariff engine in the utilities platform enabled XYZ to also tap into the various resell margins which is available to resellers of electricity in the South African market, automating the end-to-end billing process from meter reading through to revenue collection. Using a similar principle as a pre-paid mobile data subscription, the platform's Smart Wallet feature enables XYZ to collect rent as well as utilities revenue (see Section 5.5), without standing the risk of tenants going into arrears on their bills.

The IUMP not only made the utilities management process more efficient, but also enabled XYZ to unlock additional revenue streams.

6.5 An optimized utilities payment protocol

The conventional payment protocol within most residential estates determines that rent is paid in advance on the first day of every month. Electricity is normally supplied by a third party on a pre-paid basis and cold water is billed one month in arrears. The optimized payment protocol as well as consequences for non-payment if illustrated in Table 6.1.

Table 6.1: Non-payment consequence

Order of Payment	Service and payment timing	Consequence of non-payment
1.	Rent – paid one month in advance	- No access to electricity and warm water - Eviction order 7 days after month end - Normal eviction process
2.	Cold water – paid one day in arrears, accurately measured and reconciled by the smart meter.	- No access to electricity and warm water
3.	Electricity – paid one day in arrears, accurately measure and reconciled by the smart meter.	- No access to electricity and warm water

[5] Company DEF is a pseudo name to ensure confidentiality. The company has core competence in the development and configuration of utilities management software and hardware.

4.	Warm water – paid one day in arrears accurately measured by the smart meter.	- No access to electricity and warm water

The implementation of the optimized payment protocol is enabled by the IUMP as well as the Smart Payment Application (SPA) developed by DEF. The basic functioning of the mechanism is presented in Section 5.6.

6.6 Smart Payment Application

The Smart Payment Application (SPA), was developed by company DEF and is downloaded from the iStore or Googleplay store and enables XYZ to invoice the tenant for a number of services including rent, utilities (electricity, cold water and hot water), data, parking and laundry services. The tenant transfers money via the SPA to the bank account of XYZ, but has not control of the allocation of available funds.

The SPA however enables XYZ to prioritize the allocation of paid funds in line with the payment protocol presented in Table 5.1. Apart from the rent that is payable in advance, all other invoices are payable in arrears. The implication of the above prioritized payment methodology is thus far bigger control. Using a similar principle as a pre-paid mobile data subscription, the SPA regulates the tenants access to services. If no funds are available (due to for example overdue rent) in the tenant account hot water and electricity services are automatically suspended until credit is available again.

In addition to the above SPA is also a powerful communication tool.

- The tenant is notified timeously when balances run low to enable timeous fund transfer or top-up;
- The tenant can view available balances and plan further transfers;
- The tenant can view statements;
- The tenant can view all utility usage reports in real-time and is able to compare current utilization with any previous period; and
- The tenant can log maintenance issues via the SPA and receives regular feedback regarding progress.

7. FRAMEWORK EVALUATION

It was clearly stated that the solution had three design objectives, control, process optimization and sustainability. Each one of these objectives were met during the implementation and ongoing management of the solutions framework.

Design Objective	Framework Impact	Evaluation
Control	- Various strategically placed bulk meters enable more accurate control over possible leakage and possible utility theft. - The optimized payment protocol give more control over the timing of rent payment. Late payment result in no access to warm water and electricity. - The solution resulted in 100% cost recovery of electricity, hot and cold water. - Tenants has real time access to usage statistics and can compare current usage with historical usage patterns.	Success
Process Optimization	- The Smart Payment Application optimized the often cumbersome acquisition of utilities for tenants. - IUMP automates the control process and no human intervention is required. - The algorithms governing the process ensures that services are suspended and reinstated without any human intervention.	Success

Sustainability	**Environmental sustainability** - 90% of the electricity needed to generate hot water is derived from PV panels. • **Financial sustainability** - The optimized cost recovery of utilities lead to a reduction of at least 1,5% in total operating cost. This has a huge long-term impact on the investment case. - The automated process requires far less human intervention, which resulted in significant cost savings on the part of the property manager. - The margin on the sale of hot water funds the additional Capital Cost required to implement the solution. The payback period is 7,5 years.	Success

8. CONCLUSION

The objective of this paper was to illustrate how the application of FIR technology and principles during the construction and management of a residential housing development can add value and contribute to the well-being of property owners, managers and residents.

In order to achieve the objective the following aspects were presented in the paper. In Section 2 the South African Residential Housing Landscape was briefly discussed. The section concluded with an estimate of the nature and extent of urban housing demands over the next 12 years.

In Section 3 institutional investment and management of residential housing complexes were discussed. The section concluded with a presentation some of the major challenges faced by property managers within this industry. In Section 4 the Sun Village residential estate was briefly described. The solution framework was developed and deployed within the Sun Village estate.

In Section 5 the solutions framework was presented. It was explained that the framework consisted of six components namely a Centrailised Hot Water System powered by a PV plant, smart metering of three utilities (electricity, cold water and hot water), an automated shut-off mechanism for hot water, an Integrated Utilities Management Platform, an optimized utilities payment protocol and a Smart Payment Application.

Finally the framework impact was evaluated against the stated design objectives. In Section 7 it was illustrated that the application of FIR technology and principles can indeed add value to a broad spectrum of stakeholders within a residential housing development.

9. REFERENCES:

[1] **Schwab, C 2016.** The Fourth Industrial Revolution: What it means, how to respond. *World Economic Forum*.

[2] **Boston Consulting Group.** 2018. Shaping the future of construction. Future Scenarios and implications for the industry. *World Economic Forum*.

[3] **Eisenhardt**, K.M. Building theories from case study research, *Academy of management review*. 14 (1989) 532–550.

[4] **Eisenhardt, K.M., Graebner, M.K.,** Theory building from cases: opportunities and challenges, *Academy of management journal*. 50 (2007) 25–32.

[5] **Drath, R., Alexander H.,** Industrie 4.0: Hit or hype? *IEEE industrial electronics magazine*. 8, no. 2 (2014): 56-58.

[6] Stats SA 2017. Mid Year polulation estimates. 2017 P0302, Pretoria

[7] **Williams, J.J. (2006).** Community Participation. Policy Studies, vol. 27, no. 3, pp. 197–217.

[8] **Ansari, S. (2017).** The Neo-Liberal Incentive Structure and the Absence of the Developmental State in Post-Apartheid South Africa. African Affairs, vol. 116, no. 463, pp. 206–232. ISSN 0001-9909.

[9] Stats SA 2015. Midyear population estimates. 2015 P0302, Pretoria

[10] **Escobar, M., Ferrando, S., & Rubtsov, A.** (2017). Optimal investment under multi-factor stochastic volatility. *Quantitative Finance*, 17(2), 241-260.

[11] Tenant Profile Network (2018). The Rental Monitor, Q1 2018

[12] **PAS-55**. 2010. *PAS-55*. British Standards Institute.

[13] **Conran, T.** 2013. *Eco House Book*. Octopus Books. London

[14] **Chu, S., Cui, Y., & Liu, N.** (2017). The path towards sustainable energy. *Nature materials*, 16(1), 16.

[15] **Sommerfeld, J., Buys, L., & Vine, D**. (2017). Residential consumers' experiences in the adoption and use of solar PV. *Energy Policy*, 105, 10-16.

[16] **Bianco, V., Scarpa, F., & Tagliafico, L. A.** (2017). Estimation of primary energy savings by using heat pumps for heating purposes in the residential sector. *Applied Thermal Engineering*, 114, 938-947.

A CONCEPT DEMONSTRATOR FOR SELF-ORGANISING DEMAND-DRIVEN INVENTORY MANAGEMENT IN PHARMACEUTICAL SUPPLY CHAINS

M. du Plessis[1], JH van Vuuren[2]* & J. van Eeden

[1]Stellenbosch Unit for Operations Research in Engineering (SUnORE)
Department of Industrial Engineering
University of Stellenbosch
18314937@sun.ac.za

[2]Stellenbosch Unit for Operations Research in Engineering (SUnORE)
Department of Industrial Engineering
University of Stellenbosch
vuuren@sun.ac.za

[3]Department of Industrial Engineering
University of Stellenbosch
jveeden@sun.ac.za

ABSTRACT

Perennial stock-outs of essential medicines are commonplace in the pharmaceutical supply chains of developing countries. Stock-outs are mainly attributed to a general lack of collective information sharing in pharmaceutical supply chains. In this paper, a computerised agent-based simulation model concept demonstrator is proposed and demonstrated hypothetically as part of a larger drive to establish the value of leveraging information sharing in pharmaceutical supply chains with a view to enhance decision-making. The objective of this paper is to outline the prerequisite research inputs, design requirements and hypothetical implementation of the aforementioned demonstrator. The work reported on in this paper remains a work in progress.

[1] The author is enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, University of Stellenbosch
[2] The author is a professor in Operations Research at the Department of Industrial Engineering, University of Stellenbosch
[3] The author is a senior lecturer at the Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION:

Developing nations carry a considerable burden in terms of life-threatening diseases while the treatment of these diseases is significantly complicated by stock-outs and shortages of critical medicines. Stock-outs are preventable, but to successfully thwart medicine stock-outs and their damaging consequences demands a major overhaul in the management of traditional pharmaceutical supply chains of developing countries.

Recent statistics underline the scale of the global medicine stock-outs dilemma. The Global AIDS Response Progress Reporting programme [1], for example, reported that 38 of 108 low- and middle-income countries experienced stock-outs of antiretroviral medicines in 2013. In South Africa, a survey conducted in 2015 by the *Stop Stock-outs Project* consortium [2] revealed that approximately one in four health care facilities suffered from stock-outs of either antiretroviral or tuberculosis medicines during the three-month period preceding the survey. Furthermore, 70% of these stock-outs lasted longer than one month, underlining the supply chain's failure to resolve the root causes of stock-outs rapidly.

The consequences of medicine stock-outs are pervasive and are the most severe on the subsequently untreated patients. Increased drug resistance, aggravation of, or transmission of, disease and even death are some of the harrowing consequences associated with treatment failures [2,3,4]. The impact of stock-outs is particularly harsh on impoverished communities in rural areas which depend solely on public health care services. These poor patients are forced to pay frequent, and costly, visits to their local health care facilities. Regrettably, if they are confronted with stock-outs at these facilities they are turned away and compelled to visit even farther facilities, with no guarantee of medicine availability at these facilities either [5].

The prevailing reasons for pharmaceutical supply chain under-performance in public health sectors include fragmented accountability amongst stakeholders [6], superfluous supply chain complexity [6,7], funding complexities and inadequacies [6,8,9], as well as insufficient inventory management in the face of information shortages and incompetent distribution systems [9,10]. A lack of data capturing and data sharing is, however, attributed as one of the predominant obstacles toward pharmaceutical supply chain improvement in developing nations [6].

Developing countries may not have access to the resources required to implement proper information technology systems in their pharmaceutical supply chains, but the irrefutable advantages of information sharing are plain to see. Sharing supply chain information, such as demand forecasts and inventory levels, across an entire supply network allows organisations to proactively plan for disruptive events, instead of reacting (belatedly) to these events [11]. As a result, supply chains are able to better balance supply and demand, improve stakeholder accountability and ameliorate overall supply chain performance at a reduced cost [6,12,13]. It may be argued that information sharing is a suitable starting point for supply chain reform, because it allows organisations to collaborate to their mutual benefit.

Initiatives utilising the benefits of information sharing in pharmaceutical supply chains have successfully been introduced in some African countries in recent years. A study conducted in 2011 disclosed that at least 60% of stock-outs in the Senegalese contraceptive supply chain occurred at warehouses and health care facilities, despite stock availability at a national level. These problems sprouted from dismal inventory management and poor distribution practices. Upon the implementation of a new system according to which dedicated logisticians actively utilise stock data to manage inventory and curb stock-outs, these stock-outs declined to less than 2% across 140 health care facilities during the first six months [14].

The *SMS for Life* programme, established in 2009, is a web-based reporting system that allows health facility workers to report stock levels on a weekly basis by means of simple SMS messages. This practice of stock level reporting has subsequently alleviated the stock-out predicaments in Kenya and Tanzania and the system is geared for roll-out in more African countries [15,16].

The value of mobile technology in respect of information sharing in pharmaceutical supply chains is also underlined in South Africa's *Stock Visibility Solution* (SVS) programme. The SVS is a mobile phone-based reporting system that allows dispensing clinic staff to report stock levels at regular intervals [17]. The periodic capturing of stock level data allows health care facilities to purposefully manage inventory in a drive to thwart stock-outs.

This paper reports on work in progress that is aimed at, amongst others, the utilisation of information sharing in pharmaceutical supply chains with a view to enhance decision-making.

2. PROBLEM DESCRIPTION AND RESEARCH METHODOLOGY

The problem considered in this research involves the performance of conventional pharmaceutical supply chains in developing countries and how these may be improved by the adoption of demand-driven supply chain management principles. In particular, the practice of supply chain information sharing is investigated in order to establish its potential value in respect of effective inventory management. An agent-based simulation model concept demonstrator is developed for use as a test bed to evaluate the efficacy of various inventory replenishment policies within a pharmaceutical supply chain context. The simulation model also accommodates the possibility of modelling user-specified demand scenarios in order to investigate their influence on the effectiveness of inventory replenishment regimes.

The concept demonstrator embraces two modelling paradigms. The first is a *descriptive* paradigm where the model is employed to evaluate the effectiveness of a pre-specified, traditional inventory management policy explicitly embedded in the model. The second paradigm, on the other hand, follows a *prescriptive* approach. According to this paradigm, the user does not select a pre-defined policy as in the case of the descriptive paradigm. Instead, the simulation model is employed to *discover* effective inventory management protocols for the simulated pharmaceutical supply chain network. In other words, an effective inventory management policy is *prescribed* to the user.

The execution of research toward this paper is segmented into three distinct stages. The first stage comprises a brief review of the academic literature relevant to this research project. Thereafter a conceptual framework for capturing the structure of a pharmaceutical supply chain network in a format suitable for use in a simulation modelling environment is established. Finally, a hypothetical example of applying the proposed simulation model is proffered in the third and final stage.

3. LITERATURE REVIEW

The literature review in this section consists of four disparate parts, namely a review of the notion of demand-driven supply chain management (in §3.1), a review of the various basic concepts in inventory management (in §3.2), a brief overview of the concepts of self-organisation and emergence (in §3.3) and finally, a review of the machine learning paradigm of reinforcement learning (in §3.4).

3.1 Demand-driven supply chain management

A common denominator in the traditional management of supply chains is an emphasis placed on the activities involved with the downstream movement of commodities along a supply chain [18]. Organisations will, for example, streamline their production processes and distribution operations in order to improve the efficiency with which goods are moved downstream in a supply chain. Despite acknowledging the importance of these downstream management activities, advocates of the so-called *demand chain management* (DCM) notion suggest that the focus of this traditional approach is misplaced. DCM is a relatively new concept supporting the notion that end user demand should drive the upstream processes (such as manufacturing and distribution) in a supply chain [18]. As such, the end user is considered as the starting point in a supply chain as opposed to being viewed as the final destination. This particular school of thought arises from the idea that a supply chain ultimately serves to fulfil the needs of the end user. Although products flow downstream toward the end user in a supply chain, it is the end user's demand that should govern the nature of the upstream activities.

Fisher [19] proposed that any supply chain performs two distinct functions. The first is the *physical function* which embodies the physical transformation of raw materials to finished products, and the movement of these goods along a supply chain. The physical function determines a supply chain's efficiency. Manufacturing, delivery and inventory storage outlays are classified as incurring physical costs since they are part of the physical function. The second function is the *market mediation* function and its purpose is to ensure that customer demand is successfully satisfied. Market mediation costs are incurred when supply exceeds demand, or the other way around. In the case of oversupply, excessive stocks may be sold at a loss or even discarded in the case of perishables. Undersupply of stock, on the other hand, reflects lost sales opportunities. In other words, the market mediation function embodies the idea that neither a surplus nor a shortage of stock is desirable in a supply chain.

Fisher furthermore suggested that organisations may prioritise one function at the expense of the other. An organisation subject to predictable demand can, for example, deliberately plan to avoid both a surplus of stock as well as a shortage of stock. Such a position enables a firm to devote its attention to enhancing supply chain efficiency because the market mediation costs are not considered as significant. Organisations faced with unpredictable demand, on the other hand, typically prioritise market mediation costs over physical costs, because they prioritise customer satisfaction irrespective of their attained level of supply chain efficiency.

De Treville *et al*. [20] subsequently defined a *demand chain* as a supply chain in which the market mediation function predominates a supply chain's function to optimise its physical efficiency (the physical function). The adoption of a demand chain approach may seem suitable for a pharmaceutical supply chain because the need to successfully fulfil patient demand is of paramount importance. Pharmaceutical supply chains are, in fact, compelled to pursue a patient service level of 100% because failure to do so would signify the occurrence of stock-outs [21]. Organisations in pharmaceutical supply chains may, however, pursue conflicting objectives. Consider a primary health care facility, such as a clinic, which seeks to minimise medicine stock-outs so as to fully satisfy patient demand. A drug manufacturer upstream, on the other hand, may solely pursue profit maximisation with little regard for the downstream clinic's service level target. This example illustrates that a progression from a conventional pharmaceutical supply chain to a *pharmaceutical demand chain* which prioritises market mediation may not be as simple as it would seem at first.

The practice of information sharing is a powerful enabler of demand-driven supply chain management because it allows organisations to better understand customer demand and to collaborate effectively. The concept of information sharing, also called *supply chain visibility*, refers to the degree according to which supply chain organisations share information that is pivotal to their own activities and which they consider to be of mutual benefit to themselves and other firms in the supply chain [22]. Inventory levels, demand forecasts, order tracking and sales data are examples of information shared in supply chains in order to enhance their collective performance [23]. In the case of a sudden disease outbreak, for example, patient demand for a particular drug may increase considerably over a short period of time. If health care facilities do not carry enough stock to fulfil this increased demand, they set off a reverberating chain of belated, large orders along the supply chain. If the rapid demand increase is, however, made known to upstream facilities promptly through information sharing, they can increase their operations accordingly in anticipation of larger orders.

3.2 Inventory replenishment

A significant trade-off faced by inventory managers during their decision-making processes is the trade-off between supply chain responsiveness and efficiency [24]. Carrying large inventories and shortening lead times generally make a supply chain more responsive. The increased responsiveness is, however, traded for significant inventory holding costs and large transport costs, respectively [24].

Inventory replenishment policies are typically employed by inventory managers to determine reorder points and reorder quantities. Simchi-Levi *et al.* [25] identified six supply chain variables that play a role in the formulation of an inventory replenishment policy. *Customer demand* is arguably the most significant factor because organisations ultimately strive to fulfil their customers' demand. Secondly, *ordering costs* and inventory *holding costs* are of obvious financial importance. And to ensure the timely receipt of ordered goods, the reorder point should be informed by the *replenishment lead time*, which may not be deterministic. Furthermore, the order quantity may be based on the *current inventory level* of the product in question. Additionally, the *length* of the *planning period* shapes the scope and the nature of inventory management decisions. Finally, the *service level* target may be a determinant of the reorder point and the reorder quantity.

The dynamic nature of supply chains suggests that the parameters of an inventory replenishment policy should be informed by the current state of the supply chain environment with a view to making better decisions. In other words, inventory replenishment protocols should not be too rigid, for otherwise they may fail in the face of changes in the supply chain environment. Owing to the large degree of variability and uncertainty in a supply chain, the inventory management process remains an intricately complex task.

3.3 Self-organisation and emergence

The concepts of *self-organisation* and *emergence* are reviewed in order to explore their potential application to inventory management protocols in pharmaceutical supply chains.

De Wolf and Holvoet [26] describe self-organisation as a continuous process in which coordinated organisation manifests itself through the independent behaviour of systems, without any control instructions being imposed from outside the system. 'Organisation' here refers to the presence of a so-called 'structure' that can be of a spatial, temporal or functional nature. Although a self-organising process is void of external control, it does not preclude data inputs from outside the system. A fundamental property of self-organising systems is that they are considered extremely robust and adaptable because they can reproduce 'organisation' in the face of environmental changes [26,27].

The presence of self-organisation may give rise to the related phenomenon of *emergence*. Emergence materialises in a system when the local interactions between its individual constituents culminate at a higher level in the development of a structure (called 'coherent emergents') that is not explicitly represented at a lower level [26,28]. An example of self-organisation and emergence in nature is illustrated in Figure 1. When a colony of ants arrive at a gap in their path, they often use their bodies to build a living bridge without any external supervision or instructions. Each ant follows two simple rules. First, it slows down as it reaches the gap and secondly, it freezes when it feels another ant walking over it. The ants continue in this fashion until they have successfully bridged the gap. Through the ants' self-organising behaviour, a living bridge *emerges*. The bridge may be classified as an *emergent* because no individual ant is representative of the bridge. The bridge is only formed at a higher level through the local interactions between the ants at a lower level.

Figure 1: A living bridge emerges from the self-organising behaviour of ants [29,30].

It may be argued that effective, externally coordinated inventory management in a pharmaceutical supply chain is extremely difficult, or even impossible, given the myriad of supply chain variables that influence inventory management decisions. Self-organisation (a process void of external control) is therefore explored as an alternative means of coordinating inventory management. A self-organising supply chain, by implication, is void of any form of centralised control and each facility manages its own inventory exclusively. In this research project, we investigate the conjecture that local coordination between facilities may lead to the *emergence* of a greater structure where the global supply chain functions as a coordinated system in respect of inventory management.

Consider a simple example of a self-organising pharmaceutical supply chain in which each facility in the chain 'organises' itself with a view to prevent stock-outs locally. These facilities, in other words, are autonomous and actively manage their own inventory in pursuit of an 'organisation' in which stock-outs are prevented. Additionally, in an information sharing supply chain, these facilities may utilise the available information to inform their inventory management decisions accordingly. There is, however, no explicit coordination between the facilities in the supply chain. If a storage depot is, for example, perturbed by a drastic demand increase, the facility may 'reorganise' itself by increasing its order quantities. Emergence may subsequently occur in the supply chain as a set of management policies prescribing reorder points and reorder quantities in pursuit of effective inventory management.

3.4 Reinforcement learning

Reinforcement learning is a branch of machine learning where a learning agent learns behaviour in an environment through interaction with the environment [31]. The premise of reinforcement learning rests on the idea that if a particular action yields desirable results, the inclination to repeat the same action is *reinforced* [32]. This closely relates to the learning process followed by humans and animals. A new-born elephant, for example, tries many strategies and fails often before it can stand upright. Over time, the baby elephant learns to avoid the actions that caused it to fall down and it hones the skills that proved more fruitful in pursuit of its goal to stand upright.

A fundamental characteristic of reinforcement learning is that a learning agent can evaluate the desirability of its actions according to a numerical reward signal, but it is not told which actions to take in order to improve its performance [33,34]. The reward signal is expressed in terms of a pre-specified goal that is pursed by the agent. Hence, a learning agent has to attempt many different strategies by itself in order to learn what behaviour maximises its reward signal. This learning process can informally be described as learning through *trial-and-error*. The new-born elephant, for example, is said to learn through trial-and-error which strategies prove to be more successful.

The reinforcement learning approach is commonly described in terms of an *agent* and an *environment* [33,34]. The *agent* is the learning actor which interacts with its *environment* in order to learn about the environment. A *state* describes the situation in the environment at any given time instant. At discrete time steps, the agent is presented with an array of actions from which it can choose. The selected action influences the environment and the environment provides feedback to the agent in the form of a reward signal and by transitioning into a new state. The reward signal is employed to evaluate the immediate reward received for the selected action. Notably, the agent's objective in reinforcement learning is to maximise its cumulative reward and, occasionally, these rewards may be significantly delayed [33,34]. The reinforcement learning cycle repeats itself many times and, over time, the agent learns to map different situations to particular actions that yield desirable results. A schematic of this learning paradigm is shown in Figure 2. The outcome of a reinforcement learning process may be described as a *look-up table* [33]. This table maps all possible environment states to appropriate actions that have proved to maximise the agent's reward during the learning process.

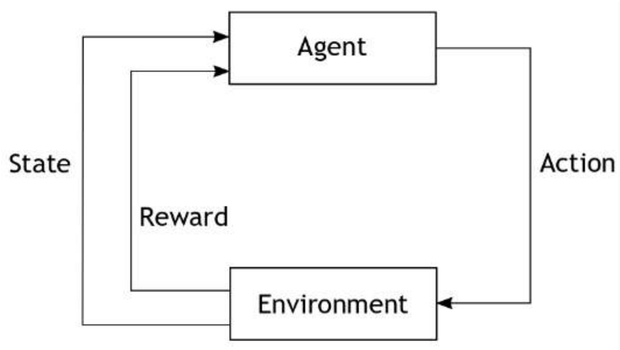


Figure 2: The reinforcement learning cycle (adapted from [33]).

Reinforcement learning forms the cornerstone of the prescriptive paradigm of the proposed concept demonstrator, as discussed in §1. Each facility type in a pharmaceutical supply chain (*i.e.* manufacturer, warehouse, clinic, *etc.*), is trained as a reinforcement learner. A unique look-up table of state-action pairs is subsequently generated for each facility type that can be utilised by the facility to inform decision-making on a daily basis aimed at improving performance indicators aligned with various management objectives.

4. CONCEPTUAL INPUT FRAMEWORK FOR PHARMACEUTICAL SUPPLY CHAIN MODELLING

As discussed in §1, an agent-based pharmaceutical supply chain simulation model is put forward in this paper. The simulation model follows a generic design so as to enhance its flexibility and potential value for decision makers in pharmaceutical supply chains. According to this generic design, the simulation model receives a user-specified supply chain structure as input. This input structure, called an *input framework*, should sufficiently capture the level of abstraction required to model a pharmaceutical supply chain mathematically, as per the purposes of this research. This section is devoted to a description of the conceptual design of such an input framework. This framework is not, however, exhaustive, but simply serves as a point of departure for the development of a comprehensive, well-rounded input framework.

The input framework should capture facility-specific information, product-specific information, as well as inventory management parameters and product demand profiles. A list of attributes that captures the high-level, facility-specific information is shown in Table 1. This information set describes the size of the supply chain network and its prevailing facility types.

Table 1: Facility-specific information provided as input.

Attribute	Description
Facility name	Common name used to identify the facility
Location	Spatial information
Facility type	The nature of a facility's operations. For example: Manufacturer, storage facility, hospital, clinic, *etc.*
Tier	Specification of the relevant echelon in the supply chain
Storage capacity	The facility's total storage capacity for the purposes of the simulation

The simulation model's generic design allows for the inclusion of user-specified pharmaceutical products. The product-specific information required extends only to the name of the product and its shelf life, as shown in Table 2.

Table 2: Product-specific information provided as input.

Attribute	Description
Product name	Commonly used product name
Shelf life (if perishable)	Shelf life duration (from date of manufacture)

A traditional *from-to* matrix can be used to capture the connections between facilities in a supply chain. These connections indicate the flow of goods between facilities. The matrix is of size $n \times n$ where n denotes the total number of facilities in the supply chain. If product units flow *from* facility i *to* facility j, the $(i, j)^{\text{th}}$ entry in the matrix adopts a value of 1, or a value of 0 otherwise. The facility names are obtained from Table 1. An example of a from-to matrix for a supply chain comprising three facilities is shown in Table 3. Facility A, for example, distributes goods to Facilities B and C. Facility B, on the other hand, distributes only to Facility C and Facility C, in turn, does not distribute any inventory to other facilities.

Table 3: An example of a 3 × 3 from-to matrix provided as input.

	Facility A	Facility B	Facility C
Facility A	-	1	1
Facility B	0	-	1
Facility C	0	0	-

As discussed in §1, the descriptive paradigm of the concept demonstrator allows the user to evaluate the performance of pre-specified inventory replenishment policies. In order to facilitate this paradigm, the user is required to specify the parameters of these policies as part of the input framework. The relevant inventory management parameters required to model the inventory management processes are presented in Table 4. According to the table, the user can specify parameters for a continuous review policy, or for a periodic review policy. Table 4 may, of course, be extended to include more inventory management policies.

Table 4: Inventory management parameters provided as input.

Attribute	Description
Ordering facility	Facility name from Table 1
Product	Product name from Table 2
Starting inventory	Product quantities available at the start of the simulation
Minimum order quantity	If applicable
Maximum order quantity	If applicable
Reorder point	For a continuous review policy
Reorder quantity	For a continuous review policy
Review interval	For a periodic review policy
Order-up-to level	For a periodic review policy
Lead time (days)	As a function of order size, may be stochastic

In order to model the manufacturing operations of manufacturers in the supply chain, the production process characteristics have to be captured in the format as shown in Table 5. This information is applicable to manufacturers only.

Table 5: Manufacturing information provided as input.

Attribute	Description
Manufacturer	Facility name from Table 1
Product	Product name from Table 2
Starting inventory	Product quantities available at the start of the simulation
Production trigger	Signal that triggers the initiation of the production process
Production rate	Expressed in number of batches per time unit
Batch size	The number of units in a single batch

Finally, forecasted demand data and actual demand data for the simulation period can be provided as input. Users may provide either synthesised data or actual data as input. The nature of the demand data required is elucidated in Table 6.

Table 6: Demand data provided as input.

Attribute	Description
Facility	Name from Table 1
Product	Product name from Table 2
Actual demand	Daily demand for each simulated day
Forecasted demand	Daily forecasted demand for each simulated day

A potential implementation of this input framework is demonstrated by means of a hypothetical example in the following section.

5. HYPOTHETICAL EXAMPLE

The objective of this section is to integrate the salient elements of the literature review in §3 with the input framework of §4 in order to demonstrate how it may be applied in practice by means of a small hypothetical example.

Consider a simple pharmaceutical supply chain comprising a single manufacturer, a single warehouse and two clinics. The supply chain facilitates the flow of *Painstill* drugs from the manufacturer to the clinics, via the warehouse. Currently all four facilities in the supply chain employ traditional continuous review replenishment policies.

The Painstill supply chain has suffered from large-scale stock-outs in recent months and it has been decided to investigate avenues for improving its inventory management practices. In particular, the value of supply chain information sharing is of interest and how it may inform effective inventory management in the supply chain with a view to minimise stock-outs. The management team has turned to the simulation model proposed in this paper to support their decision-making processes. After populating the input framework of §4, the management team decides to employ both the descriptive and prescriptive modelling paradigms.

Descriptive paradigm

According to the descriptive paradigm, a pre-defined replenishment policy is selected for each facility from a list of possible policies. The management team decides to continue with a continuous review policy for each facility. Using the simulation model, the management team can now experiment with different parameter values (reorder points and reorder quantities) for each facility in order to determine how they may improve the effectiveness of the continuous review policies. The operation of the supply chain is now simulated according to the specified parameters. The movement of Painstill units through the supply chain, and charts denoting information such as inventory levels may be displayed during the simulation model execution. At the end of the simulation run, a set of key performance indicator values are provided as output. Examples of suitable key performance indicators may include attained service levels, the number of stock-outs per facility, the average stock-out duration per facility, as well as the procurement costs and inventory holding costs incurred by each facility. An example of a graphic denoting a facility's stock level data and demand data are shown in Figure 3. Suppose that the sudden demand increase observed at day 111 is attributed to an unexpected disease outbreak. It is evident from the stock level graph that the facility carried enough stock to fulfil the increased demand initially. The stock level has, however, declined to a minimum of 140 units on day 122. The relevant decision-makers may therefore infer that it is best to increase the facility's order quantities in the face of a similar demand increase in order to negate the possibility of stock-outs.

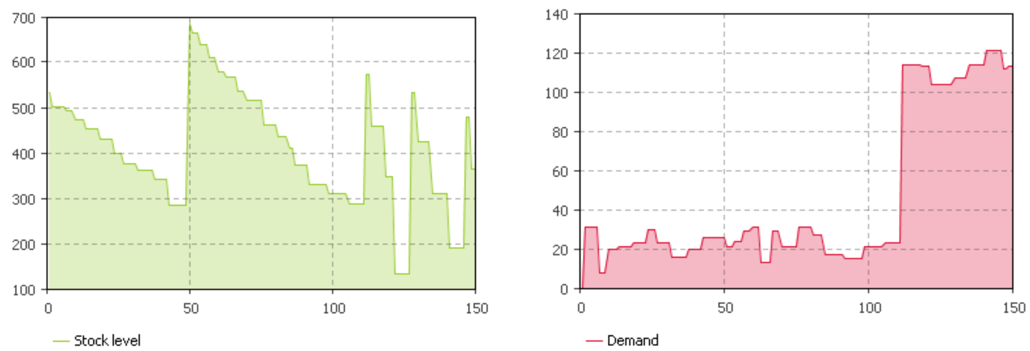


Figure 3: A graph of a facility's stock level over time (left) and the same facility's corresponding demand over time (right). A sudden demand increase is observed at day 111 due to a sudden disease outbreak.

Prescriptive paradigm

The prescriptive paradigm integrates the concepts of demand-driven supply chain management, self-organisation and reinforcement learning as a means to prescribe effective replenishment policies for the modelled supply chain network. The output of this paradigm is therefore a set of key performance indicator values, as well as a set of policies prescribing reorder points and reorder quantities for each facility. 'Self-organisation' in this context means that each facility makes its own inventory decisions, as alluded to in §3.3. The prescriptive paradigm lends itself to a large degree of scalability because of the supply chain's self-organising property. The effects of the local interactions between facilities may ripple outward until a form of organisation is achieved and maintained across the entire supply network. The size of the supply chain therefore has little influence on the model complexity.

The user should explicitly define the level and degree of information that may be shared and used by other facilities in order to facilitate their decision-making processes. The management team can, for example, explore the effect of sharing both clinics' stock level data with the warehouse. This level of visibility may presumably allow the warehouse to increase its inventory proactively should the clinics' stock levels start to decline rapidly in response to a sudden demand increase. The management team can, however, also investigate the value of sharing the clinics' stock level data with both the warehouse and the manufacturer, provided that the manufacturer also has visibility over the warehouse's stock level data. Intuitively, it may be argued that the increased level of supply chain visibility may be accompanied by improved overall supply chain performance. Comparing these two scenarios at the hand of the simulation model may elucidate whether the larger investment in information technology required for the second scenario is, in fact, justified in respect of the simulation results. It may, for example, be that the increased level of information sharing of the second scenario does not significantly improve on the effectiveness delivered in the first scenario. The simulation model may prove extremely useful for similar comparison analyses.

Once the information sharing structure has been configured in the model, reinforcement learning may be applied as a mechanism to discover self-organising replenishment heuristics. Since facilities of the same type share common operational characteristics, all the instances of a particular type of facility can use the same look-up table. Therefore, only one agent can represent each facility type, be trained and only generate one look-up table per agent. For the Painstill supply chain, a manufacturer agent, a warehouse agent and a clinic agent have to be trained as reinforcement learners, respectively. Notably, both clinics use the single look-up table generated by the clinic agent. The learning process is expected to be computationally expensive, but this is an offline process which is executed *a priori*.

The respective reward functions of these agents may be specified by the user. A typical reward function should reward desirable actions, such as the successful fulfilment of customer demand. For undesirable scenarios, such as the occurrence of stock-outs, product expirations or large inventory carrying costs, agent punishment should result in the form of a negative reward signal. Notably, the reward function need not be particular to one agent, but rewards may be shared amongst agents. Rewards may be shared between a warehouse and a clinic (the clinic orders from this warehouse), for example, in an attempt to enhance their collective performance.

A *state* should describe all the information that is visible to a learning agent at any given time instant during the learning process. The state may include a number of dimensions, such as the agent's current inventory level and its forecasted demand. If, for example, the Painstill manufacturer has visibility over the warehouse's stock level, the manufacturer agent's state space would include the warehouse's stock level data. In other words, the size of an agent's state should be informed by the level and degree of information sharing. At discrete time steps, each agent should decide whether to place a new order for Painstill drugs, or not. If the agent decides to order, the order quantity should also be selected. The simulation model should replicate many possible scenarios in the supply chain and the reinforcement learning algorithm should experiment with different order strategies during the process. Over time, each agent learns which inventory management decisions (actions) yield the most reward within a particular situation (state), and these are documented in a look-up table. In other words, each agent learns when to place a new order and how much to order, given a certain state.

The management team may use the results of both the descriptive and prescriptive paradigms to provide them with decision support in pursuit of their drive to improve the efficiency of the Painstill supply chain. The simulation results may, for example, provide insight as to which facilities suffer the most from stock-outs and why under different demand scenarios. The management team can also identify those facilities at which product expirations occur most frequently and subsequently learn which replenishment policies may prevent them. Additionally, cost-related key performance indicators may provide an indication of the financial feasibility of a particular policy. Finally, the prescriptive paradigm may elucidate which type of information should be shared, and with whom, for the best outcome in respect of the management team's various objectives.

6. CONCLUSION

The objective of this paper was fourfold. First, to establish the background context of this research project as the inventory management methodologies of demand-driven pharmaceutical supply chains. Secondly, to provide a brief overview of the relevant academic literature which serves as a basis for the work conducted in this research. Thereafter, a preliminary, conceptual input framework for pharmaceutical supply chain modelling was developed. Finally, the potential use of the planned proposed concept demonstrator was illustrated by means of a hypothetical example. It is important to stress that the work described in this paper is not concluded. This paper serves as a prelude to a larger research project in which the value of self-organisation and information sharing in the pharmaceutical supply chains of developing countries is explored and quantified. It is acknowledged that these concepts may not be readily compatible with existing supply chain infrastructure and resources in these countries. This research, however, aims to elucidate whether self-organisation is a suitable instrument for pharmaceutical supply chain reform.

REFERENCES

[1] **World Health Organization**. 2014. *Global update on the health sector response to HIV, 2014*, World Health Organization.

[2] **Doctors Without Borders (MSF), RuDASA, RHAP, TAC, SECTION27 and SAHIVSoc**. 2016. *2015 Stock outs National Survey: Third annual report – South Africa*, Stop Stock Outs Project.

[3] **Harries, A.D., Schouten, E.J., Makombe, S.D., Libamba, E., Neufville, H.N., Some, E., Kadewere, G. and Lungu, D**. 2007. Ensuring uninterrupted supplies of antiretroviral drugs in resource-poor settings: An example from Malawi, *Bulletin of the World Health Organization*, 85(2), pp. 152-155.

[4] **Nicholson, A., English, R.A., Guenther, R.S. and Claiborne, A.B. 2013**. *Developing and strengthening the global supply chain for second-line drugs for multidrug-resistant tuberculosis*, The National Academic Press, Washington (DC).

[5] **Hodes, R., Price, I., Bungane, N., Toska, E. and Cluver, L.** 2017. How front-line healthcare workers respond to stock-outs of essential medicines in the Eastern Cape Province of South Africa, *South African Medical Journal*, 107(9), pp. 738-740.

[6] **Yadav, P**. 2015. Health product supply chains in developing countries: Diagnosis of the root causes of underperformance and an agenda for reform, *Health Systems & Reform*, 1(2), pp. 142–154.
[7] **Humphreys, G**. 2011. Vaccination: Rattling the supply chain, *Bulletin of the World Health Organization*, 89(5), pp. 324–325.
[8] **Cameron, A., Ewen, M., Ross-Degnan, D., Ball, D. and Laing, R.** 2009. Medicine prices, availability, and affordability in 36 developing and middle-income countries: A secondary analysis, *The Lancet*, 373(9659), pp. 240–249.
[9] **Kangwana, B.B., Njogu, J., Wasunna, B., Kedenge, S.V., Memusi, D.N., Goodman, C.A., Zurovac, D. and Snow, R.W**. 2009. Malaria drug shortages in Kenya: A major failure to provide access to effective treatment, *American Journal of Tropical Medicine and Hygiene*, 80(5), pp. 737–738.
[10] **Bateman, C.** 2013. Drug stock-outs: Inept supply-chain management and corruption, *South African Medical Journal*, 103(9), pp. 600–602.
[11] **Griffin, P.M., Nembhard, H.B., DeFlitch, C.J., Bastian, N.D., Kang, H. and Munoz, D.A.** 2016. *Healthcare systems engineering*, John Wiley & Sons, Hoboken (NJ).
[12] **Frohlich, M.T. and Westbrook, R.** 2002. Demand chain management in manufacturing and services: Web-based integration, drivers and performance, *Journal of Operations Management*, 20(6), pp. 729–745.
[13] **Ripin, D.J., Jamieson, D., Meyers, A., Warty, U., Dain, M. and Khamsi, C**. 2014. Antiretroviral procurement and supply chain management, *Antiviral Therapy*, 19(3), pp. 79–89.
[14] **Daff, B.M., Seck, C., Belkhayat, H. and Sutton, P**. 2014. Informed push distribution of contraceptives in Senegal reduces stockouts and improves quality of family planning services, *Global Health: Science and Practice*, 2(2), pp. 245–252.
[15] **Barrington, J., Wereko-Brobby, O., Ward, P., Mwafongo, W. and Kungulwe, S**. 2010. SMS for Life: A pilot project to improve anti-malarial drug supply management in rural Tanzania using standard technology, *Malaria Journal*, 9(1), pp. 1–9.
[16] **Githinji, S., Kigen, S., Memusi, D., et al**. 2013. Reducing stock-outs of life saving malaria commodities using mobile phone text-messaging: SMS for Life study in Kenya, *PLOS One*, 8(1), pp. 1–8.
[17] **Mezzanine**. 2018. *Stock Visibility Solution*, [Online], [Cited June 2018], Available from: https://www.mezzanineware.com/svs/.
[18] **Langabeer, J.R. and Rose, J**. 2002. *Creating demand driven supply chains: How to profit from demand chain management*, Spiro Press, London.
[19] **Fisher, M**. 1997. What is the right supply chain for your product?, *Harvard Business* Review, 75(2), pp. 105–116.
[20] **De Treville, S., Shapiro, R.D. and Hameri, A.P**. 2004. From supply chain to demand chain: The role of lead time reduction in improving demand chain performance, *Journal of Operations Management*, 21(6), pp. 613–627.
[21] **Uthayakumar, R. and Priyan, S**. 2013. Pharmaceutical supply chain and inventory management strategies: Optimization for a pharmaceutical company and a hospital, *Operations Research for Health Care*, 2(3), pp. 52–64.
[22] **Barratt, M. and Oke, A**. 2007. Antecedents of supply chain visibility in retail supply chains: A resource-based theory perspective, *Journal of Operations Management*, 25(6), pp. 1217–1233.
[23] **Lee, H.L. and Whang, S**. 2000. Information sharing in a supply chain, *International Journal of Manufacturing Technology and Management*, 1(1), pp. 79–93.
[24] **Chopra, S. and Meindl P**. 2013. *Supply chain management: Strategy, planning and operation*, 5th Edition, Pearson Education Limited, Essex.
[25] **Simchi-Levi, D., Simchi-Levi, E. and Kaminsky P**. 2000. *Designing and managing the supply chain: Concepts, strategies, and cases*, McGraw-Hill, New York (NY).
[26] **De Wolf, T. and Holvoet, T.** 2005. *Emergence versus self-organisation: Different concepts but promising when combined*, pp. 1-15 in **Brueckner, S.A., Serugendo, G.D.M., Karageorgos, A. and Nagpal, R. (Eds).** *Engineering self-organising systems: Methodologies and applications*, Springer, Berlin.
[27] **Heylighen, F**. 2001. The science of self-organization and adaptativity, *Encyclopedia of Life Support Systems*, 5(3), pp. 253-280.
[28] **Odell, J**. 2002. Agents and complex systems, *Journal of Object Technology*, 1(2), pp. 35–45.
[29] **Hartnett, K**. 2018. *The simple algorithm that ants use to build bridges,* [Online], [Cited June 2018], Available from: https://www.quantamagazine.org/the-simple-algorithm-that-ants-use-to-build-bridges-20180226/.
[30] **Manohar, V**. 2018. *Unity is strength*, [Online], [Cited June 2018], Available from: https://www.shutterstock.com/image-photo/unity-strength1011406435?src=BbM2Y7v-2FfNGqgiTyr5TKw-1-74.
[31] **Kaelbling, L.P., Littman, M.L. and Moore, A.W**. 1996. Reinforcement learning: A survey, *Journal of Artificial Intelligence Research*, 4, pp. 237–285.
[32] **Sutton, R.S., Barto, A.G. and Williams, R.J**. 1992. Reinforcement learning is direct adaptive optimal control, *IEEE Control Syst*ems, 12(2), pp. 19–22.
[33] **Marsland, S**. 2009. *Machine learning: An algorithmic perspective*, CRC Press, Boca Raton (FL).
[34] **Sutton, R.S. and Barto A.G**. 1998. *Reinforcement learning: An introduction*, The MIT Press, Cambridge (MA).

TOWARDS A MORE EFFICIENT AND EFFECTIVE PIPELINE OF TUBERCULOSIS MEDICATION: THE VALUE OF IDENTIFYING TRENDS AND INFLUENCING FACTORS

N. Hanekom[1*], L. Bam[2] & I.H. de Kock[3]

[123]Health Systems Engineering and Innovation Hub
Department of Industrial Engineering
Stellenbosch University
[1]18293700@sun.ac.za, [2]louzanne@sun.ac.za, [3]imkedk@sun.ac.za

ABSTRACT

Tuberculosis poses a significant risk to global health with estimated 1.7 million deaths worldwide in 2016. One key issue in tuberculosis management relates to the drug pipeline, with drug development not keeping pace with the rate at which the disease expands and changes. Identifying and addressing factors that inhibit tuberculosis research and development is essential. Research to identify trends in the drug pipeline and evaluate the relations between these trends and other influencing factors will strengthen the existing body of knowledge, enabling improved decision-making on investment in drug research and development, and structuring incentives to encourage investment.

[1] The author was enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

Tuberculosis (TB) is the ninth biggest cause of death worldwide, affecting the lives of over 10,4 million people in 2016 [1]. An estimated four million people (40% of TB cases) are undiagnosed each year; this figure increases to 77% among individuals with drug-resistant TB (DR-TB) [2]. In most developing countries, diagnosed patients face drawn-out delays (up to 28–30 days) in receiving feedback confirming their TB status, thus prolonging the time during which they are without appropriate treatment or until they undergo the necessary drug-susceptibility testing (DST) [1]. Diagnosing TB is the first step in treating affected people and preventing the transmission of the disease [2]. According to the World Health Organisation, infection rates will only decrease if, firstly, the country and local-level uptake of the diagnostic tools available increases significantly, and, secondly, investment in basic scientific research and the diagnostic pipeline is increased considerably [2].

DR-TB is a growing threat: in 2016, 600 000 new cases of resistance to rifampicin (the most effective first-line TB drug) were diagnosed, and 490 000 of these cases were diagnosed as multidrug-resistant TB (MDR-TB) [1]. The 2017 treatment outcome data show a global success rate of 83% for TB treatment and 54% for DR-TB [1]. DR-TB treatment has a noticeably lower success rate than that of TB. The need for further advancement in the development of more effective, safer, affordable, accessible, shorter course and better tolerated treatment regimens is evident, and is being investigated by stakeholders and researchers as this is of paramount importance to reduce the global prevalence of this disease [3].

Despite research suggesting that the research and development (R&D) of TB drugs over the past few decades has not yielded any major breakthroughs, there has been noticeable progress in the TB treatment drug pipeline recently. Three drugs are currently in phase 3[4] of clinical development trials and are being distributed on a small scale to DR-TB patients [1]. Unfortunately, access to these drugs is restricted in developing countries, where the disease is most prevalent [3].

Dowdy [23] describes TB as an “epidemic at a crossroads”. According to Dowdy, “strains of DR-TB will emerge that are more transmissible and more difficult to treat”, yet there are more drugs and technologies to use against the disease than ever before [23]. There is evidence that DR-TB can effectively be reversed, unlike most other DR pathogens. Thus, the next decade can either be one in which the TB epidemic becomes an “unprecedented global epidemic” or one in which the “global burden will be unprecedentedly reversed” [23]. According to Dowdy, the difference between these two outcomes lies with the global TB control community, and whether there is enough political will to prioritise the mitigation of the disease [23].

The *end TB strategy* is related to the Sustainable Development Goals and aims to “eliminate TB by 2030” [4]. In order to achieve the global goal of ending TB, it is estimated that an annual amount of $2 billion should be invested in TB R&D globally. Currently, approximately $650 million is spent on global TB R&D per year, which indicates that there is a significant investment gap for the development of TB drugs [4]. According to the 2016 report on TB research funding, the lack of funding for TB is now a human rights issue in need of a political solution [2] & [4]. Lucica Ditiu, executive director of the Stop TB Partnership, states that the lack of funding for TB R&D is impacting everything related to the development of new drugs and technologies, with specific reference to: i) the rate at which new technologies become available to the market; ii) the state of the TB drug pipeline; iii) advocacy efforts; and iv) the possibility of reaching targets in global plans to end TB [4].

There is consensus in literature that investment in TB R&D is needed to increase the number of drugs that progress through the drug pipeline effectively [1] & [2]. Literature does not specify, however: i) why the private sector shows less interest in investing in TB compared to other diseases; ii) why TB drugs progress through the pipeline at a slower rate than drugs for other diseases; iii) the trends of drug R&D; and iv) what the major factors are that affect the drug movement of the TB drug pipeline. It is therefore argued in this paper that addressing the factors that affect the TB drug pipeline can contribute towards the advancement of effective progress of TB drugs through the R&D pipeline.

This article aims to investigate the trends and factors affecting the advancement of drugs in the drug pipeline and the TB drug pipeline in particular, in order to: i) highlight the value of identifying factors causing a loss of efficiency in the pharmaceutical R&D process; ii) enable informed decision-making about advancing drug development; and iii) facilitate discussions concerned with directing investment in drug R&D to contribute towards increasingly effective and efficient output of drugs from the TB drug pipeline. The relationships between pipeline trends and factors affecting the pharmaceutical and TB drug pipeline is discussed and analysed to determine similarities or differences between the pipelines. Potential areas for future research are identified. Lastly, the value of future research in this field for the existing body of knowledge is briefly investigated to

[4] Clinical development phases (1, 2 and 3) refer to the research studies, forming part of the drug R&D process, that determine whether a drug is safe and effective for human consumption. Each phase has different outcomes. The more advanced a drug is in the clinical trial phases, the higher the possibility for the drug to be accepted by regulatory agencies.

establish whether an understanding of the pipelines might improve the effectiveness and efficiency of, specifically, the TB drug pipeline.

2. METHOD

In order to address the aim stated above, this study investigates two drug pipelines; firstly, the pharmaceutical drug pipeline (diseases non-specific) and secondly, the TB drug pipeline (disease specific). These two drug pipelines will be addressed separately and subsequently compared to establish similarities and differences. Figure 1 depicts the research inquiry process followed to investigate each pipeline in order to ultimately establish the value of identifying those factors and trends that affect the pharmaceutical pipeline.

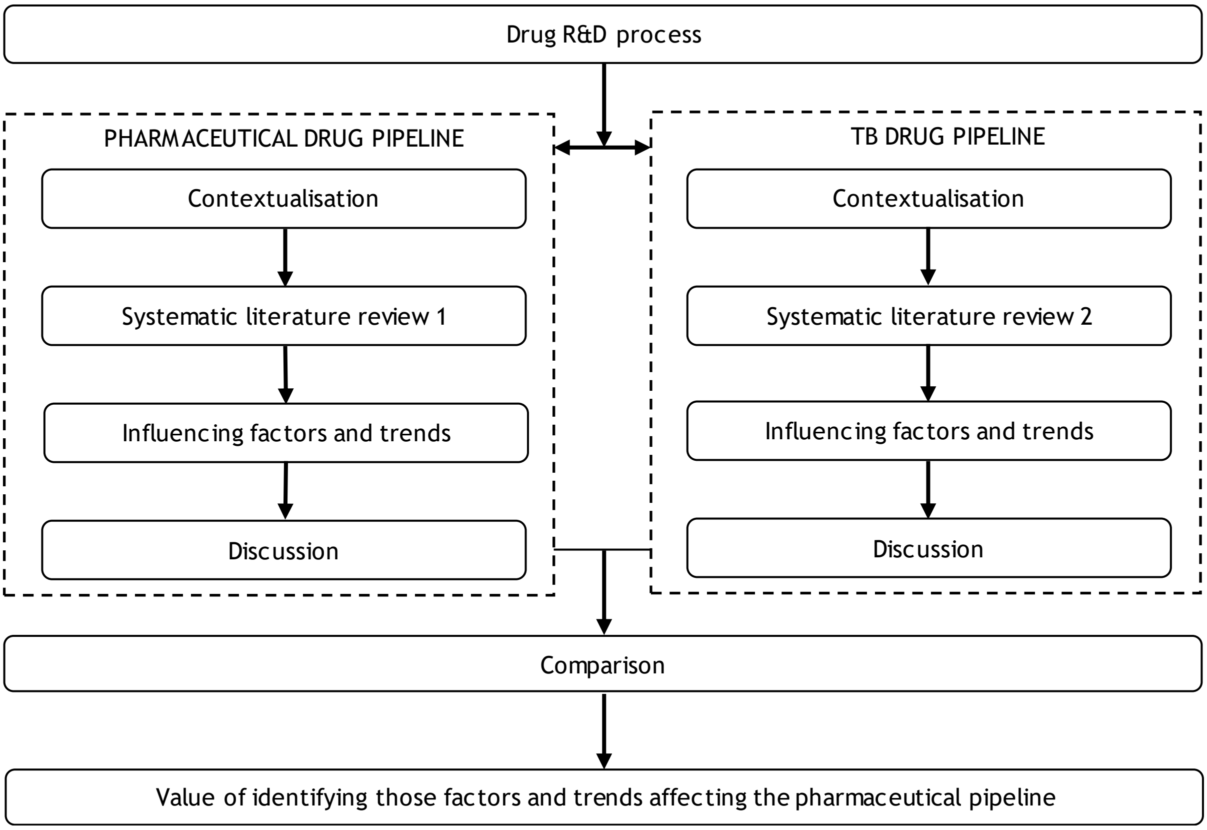


Figure 1: Research study method

As can be seen in Figure 1, the drug R&D process will be discussed first. Subsequently, both the pharmaceutical and TB drug pipelines will be discussed. Distinct systematic reviews of both pipelines (systematic review 1 and 2) will be conducted, resulting in the identification of two sets of influencing factors and trends relating to drug pipelines. A separate analysis of each pipeline and a comparison of the two pipelines will follow to establish relationships within the pipelines and similarities and differences between the two pipelines. The analysis will result in understanding the impact of various factors on the pipelines, and possible ways to improve pipeline efficiency and effectiveness, based on pipeline trends. Finally, the value of the study regarding the improvement of the TB pipeline will be presented, followed by an evaluation of how this information might be used to improve the impact on the pharmaceutical or TB drug pipeline.

3. DRUG RESEARCH AND DEVELOPMENT PROCESS

Drug R&D refers to the process followed from product discovery to the successful development of a drug. It includes all research conducted and review processes completed up to the introduction of the new drug.

The development of pharmaceutical drugs is an iterative, time consuming and costly process [5]. The process alternates between several factors, including: i) theoretical biology; ii) the appropriate use of animal assays to determine a compound's biological activity in the body; and iii) optimising the chemical compounds with medicinal chemistry [5]. The cost of R&D for each successful new drug was estimated to be $2.6 billion in 2016, as opposed to $1 billion in 2000 [6]. Costs include the cost of failures, thus the cost of all the drugs screened, tested and assessed but not necessarily approved. PhRMA [5] states that the development costs of drugs are still rising and become even higher when the cost of research after drug approval is considered. Drug development is a lengthy process, lasting up to fifteen years from initial discovery to product launch. The duration of the drug development process can be influenced by various factors, including the testing and analysis of the drug for safety and efficacy.

Drugs require approval from a recognised pharmaceutical regulatory agency, authorising the drug to be launched provided that it adheres to the international guidelines and standards set out for drugs [6]. One of the most well-known regulatory agencies is the Food and Drug Administration (FDA), the federal agency of the United States Department of Health and Human Services responsible for ensuring that organisations in the US, and a number of other countries, adhere to regulatory frameworks [7]. When a drug is FDA approved, it means the potential risk of the item is outweighed by its benefits, thus making it, legally, safe to use [7].

To discover and develop a new drug, researchers must understand the basic causes of a disease in terms of proteins, genes and cells [8]. The information emerging from the disease-cause analysis is known as 'targets' and identifies the potential factors that can be affected by drugs to diagnose, prevent or treat a disease [8]. The validation of the identified targets, discovering the right molecule to interact with the target, and testing for safety and efficacy are only a few of the tasks to be completed [8]. The drug development process occurs in five distinct phases, namely: i) drug discovery; ii) the preclinical phase; iii) clinical trials; iv) the review phase; and v) post-marketing surveillance. Each stage contributes to fine-tuning the developing drug so that it is in the best possible state for the target disease. Figure 2 illustrates the phases in drug R&D, the stipulated number of compounds (drug candidates) per phase, as well as the average duration of each phase.

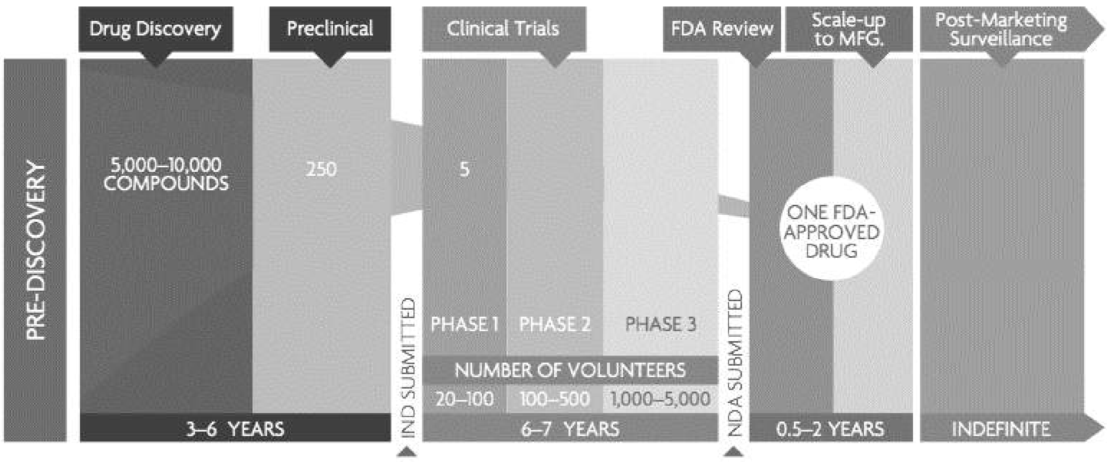


Figure 2: Drug R&D process ***[8]***

The primary aims of the review process in drug R&D are to ensure that: i) drugs are safe for human consumption; ii) drugs are effective in treating the disease targeted; iii) drugs are affordable for users; and iv) the benefit of the new drug outweighs the potential risk [8]. The benefit versus risk ratio is determined by the FDA, or another regulatory agency, scrutinising the data collected in the preclinical and clinical findings. The drugs need to meet the safety and efficacy standards set by regulatory bodies. Currently only 12% of candidate drugs (drugs in the R&D process) receive FDA approval [9].

Although scientific advances enable a greater understanding of diseases at molecular level, it is evident that scientific, technical and regulatory challenges still exist in the R&D process [5]. According to PhRMA [5], an in-depth analysis of the R&D process could clarify why the successful development of drugs takes so long.

4. PHARMACEUTICAL DRUG PIPELINE

This section focuses on the pharmaceutical sector. Firstly, it discusses the context of the pharmaceutical pipeline. Then a systematic literature review is conducted, and factors and trends that influence the pipeline are identified. Finally, any similarities or differences between the factors and trends are analysed and evaluated.

4.1 Contextualisation: The pharmaceutical drug pipeline

The drug pipeline refers to the set of drugs that a pharmaceutical company, or the entire pharmaceutical industry, in the process of discovering, research or development at a given point in time [10]. The drug pipeline encompasses the amount of active R&D taking place, thus serving as a form of reference to the extent of interest, investment and resource allocation in a specific drug or disease [11]. In the pharmaceutical industry, the drug pipeline includes all the processes from initial drug discovery to the introduction of the product for public consumption [10].

The drug pipeline does not end when the drug development process has been completed and the drug approved for launch; ongoing research and data collection form part of post-approval studies [9]. These studies are conducted for as long as the product is used by patients and include the examination of the drug and its effects on drug users; these insights can also be used to expand treatment options in future drug development [9].

The pharmaceutical pipeline is under tremendous pressure when the significant number of events, processes, stakeholders, circumstances and regulations influencing the outcome are considered. Great advances in science, technology and management practices in drug development have been made over the past 60 years; yet the number of new drugs approved per billion US dollars spent on drug development has decreased about 80-fold [12]. The time and cost challenges are well known for their impact on the drug industry. Another contributor to the loss of efficiency in the drug pipeline is the 'curse of attrition' [13]. This refers to the considerable number of drugs being rejected in clinical trial phases, as the drug progresses through the compulsory trials and processes [13]. The low success rate of compound development is further impaired by the amount of funds lost once a drug is rejected at such an advanced stage of development [13].

The pharmaceutical industry strives to decrease the number of drug compounds exiting the R&D system without being approved, thereby minimising lost investment costs, research effort and time. An ideal scenario would entail a more extensive safety and efficacy test being initiated in earlier phases of the drug R&D process. The elimination of unsafe and ineffective compounds in an earlier phase would mean that less development effort goes to waste.

4.2 Systematic literature review

Numerous research studies have been aimed at pinpointing factors that contribute to the loss of efficiency in the pharmaceutical R&D process [27]. The value of identifying these factors lies within the opportunity to potentially address the identified factors in the R&D process, thus limiting the negative effect that it might have on the pharmaceutical pipeline. This study, however, aims to identify the factors that have a direct influence on the drug R&D process, thus affecting the state of the drug pipeline. A systematic literature review has been used to determine such factors.

4.2.1 Systematic literature review method

The literature review search was done in the Scopus[5] literature database. The objective was to establish the factors that lead to a lack of efficiency in the pharmaceutical drug pipeline. Answering the following two research questions (RQs) will contribute to addressing the primary objective of the study effectively.

- *RQ 1: What factors influence the overall drug pipeline of the pharmaceutical industry?*
- *RQ 2: What trends can be identified in the development of drugs over the past 10 years?*

Keywords for the search were derived from the two research questions and arranged in a logical manner. A search was completed with the search line: ("clinical trial" OR ((pharmaceutical OR drug*) W/5 (r&d OR pipeline OR development)) W/5 (factor* OR challeng* OR influenc* OR improv* OR affect*)).

4.2.2 Systematic literature review results

The search, using the keywords mentioned above, gave an output of 16 309 possibly relevant documents. The document set was further limited by type to journal sources only, excluding 1 017 articles. The document type was limited to articles, leading to 8 623 articles in total. The publication date was limited to a range from 2008 to 2017, resulting in a total of 5 504 documents. Finally, all articles written in languages other than English were excluded, resulting in set of 5 099 documents.

In order to reduce the number of documents in the document pool further, it was decided to use the top 200 cited documents from the set of 5 099. To correct for the bias inherent in only selecting the top cited documents, all documents published from 2015 to 2017 were also included in the document pool. This resulted in a document pool of 200 (top cited) + 2 049 (published 2015-2017). Eleven documents were duplicates in the two sets, which gave a total of 2 238 preliminary, relevant documents.

The titles of all 2 238 documents were scanned for relevancy to the two research questions. Consequently 147 documents were deemed relevant. The abstracts of these 147 documents were reviewed and resulted in the final selection of 97 documents with information relevant to both RQ1 and RQ2. The abstracts of these documents were analysed from two perspectives. Firstly, the abstracts were comprehensively investigated to establish factors that correlate with RQ1 (see Section 4.3). Secondly, the abstracts were explored to identify trends that correspond with RQ2 (see Section 4.4).

[5] Scopus is the database of Elsevier, and the world's largest abstract and citation database of peer-reviewed literature. Scopus provides global interdisciplinary and scientific information across all research fields. Scopus is free to use [27].

4.3 Results: Influencing factors

The preliminary identification of influencing factors was conducted by investigating the abstracts of the 97 documents and identifying factors that correlate with RQ1. In total, 39 factors were identified. The range of occurrence varied from a single occurrence to 13 occurrences across the 97 articles. Table 1 shows the most prominent factors present in the articles included in the dataset.

Table 1: The top occurring influencing factors identified in the preliminary analysis of the document pool

No.	*Influencing factor*	*Occurrence*
1.	Policy & regulatory issues	13
2.	Set-up of clinical trials; randomisation in trials (RCT); trial methodology & choice of metrics	13
3.	Recruitment and retention of participants; enrolment & minority representation; little clinical trial awareness of potential participants	11
4.	Complexity of trials; deal with multiple endpoints; better operational framework; clinical trial activation difficulty	10
5.	Clinical trial risk	7
6.	Lack of transparency; accountability; accessibility of clinical trial information	7
7.	Quality of clinical trial; improve use of innovative clinical trial tools; quality of pre-clinical trials	7
8.	Physician participation; relationships between stakeholders; collaboration; data sharing & intellectual property	6
9.	Lack of capacity and funding; lack of ROI	5
10.	Ethical obstacles and issues	5

The top four factors found to influence the pharmaceutical pipeline occurred in 10% or more of the document pool selected for this systematic review. All four factors are briefly discussed below.

1. Policy and regulatory challenges refer to any challenge encountered in ensuring, establishing or completing the regulations laid out by the regulatory drug agencies of the pharmaceutical industry. These challenges might exist because of national or international policies and are often influenced or enforced by government.
2. Clinical trial set-up refers to the way in which the clinical trials are organised, planned or arranged. The set-up determines how the activities of the trial phases will operate and what each step will entail.
3. Participants of clinical trials refers to the patients on whom tests are being conducted. Participants usually volunteer for clinical trials and might be provided with some sort of incentive to participate.
4. The complexity of clinical trials refers to the difficulty of completing and performing the actions required for the trials. It refers to the operational challenges experienced in carrying out the necessary protocols in all aspects of the clinical trials.

4.4 Results: Pharmaceutical pipeline trends

Trends in the pharmaceutical pipeline indicate a general direction in the development of or changes to the pipeline. Of the 97 abstracts reviewed, eight mentioned trends in the pharmaceutical drug pipeline. Four trends in drug R&D and pipelines are identified and investigated in this section, namely: i) R&D productivity; ii) investment capital and returns in the pharmaceutical sector; iii) clinical trial registration; and iv) the cost of clinical trials.

4.4.1 R&D productivity

The productivity of pharmaceutical R&D can be measured by various methods. According to Lendrem [14], productivity is measured by evaluating the number of new therapeutic drugs (NTDs) per billion dollars R&D spent per annum; Schulze et al. [15] evaluated the number of peak sales value of NTDs instead. The method of measurement used by Landrem [14] includes the effect of inflation-adjusted R&D costs.

The productivity evaluation, as mentioned by Landrem [14], concluded that escalating R&D costs is a dominant feature influencing the productivity of R&D during the period 1990 to 2013. The rise in operating costs, according to Hammer and Champy [16], might be a result of a change in focus during the 1990s to maximise the development speed of drug R&D. The cycle times of successful molecules were halved from 1990 to 2001, but this led to an immense increase in development costs, ultimately affecting the entire drug development process. The productivity of R&D remained relatively constant over the period 1990 to 2013, but decreased drastically when inflation was considered. The increase in the inflation-adjusted R&D costs offered an explanation for the market decline in overall R&D productivity [14].

4.4.2 Investment capital and returns in the pharmaceutical sector

The investment capital in this sector has decreased over time in response to many factors. These factors include preclinical scientific breakthroughs [17], clinical trial data, regulatory oversight, healthcare policies, pricing, technology and other economic changes related to drug discovery and development [18]. According to Thakor et al. [18], the most direct driver of capital flow in and out of the industry is the performance of pharmaceutical investments, thus providing attractive returns on the investments made. Some sources state, however, that not all pharmaceutical companies are struggling to realise returns, and that healthcare venture capital outperformed all other venture sectors over the past decades [18]. The annual returns of the pharmaceutical sector for the period 1980 to 2015 exceeded that of the stock market by 3%. The pharmaceutical portfolio also outperformed the market portfolio, where $1 invested in pharmaceutical companies in 1980 would be worth $114, compared with $44 if invested in the market at the same time [18].

Each investment holds a certain amount of risk and volatilities in returns [18]. The Sharpe ratio, a measure of an investment's return per unit of total risk, for the pharmaceutical sector was higher than that of the average market. The high Sharpe ratio indicates that the risk-adjusted returns of the pharmaceutical sector were better than the average market for the period 1980 to 2015.

4.4.3 Clinical trial registration

The registration of clinical trials is necessary to increase their ethical and scientific value [19]. More than half of clinical trials are never published and are reported selectively, resulting in a waste of resources and decision-making based on biased evidence, such as exclusive groups of patients used to participate in trials [19]. According to Viergever and Li [19], the number of registered clinical trials increased substantially between 2004 and 2013, from 3 297 to 23 384. Table 2 indicates the number of clinical trials registered, based on regional income groups.

*Table 2: Clinical trial registrations based on income groups, adapted from **[19]***

Region (country income groups)	***Number of trials registered in 2005-2013***	***Percentage of all clinical trials registered in 2005-2013***
High-income countries	143 137	82.5
Upper middle-income countries	24 937	14.4
Lower middle-income countries	8 229	4.7
Low-income countries	1 433	0.8
Not specified	6 319	3.6

It is evident from the information presented in Table 2 that high-income countries have the highest number of registered trials, representing 82.5% of all the clinical trials registered globally. In comparison, low-income countries conduct only 0.8% of the total number of clinical trials registered.

The registration of clinical trials has improved transparency in pharmaceutical research by increasing access to information across the globe [19]. Challenges still exist though [1]. These include: i) the quality of data available; ii) the accessibility of all clinical trial data; and iii) data searchability, data aggregation and linking data [19].

4.4.4 The cost of clinical trials

The cost of each clinical trial completed is influenced by a range of factors. The factors identified in Section 2.3, amongst other things, affect the cost of the trial. Sertkaya et al. [20] evaluated all direct cost components and constructed a list. Their study [20] established that the average cost of each of the various stages are as depicted in Table 3.

Table 3: Average cost per clinical trial phase, adapted from ***[20]***

Phase number	***Average range of clinical trial cost***
Phase 1	$1.4 million to $6.6 million
Phase 2	$7 million to $19.6 million
Phase 3	$11.5 million to $52.9 million

The top three cost drivers, as established by Sertkaya et al. [20], were clinical procedure costs (15-22%), administrative staff costs (11-29%), and site monitoring costs (9-14%). It is important to note that these findings are based on trials funded by pharmaceutical and biotechnological organisations and not governments, academic institutions or other organisations [20].

4.5 Discussion of results

The factors and trends identified can be evaluated based on the effect that they have on one another, and on how certain factors influence the trends within the pipeline.

The productivity of drug R&D is a result of several factors, including the cost of R&D (as mentioned in Section 4.4.1). Policies and regulations (influencing factor 1) potentially reduce the amount of NTDs introduced into the market. This does imply, however, that the process of eliminating unsafe and ineffective compounds from the pipeline is meticulous, resulting in reduced risk for potential drug users. The complexity and difficulty of trials (factor 4) mean that more time is required to conduct accurate studies. The quality of the trials also plays a role in the time it takes to complete the necessary procedures and whether it is necessary to repeat the study because of inadequate, inaccurate or insufficient data. The recruitment and retention of participants (factor 8) in the drug development process has a direct effect on the length of a study. It might take longer than planned to recruit all the participants necessary for the study, or the participants might be unable to complete the study, making the study unacceptable to regulatory authorities.

The amount of investment capital and the returns of the pharmaceutical industry relies on the amount of risk (factor 5) involved in the drug development process. For each disease this risk differs. Clinical trial registration requires organisations to be transparent (factor 7) about the procedures of the clinical trials and the trial outcome and information. Lastly, the cost of clinical trials is affected by almost all aspects of the drug development process. The longer the process, the higher the cost of drug development. The attrition of drug compounds during the R&D process also plays a major role in the cost – funds are lost when compounds pursued for many years fail to qualify as safe and effective drugs.

The state of the pharmaceutical pipeline is undoubtedly dependent on all the factors mentioned in Table 1. The trends identified in Section 4.4 discusses the impact of those factors with one another and the status quo of the pipeline.

5. PHARMACEUTICAL TB DRUG PIPELINE

This section describes the pharmaceutical TB drug pipeline. Firstly, various views on the TB drug pipeline are discussed. The current state and number of drugs in the pipeline are evaluated. A systematic literature review was conducted to determine influencing factors and trends in the TB drug pipeline. The results of the literature review are analysed at the end of the section.

5.1 Contextualisation: The pharmaceutical TB drug pipeline

Inspecting the drug pipeline of a specific disease differs from viewing the overall pharmaceutical pipeline. An overview of the TB drug pipeline incorporates characteristics of the disease and how those characteristics affect the drug development process.

5.1.1 Different views of the TB drug pipeline

According to the *2017 Pipeline Report* [2], the overall TB drug pipeline can be divided into four explicit pipelines. The four pipelines are the: i) diagnostics; ii) prevention; iii) treatment; and iv) diagnostics and treatment for children pipelines. Each of the pipelines has different aims and objectives in the way it addresses the TB epidemic. These various TB drug pipelines are discussed below.

I. Diagnostics pipeline

The TB diagnostics pipeline includes the development of all technologies, tools and tests to identify and specify TB and the type of TB of the person tested. TB diagnostic tests include patient diagnosis and drug-susceptibility testing [2]. TB diagnostics also include treatment monitoring technologies, which are necessary to establish whether treatment regimens are effectively improving or worsening the patient's condition.

II. Prevention pipeline

Mycobacterium tuberculosis (MTB) is the causative agent of TB. According to a study completed in 2015 [21], 1.7 billion individuals are infected with MTB. However, it is only when the MTB infection progresses to the active, transmissible state that a person is said to have TB. For most people, an MTB infection will not progress to the active state; for others, certain life events (aging, pregnancy) or immune-compromising conditions might lead to an active TB infection [2].

The TB prevention pipeline primarily entails the R&D of vaccines or other innovation therapies. According to Behr et al. [22], the two primary goals of TB vaccine development is to, firstly, find a vaccine to boost the current vaccination available to prevent active TB in adults infected with MTB and, secondly, to find a novel vaccine to replace existing vaccination for infants. Behr et al. [22] consider it important to pursue the parallel development of vaccines not only to prevent active TB from developing, but also to develop a vaccine to prevent sustained infection by MTB.

III. Treatment pipeline

The bacterial species MTB is treated with a combination of antibiotics [2]. Several antibiotics are used as first-line medicines to treat the infection. The extensive overuse and misuse of antibiotics since their introduction in the 1950s have led to a rise in DR-TB infections [23]. The DR infections occur because of inadequate or interrupted treatment or can be transmitted from one person to another.

IV. Diagnostics and treatment for children pipeline

Childhood TB has unique challenges with regard to diagnosis, prevention and treatment, as the disease differs considerably from TB in adults [24]. The BCG vaccine is primarily used by to prevent TB in children but is not fully protective and is not recommended for HIV-infected children. According to the 2017 WHO report, only 38% of TB-infected children are diagnosed and reported to national authorities each year [1]The diagnosis of childhood TB is complicated by the disease having non-specific symptoms. In most cases, children present smear-negative results, even though they are infected with the disease [24].

This research study focuses on the treatment pipeline: the current state of drugs in R&D intended to treat TB. This study does not distinguish between drugs aimed at treating adults or children, but rather views the pipeline from a systems perspective.

5.1.2 The current TB pipeline

There are a significant number of challenges in the management of TB on a global scale [3]. A considerable number of organisations, stakeholders, researchers and advocates are working on developing drugs to mitigate the effect that TB has on patients. Yet a more effective, shorter, safer and more easily tolerable treatment regimen is needed to manage the disease effectively [3]. The current TB drug pipeline, as reported in March 2018, consists of 14 candidate drugs for drug-susceptible, drug-resistant and latent TB. The 14 drugs are all in clinical stages of development; two drugs are in early stage development, nine are novel and in clinical trial phases 1 and 2, three have been approved and are in clinical trial phase 3 [3]. In addition to these drugs, other immune-based and host-directed treatment arrays are also under development [3]. The drug development industry has made extraordinary progress with the approval of two novel anti-TB drugs, bedaquiline and delamanid, in 40 years [25]. Figure 3 depicts the recent drug development pipeline for medicines currently progressing through clinical trials.

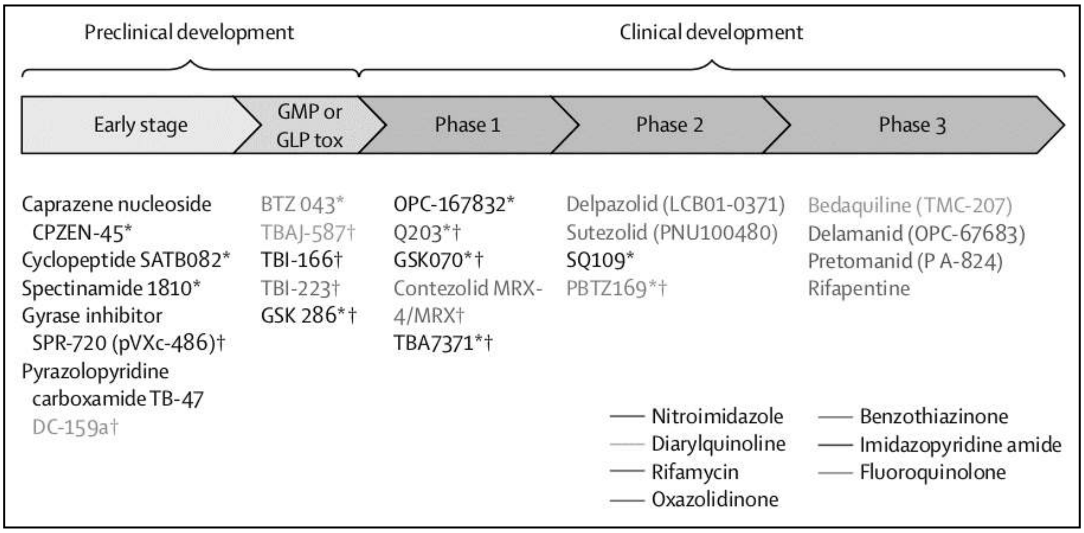


*Figure 3: New global TB drug development pipeline **[3]***

5.2 Systematic literature review process

The objective of this literature review was to establish what existing literature says about factors that lead to the lack of efficiency in the TB drug pipeline, in particular. The trends experienced in the TB drug pipeline were investigated to determine the state of the current pipeline. This systematic review was conducted to replicate the platform created with the previous systematic review, making a comparison between the two pipelines possible.

5.2.1 Systematic literature review method

As in the systematic literature review described in Section 4.2, the document search was completed in Scopus. The RQs presented in Section 4.2.1 were adjusted to form new RQs that would apply to the TB drug pipeline instead of the pharmaceutical drug pipeline. Addressing these two RQs will assist in comprehending the TB drug pipeline.

- *RQ 3: What factors influence the TB drug pipeline?*
- *RQ 4: What trends can be identified in the development of TB drugs over the past 10 years?*

The search line was as follows: (Tuberculosis OR tb) AND ("clinical trial" OR ((pharmaceutical OR drug*) W/5 (r&d OR pipeline OR development)) W/5 (factor* OR challeng* OR influenc* OR improv* OR affect*)).

5.2.2 Systematic literature review results

The search resulted in a total of 230 documents. The sources were limited to journals and conference proceedings, leading to 219 documents in the pool. Conference proceedings were added to this systematic review, and not to the systematic review in Section 4, because the document pool for this search was much smaller, since including TB narrowed the search extensively. The document type was then limited to articles and conference papers, leading to 121 documents. The publication date, as in the previous search, was limited to the last ten years (excluding 2018), resulting in 80 documents. The document pool was finally restricted to only English-language publications, leading to a total of 77 documents. The titles of the documents were scanned to find articles that mention anything that relates to RQ3 & 4. A total of 33 documents were found to be applicable to the RQs. The abstracts of these 33 documents were screened from two perspectives; firstly, to determine the factors influencing the TB drug pipeline (RQ3) and, secondly, to identify trends in the TB drug pipeline (RQ4).

5.3 Results: Influencing factors

With the first review of the 33 article abstracts, a total of 24 factors influencing the TB drug pipeline, thus relevant to RQ3, were identified. Table 4 shows the eight most frequently occurring factors identified.

Table 4: Top 8 factors influencing the TB drug pipeline

No.	***Influencing factor***	***Occurrence***
A.	Drug resistance	6
B.	Developing countries	5

C.	Resource-constrained (limited) setting	5
D.	Complex structure of MTB within host; different physiological states of bacterium	4
E.	Access to drugs; implementation of current & new treatment; low exposure	3
F.	Lack of finances	2
G.	Not enough DST technology	2
H.	Uncertain predictive value of preclinical animal models	2

The top three factors are briefly discussed below:

A. For TB, drug resistance refers to the characteristic of the disease that develops in many patients as a result of their not taking the medicine as scheduled, or by the DR-TB being transmitted from one person to another. The effect of drug resistance on the TB drug pipeline was mentioned in six of the 33 articles (18%) scanned in the systematic review. Patients infected with DR-TB are resilient to the first-line drugs used to treat the disease, as mentioned in Section 1.

B. Developing countries refers to the setting in which the disease occurs most often. Treating diseases in developing countries present challenges that do not necessarily occur in developed countries, thus making it a valid characteristic to consider with regard to the state of the drug pipeline.

C. Resource-constrained setting refers to the capability of the region in which TB occurs to perform the necessary actions and allocate the required funds to have a positive influence on the development of drugs.

5.4 Results: TB drug pipeline trends

With the second review of the 33 article abstracts, three major trends surfaced in two articles. Additional literature was included to describe the trends in the TB drug pipeline effectively. Each of the three trends are described in this section.

5.4.1 Funding trends by research category

TB is being addressed by organisations globally. Unfortunately, the funding available to combat TB does not seem to receive the attention needed to effectively mitigate the disease. The global plan to stop TB recommends an annual $2 billion investment to accomplish the goal of ending TB by 2030 [2]. At present, the effective investment in TB R&D is approximately $650 million per year. Figure 4 indicates the total TB R&D funding for the period 2005 to 2015.

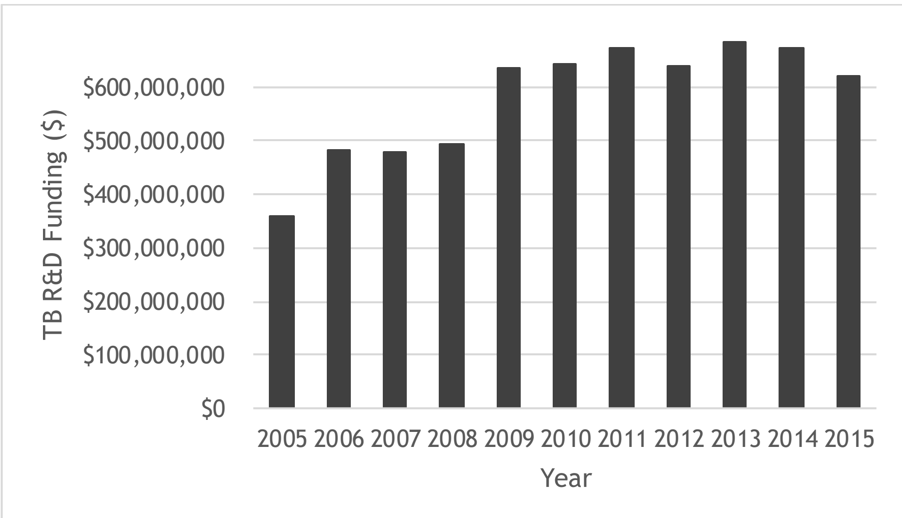


Figure 4: Total TB R&D funding, 2005-2015, ***[4]***

Funding per research category per year gives an indication of the focus of investors in TB R&D over the past 10 years. The drug development received the most funding, achieving a maximum of $268 million in 2013, but decreasing with over $30 million to a less desirable $232 million in 2015 [4]. Diagnostics reached its funding peak with $111 in 2014 million, but decreased to $81 million in 2015; operational research achieved a maximum of $88 million in 2011, decreasing to $61 million in 2015 [4]. These figures indicate the decreased market interest in TB.

5.4.2 Number of drug approvals

The FDA approval of drugs for treating TB has been notably slow over the past 40 years [3]. The exception, however, is the approval of the two novel drugs bedaquiline and delaminid. At the time of writing, a third drug, pretomanid, is undergoing phase 3 clinical trials along with the two drugs mentioned previously [3].

5.4.3 TB disease burden

Measuring the effectiveness of the drugs, innovations and technologies launched is challenging. One way in which the success of the strategies used to end TB can be measured is by assessing the overall disease burden. Globally the TB mortality rate (measured per 100 000 population) decreased by 37% between the years 2000 and 2016 [1].

5.5 Discussion of results

The amount of funding invested in TB drug R&D and the overall drug pipeline depends greatly on the return on investments in TB drug R&D [4]. The number of drug approvals for TB is influenced by the complex structure of MTB within the host (influencing factor D). The different physiological states of the bacterium (factor D) also lead to a greater variety of scenarios in which the drug compounds in development should act. The number of drug approvals is also affected by the strict policies and regulations put in place to ensure that only safe and effective drugs are introduced into the market.

The disease burden of TB is subject to most of the factors listed in Table 4. Drug resistance (factor A) causes the disease to be less treatable with the drugs on the market. Because the disease is most prevalent in developing countries (factor B) with a constrained amount of resources (factor C), it decreases the likelihood that the impact of the disease can be mitigated effectively. When the access to drugs of patients infected with TB is limited (factor E), the effective treatment of those patients becomes unviable. The disease burden cannot be lessened if there is no easy access to the right types of drug for patients. Improving the availability of DST technology (factor G) will improve the ability of clinics to determine whether someone is infected with DR-TB or not, thus providing the patient with the right type of medicine.

6. COMPARISON BETWEEN PHARMACEUTICAL AND TB DRUG PIPELINES

It is apparent that there are similarities and differences between the pipeline of the pharmaceutical industry and the TB drug pipeline. It is important to note that the pharmaceutical pipeline serves as a reference to which the TB drug pipeline is compared.

6.1 Evaluation of influencing factors

The top eight factors in the pharmaceutical sector focus mainly on improving the efficiency and productivity of the drug pipeline by ensuring that the drugs progressing through the clinical trial phases are approved as fast as possible and identifying reasons if this does not happen. The factors that occurred the most in the TB drug pipeline indicate something about the nature of the disease, the complexity of the disease itself and the state of the setting in which most TB cases occur. The only factor found in both pipelines is the lack of funding (factors 9 and F in tables 1 and 3, respectively).

It is evident that the TB drug pipeline cannot be examined without taking the circumstances and characteristics of the disease into account. These circumstances and characteristics include: i) the number of cases reported per year; ii) the success rates of drug treatments; iii) the regions in which this disease is most prevalent; iv) the complexity of treatment; and v) the length of the treatment course.

6.2 Evaluation of pipeline trends

Although there is a lack of funding and investment in both pipelines, it is evident that the funding gap in the TB drug pipeline is much greater than the funding gap in the general pharmaceutical industry, indicating that the investment gap is most likely not as prominent for the R&D of other drugs.

The productivity of the development of TB drugs differs significantly compared to productivity in the pharmaceutical pipeline. The number and rate of TB drug approvals are much lower and slower than those in the pharmaceutical drug pipeline. This indicates the difficulty in finding an effective cure for TB and DR-TB, as a result of disease complexity and drug resistance.

The prevalence of TB in developing countries gives rise to smaller returns on investment compared with the R&D for other diseases. It is assumed that pharmaceutical organisations or private investors aim to maximise their returns on investment. Consequently, TB R&D investment is not seen as a first option and funds are rather invested elsewhere.

7. THE VALUE OF THE IDENTIFICATION AND ANALYSIS OF TRENDS AND INFLUENCING FACTORS IN THE PHARMACEUTICAL AND TB DRUG PIPELINES

The value of the factors and trends identified in this study, and ways in which the identification of influencing factors and trends in drug pipelines might contribute towards more effective investment and better decision-making, are discussed in this section.

7.1 The ability of trend analysis to enable decision-making in TB R&D investment

The trend and influencing factor analysis of the general pharmaceutical industry identified characteristics of the industry and sector in terms of market attractiveness, productivity and costs that would not be possible otherwise. Comparing the TB and general pipelines identifies gaps in the TB drug pipeline that do not exist in the general sector. The identification of these gaps serves as a platform to improve the TB drug pipeline. The TB drug pipeline undoubtedly requires more funding and investment to mitigate the disease effectively. Questions arise on how additional funding options will be made available or how investors will be encouraged to invest in this disease – keeping in mind that TB mostly occurs in developing countries with limited resources and relatively low returns on investment, and with the infection evolving day by day into an even more drug-resistant strain.

It should be noted that all drug pipelines are similar, but that variance is caused by the disease itself. The disease, its characteristics, most popular region of occurrence, challenges, complexity, and all other factors affect the drug pipeline. TB has exceptional characteristics, as it is currently the infectious disease that kills the most people in the world.

Identifying trends in the TB drug pipeline not only identifies the characteristics of the disease, but also provides a platform to present to stakeholders and potential investors, encouraging contributions and influencing future investment decisions.

7.2 Encouraging investment in TB R&D with the use of incentives

Decision-making on investment in TB drugs includes the consideration of incentives [25]. Burki [25] remarks that, even if funds were to be sufficient, other incentives should be available to keep the sector sustainable, as the high development costs will not lead to profitable returns. Parys suggests that push (grants) and pull (the transferable priority review voucher scheme run by the FDA and advancement prizes for achieving a specified goal) mechanisms are the best way to encourage private companies to invest in the disease [25].

Médicins Sans Frontières (MSF) proposed a new model for TB R&D based on push and pull mechanisms, with pooled intellectual property as an addition. The model is primarily aimed at separating the selling price of drugs from the R&D process costs [25]. The proposed MSF model, is intended to reduce the duplication of R&D efforts by different stakeholders, ensures open collaborative research, reduces the risk of compounds being rejected, and accelerates drug combination development [25] & [26]. To MSF's disappointment, the WHO rejected the proposal in December 2013, and it was thus not pursued any further.

7.3 The benefit of an in-depth understanding of the TB drug pipeline

Understanding the drug pipeline, and more specifically the TB drug pipeline, creates the opportunity to forecast and strategise decision-making more accurately. Forecasting in pharmaceutical pipelines enables stakeholders to make more informed predictions of the unpredictable market, and to allocate funds more efficiently to areas lacking the required growth. The complete understanding of the drug pipeline presents the opportunity to adjust the overall drug development strategy to align with the goals of effectively mitigating the disease.

8. CONCLUSION AND FURTHER RESEARCH

The drug pipeline is a complex and interconnected system. Thirty-nine factors influencing the pharmaceutical drug pipeline have been identified through a systematic literature review. The foremost factors are regulations and policy, the set-up of clinical trials, and circumstances regarding the participants of clinical trials. Trends identified in the pharmaceutical industry include the relative stagnation of productivity (in terms of NTDs launched) and a decline in capital investment in the sector. The transparency of clinical trials has increased, as more trials have been registered over the past decade. The cost of clinical trials has sky-rocketed – clinical procedures, administrative staff and site monitoring costs are responsible for between 37% and 65 % of total trial costs.

The complete / total TB drug pipeline can be divided into four distinct sub-pipelines. This study focused on the treatment pipeline only. The R&D in TB drugs has only recently made progress, with the approval of two novel drugs for the first time in 40 years. Twenty-four factors influencing the TB drug pipeline have been identified. The foremost factors are the increasing drug resistance of the infection, and the fact that disease occurring primarily in developing countries with limited resources. The trends most visible in the TB drug pipeline include the seeming shortage of private investment in TB R&D, and a lack of productivity (when based on drug approvals).

The conclusion is that the realisation of returns in the pharmaceutical industry depends on the type of drug being developed and on certain characteristics of the disease being targeted. In the case of TB, the characteristics include the high prevalence of the disease in developing countries, making it unattractive as a potential investment. Looking at investment trends, it can be deduced that investments in diseases occurring mostly in resource-constrained countries hold greater risks for stakeholders, compared to diseases in developed countries with more funds available.

Future research opportunities exist within the pharmaceutical R&D pipeline to identify characteristics that improve the market attractiveness of certain diseases in comparison with others. Research should be done to determine the features of diseases and the similarity of those features in various diseases that currently experience a lack of capital investment. Current and alternative finance mechanism schemes should be investigated to potentially improve the interest of private investors in diseases with unattractive characteristics.

9. REFERENCES

[1] **WHO.** 2017. *Global Tuberculosis Report 2017*. Geneva: World Health Organization.

[2] **Frick, M., Gaudino, A., Harrington, M., Horn, T., Jefferys, R., Johnson, J.** 2017. *2017 Pipeline Report*. New York: Treatment Action Group.

[3] **Tiberi, S., Du Plessis, N., Walzl, G., Vjecha, M.J., Rao, M., Ntoumi, F..** 2018. Tuberculosis: Progress and advances in development of new drugs, treatment regimens, and host-directed therapies, *Lancet Infectious Diseases*, 2018 Mar 23. pii: S1473-3099(18)30110-5. doi: 10.1016/S1473-3099(18)30110-5.

[4] **Frick, M.** 2016. 2016 Report on tuberculosis research funding trends, 2005–2015: No time to lose. New York: Treatment Action Group.

[5] **PhRMA.** 2015. Biopharmaceutical research & development: The process behind new medicines. Washington, DC: PhRMA.

[6] **Bujar, M., McAuslane, N. & Liberti, L.** 2017. R&D Briefing 65: New drug approvals in six major authorities 2007–2016: Focus on the internationalisation of medicines. London: Centre for Innovation in Regulatory Science.

[7] **Rottenstein Law Group.** n.d. What does "FDA approval" mean? [Online]. Available: http://www.rotlaw.com/legal-library/what-does-fda-approval-mean/. [Accessed March 2018].

[8] **PhRMA.** 2007. Drug Discovery and Development: Understanding the R&D process. Washington, DC: PhRMA.

[9] **PhRMA.** 2016. 2016 Biopharmaceutical Research Industry Profile. Washington, DC: PhRMA.

[10] **Surowiecki, J.** 2014. The pipeline problem, *The New Yorker*, 16 February 2004. [Online]. Available: https://www.newyorker.com/magazine/2004/02/16/the-pipeline-problem. [Accessed March 2018].

[11] **Farlex and Partners.** 2009. *Medical Dictionary*. [Online]. Available: https://medical-dictionary.thefreedictionary.com/pipeline. [Accessed 20 March 2018].

[12] **Scannell, J.W., Blanckley, A., Boldon, H. & Warrington, B.** 2012. Diagnosing the decline in pharmaceutical R&D efficiency, *Nature Reviews*, 11(3), pp. 191-200.

[13] **Bunnage, M.E.** 2011. Getting pharmaceutical R&D back on target, *Nature Chemical Biology*, 7, pp. 335-339.

[14] **Lendrem, D., Senn, S.J., Lendrem, B.C. & Isaacs, J.D.** 2014. R&D productivity rides again?, *Pharmaceutical Statistics*, 14(1), pp. 1-3.

[15] **Schulze, U., Baedeker, M., Chen, Y.T. & Greber, D.** 2014. R&D productivity: On the comeback trail, *Nature Reviews Drug Discovery* 13(5), pp. 331–332.

[16] **Hammer, M. & Champy, J.** 1993. Reengineering the corporation: A manifesto for business revolution, *Business Horizons*, 36(5), pp. 90-91.

[17] **Cortright, J. & Mayer, H.** 2002. Signs of Life: The growth of biotechnology centers in the U.S. Washington, DC: Brookings Institution Center on Urban and Metropolitan Policy.

[18] **Thakor, R.T., Anaya, N., Zhang, Y., Vilanilam, C., Siah K.W., Wong C.H. & Lo, A.W.** 2017. Just how good an investment is the biopharmaceutical sector?, *Nature Biotechnology*, 35(12), pp. 1149-1157.

[19] **Viergever, R.F. & Li, K.** 2015. Trends in global clinical trial registration: An analysis of numbers of registered clinical trials in different parts of the world from 2004 to 2013, *BMJ*, 5(9), pp. 1-5.

[20] **Sertkaya, A., Wong, H.H., Jessup, A. & Beleche, T.** 2016. Key cost drivers of pharmaceutical clinical trials in the United States, *Clinical Trials*, 13(2), pp. 117-126.

[21] **Houben, R.M.G.J. & Dodd, P.J.** 2016. The global burden of latent tuberculosis infection: A re-estimation using mathematical modelling, *PLoS Medicine,* 13(10), pp. 1-13.

[22] **Behr, M., Ellis, R., Ernst, J., Evans, T., Flynn, J., Hatherill, M..** 2015. Developing vaccines to prevent sustained infection with Mycobacterium tuberculosis: Conference proceedings. National Institute of Allergy and Infectious Diseases, Rockville, Maryland USA, November 7, 2014, *Vaccine*, 33(26), pp. 3056-3064.

[23] **The Lancet Respiratory Medicine.** 2017. Global rise of multidrug resistant tuberculosis threatens to derail decades of progress. [Online]. Available: https://medicalxpress.com/news/2017-03-global-multidrug-resistant-tuberculosis-threatens.html. [Accessed 11 May 2018].

[24] **Whittaker, E., Jones, C. & Kampmann, B.** 2012. TB in children, *European Respiratory Monograph*, 58, 2012, pp. 206-218.

[25] **Burki, T.** 2014. Improving the health of the tuberculosis drug pipeline, *Lancet Infectious Diseases,* 14(2), pp. 102-103.

[26] **Brigden, G**. 2015. 3P TB R&D Proposal – Push, Pull, Pool. Médecins Sans Frontières Access Campaign TB Symposium, Yerevan.

[27] **Naci, H., Carter, A.W., & Mossialos, E**. 2015. Why the drug development pipeline is not delivering better medicines. BMJ. h5542, pp. 1756-1833.

THE RELIABILITY OF COMMUTER RAIL SERVICE IN THE CITY OF CAPE TOWN: A GROUNDED THEORY APPROACH FOR IDENTIFYING THE FACTORS AFFECTING THE PROVISION OF A RELIABLE COMMUTER RAIL SERVICE

P.N. Zincume[1*], T. Ryan[2] & C. Shaw[3]

[1]Department of Industrial Engineering
University of Stellenbosch, South Africa
philaniz@sun.ac.za

[2]Graduate School of Business
University of Cape Town, South Africa
tom.ryan@gsb.uct.ac.za

[3]Department of Mechanical Engineering
University of Cape Town, South Africa
Corrinne.Shaw@uct.ac.za

ABSTRACT

The 2013 survey conducted by the National Household Travel Survey found that the majority of the commuter rail service users are dissatisfied with the quality of the service due to the lack of reliability characterised by the increase in the percentage of train cancellations and delays. This paper uses the Grounded Theory as its main methodology and develops the theory of commuter rail provision using Systems Archetypes and presents it in the form of a Causal Loop Diagram. The developed theory is then used to recommend a solution loop that seeks to improve the quality of commuter rail service.

[1] The author was enrolled for an M Phil (Engineering Management) degree in the Department of Mechanical Engineering, University of Cape Town *Corresponding author

[2] Graduate School of Business, Developing strategic capability, Executive development and Systems thinking, University of Cape Town

[3] Department of Mechanical Engineering, Engineering Management, Systems thinking and practice, University of Cape Town.

1. INTRODUCTION

The improvement of public transport system in South Africa has been one of the strategic objectives for the government since the beginning of the post-apartheid era [1]. During the apartheid era, many South Africans were dispersed and relegated to poor areas away from job opportunities, health and educational facilities and economic participation [2]. This social and economic exclusion is still prevalent today, with millions of South Africans still requiring means of travelling to access economic opportunities [3]. This implies that the government is faced with a responsibility of public transport that is reliable, safe, efficient and cost-effective.

The public transport system in Cape Town comprises (primarily) of three modes i.e. taxis, buses and commuter rail service. Contrary to other major South African cities, in Cape Town, the commuter rail service dominates the public transport system [4]. This has led to the Department of Transport (DoT) labelling the commuter train service as the backbone of transport in the City of Cape Town [5][6]. However, the Metrorail (a subsidiary of the Passenger Rail Agency of South Africa (PRASA)) timekeeping document has shown that the commuter rail service has worsened from 2013 [7]. Metrorail is the sole provider of commuter rail services in Cape Town (and the other regions of the country). The 2013 survey conducted by the National Household Travel Survey (NHTS) revealed that the majority of the commuter rail service users are dissatisfied with the quality of the service due to the lack of reliability, characterised by the increase in the percentage of train cancellations and delays [8].

1.1 Train cancellations and delays

The quality of commuter rail service refers to factors including but not limited to reliability, availability, maintainability, safety and cleanliness [9][10]. This paper aims to improve the quality of commuter rail service by focusing on reliability. Metrorail operates an ageing fleet of trains, some in operation since 1958, and they predominantly make use of the percentage of cancellations and delays as the empirical indicator to measure the reliability of their fleet [11]. The train cancellations and delays are at the heart of Metrorail's activities because they indicate the ability of Metrorail to provide a reliable service. Therefore, the concern variable (CV) for this study is the *percentage of trains cancelled and delayed* and this is shown in Figure 1.

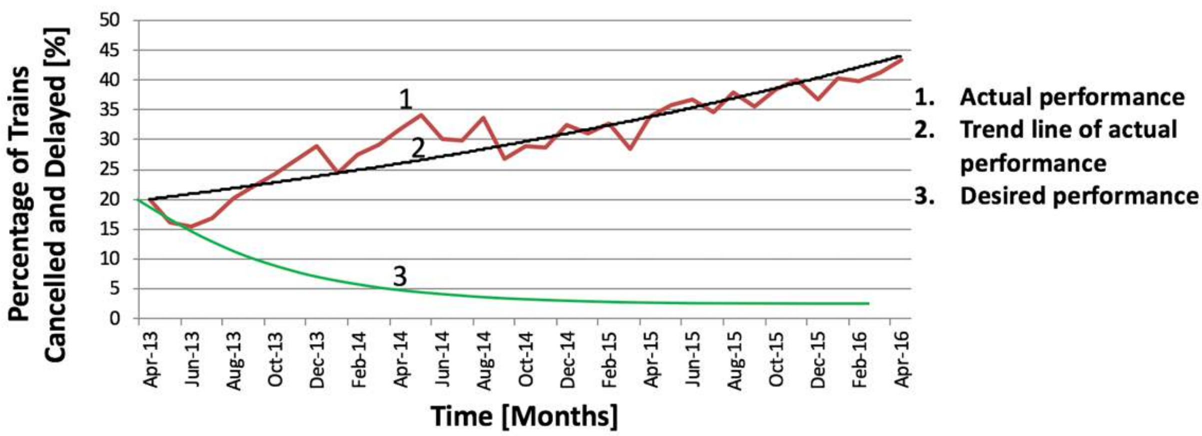


Figure 1: Percentage of Trains Cancelled and Delayed Over Time [7]

Figure 1 shows that train cancellations and delays have been increasing since 2013 and if there is no intervention, the service will continue to worsen. Although there has been an improvement in some of the months, the trend line shows that the overall result is a deteriorating train service. The data that was used for Figure 1 was obtained from the Metrorail time keeping document [7].

1.2 Objective of this paper

Researchers in the area of public transport tend to focus their studies on the issues faced by the commuters [12][13][14][15]. In studying the reliability of commuter rail in the city of Cape Town (within a region known as Western Cape), this paper takes a slightly different approach and focuses on the factors that affect the provider of commuter rail. In addition to identifying these factors, this paper seeks to develop a theory that describes the causal relationships between these factors and use this theory to inform recommendations for improvement.

1.3 Research Questions

Maxwell [16] points out that research questions are useful for explaining what the study is intended to learn and understand. The research questions are useful in determining what data to collect and what methods to be used in the analyses. The following research questions were structured and used in this study:

Question 1: What are the factors affecting the provision of a reliable commuter rail service in the City of Cape Town?

Question 2: How can these factors inform recommendations to improve the commuter rail service?

Based on the objective of the study, an appropriate research methodology was chosen to answer the research questions and realise the study's objective and is described in section 2 below.

2. THE RESEARCH METHODOLOGY

As noted by Maxwell [16], the primary difference between a qualitative and a quantitative study is that the qualitative study involves the collection of data that is mainly in the form of words whereas the quantitative study involves the data that is either in the form of (or can be expressed as) numbers. Cresswell [18] emphasises that the quantitative research focuses on variables, measuring objective facts, applies statistical analysis and turns to ignore context. He further adds that this approach is useful in testing whether the predictive generalisations of an existing theory hold true. However, the qualitative approach primarily focuses on understanding social life and the meaning that the people attach to it. The qualitative approach was chosen for this study because it is an approach well suited for developing theories and understanding phenomena [16].

This qualitative study used Grounded Theory (GT) as a research methodology, developed by Glaser and Strauss [17]. The GT methodology was chosen because it is useful for developing theory that is grounded in data, allowing the theory to evolve from the research process as a product of an interplay between the data collection and the data analysis [17][19]. The use of GT in the research process is useful in preventing the researcher from forcing data into already existing theories, while ensuring that a new theory in the provision of commuter rail service is generated. Easterby-Smith, Thorpe and Lowe [20] point out that through the use of GT, the structure is derived from the data by systematically analysing it to identify themes, patterns and categories. This means that the study does not begin with a preconceived theory but the theory is allowed to inductively emerge from themes, patterns and categories.

As mentioned in section 1, Metrorail is the provider of commuter rail services in Cape Town, therefore, the research is conducted using Metrorail- Western Cape managers as study participants. All departments were included and the details of data collection are given in section 3.

The outcome of the data collection cycles and the GT process are the categories that are used as the building blocks for constructing the theory that reflects the factors and relationships involved in the provision of commuter rail service. Figure 2 illustrates the dynamic interplay between the four data collection cycles and analysis. GT demands that data collection and analysis occur concurrently rather than in a linear sequence [21][22]. The research process iterates between the commuter rail sector (cycles 1-4) and the theoretical world (GT methodology).

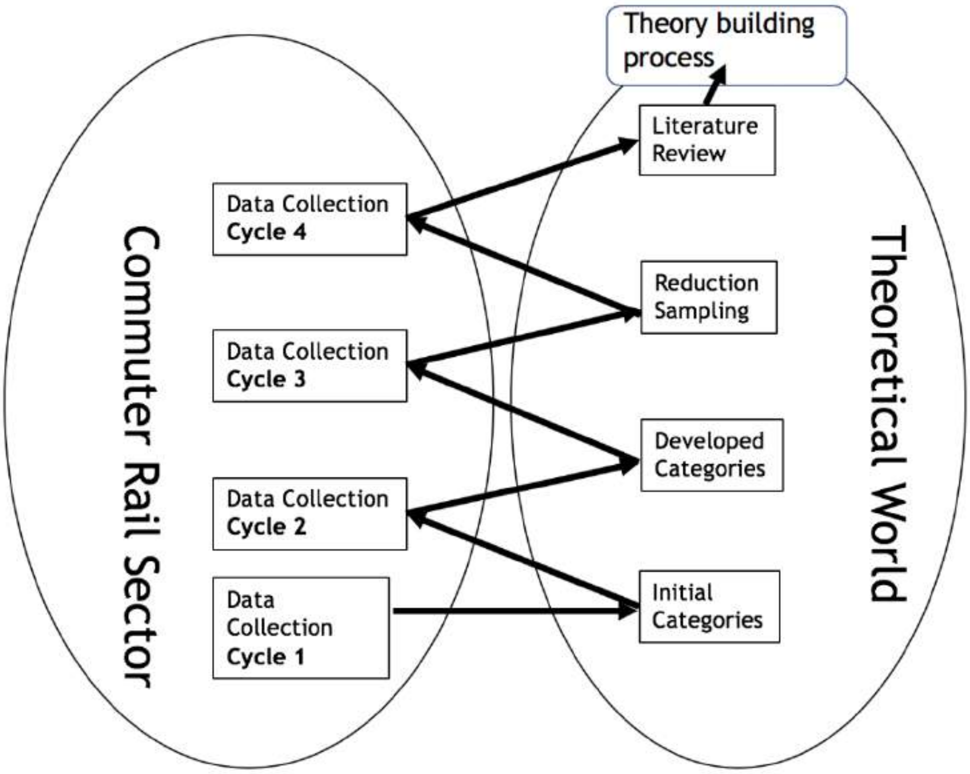


Figure 2: The research process

3. CONCURRENT DATA COLLECTION AND ANALYSIS

This section follows the research process presented in section 2. The dynamic interplay of data collection and analysis is presented in preparation for the theory building process which is presented in section 5.

3.1 Data Collection and Analysis of Cycles 1-3

Semi-structured face to face interviews were conducted as the primary means of data collection for cycle one. Three Metrorail managers from the Training and Development, Train Operations and Recruitment and Selection departments were interviewed for cycle one. The collected data was coded into propositions (open coding) and a total of 80 propositions had been collected by the end of cycle one. From these propositions, 16 categories emerged.

For cycle two, three more managers from Business Development, Finance and Supply Chain Management were interviewed. The data collection of cycle two was not only limited to interviewing but was extended to include documentary research [23] using the rail-related practitioner journals as the source of data together with newspaper articles and company documents such as the PRASA corporate plan, financial reports and recovery documents. In addition, data was collected by means of participant observation. Five PRASA management meetings were attended and recorded to be used as a source of data. New categories emerged and some of the new propositions were coded for existing categories. At the end of cycle two, seven more categories had been added and two categories modified, which led to a total of 23 categories and 130 propositions.

For cycle three, the last three managers were interviewed from the Infrastructure, Rolling Stock and Customer Service departments. Fifty more propositions were added after conducting these interviews and attending Metrorail management meetings. This contributed five more categories and a re-labelling of some of the existing categories. The addition of five more categories increased the number of categories from 23 to 28 categories. The participants for research cycles one to three were selected such that all the departments were included.

The data gathering process from cycle one to cycle three increased the researcher's understanding of the situation as more insight was gained. This allowed the categories to be developed as more data was collected. Some of the categories were re-labelled to give a more comprehensive naming of the concepts contained in those categories. An example of this is the category that had initially been labelled as *'shortage of personnel'* in cycle one and was developed to *'internal capacity'* in cycle two, the same category was re-labelled as *'availability of resources'* in cycle three. Each category was reviewed as more propositions (props) were collected and some categories remained the same as others were re-labelled. The results of the first three cycles are presented in Table 1.

Table 1: Categories after three cycles

Category	*Props*	*Category*	*Props*	*Category*	*Props*
Availability of Resources	14	Training	5	Technology Management	6
Internal Processes and Systems	13	Communication with Commuters	7	Decentralisation	4
Supply Chain Management	7	Ageing Assets	9	Profitability	3
Vandalism	11	Management skills within PRASA	8	Management of Contractors	6
Rail Modernisation	8	Political Interference	3	Maintenance Execution	1
Availability of Monetary Resources	13	Stakeholder Integration	5	Asset Protection	1
Passenger Experience	9	Rail Regulation	1	Employee Wellbeing	2
Public Transport Modal Integration	3	Interdepartmental Relations	6	Misalignment of functions and Departments	3
Influence of Labour Unions	7	Service Demand	11	Effectiveness of Management	9
Employee Performance	5				
Total propositions = 180					

3.2 Reduction Sampling

After the three cycles of concurrent data collection and analysis, the emergent categories were reduced to the strongest categories (referred to as core categories) through reduction sampling. The reduction sampling process was not an easy process because the validity of the study had to be maintained by ensuring that categories are not forced to labels where they do not belong. Often, the researcher had to go back to the original propositions that were contained in the category to ensure that the concepts that emerged from data are not lost during the reduction sampling process. This was a rigorous and iterative process and it broadened the understanding of what is truly going on.

Holton [24] emphasises that the category labels chosen as core categories should have more explanatory power and should account for most of the variation around the focus of the study. This study employed the use of an affinity diagram (AD) to do the reduction sampling [25]. During the reduction sampling process, seven categories emerged as core categories and these categories were used as the building blocks for theory formation in the theory building process. Table 2 illustrates the reduction sampling process using an AD and the seven emergent core categories are shown (in bold) with the remaining categories subsumed under these core categories.

Table 2: The AD used for reduction sampling

1. Availability of Resources	**2. Internal Processes and Systems**	**3. Vandalism**	**4. Quality of Management**
Availability of Resources	Internal Processes and Systems	Vandalism	Employee Performance
Availability of Monetary Resources	Supply Chain Management	Asset Protection	Effectiveness of Management
Training	Communication with Commuters	Passenger Experience	Employee Wellbeing
Profitability	Maintenance Execution		Management skills within PRASA
	Rail Modernisation		Management of Contractors
	Technology Management		Decentralisation
			Political Interference

5. Interdepartmental Relations	**6. Service Demand**	**7. Ageing Condition of Assets**	
Stakeholder Integration	Service Demand	Ageing Assets	
Interdepartmental Relations	Public Transport Modal Integration	Rail Regulation	
Misalignment of functions and Departments			
Influence of Labour Unions			

3.3 Selective Sampling (Cycle 4)

The milestone for cycle four was to attempt to saturate the core categories. Saturation refers to the state where the criterion given by Corbin and Strauss [26] is met. This criterion was used to saturate the core categories and is discussed in this section. Cycle four was a selective sampling cycle because the data collection was informed by the core categories that emerged in the first three cycles. Based on these core categories, decisions were made with regards to where to collect data, what data to collect and from whom.

(i) No new or relevant data seem to emerge regarding a category.

In an attempt to saturate the core categories, the goal was to collect between seven to 10 propositions with the hope that the categories can be saturated. Six out of the seven categories were saturated within seven propositions as no new relevant data seemed to emerge. The exception was the *Availability of Resources* which could not be saturated in less than seven categories and required ten propositions to be saturated.

(ii) The category is well developed in terms of its properties and dimensions.

Categories were constantly compared and checked to ensure that they are well developed in terms of their properties and dimensions. This also ensured that all the categories are at the same level in the ladder of abstraction.

(iii) The relationships among categories are well established and validated.

The relationships among the core categories are established using an interrelationship diagram (ID) adapted from [25] and validated through the collection of data in cycle four. The usage of an ID is discussed in detail in section 5.2.

4. LITERATURE REVIEW

A comprehensive literature review is usually the first step in most research studies. However, one of the unique features of the GT methodology is the timing of the literature review. The GT literature review is based on the grounded theory results and therefore focused on those results. It is meant to be relevant to the situation at hand rather than a general review of the current literature. Even-though a quick literature scan is done at the beginning of the study, a comprehensive literature review is only conducted after the core categories have emerged from data and these categories are used to direct the review of literature.

As pointed out by Glaser and Strauss [17], conducting the literature review after the data collection allows the categories to emerge naturally from the empirical data during the analysis. The categories emerging from the fourth cycle were used to conduct the literature review. The insights collated from the literature review were woven into the existing categories in preparation for the theory building step as shown in Figure 2. The detailed grounded literature review based on the emergent categories has not been included in this paper because of the length constraints and also because of the focus being on the results of the GT process. However, an overview of scholarly work has been included in this section.

According to the PRASA Corporate Plan [9], Metrorail operates in Gauteng, KwaZulu-Natal, Western Cape and Eastern Cape regions and has a network covering more than 15% of South Africa's rail network. Metrorail provides more than 2.2 million passenger trips every weekday to the South African public. Metrorail operates a total of 698 train trips every weekday, 360 trains on Saturdays and 225 trains on Sundays in North, South and Central suburbs of Cape Town [9]. According to the NHTS [8], more than sixty percent of the working population in Western Cape region relies on commuter rail as the primary mode of public transport. This high percentage has led the Department of Transport [3], City of Cape Town [5] and many other transport authorities referring to commuter rail as the "backbone" of public transport in the city of Cape Town. Unlike the rest of South Africa where the commuter rail has only just about 10% of the public transport market share; in the Western Cape, the commuter rail enjoys over 60% market share of the public transport as shown graphically in Figure 3 and discussed in detail by Clark and Crous [4].

Whilst the commuter rail enjoys the majority of the public transport market share, various authors have pointed out elements of poor commuter rail service delivery provided by Metrorail. Clark and Crous [4] discussed one of the greatest problems faced by the commuters; the problem of overcrowding. Ryneveld [13] has shown that over seventy percent of commuters have voiced their concern with the issue of overcrowding. Metrorail has attributed this problem to the majority of their fleet running with fewer coaches than required and this has often led to fatalities [9]. Moreover, there is an increasing rate of injuries on the commuter rail service due to violence and crime which threatens the security of commuters [4]. The increasing rate of crime on the commuter rail system and other safety related issues have added to the growing negative user perception as pointed out by the NHTS [8].

According to the NHTS [8], the availability and reliability of train service remains the greatest problem and cause of frustration for the majority of the commuter rail users. Ryneveld [13] has pointed out that the frustration of commuters due to the availability and reliability of train service has resulted to a number of serious instances of vandalism and cases of arson with whole trains being burnt.

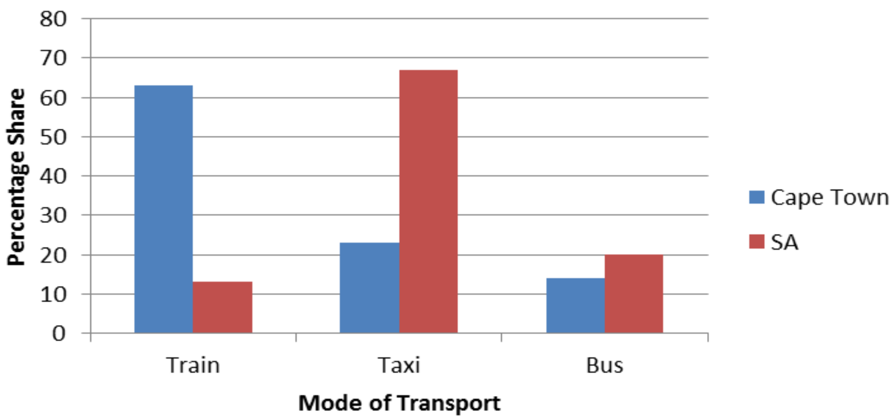


Figure 3 Comparison of public transport market share (adapted from Clark and Crous [4])

The literature review of commuter rail transportation revealed that there is limited scholarly work available on the topic of commuter rail transportation in South Africa. It is probable that the reason for this limitation is the monopoly that exists in the field of commuter rail as PRASA is the sole provider of rail transport (excluding the luxury rail transport providers).

5. THEORY BUILDING PROCESS

This section used the core categories developed in the previous sections to build a theory that answers the research questions and realises the aim of the study.

For this paper, the theory was built using the modelling process described by Beer [27]. Beer [27] asserts that a managerial theory must be built by first constructing a well-established and accepted scientific model. Figure 4 is a graphic representation which was adapted from Beer's work on modelling managerial situations using scientific analogies. Beer's modelling process was broken down into seven steps that are used to convert the core category labels into a theory that answers the research questions.

5.1 Step 1: Framing Core Categories as Variables

The first step in the process was to frame the core category labels as variables. This was achieved by first doing a concept analysis of each core category label. The concept analysis focused on three aspects i.e. antecedents, defining attributes and consequences. Once the concept analysis of the core categories had been completed, the concern variable was used to decide on which of the three aspects was most strongly related to the concern variable and to use that aspect to transform the core category label into a variable. This ensured that all the variables used in the final theory were relevant to the concern variable. The final results of step one are presented in Table 3.

5.2 Step 2-7: Using the resultant variables to finalise the theory in a CLD

Step two through to step seven built upon the results of step one. The resultant seven variables together with the CV were used as the building blocks for the final theory. Initially, the ID was constructed using the seven resultant variables from step 1 and the CV. The construction of the ID (step 2) was done using the question; "Does the change in variable A cause a change in variable B or does the change in variable B cause a change in variable A?" The ID took into account the direct relationships between the variables and ignored the indirect relationships. The resulting ID is presented in Figure 5.

Table 3: Summary of the core categories framed as variables

Framing Categories as Variables	
Original Category	***Framed as a Variable***
Availability of Resources	Availability of spare parts
Internal Processes and Systems	Efficiency of Internal Processes and Systems
Vandalism	Level of Rolling Stock and Infrastructure Vandalism
Quality of Management	Management of maintenance execution
Interdepartmental Relations	Departmental Cohesion
Service Demand	Service Demand
Ageing Condition of Assets	Impact of asset age

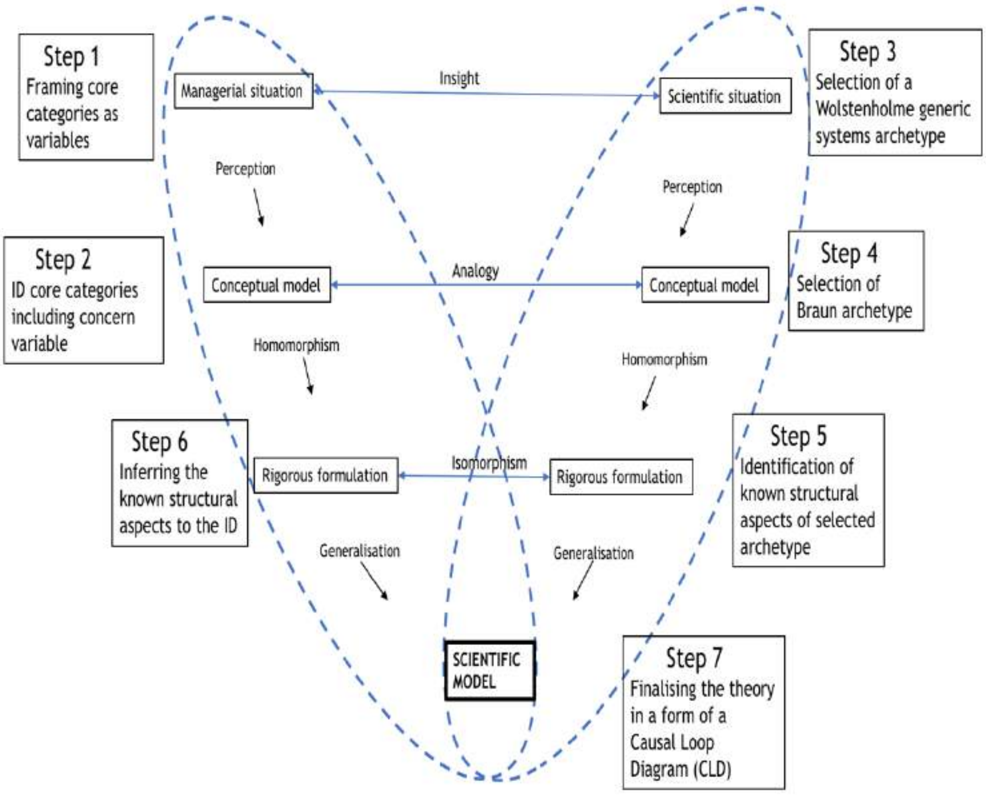


Figure 4: The modelling process (adapted from Beer [27])

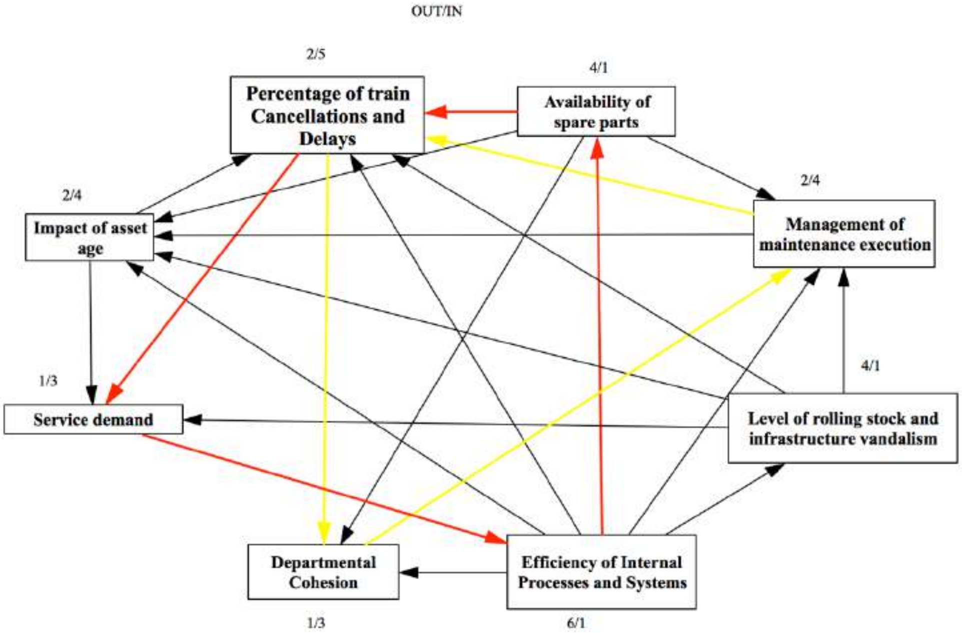


Figure 5: The ID of the 7 core categories with the CV (adapted from Brassard [25])

The CV emerged as the outcome whilst the driver is the *Efficiency of Internal Processes and Systems*.

The percentage of cancellations and delays is the result of what happens in the system and has a direct relationship with all the variables involved. This means that the train cancellations and delays are just a result of what happens elsewhere in the system. An example of this is the influence which the availability of spare parts has on the outcome. The availability of spare parts has an influence on the number of train-sets that operate in-service with defective components (the terminology used for this condition is *cut-outs in service*). When there are no spare parts available, the cut-outs in service will increase and this increase of cut-outs in service compromises the level of reliability of trains.

The significance of the concern variable emerging as an outcome is in that the focus has to be put on the other parts of the system to effectively impact on the concern variable. Focusing on the availability of spare parts or the management of maintenance execution or the efficiency of the internal processes and systems will have a far-reaching impact than to focus on the delays and cancellations in themselves.

The *Limits to Growth* generic archetype was chosen as the archetype for the theory building process because the research focus of this study is concerned with the issue of growth/decline (in the percentage of train cancellations and delays) which is used as an empirical indicator of the quality of commuter rail service. Braun [28] points out that the *Limits to Growth* archetype is used when the concern is about growth/decline. This archetype was originally introduced in the 1970's and is discussed by Braun [28] in detail. Based on the structural aspects of the Limits to Growth generic archetype and inferring these structural aspects (step 6) to the ID of Figure 5, the resulting causal loop diagram (CLD) is presented in Figure 6.

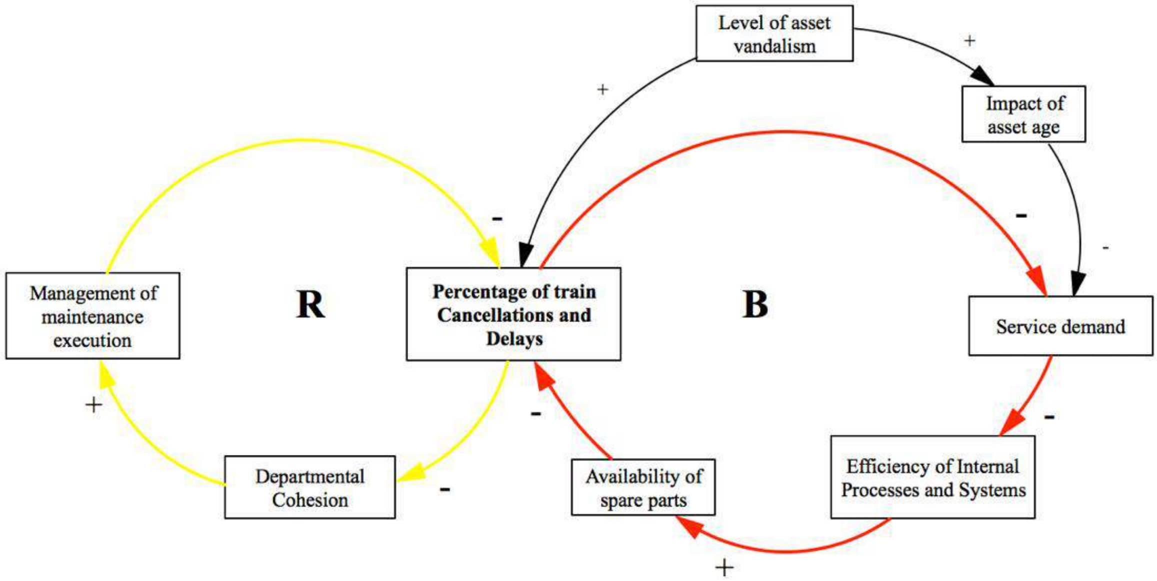


Figure 6: The emergent theory presented as a CLD

The emergent theory presented in Figure 6 represents the theory of providing the commuter rail service in the city of Cape Town; it is a depiction of the underlying system structure. As suggested by Clayton and Michael [29], a theory has an ability to predict which actions will lead to what results and can also assist in interpreting the prevailing situation by giving insights to what is happening and why it is happening.

The reinforcing loop (R) shows that the percentage of train cancellations and delays is closely linked to the ability of management to implement the correct maintenance strategies and to manage the execution of maintenance. As discussed in detail by Swanson [30], the management of maintenance execution is critical to the operations as it leads to high availability and reliability and conversely, the poor management of maintenance execution may lead to failures, poor utilisation of equipment and delays. Based on the interviews conducted, the low cancellations and delays of trains are viewed as successful results and lead to high morale within the departments. This improves the ability of departments to work together to achieve even greater success and when the departments are working in cohesion, the management of maintenance execution improves even further. This creates a reinforcing loop which yields the desired result of reducing the percentage of train cancellations and delays and thereby improving the reliability of commuter rail service.

The balancing loop (B) shows that an improvement in the percentage of train cancellations and delays leads to a high demand of train service and conversely, deterioration in cancellations and delays repels commuters. Clark and Crous [4] found in their study of public transport in Cape Town that due to the decline in the reliability of the train service, those commuters who can afford alternatives have increasingly shifted away from rail transportation. As shown in the theory, the limiting condition in the provision of commuter rail is the impact that the age of the asset has on the service demand. This limiting condition creates a slowing action in the system; PRASA has revealed that the network of 2 000 km is currently subject to 103 km of speed restrictions where trains are operating at lower than normal speed due to the poor condition of the track. The rolling stock and infrastructure are in an ageing condition and the track components are severely worn out and have reached the end of their life cycle [9].

The *efficiency of internal processes and systems* refers to the ability of the internal process and also the internal systems to respond to the business requirements and how efficient the response is. The ID that was presented in Figure 5 showed that this variable is the driver in the system. This means that the internal processes and systems offer an area of the most significant leverage if one wants to make improvements in the system. It is easy to understand why the efficiency of the internal processes and systems emerged as the driver; all the other variables depend on how the processes and systems that have been put in place respond to the business requirements. If the procurement process (which is part of the internal processes) can respond to the high demand of spare parts, the percentage of cancellations and delays can be reduced and conversely, if the demand of spare parts is not met, the percentage of cancellations and delays increases.

The emergent theory is used to inform recommendations to improve the commuter rail service. Based on the work of Wolstenholme [31] on system archetypes, a solution archetype can be constructed by using some element of the achievement action to minimise the reaction in other parts of the organisation/system. The proposed solution to resolve the concern of the increasing percentage of cancellations and delays is to introduce a reinforcing loop in parallel with the intended consequence reinforcing loop and this is done to unblock the resource constraint [32]. This is illustrated in Figure 7.

The efficiency of the internal processes and systems emerged as the driving variable in the research results. All the other variables depend on how the processes and systems that have been put in place respond to the business requirements. This result implies that improving the efficiency of the internal processes and systems offers a leverage point, i.e. high returns can be gained by focusing on this area.

Figure 7 shows the solution loop that was derived from the research results. An element of achievement action (*Management of maintenance execution*) is used to introduce a reinforcing loop (R1) in parallel with the intended consequence reinforcing loop (R) and this is done to unblock the resource constraint. It is recommended that management does not only focus on managing maintenance execution but also focus on the efficiency of internal processes and systems.

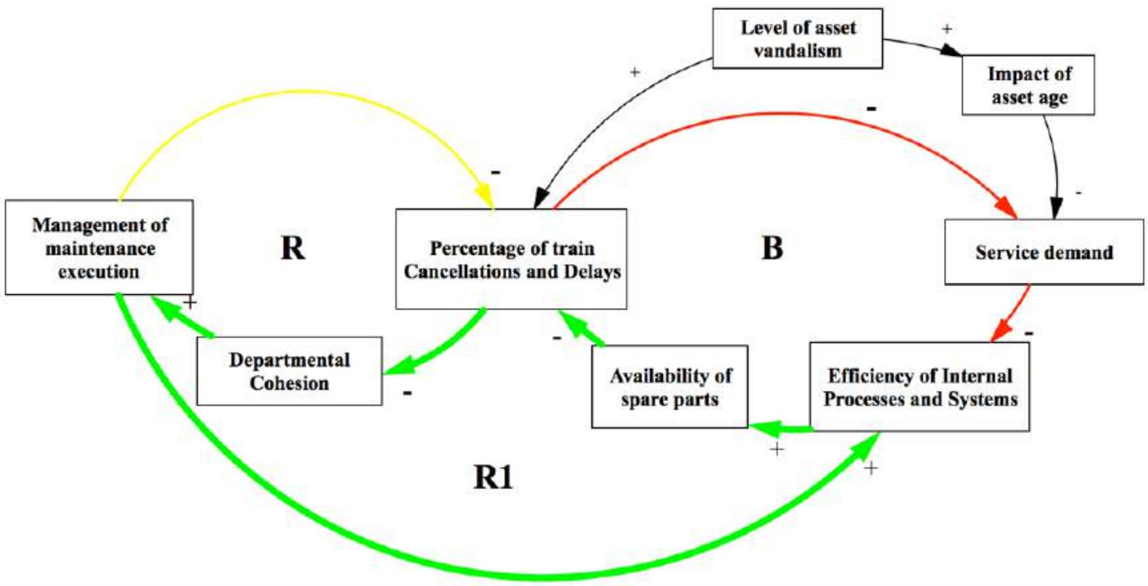


Figure 7: The solution loop for the provision of commuter rail service

6. CONCLUSION AND RECOMMENDATIONS

This paper investigated the concern of the reliability of commuter rail service in the Western Cape characterised by the increasing percentage of delays and cancellations of the train service. GT was used as the main methodology for the study and the commuter rail service provision theory was developed using Systems Archetypes and presented in the form of a CLD. Four cycles of data collection seeking categories, building on previous cycles as analyses were conducted after each cycle. The theory was built after conducting a comprehensive literature review and combining the theoretical aspects of Beer's [27] work to further develop core categories as variables and present the theory as a CLD, illustrating and recommending that management of the rail service not only look at maintenance execution but other factors which include the efficiency of the internal processes and systems. It is further recommended that future studies may explore international commuter rail service providers and compare the results of this study to international rail providers.

REFERENCES

[1] Barrett, J. 2003. Organizing in the Informal Economy: A Case Study of the Minibus Taxi Industry in South Africa. Seed Working Paper Number 358158.

[2] Coggin, T. and Pieterse, M. 2015. A right to transport? Moving towards a rights-based approach to mobility in the city. *South African Journal on Human Rights, 31(2),* pp.294-314

[3] Walters, J. 2013. Overview of Public Transport Policy Developments in South Africa. *Research in Transportation Economics, 39(1),* pp.34–45.

[4] Clark, P. and Crous, W. 2002. Public transport in metropolitan Cape Town: Past, present and future. *Transport reviews*, *22*(1), pp.77-101.

[5] City of Cape Town, 2012. City Of Cape Town- Integrated Development Plan (2012 – 2017). Available: http://www.capetown.gov.za. [Accessed on 4 June 2016]

[6] Branch, M., 2007. Department of Transport, City of Cape Town, 2012. City of Cape Town- Integrated Development Plan (2012 - 2017).
[7] Time Keeping Document – Train Operations Department, Metrorail Western Cape. (2016)
[8] NHTS, 2013. National Household Travel Survey. Available: http://www.statssa.gov.za/publications/P0320/P03202013.pdf [Accessed on 25 February 2016]
[9] PRASA, 2016. Passenger Rail Agency of South Africa Corporate Plan 2016-2019. Available: http://www.prasa.com [Accessed on 25 February 2018]
[10] Fourie, C. J. & Tendayi, T. G. 2016. A decision-making framework for effective maintenance management using life cycle costing (LCC) in a rolling stock environment. *South African Journal of Industrial Engineering*, *27*(4), 142-152.
[11] Conradie, P.D.F., Fourie, C. J., Vlok, P. J. & Treurnicht, N. F. 2015. Quantifying system reliability in rail transportation in an ageing fleet environment. *South African Journal of Industrial Engineering*, *26*(2), 128-142.
[12] Matzopoulos, R. and Lerer, L.B., 1998. Hours to hell and back: A social epidemiology of railway injury in a South African city, 1890–1995. *Social science & medicine*, *47*(1), pp.75-83.
[13] Ryneveld, P. Van, 2008. 15 Year Review of Public Transport in South Africa with emphasis on metropolitan areas.
[14] Van der Meulen, R. D. (2010). How might the railway Renaissance evolve in South Africa?. In *Proc. 29th Annual Southern African Transport Conference* (pp. 630-639).
[15] Wilkinson, P. 2001. Towards a new rationality in the planning and management of passenger transport systems in South African cities: A critical analysis of contemporary policy discourse. University Of Cape Town.
[16] Maxwell, J.A., 2013. Qualitative Research Design: An interactive approach. 3rd Edition. London New Delhi: Sage Publication
[17] Glaser, B. & Strauss, A. 1967. The discovery grounded theory: Strategies for qualitative inquiry. Aldin, Chicago.
[18] Creswell, J.W., 1994. Research Design: Qualitative & Quantitative Approaches. Thousand Oaks, California: Sage Publications
[19] Pandit, N. R., 1996. The creation of theory: A recent application of the grounded theory method. *The qualitative report, 2(4),* pp. 1–15.
[20] Easterby-Smith, M., Thorpe, R. & Lowe, A. 2002. *Management Research: An Introduction*. 2nd edition. London: Sage Publications
[21] Payne, S. 2007. *Analysing qualitative data in psychology,* London: Sage
[22] Charmaz, K., 2014. Constructing Grounded Theory. 2nd Edition. London: Sage Publications (Ltd).
[23] Mogalakwe, M., 2006. The use of documentary research methods. *African sociological review, 10(1),* pp.221–230.
[24] Holton, J. A. 2008. Grounded theory as a general research methodology. *The grounded theory review*, *7*(2), 67-93.
[25] Brassard, M. 1989. The Memory Jogger Plus+ TM: Featuring the seven management and planning tools. Goal/Qpc.
[26] Corbin, J. & Strauss, A. 2008. Basics of qualitative research: Techniques and procedures for developing grounded theory.
[27] Beer, S. 1994. Decision and Control: The meaning of operational. *The meaning of operational research and management cybernetics*.
[28] Braun, W. 2002. The System Archetypes. *The systems modelling workbook*, pp.1–26.
[29] Clayton, C. & Michael, R. 2003. Management theory. *Harvard Business Review*, (September), pp.66–74.
[30] Swanson, L. 2001. Linking maintenance strategies to performance. *International journal of production economics*, *70*(3), pp.237-244.
[31] Wolstenholme, E.F. 2003. Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review, 19(1),* pp.7–26.
[32] Wolstenholme, E.F. & Corben, D.A., 1993. Towards a Core Set of Archetypal Structures in System Dynamics. *Proceedings of the 1993 International System Dynamics Conference*, p.583-588.

LEAPFROGGING TO INDUSTRY 4.0: LESSONS FROM THE HEALTHCARE INDUSTRY

W. Maisiri[1]* & L. van Dyk[2]

[1]Department of Industrial Engineering
North West University, South Africa
wmlisper27@gmail.com

[2] Department of Industrial Engineering
North West University, South Africa
Liezl.VanDyk@nwu.ac.za

ABSTRACT

The first three industrial revolutions, namely Industry 1.0 (mechanization), Industry 2.0 (mass production through electrification) and Industry 3.0 (automation), lasted several hundreds of years. Now the era of Industry 4.0 (the fourth industrial revolution) has come. Industry 4.0 encompasses a paradigm shift from automated systems to intelligent systems, with the objective being to optimize and achieve sustainable systems. Industry 4.0 is becoming standard practice in developed countries. In contrast, developing countries still have to catch up with the industrial revolution phases that have played out in developed countries already. This presents a leapfrogging opportunity for developing countries to go straight into Industry 4.0. The impact of this could be far-reaching, as it could enable developing countries to attain the goal of inclusive and sustainable industrial development. The purpose of this article is to critically investigate the opportunities and challenges of Industry 4.0 for developed and developing countries using the private and public health sector as a point of investigation. The article presents a comprehensive literature review on Industry 4.0 focusing on healthcare systems. Applying this to the private and public healthcare sectors could aid the identification of leapfrogging opportunities beyond the healthcare industry.

[1] The author is enrolled for an PhD degree in the Department of Industrial Engineering, North West University, South Africa
*Corresponding author

1. INTRODUCTION

In the course of history, the world has seen three major industrial revolutions, namely mechanization using water and steam (Industry 1.0), mass production in assembly lines using electric power (Industry 2.0) and automation using electronics and information technology (Industry 3.0) [1-4]. These three industrial revolutions each lasted several years, and now the era of Industry 4.0 (I4.0), the fourth industrial revolution, has come.

Development touches all areas of human existence. As such, the healthcare sector has also evolved, although in its own fashion. One could analogously identify the revolutions in healthcare Health 1.0, Health 2.0 and Health 3.0, with Health 4.0 just having been introduced.

I4.0 comprises a paradigm shift from automated systems to intelligent systems, with the objective being to promote sustainable production and agile supply chains [3]. I4.0 is becoming standard practice in developed countries. Major strides in advocating and promoting I4.0 has been witnessed in countries such as Germany, the United States, Japan and Korea [3].

United Nations Sustainable Development Goal number nine promotes inclusive and sustainable industrialization [5]. However, there is a noticeable industrialization gap between developed and developing countries. The current status shows that developing countries have to catch up with phases of industrial revolution that have been surpassed in the developed world. This presents leapfrogging opportunities for developing c ountries to go straight to Industry 4.0 to achieve the goal of inclusive and sustainable industrialization.

The principles of I4.0 are equally applicable to the healthcare systems [2]. The South African healthcare sector comprises of private healthcare, which is well resourced and advanced, and public healthcare, which has noticeable shortages of resources and is under-developed [6-8]. In this study, the private and the public health sectors are used to represent the developed and developing countries respectively.

The purpose of this paper is to critically investigate the opportunities and challenges of I4.0 for both developed and developing countries. A strengths, opportunities, aspirations and results (SOAR) analysis is applied to identify possible leapfrogging opportunities.

The paper is organized as follows: first, Section 2 presents a comprehensive literature review on I4.0 and the healthcare system. Section 3 discusses the South African healthcare system. The application of the SOAR analysis to identify the leapfrogging opportunities and challenges that Industry 4.0 presents is performed in Section 4. Section 5 examines I4.0 opportunities and challenges for developing and developed countries. Section 6 presents a discussion on leapfrogging opportunities to I4.0 in developing countries and Section 7 presents the conclusion and suggestions for further work.

2. INDUSTRY 4.0 AND THE HEALTHCARE SYSTEM

Industry 4.0 design principles are focused on interoperability, virtualization, decentralization, real-time capability, service orientation and modularity [9]. This section presents the overview of I4.0 in Section 2.1 and Health 4.0 in Section 2.2. The application of I4.0 in healthcare system (Section 2.3) is presented last.

2.1 Overview of Industry 4.

Figure I shows the phases of the industrial revolution, including the recently advocated I4.0. The ideas at the base of I4.0 emerged in Germany as a high-tech strategy to promote digitization of its manufacturing industry with the goal to remain competitive [10, 11]. The 2013 Hannover industry fair resulted in the coining of the term I4.0, now widely used in European countries and accepted in some Asian countries such as China [1, 11]. In other countries, such as the United States, the term Internet of Things (IoT), Internet of Everything (IoE) or Industrial Internet is used instead [1, 11].

McKinsey [12] defines I4.0 as “digitisation of the manufacturing sector, with embedded sensors in virtually all product components and manufacturing equipment, ubiquitous cyber physical systems, and analysis of all relevant data.”

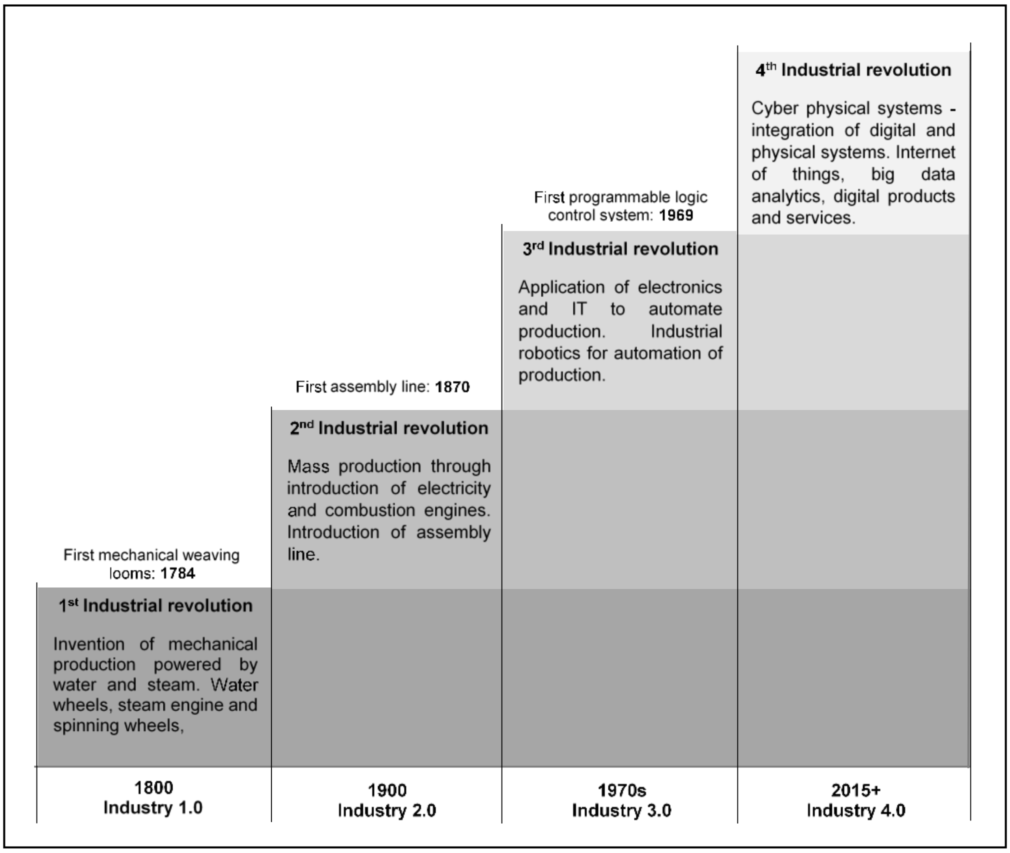


Figure I: Phases of Industrialization [10, 11, 13]

I4.0 in essence entails integrating a set of technologies to permit ecosystems of intelligent, autonomous and decentralized factories and integrated products and services [2]. I4.0 is driven by technologies such as cyber-physical systems (CPS), IoT, 3D-printing, big data analytics (BDA), autonomous systems, cloud computing and mobile solutions [2, 14, 15].

For the purposes of this discussion, the terms I4.0 and digitization are used interchangeably, while CPS is interchanged with digitization of production. Embedded software is used to allow communication between all production components to produce smart products. These components know how they are made and what they will be used for. A digital factory is characterized by CPS, smart robots and machines; a new quality of connectivity; big data; energy efficiency; decentralization and virtual industrialization [15].

I4.0 can be understood as the application of the generic concept of CPS to the interaction between humans, machines and products during production processes enabled by cyber-physical production systems (CPPS) [16]. Thoben simplifies I4.0 by stating that it is a paradigm shift from automated systems to intelligent systems [3].

MacDougall clearly points out that the term smart industry can be used interchangeably with I4.0. I4.0 can be described as the technological evolution from embedded systems to CPS, and this will radically transform industry and the production value chain and business model. I4.0 drives a shift from centralized to decentralized production, which constitutes a reversal of convectional production process logic [17].

I4.0 strategies seek to shift industries from being manufacturers to service providers that permit individualization and personalization for consumers [9]. The influence of I4.0 on individualization and virtualization stretches across different industrial fields [9].

According to Rüßmann et al., I4.0 is powered by nine technological advances, namely autonomous robots, simulation, horizontal and vertical system integration, the industrial IoT, cybersecurity, the cloud, additive manufacturing, augmented reality and big data and analytics [4]. Thuemmler and Bai further state that the primary elements of Industry 4.0 are CPS, IoT, Internet of Services (IoS) and smart factories [9].

Interoperability, virtualization, decentralization, real-time capability, service orientation and modularity are the design principles of I4.0. Thuemmler and Bai note that I4.0 design principles are highly applicable in the healthcare industry [9].

2.2 Health 4.0

The healthcare system has transformed from Health 1.0 through to Health 4.0 in a manner that is similar to industry, but not identical. Industry 1.0, Industry 2.0 and Industry 3.0 derived technologies do not directly correlate with Health 1.0, Health 2.0 and Health 3.0 respectively. However, there is direct core relationship between Health 4.0 and I4.0 with Health 4.0 being an I4.0 strategic concept applied to the healthcare systems domain. Figure II shows the paradigm shift in the healthcare system over the years.

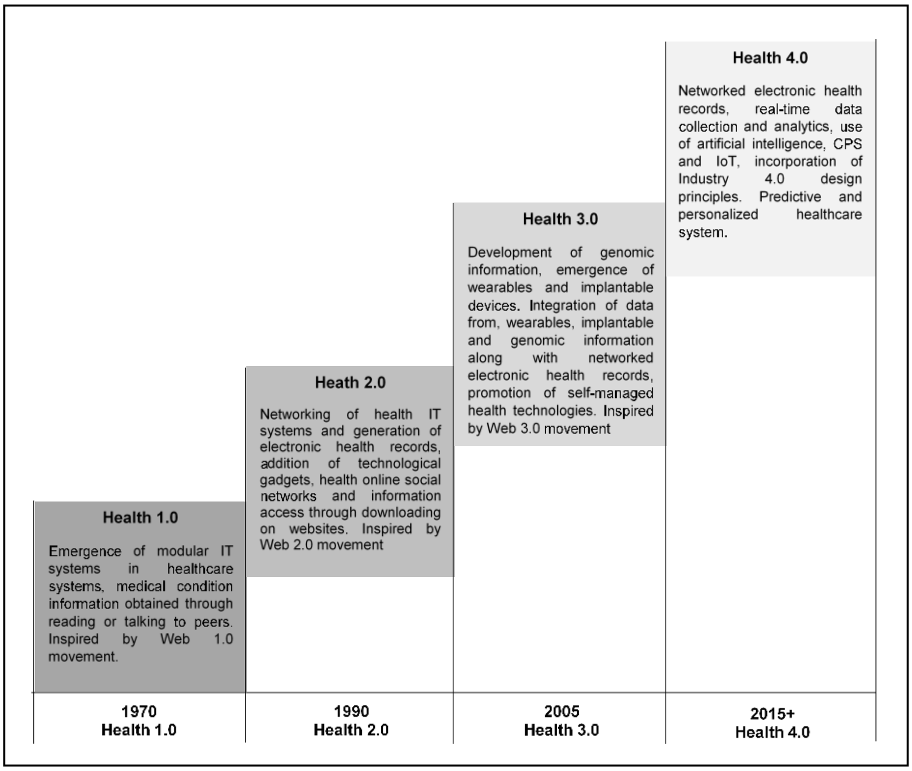


Figure II: Healthcare system paradigm shifts (Adopted from [9, 18-22])

Health 1.0 to Health 3.0 were inspired by Web 1.0 to Web 3.0 respectively [19]. Health 1.0 is the era of the introduction of modular IT systems in the healthcare industry [22]. Knowledge about medical conditions came from reading or talking to family and friends and the medical provider was the final reference [18, 19].

Health 2.0 allowed people to access information to discuss with their caregivers by means of websites [18]. Technological gadgets such as wristbands, back bands and sleep monitors were added to ensure monitoring and quantifying of and reporting on health matters [20]. Patients gained the privilege of reading online social network sites and getting input on how their peers have managed their sickness [21].

Health 3.0 concentrates on the consumer patient model [19]. Health 3.0 focuses on ensuring a better quality of life and promoted networking between people with related health matters using internet platforms [19]. People are empowered to take responsibility for their health and health treatments with the availability of secured technologies and support groups [21]. Health 3.0 sought to destroy patient dependency syndrome, promote patient independence and make the patient part of the medical team [21]. Health 3.0 promoted communication and self-managed health technologies [21].

The concept of Health 4.0 is derived from I4.0 as a strategic concept for the healthcare systems domain. Progressive virtualization to enable the automation and personalization of health care in next to real time for patients, professionals and formal and informal carers is attainable through Health 4.0 [9, 23]. The use of CPS, IoT, IoS, IoP, cloud computing and the development of 5G mobile communication networks will enable the personalization of healthcare [9]. Further to this, Industry 4.0 design principles stand to be the main goal of Health 4.0 [23].

Health 4.0 facilitates a digitally connected healthcare system that provides efficient healthcare services and produces sustainable improvements in medical care for patients and increased profitability for care providers [24].

Health 4.0 is driven by electronic health records, artificial intelligence (AI), real-time data from wearable devices and data analytics [22]. A predictive and personalized healthcare system can be accomplished through emphasis on collaboration, coherence and convergence [22]. The availability of real-time data and data analytics permits timely and innovative diagnoses and medical response [22]. Health 4.0 aims to shift the entire healthcare system to a value-based system with measurable outcomes and provides proactive prevention[22].

2.3 Application of Industry 4.0 in the healthcare systems

CPS is at the core of I4.0 and enables the linkage between the physical world and the virtual world to create the IoT [9]. Medical CPS is an approach that is positioned to permit smart healthcare systems to monitor, process and make independent decisions in the absence of a healthcare provider [23]. An example of the application of CPS in the health domain is "the connection of body area networks and sensors in smart pharmaceuticals to disease management platforms with either autoregulatory feedback loops or feedback via accessories such as smart phones" [9].

IoT, cloud computing and BDA are key I4.0 technologies that can universally facilitate efficient and enhancing sustainable healthcare systems through the personalization of healthcare and improved basic nursing care [25-28].

IoT is the connection of unlimited, smart, physical and virtual objects with distinct identities with internet to generate an infinite CPS framework [25, 29-31]. IoT solutions involve capturing, monitoring and transmitting data to a public or private cloud to facilitate accessible and efficient automation [29].

Diverse medical sensors, advanced devices such as heart pressure watches, smartphones, imaging devices, personal digital assistants and electronic health records (EHR) are integral elements of the healthcare IoT system [29, 30]. Healthcare providers can noticeably improve the quality of care and health outcomes by using health data monitored and collected by these devices [29, 31].

Figure III shows a typical healthcare IoT for a hospital. A patient can access their EHR and medical histories stored securely in a cloud by scanning their identity card. At the other end, healthcare professionals can access similar records on their mobile medical devices [31]. This will facilitate efficiency and effectiveness in the healthcare sector.

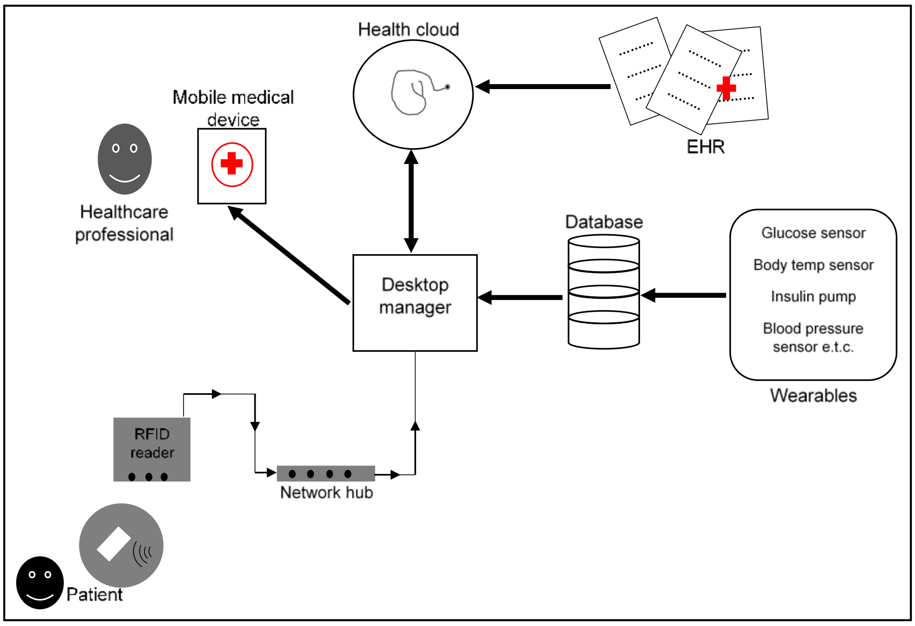


Figure III: Schematic diagram for a typical IoT healthcare system [31]

IoT and cloud computing application in the healthcare systems leads to significant improvements in the provision of patient-centric healthcare and reduces healthcare cost, leading to a sustainable healthcare industry [25, 30]. Medical IoT can noticeably assist in improving agility in remote health deliverables [29].

IoT promotes remote health monitoring, remote diagnostics, chronic diseases management and independent care for the elderly [29, 31]. The challenge of patients not complying with medication and treatment by healthcare providers can potentially be addressed by the application of IoT. Further to this, IoT will significantly add value and change the landscape in aspects of authenticating medicine, monitoring drug supplies and providing efficient scheduling of available resources to ensure best use for more patients [29].

Challenges associated with medical IoT include the need to detect and manage mobile medical sensors at any time and maintaining robust protection and security of patients' information [30, 32]. Other issues to be considered in healthcare IoT is reliability and quality of care [30]. Medical IoT faces the challenge of non-existent standardization of proprietary protocols to facilitate communication between sensors from different manufactures [31].

The emergence of wearable medical devices with sensors resulted in the generation of large volumes of medical data on human physical and mental health [26, 29, 30, 32]. Further to this, digital devices such as smartphones and electronic health records (EHR) add to the available medical data [27, 31]. The large volumes of medical data of high velocity, intricacy and of a wide variety requiring advance techniques and technologies to analyse can be referred to as big data [27, 30, 32].

The big data explosion in the healthcare sector provides an opportunity for BDA, an I4.0 concept [3, 27, 31]. BDA provides noticeable possibilities to improve healthcare, save lives and minimize costs by evaluating medical data to attain an in-depth understanding of different views on human life [27].

The application of BDA in healthcare can result in providing vital support for creating individually customized healthcare where an individual receives the right health intervention and the health problem can be identified in an evidence-based manner [27, 29]. BDA promotes evidence-based decisions, thorough and perceptive diagnoses and treatments leading to improved quality of healthcare [27].

BDA has assisted healthcare players with tracking adherence to treatments and monitoring trends that promote wellness in individuals and populations [27, 29]. Further to this, BDA can facilitate identifying diseases at an early stage and exposing healthcare fraud effectively and efficiently [27].
The implementation of BDA has been noticeably slow in the healthcare industry due to diverse and interlinked technological, legal and ethical problems associated with the industry [29, 31]. Data assurance, known as veracity, is potentially a problematic factor in healthcare BDA. The quality of BDA and its outcomes are major concerns in healthcare since life or death decisions are made [3, 27]

Patients' lack of trust in sharing personal data stands to be one of the obstacles to implementing BDA. Latif et al. [29] suggest a patient-centric big data-enabled healthcare system, shown in Figure IV, which will give patients control over their personal information [29]. The suggested patient-centric model can promote information sharing in the healthcare system, thus enabling the BDA.

Healthcare BDA is faced with the challenge of implementing real-time analysis, which is an essential aspect in this sector. The gap that exists between data collection and data analysis has to be addressed [27].

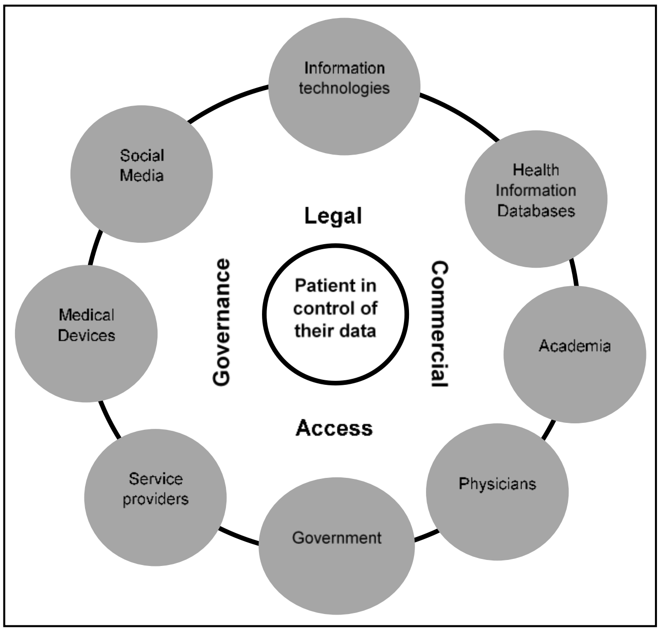


Figure IV: Patient-centred big data-enabled healthcare ecosystem [29]

Inadequate knowledge of patients' historical health and personal preferences result in the medical practitioners offering inferior healthcare services. Artificial intelligence (AI) and machine learning (ML) are the cornerstones of systems that are driven by scientific and statistically driven data that will result in optimal healthcare outcomes [2, 29]. Predictive AI algorithms can be used to analyse data obtained from patients' records, thus minimizing re-admission rates [31]. AI and automated systems in healthcare systems could provide vital assistance to doctors and this would significantly complement and augment their expertise [2, 29]. In this way accuracy, effectiveness and efficiency in the healthcare system can significantly improve as a nurse or physician would be able to perform at a level of a specialist [29]. AI will positively influence disease control and medicine by permitting early detection [2].

The use of mobile apps in healthcare is on the rise and they are helping patients to manage their health outcomes, to find healthcare providers and to improve the health of their lifestyle [27, 31]. Mobile devices and advanced wireless technologies can stimulate various healthcare solutions for personalized care. Information provided by the mobile devices provide data to be used in big data analytics and AI to ensure smart healthcare solutions. Further to this, smartphones and tablets can be used as IoT terminals that monitor healthcare data in real time and communicate with physical sensors in the medical IoT network [29].

The integration of IoT, cloud computing, BDA, mobile solutions and wireless devices will significantly contribute in mitigating problems of uneven distribution of healthcare resources, healthcare disparities, and the growing number of patients with chronic diseases and increasing medical expenses.

3. THE SOUTH AFRICAN HEALTHCARE SYSTEM

The South African healthcare system divides into a private and public healthcare sector. A significant equality gap exists between the private and public healthcare sectors in terms of resources [6, 7]. Private healthcare is perceived to be over-serviced while covering a minority population, while the public healthcare is under-serviced with majority coverage [8].

The private healthcare sector is driven by profit-making [6, 7]. Practitioners and institutions charge noticeably more than the public sector for the same medical procedures and medication [8, 33]. Private healthcare is furthermore characterized by considerable resources and mainly accommodates the affluent population who can afford the services [6, 7, 34, 35]. The private healthcare sector is perceived to provide good quality of care due to good financial accessibility and human resources [6, 7]. It is approximated that 80% of South Africa's health specialists are in the private healthcare sector [35].

In contrast, the public health sector is under-resourced with old facilities and infrastructure and financial problems. This sector serves the majority of South Africans [6, 7, 34]. Inadequate quality of service and a shortage of qualified human resources stands out in the public healthcare sector [7, 8].

Users of public healthcare sector are reported to have noticeable challenges with drug availability, incorrect diagnoses, the cleanliness of facilities, rude staff, long waiting times and rushed appointments [6, 7]. In contrast, the private healthcare sector is characterized by short waiting times, thorough consultation and proper disease control and prevention [6, 7].

The South African healthcare sector spends 3.5% more of its GDP than the recommended World Health Organization (WHO) threshold [33, 36]. Figure V shows how the South African healthcare GDP expenditure compares to the WHO recommended threshold.

Figure VI further compares the private and the public healthcare sector GDP expenditure to the population serviced. Note that the private healthcare sector spends 4.4% of the GDP and services 16% of the population. The public healthcare sector spends 4.1% of the GDP and services 84% of the population [33, 36]. This points to a gross inequality in how resources are shared between the private and public healthcare sectors.

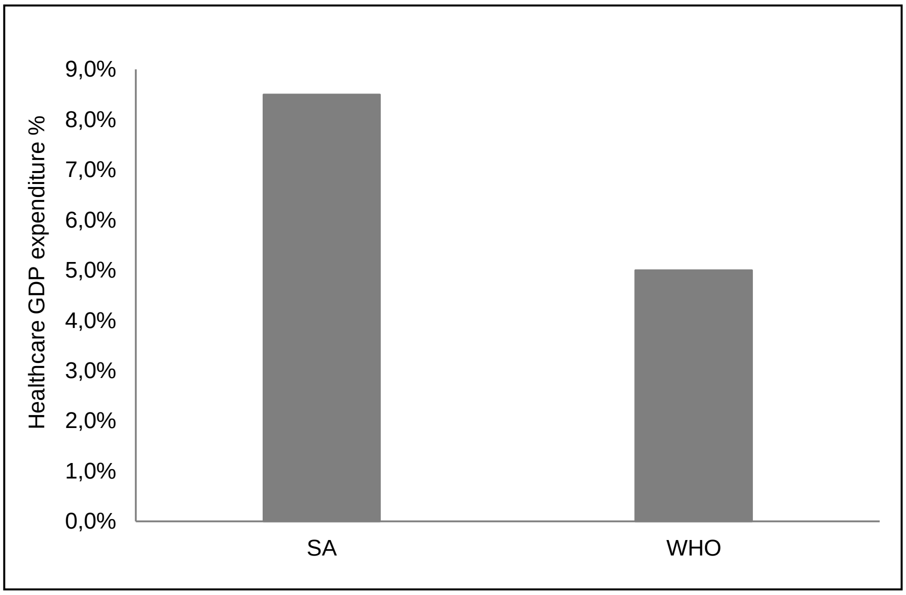


Figure V: GDP expenditure comparison (Adopted from [36])

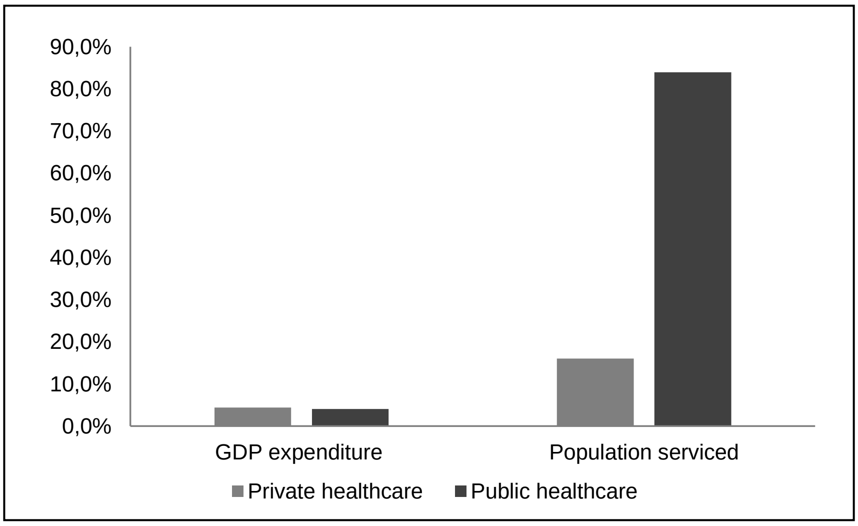


Figure VI: Comparison of private and public healthcare sectors (Adopted from [36])

In a move to transform the healthcare sector in South Africa, the government announced in 2011 that they intend to introduce National Health Insurance (NHI). A policy that introduces NHI could revolutionize healthcare delivery in South Africa by ensuring universal health coverage [37]. The envisioned policy to provide NHI is currently in the drafting phase, with several draft policies and pilot projects currently on the table. Although the policy would affect medical aids, the government insists that the introduction of NHI is not intended to destroy the private healthcare sector, but to ensure quality healthcare for the majority [35].

A key component of the NHI implementation plan is to strengthen the public health sector in South Africa [37]. The government is aiming to create an integrated healthcare system that is affordable and reachable to the majority of South Africans, thus addressing the existing inequalities [7].

A partnership between the private and public healthcare sectors is inevitable if we want to address the challenges encountered in the healthcare sector [34]. The resources of the private and public healthcare sectors would have to be combined to achieve the goal of equal access to quality healthcare in South Africa [34, 36]. The move will result in a leap to digitization of the healthcare sector in South Africa, thus addressing the current quality of service in the public health sector.

4. LEAPFROGGING TO INDUSTRY 4.0: OPPORTUNITIES AND CHALLENGES

4.1 Overview of the leapfrogging concept

The concept of technological leapfrogging entails that developing countries skip investment in obsolete technologies and investing in future technologies [38]. Perkins states that "leapfrogging implies a development strategy for industrialising countries to bypass the dirty stages of economic growth through the use of modern technologies that use fewer resources and/or generate less pollution" [39]. It can be argued that leapfrogging is an advancement strategy for developing countries skip certain phases of industrial revolutions jump to I4.0.

Technological leapfrogging can also mean skipping ahead to become a forerunner rather than just skipping over generations of technology [38]. The problems and challenges developing countries face present innovation opportunities to address the challenges [40]. Sarabhai further argues that the available knowledge presents leapfrogging opportunities for developing countries [41]. The country's capacity to technologically leapfrog is significantly influenced by human capital and economic status. Skills development is therefore a vital component in facilitating technological diffusion [38].

The strategic thinking process that involves SOAR analysis focuses on creating future aspirations and desired results, thus it can be applied in analysing technological leapfrogging opportunities. In this study, the SOAR analysis model was applied to analyse the opportunities and challenges of leapfrogging to Industry 4.0 using the South African healthcare system as a point of reference.

4.2 Strengths, opportunities, aspirations and results (SOAR) analysis

SOAR analysis is a dynamic strategic thinking tool used by organizations to identify their strengths and opportunities when they plan for the future [42-44]. The SOAR analysis uses an appreciative inquiry approach to concentrate on what works [42-46]. This model is driven by the desire to remain focused on the targeted positive outcomes by translating challenges into opportunities [43, 44].

The SOAR analysis can be used as an alternative to the strength, weakness, opportunities and threats (SWOT) analysis [45]. In contrast to a SWOT analysis, a SOAR analysis focuses on co-producing the desired future by means of a process of inquiry, imagination, innovation and aspiration [42, 43, 47]. SOAR remodels the SWOT analysis into a progressive framework and creates a transformational strategic thinking process [45].

The SOAR analysis model can be used to investigate initiatives, develop strategic plans and to focus and redirect efforts and resources [42, 45]. It does not entirely leave out dealing with challenges, but rather reframes the negative issues as opportunities [45].

The SOAR analysis has significant capacity to create and manage change and map organizational strategic planning [45]. This model can be applied to all industries and has been used by corporations, municipal governments and the healthcare sectors [42, 45].

Table 1 shows a graphical presentation of the SOAR analysis model. The process of SOAR analysis begins by understanding strengths and formulating a conception of the potential opportunities, followed by rising innovation in the future by aspirations and finally thinking inspiration to realize results [45, 48].

Table 1: Graphical presentations of SOAR analysis [42, 44-46]

Strategic inquiry	**STRENGTHS** What the organization does best, key assets, resources, capabilities and accomplishments.	**OPPORTUNITIES** Conditions that can be used to influence success. These include partnerships that can lead to success, threats that can be reframed as opportunities.
Appreciative intent	**ASPIRATIONS** Preferred future – build on the current strengths and challenge the current situation.	**RESULTS** Measurable outcomes that prove that goals and aspirations have been achieved.

4.3 SOAR analysis of the South African healthcare system

Digital transformation trends are becoming global and universal and cannot be overlooked if organizations and companies have to match the rest [11]. This presents the need for developing countries to formulate technological leapfrogging strategies.

This section applies the principles of a SOAR analysis in evaluating the opportunities and challenges of leapfrogging to I4.0 using the South African healthcare system as a point of reference. The strength and opportunities were derived from the private and public healthcare sectors, which represents developed and developing countries respectively. Table 2 gives an overview of the SOAR for the South African healthcare system.

Table 2: Overview of the South African healthcare system SOAR analysis

STRATEGIC INQUIRY	
Strengths	**Opportunities**
• A well-resourced private healthcare sector [7, 33-35]. • Provision of good quality of care due to good financial accessibility in South African private health sector [7, 33]. • Innovations such as an electronic bed management system for the public hospitals in the Gauteng province [49]. • Introduction of the NHI in South Africa [7, 33, 36, 37]. • Significant GDP expenditure in healthcare systems higher than the WHO threshold [33, 36]. • Positive response from consumers to digital services as demonstrated by tendency to access services through mobile devices [50]. • Wide use of mobile technology [2].	• Inadequate quality of service and shortage of qualified human resources in the public healthcare sector [7, 8]. • Noticeable challenges with drug availability, incorrect diagnoses, cleanliness of facilities, rude staff, long waiting times and rushed appointments [6, 7]. • Old facilities and infrastructure in public healthcare sector coupled with serving the majority of South Africans [6, 7, 34]. • Availability of standalone mobile health with the growing use of smartphones among clinicians and patients [50]. In South Africa use of mobile messaging platforms such as MomConnect has seen significant adoption by users [50]. • Potentially disruptive technologies and apps are available in standalone bases and are not connected to medical practitioners, resulting in technological progression occurring in fragmentation and silos [50].
APPRECIATIVE INTENT	
Aspirations	**Results**
• Ensuring healthy lives and promoting well-being for every person of every age [5]. • Universal and quality healthcare coverage for every individual [35, 37]. • Reduction of the inequality of resource sharing in the private and public healthcare sectors[7]. • Achieving a patient-centric healthcare system. • Agile remote healthcare deliverables.	• Partnership between the private and public health care sector [34, 36, 50]. • Complete integration of fragmented apps and technologies [50]. • Monitored healthcare in the homes of patients through personalized healthcare system [50]. • Affordable healthcare for every person. • Universal access to healthcare by all people [5]. • Integrated EHR.

5. THE OPPORTUNITIES AND CHALLENGES THAT INDUSTRY 4.0 HOLDS FOR DEVELOPED AND DEVELOPING COUNTRIES

I4.0 presents significant opportunities for both developed and developing countries. However, I4.0 could pose significantly more challenges in developing countries than in developed countries [2]. Table 3 compares the generic opportunities and challenges for both developed and developing countries.

Table 3: I4.0 opportunities and challenges

Industry 4.0 in developed countries	
Opportunities	**Challenges**
• The implementation of I4.0 can help to offset excessive production costs and solve the issue of an ageing population [2]. • I4.0 enables interconnected, flexible and efficient supply chain processes [2]. • I4.0 offers more extensive and intelligent solutions to the global demand mass	• Data security and privacy concerns arising from the need for data sharing and smooth connectivity in creating a digital ecosystem [1-3, 10, 27, 53]. • Problems of I4.0 standardization to permit communication and seamless data sharing in the IoT system [1, 2, 31].

customization in a bid to increase quality and efficiency [1, 2] • The use of BDA results in identifying faults and shortcomings leading to systems and process optimization. This results in efficient use of resources [1]. • I4.0 facilitates moving production plants offshore and manufacturing goods locally due to advanced automation technologies results [51]. • I4.0 facilitates sustainable economic, social and environmental industrial value creation [52].	• Need for intense workforce transformation to meet the I4.0 skills level requirements due to significant penetration of digital labour such as intelligent machines and advanced robots [1, 53]. • Awareness and readiness challenges due to uncertainty issues. Business assumptions should be challenged and tested [53].
Industry 4.0 in developing countries	
Opportunities	**Challenges**
• I4.0 will result in sustainable manufacturing and consumption trends in developing countries [2]. • Independent industry due to reduction in the delocalization pattern of manufacturing companies by developed countries [51]. • Minimum resistance to disruptive technology due to the absence of infrastructure legacy [51]. This presents an opportunity to introduce I4.0 ready infrastructure [11]. • Unique opportunity for technological leapfrogging like for example the adoption of mobile phones in Africa [2, 51]. • Rising need of alternative sources of electrical power can result in leapfrogging the production of clean energy [51]. • Exposure to cutting-edge innovations generates an opportunity for skills and knowledge transfer [41, 51]. • Industry and government has an opportunity to participate in training and developing the digital workforce of the future [11].	• Shortage of professionals with required set skills for I4.0 [2, 11]. I4.0 requires skillsets such as robotic programming and big data analytics. These are available in pouches in developing countries [53]. • The current economic challenges that developing countries in Africa face divert attention from innovation and focus on cost saving [2, 11]. • Noticeable problems of monetary funding and business management buy-in which are fundamental to sign up to I4.0 roadmaps [53]. • Significant challenges related to connectivity and accessibility due to old and disparate IT infrastructure, mainly in Africa [2, 11]. • Unreliable and unstable electricity network impairs implementation of I4.0 strategies [2]. • Absence of regulatory framework that facilitate implementation of I4.0 [2]. • Job security and inequality due to advancement in artificial intelligence and advanced robotics [2]. Adoption of I4.0 will require a fundamental change of employment and skills. • A lack of awareness of products and how they function contribute to public resistance to the adoption of new technologies [38].

Successful implementation of I4.0 can fast-track the attainment of inclusive and sustainable industrialization and embrace innovation in developed and developing countries [2]. Collaboration between all stakeholders and joint action between developed and developing countries will lead to successful and inclusive implementation of I4.0.

6. DISCUSSION: DEVELOPING COUNTRIES LEAPFROGGING OPPORTUNITIES TO INDUSTRY 4.0

The opportunities, aspirations and results cited in Table 2 present the need to leapfrog to I4.0. Translating the public healthcare challenges into opportunities and viewing the aspirations and results desired in the healthcare would provoke innovation in the direction of achieving Health 4.0, an I4.0 health domain.

I4.0 technologies such as IoT, BDA and AI are ideal for addressing the opportunities and achieving the aspirations and results presented in Table 2. Combining resources in the private and public healthcare sector offers the main opportunity with respect to achieving a technological leap in healthcare. This will drive the country towards achieving the sustainable development goal in the healthcare industry.

Technological leapfrogging in healthcare systems requires private sector, public sector and all other stakeholder collaboration and broader acceptance of new technologies [50]. Complete integration of fragmented apps and technologies will result in a leap towards digital health, which will significantly improve healthcare systems and affordable access to healthcare by the majority [50].

Over the past years, technological advancement has been witnessed in African countries. The increasing use of the cloud, telemedicine and virtual reality in healthcare systems shows the potential for technological leapfrogging. Further to this, there have been pockets of IoT applications on the African continent, showing the continent’s ability to attain technological advancement.

Perceiving challenges as opportunities can be a tool to inspire innovation, thus achieving technological leapfrogging that can be applied beyond the healthcare systems. For example, the challenge of old infrastructure can be viewed as an opportunity as it means that healthcare is not burdened by infrastructure legacy issues, this in turn leading to minimum resistance to embracing change [11]. The challenges that developing countries face in relation to implementing I4.0 presented in Table 3 can be translated into opportunities that can drive leapfrogging to I4.0.

Collaboration between government, industries and research institutions provides significant potential to directly implement specific I4.0 applications and to build distinctive local digital products and services that will surpass global competitors in the future [11].

7. CONCLUSION AND SUGGESTIONS FOR FURTHER WORK

The purpose of this paper was to investigate the opportunities developing countries have to leapfrog to I4.0 using healthcare system as a point of reference. The article presented a comprehensive literature review on I4.0 focusing on the healthcare systems. Thereafter a SOAR analysis was performed on the South Africa health system. This leads to the conclusion that the challenges in the healthcare system and other sectors can be translated into opportunities that drive innovation towards I4.0 in developing countries. I4.0 technologies have significant potential to drive sustainable healthcare systems and to be applicable to other domains.

Further recommended work includes evaluating the awareness of I4.0 in healthcare systems in developing countries using South Africa as point of reference and developing an I4.0 leapfrogging maturity model for the South African healthcare system.

REFERENCES

[1] **Finance, A.T.C.C.** 2015. Industry 4.0 Challenges and solutions for the digital transformation and use of exponential technologies.

[2] **United Nations Industrial Development Organization**. 2017. Industry 4.0: Opportunities behind the challenge. Austria.

[3] **Thoben, K.-D., Wiesner, S. & Wuest, T**. 2017. “Industrie 4.0” and smart manufacturing: A review of research issues and application examples, *International Journal of Automation Technology*, 11(1) pp 4 - 16.

[4] **Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P. & Harnisch, M. 2015**. Industry 4.0: The future of productivity and growth in manufacturing industries. Boston Consulting Group.

[5] **United Nations.** 2015. RES/70/1. Transforming our world: The 2030 agenda for sustainable development. Seventieth United Nations General Assembly, New York, 25.

[6] **Ranchod, S., Adams, C., Burger, R., Carvounes, A., Dreyer, K., Smith, A., Stewart, J. & van Biljon, C**. 2017. South Africa’s hospital sector: Old divisions and new developments, *South African Health Review*, 2017(1), pp 101-110.

[7] **Young, M**. 2016. Private vs. public healthcare in South Africa.

[8] **Gqirana, T.** 2016. Private healthcare in South Africa a 'sick situation', says Motsoaledi [Online]. Available: https://www.timeslive.co.za/news/south-africa/2016-03-12-private-healthcare-in-south-africa-a-sick-situation-says-motsoaledi/ [Date of access: 9 April 2018]

[9] **Thuemmler, C. & Bai, C.** 2017. Health 4.0: Application of Industry 4.0 Design Principles in Future Asthma Management. Health 4.0: How Virtualization and Big Data are Revolutionizing Healthcare. Springer. Pp 23-37.

[10] **Zhou, K., Liu, T. & Zhou, L.** 2015. Industry 4.0: Towards future industrial opportunities and challenges. *In Fuzzy Systems and Knowledge Discovery (FSKD), 2015 12th International Conference on organised by: IEEE*. Pp 2147-2152.

[11] **Deloitte.** 2016. Industry 4.0: Is Africa ready for digital transformation? South Africa.

[12] **Wee, D., Kelly, R., Cattel, J. & Breunig, M**. 2015. Industry 4.0—How to navigate digitization of the manufacturing sector. McKinsey & Company.

[13] **Schrauf, S. & Berttram, P**. 2016. Industry 4.0: How digitization makes the supply chain more efficient agile and customer-focused. PWC.

[14] CGI-**Group.** 2017. Industry 4.0: Making your business more competitive. CGI Group.

[15] **Berger, R**. 2016. Digital factories. The renaissance of U.S automotive industry. In Think Act beyond Mainstream.

[16] **Albers, A., Gladysz, B., Pinner, T., Butenko, V. & Stürmlinger, T**. 2016. Procedure for defining the system of objectives in the initial phase of an industry 4.0 project focusing on intelligent quality control systems. *Procedia Cirp,* 52. pp 262-267.

[17] **MacDougall, W**. 2014. Industrie 4.0: Smart manufacturing for the future: Germany Trade & Invest.

[18] **Nash, D.B.** 2008. Health 3.0. *Pharmacy and Therapeutics*, 33(2) pp 69.

[19] **Deloitte.** 2013. Healthcare 3.0: Healthcare for the new normal [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/sg/Documents/risk/risk-sea-healthcare-3.0-healthcare-for-the-new-normal.pdf [Date of access 12 April 2018]

[20] **Clemensen, J., Danbjørg, D.B., Syse, M. & Coxon, I.R.** 2016. The rise of patient 3.0: The impact of social media, *8th International Conference on e-Health (EH 2016)*.

[21] **Gagnon, S. & Chartier, L.** 2012. Health 3.0—the patient-clinician "arabic spring" in healthcare, *Health*, 4(02) pp 39.

[22] **Sameer, K.** 2017. The Health 4.0 Revolution [Online]. Available: https://health.economictimes.indiatimes.com/news/health-it/the-health-4-0-revolution/59187378 [Date accessed: 6 April 2018].

[23] **Huertas Celdrán, A., Gil Pérez, M., García Clemente, F.J. & Martínez Pérez, G.** 2018. Sustainable securing of Medical Cyber-Physical Systems for the healthcare of the future, *Sustainable Computing: Informatics and Systems* [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210537917304675 [Date of access: 8 April 2018]

[24] **Exceet Secure Solutions.** 2015. Health 4.0: The digitally connected healthcare system of the future: Practical Uses [Online]. Available: www.exceet.com/fileadmin/exceet/.../e-Health_Telematics_infrastructure_ENG.pdf [Date of access: 16 March 2018].

[25] **Papa, A., Mital, M., Pisano, P. & Del Giudice, M.** 2018. E-health and wellbeing monitoring using smart healthcare devices: An empirical investigation. *Technological Forecasting and Social Change* [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0040162517312696 [Date of access: 20 March 2018]

[26] **Elhoseny, M., Abdelaziz, A., Salama, A.S., Riad, A.M., Muhammad, K. & Sangaiah, A.K.** 2018. A hybrid model of Internet of Things and cloud computing to manage big data in health services applications. *Future Generation Computer Systems* [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0167739X17322021 [Date of access: 20 March 2018]

[27] **Raghupathi, W. & Raghupathi, V.** 2014. Big data analytics in healthcare: promise and potential, *Health information science and systems*, 2014 2(3) pp 1-10.

[28] **Mieronkoski, R., Azimi, I., Rahmani, A.M., Aantaa, R., Terävä, V., Liljeberg, P. & Salanterä, S.** 2017. The Internet of Things for basic nursing care—a scoping review, International *Journal of Nursing Studies*, 69, pp 78-90.

[29] **Latif, S., Qadir, J., Farooq, S. & Imran, M.** 2017. How 5G Wireless (and Concomitant Technologies) Will Revolutionize Healthcare?, *Future Internet*, 9(4), pp 93.

[30] **Manogaran, G., Thota, C., Lopez, D. & Sundarasekar, R. 2017.** Big data security intelligence for healthcare industry 4.0. Cybersecurity for Industry 4.0. Springer, pp 103-126.

[31] **Dimitrov, D.V.** 2016. Medical internet of things and big data in healthcare, *Healthcare informatics research*, 22(3), pp 156-163.

[32] **Manogaran, G., Varatharajan, R., Lopez, D., Kumar, P.M., Sundarasekar, R. & Thota, C.** 2018. A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system, *Future Generation Computer Systems*, 82, pp 375-387.

[33] **eNCA.** 2017. Motsoaledi outlines anticipated NHI paper [Online]. Available: https://www.enca.com/south-africa/watch-motsoaledi-outlines-anticipated-nhi-paper [Date of access: 9 April 2018].

[34] **Your future now.** 2015. Public and Private Healthcare in SA [Online]. Available: https://yourfuturenow.co.za/public-and-private-healthcare-in-sa/ [Date of access: 9 April 2018].

[35] **Mpho, R.** 2016. NHI not designed to destroy private healthcare, says Motsoaledi [Online]. Available: https://www.fin24.com/Companies/Health/nhi-not-designed-to-destroy-private-healthcare-says-motsoaledi-20160628 [Date of access: 9 April 2018].

[36] **Khaya, K.** 2017. NHI to deliver affordable healthcare despite opposition – Motsoaledi [Online]. Available: https://www.iol.co.za/news/politics/nhi-to-deliver-affordable-healthcare-despite-opposition-motsoaledi-9136471 [Date of access: 9 April 2018].

[37] **Gray, A. & Vawda, Y.** 2017. Health policy and legislation, *South African health review*, 2017(1), pp 13-24.

[38] **Amankwah-Amoah, J.** 2015. Solar Energy in Sub-Saharan Africa: The Challenges and Opportunities of Technological Leapfrogging, *Thunderbird International Business Review*, 57(1), pp 15-31.

[39] **Perkins, R.** 2003. Environmental leapfrogging in developing countries: A critical assessment and reconstruction, *Natural Resources Forum organised by: Wiley Online Library*, pp 177-188.

[40] **James, J.** 2014. Relative and absolute components of leapfrogging in mobile phones by developing countries, *Telematics and Informatics*, 31(1), pp 52-61.

[41] **Sarabhai, K. & Vyas, P.** 2017. The leapfrogging opportunity: The role of education in sustainable development and climate change mitigation, *European Journal of Education*, 52(4), pp 427-436.

[42] **Group map.** 2018. What is SOAR analysis? [Online]. Available: https://www.groupmap.com/map-templates/soar-analysis/ [Date of access: 10 April 2018].

[43] **Silbert, J.H. & Silbert, T.** 2007. SOARing from SWOT: Four lessons every strategic plan must know. *International Journal of AI Best Practice*, pp 1-4.

[44] **Stavros, J.M. & Cole, M.L.** 2014. SOARing towards positive transformation and change, ABAC *ODI JOURNAL VISION. ACTION. OUTCOME,* Vol 1(2)

[45] **Swafford, S.W.** 2013. Applied experiences of the SOAR framework by association management and foundation executives: Pepperdine University.
[46] **Ponea, S.** 2010. The Implementation of Appreciative Management in the Recruiting Process and the Integration of the New Employees, *Postmodern Openings*, 3, pp 5-22.
[47] **Bester, M.** 2011. An appreciative inquiry approach to curriculum change, *Curriculum inquiry in South African higher education: Some scholarly affirmations and challenges*, pp 389-410.
[48] **Kusharsanto, Z.S. & Pradita, L.** 2016. The important role of science and Technology Park towards Indonesia as a highly competitive and innovative nation, *Procedia-Social and Behavioral Sciences*, 227, pp 545-552.
[49] South Africa.info reporter. 2016. Health care goes high tech in South Africa [Online]. Available: https://www.brandsouthafrica.com/south-africa-fast-facts/health-facts/electronic-bed-management-210416 [Date of access: 4 April 2018].
[50] **Deloitte.** 2015. Digital health may provide solutions for SA healthcare consumers, but are SA Life Sciences and Healthcare organisations ready to adopt it? Asks Deloitte [Online]. Available: http://www.engineeringnews.co.za/article/digital-health-may-provide-solutions-for-sa-healthcare-consumers-but-are-sa-life-sciences-and-healthcare-organisations-ready-to-adopt-it-asks-deloitte-2015-08-26/rep_id:4136 [Date of access: 4 April 2018].
[51] **Action, A.** 2018. Industry 4.0: Opportunities and Challenges for the Developing World [Online]. Available: https://www.wwc.co.za/2018/01/30/industry-4-0-opportunities-challenges-developing-world/ [Date of access: 10 April 2018].
[52] **Stock, T. & Seliger, G.** 2016. Opportunities of sustainable manufacturing in industry 4.0, *Procedia Cirp*, 40, pp 536-541.
[53] **United Nations Industrial Developing Organization.** 2016. Industry 4.0: Opportunities and challenges of the new Industrial Revolution for Developing Countries and Economies in Transition. Austria.

ENVIRONMENTAL DATA BENCHMARKING REPORTING STRUCTURE FOR THE SOUTH AFRICAN GOLD MINING INDUSTRY

H.M. Janse van Rensburg[1*] & Dr H.G. Brand[2] & Dr J.C. Vosloo[3]

[1] North-West University, CRCED-Pretoria, Pretoria, South Africa
mjvrensburg@researchtoolbox.com

[2] North-West University, CRCED-Pretoria, Pretoria, South Africa
hgbrand@researchtoolbox.com

3 North-West University, CRCED-Pretoria, Pretoria, South Africa
jcvosloo@researchtoolbox.com

ABSTRACT

In recent years legislative and economic pressure have placed pressure on the gold mining industry to be pro-active in sustainable development. Companies should therefore not only report but manage their environmental impact.

Most mining companies have implemented an environmental data management system which collects the data for reporting. However, companies lack the reporting structure to identify environmental optimisation opportunities. This paper develops a simplified environmental report structure that can be used by top management for effective environmental management.

The methodology focuses on identifying environmental standard requirements followed by benchmarking each business unit in terms of best practices and previous performance. Each business unit will then be visually displayed against set benchmarking targets to identify underperformance. The main goal is to structure the knowledge obtain from the operations information in a manner to drive optimisation opportunity identification.

The methodology was implemented on a large gold mining company. Underperforming business units were identified. Further investigation identified electricity and water use saving opportunities at one of these operations.

This work was sponsored by ETA Operations (Pty) Ltd

[1] Mrs. HM Janse van Rensburg holds a Masters in Mechanical Engineering from the North-West University. She is currently enrolled for PhD at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[2] Dr Hendrik Brand is a registered professional engineer and holds a PhD in Mechanical engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

[3] Dr JC Vosloo is a registered professional engineer and holds a PhD in Computer engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

The South African mining industry faces many challenges due to the impact that mines pose on the livelihood of local communities and the environment [1]. In 1998 the CEO's of the largest mining companies came together at the World Forum in Daves to address the sustainable issues facing the mining industry. They decided to launch the Global Mining Initiative (GMI). The main role of the initiative was to define sustainable development and ensure long-term sustainability goals for the industry [2]. The GMI approached the World Business Council for Sustainable Development (WBCSD) to assist in the development of the definition [2].

They in turn asked the International Institute for Environment and Development (IIED) to assist in a global research effort [2]. A report was issued in 2002 named "Breaking New Ground" which defined sustainable development as follow [3]:

"One of the greatest challenges facing the world today is integrating economic activity with environmental integrity, social concerns, and effective governance systems. The goal of that integration can be seen as 'sustainable development'. In the context of the minerals sector, the goal should be to maximize the contribution to the wellbeing of the current generation in a way that ensures an equitable distribution of its costs and benefits, without reducing the potential for future generations to meet their own needs. The approach taken to achieve this has to be both comprehensive including the whole minerals chain and forward-looking, setting out long-term as well as short-term objectives."[3]

Since then the mining industry has come a long way in improving their sustainability [4]. Most mining companies have adopted a form of sustainability reporting. This reporting is based on numerous standards like the Global Reporting Initiative (GRI) standards and the ISO 140001:2015 standard. However, in 2018 M Tost et. al. [4] did a comprehensive literature review to evaluate how the mining industry has considered its environmental sustainability approach. They conclude that at this stage the mining industry is not setting on the wrong sustainability path. However, they are at risk of falling behind other industries leaders on natural capital considerations as well as societal expectations on climate change [4].

Additionally, to ensure effective environmental governance systems in South Africa numerous legislation were promulgated. This, however resulted in the mining industries focusing more on compliance, instead of focusing on optimisation opportunities identification.

The ISO 14031:2013 standard states guidelines for companies to evaluate their environmental performance. An environmental performance evaluation (EPE) is a useful tool to identify key environmental performance indicators (EPIs). EPIs are used to measure, evaluate and communicate a company's environmental performance. The guidelines are based on the Plan-Do-Check-Act (PDCA) cycle. Figure 1 shows the ISO 14031:2013 guidelines in the context of this cycle. [5]

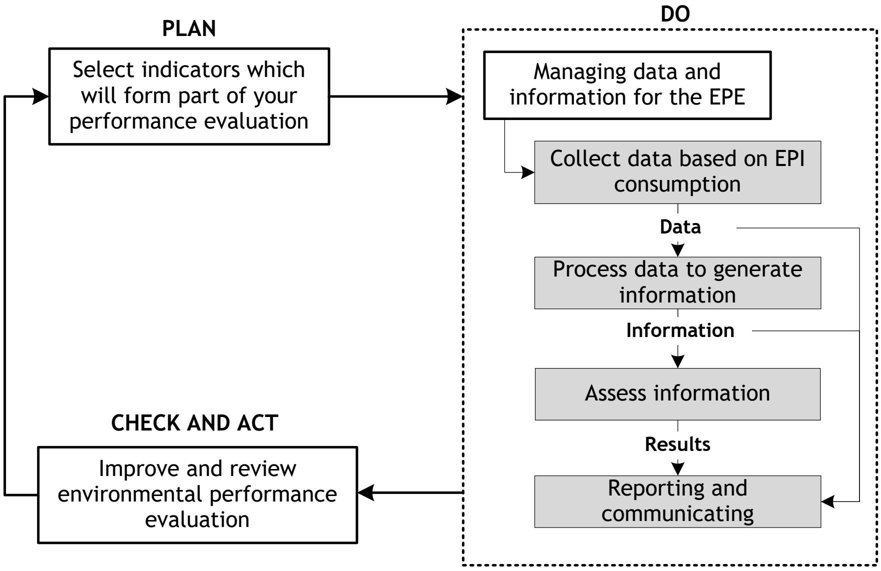


Figure 1: ISO 14031:2013 PDCA Cycle for EPEs (adapted from [5])

The first step in the cycle, planning, involves selecting the EPIs that will from part of an organisations EPE. The EPIs are a manner of converting quantitative and qualitative data into an understandable format. This format conveys the success management efforts have in influencing the company's environmental impact. The EPIs selected can be from existing indicators or new indicators. [5]

The do element specifies the direct actions that should be followed to complete an EPE. These actions are highlighted in the grey in Figure 1. This process generates three key elements namely data, information and results. This is the first three elements of the Data, Information, Knowledge and Wisdom (DIKW) hierarchy. This hierarchy is an integral part of information sciences and each element can be defined as follow [6], [7]:

- Data: A measurement of an observation
- Information: Processing data in the correct context to assist in answering interrogative question like who, what and where.
- Knowledge: Is knowing how to create a conceptual framework of multiple sources of information to give instruction. Knowledge is understanding the patterns within information.
- Wisdom: Gives knowledge a frame of reference through a set of principles. It is the inherent framework of knowing good from bad.

The check and act action are important for continual improvement of the EPIs performance. The check action will include internal and external reporting whereas the act action focusses on a periodically review of the EPE system for improvement identification. [5]. Figure 2 integrates the ISO 14031:2013 PDCA cycle into the concept of the DIKW hierarchy.

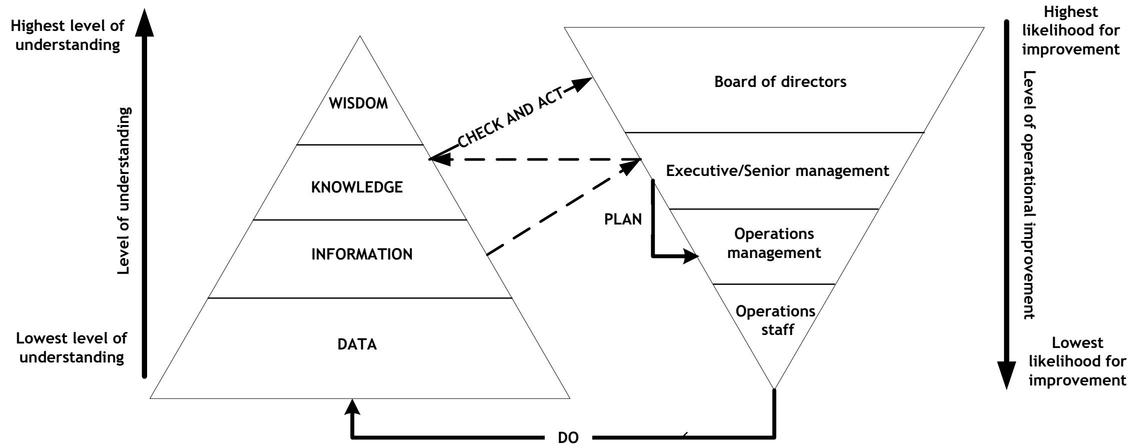


Figure 2: DIKW hierarchy and PDCA cycle for operational improvement

As is seen from the Figure 2, executive/senior management will specify the EPIs that needs to be used for an organisations EPE. These EPIs will be communicated to each operation. From here operations management and staff will collect the data and generate information. All the information from the different operations is then converted into a report by executive/senior management.

This report will be communicated to the board by executive/senior management in the form of a meeting. The knowledge obtained by the report will be discussed in this meeting. However, if the report already presented knowledge in a manner to assist the board with increased wisdom the likelihood of operational improvement is higher.

This paper will develop a reporting structure/layout that can be used to identify EPIs which are underperforming on facility level. The main goal is to structure the knowledge obtain from the operations information in a manner to drive EPI opportunity identification.

2. SUSTAINABLE REPORTING IN THE MINING INDUSTRY

2.1 Introduction

To develop a report structure for board level reporting which will drive improved EPI performance. It is important to understand which EPIs currently from part of the industry reporting chain and the data management principles pertaining to these EPIs. Furthermore, to improve the effectiveness of the report structure, methodologies which can drive EPI improvement should be investigated. These three key aspects will be discussed further in the proceeding sections.

2.2 Environmental Performance Indicators (EPIs)

Certain operational performance indicators (OPIs) directly relates to the impact that a company has on the environment. These OPIs are also known as EPIs. A company's OPIs can be determined by listing the organisation's inputs against its operational processes and relevant outputs. Figure 3 shows a general overview of a typical organisations OPIs. In a EPE a company needs to select their EPIs from these OPIs. [5]

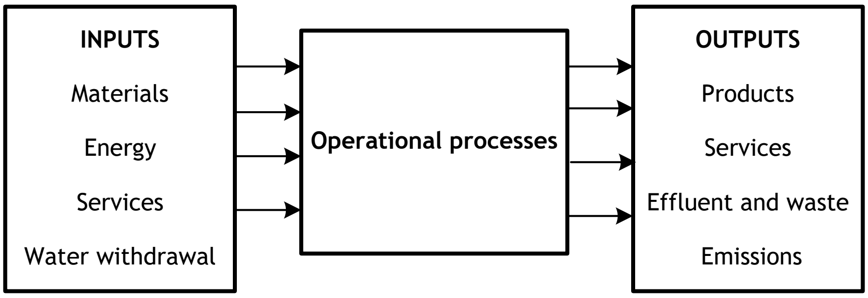


Figure 3: ISO 14031:2013 general overview an operations OPIs in terms of the environment (adapted from [5])

Based on their annual integrated report, the top 5 biggest JSE listed gold mining companies in South Africa structure their EPE and equivalent EPIs against the Global Reporting Initiative (GRI) Standards. These standards are a set of modular, interrelated standards that outline global best practice in terms of sustainability reporting. They guide companies on which data and information is important to collect and report for each EPI. The 300 set of standards are relating to environmental and climate change reporting. Table 1 define the EPIs that need to be monitored by an organisation. [8]

Table 1: GRI 300 set of standards [8]

GRI 300	GRI 301: Materials 2016
	GRI 302: Energy 2016
	GRI 303: Water 2016
	GRI 304: Biodiversity 2016
	GRI 305: Emissions 2016
	GRI 306: Effluents and waste 2016
	GRI 307: Environmental Compliance 2016
	GRI 308: Supplier Environmental Assessment 2016

As can be seen from both the ISO 14031:2013 standard and the GRI set of standards Materials, Energy, Water, Emissions and Effluent/Waste are important indicators for the evaluation of a company's environmental performance.

South Africa is a water scarce country. Gold mines does not consume large amounts of water compared with other industries like agriculture, but the mining industry has a significant impact on water quality [9]. Water is a crucial part of society and is closely monitored by government [10]. Water withdrawal and discharge therefore has a legislative risk [10].

Additionally, mines operational costs are increasing due to energy prices and the possibility of the taxation of greenhouse gasses emissions due to the consumption of energy [11][12]. Opportunities identification in energy and water is therefore crucial.

2.3 Environmental data management systems

The quote by Lord Kelvin *"to measure is to know"* accentuates the importance of data. However, if you have large amounts of data (big data) that is either inaccurate, transferred incorrectly or contains unauthorised alterations. The knowledge obtained from this data becomes useless. Therefore, effective data management should be included in an EPE.

Data management consists of multiple fields of study. Focus should be placed on data quality and integrity for reporting purposes. Data quality is the evaluation of ensuring that the data is "fit for use" and accurate [13]. Data integrity on the other hand is the prevention of the unauthorised alterations of data [14]. Figure 4 shows the data collection for reporting process in a typical mine.

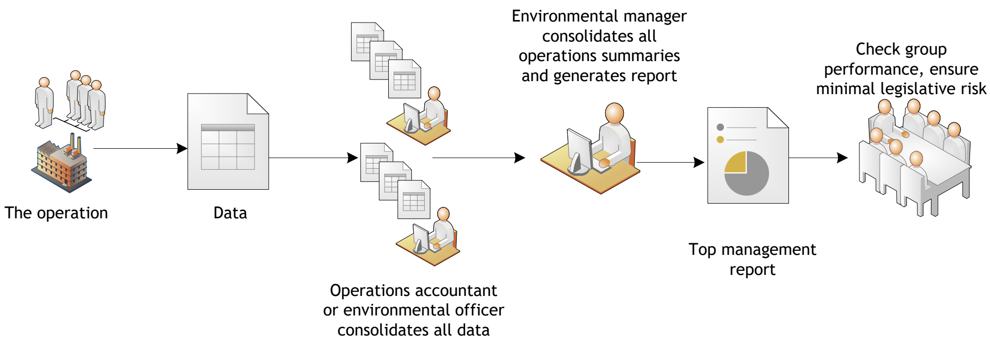


Figure 4: Environmental data collection for reporting in a gold mine

A mine's data generation, collection, processing and reporting is handled by numerous personnel as is illustrated by Figure 4. This results that data management is not so effectively implemented. Dr M van Heerden [15] identified this problem in gold mines. She developed a web-based environmental data centralisation and management system. This system was implemented on a gold mine. The average error percentage for this mine's reporting chain went from 20% to 4%.

2.4 Sustainable improvement through benchmarking and reporting

The evaluation of a mining operation's current performance against an operational target or reference performance is known as benchmarking [16], [17]. Benchmarking can be used to identify the misuse of resources [18]. An intensity analysis is a method that is used to benchmark operations [11],[18]. This methodology consists of dividing an EPI usage with its operational driver which can be tonnes of product or services rendered.

The Mining Association of Canada in cooperation with other parties issued a series of publication relating to the benchmarking of energy consumption for the international mining industry. The methodology included a breakdown of the production flow into key operating sections like exploration, mining, crushing, processing and refining. An intensity analysis of the cost of the energy consumed against multiple product related drivers was done. The energy cost intensity of each operation is then compared to identify underperforming facilities. [11] [12]

Dr L.F van der Zee [11] expanded on this research by noting that this simple comparison can give a false representation that a high production mine is energy efficient. He therefore defined additional variables that need to be considered in terms of benchmarking mines with each other. These included mine operation size, mine profit contribution, mining technology and mining depth.

The study developed a table where each mine's defined variables are compared to each other. Mines with the same parameters are grouped together for comparison. Dr C. Cilliers [18] however found that the only important variables to benchmark deep-level gold mines electricity intensity is the mining depth and tonnes rock mined. Each electricity intensity system will have an additional variable like ambient air condition and the hydrogeology of the mining area.

Dr J.I.G. Bredenkamp [21] argued that it is not always the operation with the highest electricity intensity that will result in the highest potential for cost savings. The thesis methodology simply multiplied each mine's operations electricity consumption with a percentage of theoretical improvement. The operation with the highest theoretical improvement value should be investigate first for improvement projects. Dr L.F van der Zee, Dr C Cilliers and Dr J.I.G Bredenkamp studies specifically focused only on benchmarking South African deep-level mines electricity usage.

In 2016 the South African Department of Water and Sanitation published a report that defines national benchmarking targets for the gold mining industry's water intensity. The definition for consumptive specific water use in this report is like the GRI Standards. Water intensity for this indictor was benchmarked against total gold-rich rock (also known as ore) mined. This amount excludes the total waste rock mined. The average South African gold mine's water intensity is 2.02 m^3/t. [22]

A combination of all these methodologies and targets will be used to define the benchmarking parameters within the reporting structure.

3. ENVIRONMENTAL REPORTING STRUCTURE

3.1 Introduction

To drive EPI improvement though reporting, the report structure should display knowledge in a manner that will stimulate increased wisdom. This will assist to ensure the correct decisions for improvement are made. Figure

5 illustrates the PDCA cycle displayed in Figure 1 in a new context of effective data management and improved report structure.

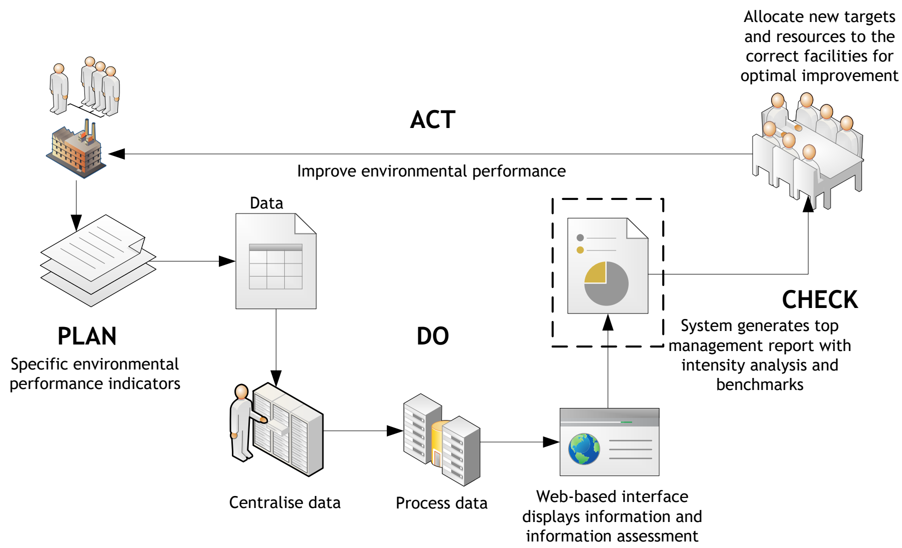


Figure 5: Environmental reporting framework

As can be seen in Figure 5, the top management report is an integral part of the reporting chain. Board level reporting should be simplified and not overly complex. The top management report structure should only indicate facility level performance. As discussed in section 2.4, to drive sustainability improvement through benchmarking, the report therefore will have the following key analysis namely operations classification, intensity analysis/benchmarking and theoretical improvement opportunity identification. Each of these sections will be discussed in more detail below.

3.2 Operational classification

To benchmark the mining facilities against each other to identify underperformance. Facilities with similar characteristic should be compared. Multiple OPIs can be used for a benchmarking comparison [11]. However, for this analysis the following three will be used.

3.2.1 Mining method

The Mining Association of Canada publications separated all the benchmarking methodologies into the different mining methods [20]. A breakdown was then done in terms of production flow. This same principle will be used for the report framework as the deep-level mines will be separated form open-pit mines. The gold processing facilities will also be separate from the mining operations.

3.2.2 Mining depth

As indicated by Dr C Cilliers [18] a mine's depth can have a significate influence on the electricity intensity of deep-level mines and will therefore be used as part of the classification.

3.2.3 Volume of material

The volume of material generated or used will help to accurately classify the type of operation. In terms of mining this variable is important because older mine's production will be lower than newer mines due to depleted resources. Gold processing plants will process two types of materials namely tonnes ore treated or old waste material each of these materials also have a significant impact on the processing intensity of the operations

3.3 Intensity analysis and benchmarking

Benchmarking is defined in section 2.3 as comparing the current operational performance with previous performance or against a set target. The intensity analysis will display the previous financial year intensity against current performance. This will immediately show if a mine is improving or declining in performance. If there are similar operations in the company, as identified by the classification, the mines current performance can also be compared. Finally, if there is a company, national or international benchmarking target for a EPI this will be displayed on the analysis for improved decision-making.

3.4 Theoretical improvement

After an overview of the current operations have been given. A table or graph should be added to guide top management to allocated resources to the mine with the most theoretical level of improvement. This will be done by multiplying the EPI consumption with a company's optimisation target.

4. APPLICATION OF STRUCTURE FOR A GOLD MINING COMPANY

4.1 Introduction

The purposed environmental report structure was based on Dr M van Heerden's implemented web-based data management system. For the reasons mentioned in section 2.1 continual optimisation is crucial in EPIs water and energy. The top management report for this company is developed on a quarterly basis. The company consist of 11 mining operations, two independent gold processing plants and one retreatment facility. The mining company in same cases will define a mine with both its shafts and gold plant as can be seen in Table 2.

This is due to the limitation the company has with data collection and metering on systems level. The main problem is that when the operational inputs, energy and water, arrive at the operations it is difficult to allocate the consumption to the processing facility or the mining operations. The mine classification technique, intensity/benchmarking analysis and operational improvement is shown in the proceeding sections.

4.2 Operational classification

The mining operations were classified against the amount of rock that they mine and the mining depth. The gold plants were classified by the type and amount of product they treat. The product that a gold plant can treat is either ore milled (also known as tonnes treated) or old process waste, also known as tailings, that still have traces of gold. Table 2 outline the gold mining company's classification. Due to the sensitivity of the processed volumes data. This variable will be classified as the following: small 0 – 700 000, medium 700 000 – 1 000 000, large 1 000 000 – 1 300 000 and mega > 1 300 000.

Table 2: Mine classification for South African Gold Mining Company

Company operations	Mining method	Mining depth	Previous FY volumes	Product variable	Classification group
Mine 1	Deep-level shafts	3,388	Medium	Rock mined	A
	Gold plant	N/A	Medium	Tonnes ore treated	
Mine 2	Deep-level shafts	1,978	Medium	Rock mined	B
	Gold plant	N/A	Large	Tonnes ore treated	
Mine 3	Deep-level shafts	1,452	Small	Rock mined	C
	Gold plant	N/A	Small	Tonnes ore treated	
Mine 4	Deep-level shafts	2,945	Medium	Rock mined	A
	Gold plant	N/A	Large	Tonnes ore treated	
Mine 5	Deep-level shafts	2,161	Large	Rock mined	D
Mine 6	Deep-level shafts	2,349	Mega	Rock mined	
Mine 7	Deep-level shafts	2,365	Small	Rock mined	E
Mine 8	Deep-level shafts	2,153	Small	Rock mined	
Mine 9	Deep-level shafts	2,050	Medium	Rock mined	
Mine 10	Open pits 1	N/A	Mega	Rock mined	F
	Gold plant	N/A	Mega	Tonnes ore treated	
Mine 11	Open pits 2	N/A	Mega	Rock mined	G
	Gold plant	N/A	Mega	Tonnes ore treated	
Gold plant A	Gold plant	N/A	Mega	Tonnes ore treated	H
Gold plant B	Gold plant	N/A	Mega	Combination (Tailings and ore)	I
Retreatment facility	Gold plant	N/A	Mega	Tonnes tailings retreated	

4.3 Intensity analysis and theoretical improvement

The Department of Water and Sanitation report [22] and the mining company use the tonnes of ore treated/ore milled for calculations and reporting. This means that this product's information is available for each operation. The intensity analysis was therefore done against this product. The total energy consumed by the mining company's operations consists of electricity, diesel, petrol, polyfuel diesel and explosives. All these energy types were converted into the same energy unit namely gigajoules (GJ). The total water withdrawn only consists of two variables namely potable water and non-potable water. These inputs are both measured in kilolitres/cubic meters (m^3). Figure 6 illustrates the OPIs against the mining operations processes.

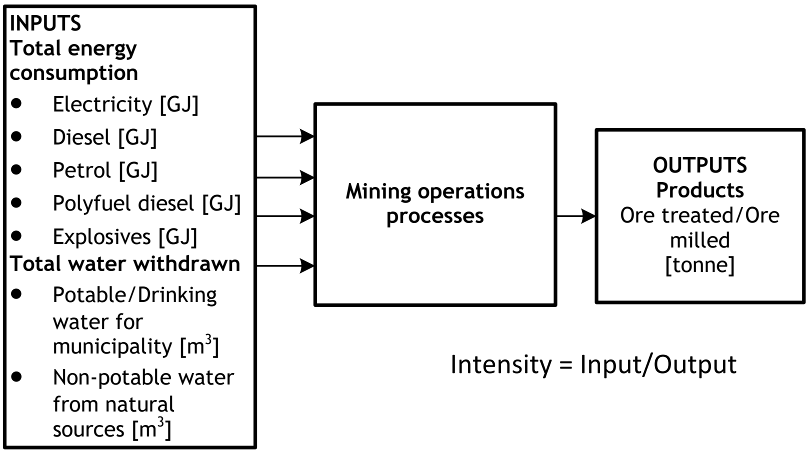


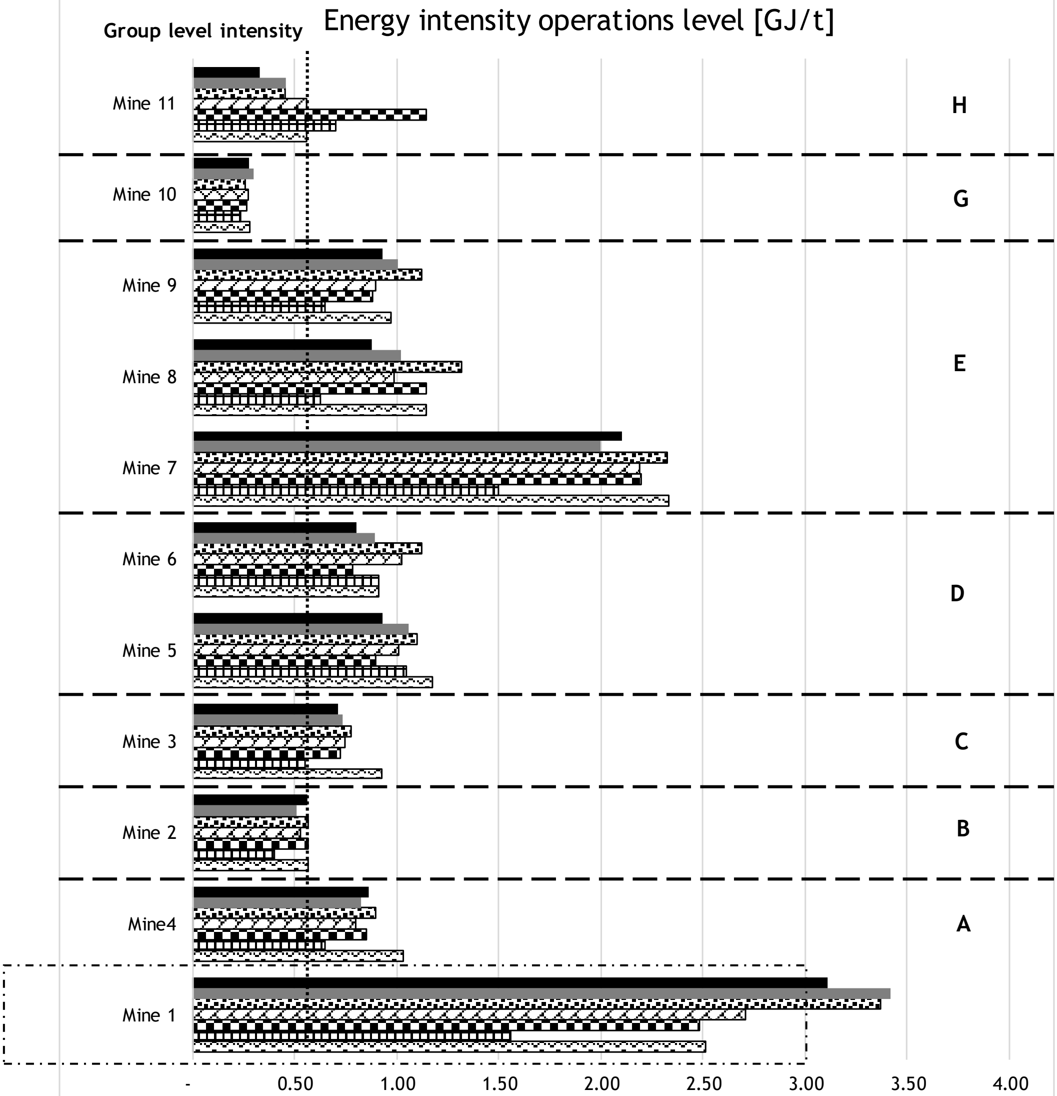
Figure 6: Breakdown of energy and water consumption sources in mining company

The key observation of the intensity analysis is summarised in Table 3.

Table 3: Key observations made from intensity analysis

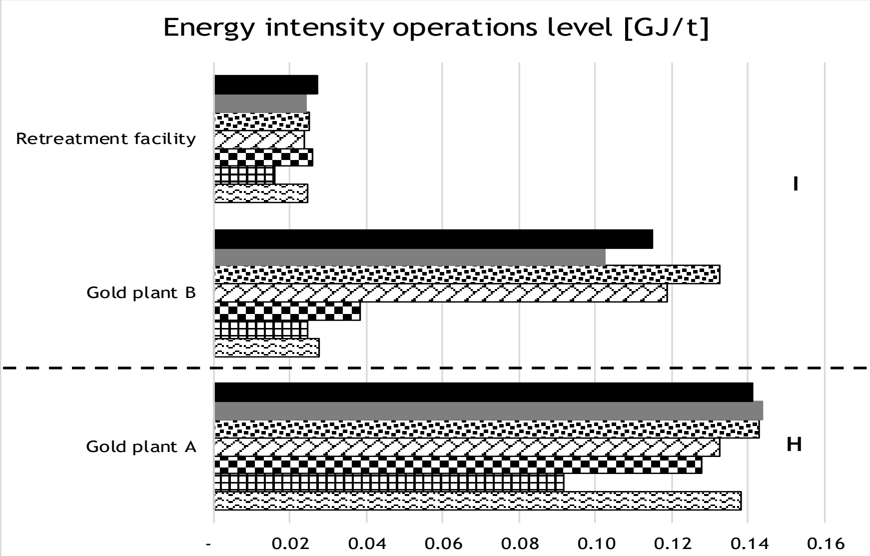
Analysis	Key observations
Figure 7 and 8	• Open pit mines are less energy intensive • Only mine 2 and mine 10 operates below the group average energy intensity • All the mines, except mine 11 had an increase in energy intensity in the first quarter of 2018 • Retreatment of tailings is less energy intensive than ore milled • It can be observed that the intensity for all the operations are irregular and does not show consistency or a consistent decline. This is an indicated that not effective management is implemented.
Figure 10 and 11	• Mine 2 water intensity significant decrease due to the installation of a water treatment plant • Mine 1, 6 and 7 operates higher that the national benchmark • Mine 7 is 2 times more inefficient that its closes competitor. • It can be observed that the intensity for all the operations are irregular and does not show consistency or a consistent decline. This is an indicated that not effective management is implemented.

The company defined a 5% improvement for the company operations in terms of water and energy consumption. Figure 9 and Figure 12 illustrates the theoretical improvement that each of the operations can achieve if their consumption is multiplied by the 5% reduction target. As is seen in these figures Mine 1 has the highest potential for improvement for both water and energy consumption. Mine 1 should therefore be firstly investigated for improvement opportunities.



	Mine 1	Mine4	Mine 2	Mine 3	Mine 5	Mine 6	Mine 7	Mine 8	Mine 9	Mine 10	Mine 11
Q1 - FY17	3.11	0.86	0.56	0.71	0.93	0.80	2.11	0.88	0.93	0.27	0.33
Q2 - FY17	3.42	0.83	0.51	0.74	1.06	0.90	2.00	1.02	1.00	0.30	0.46
Q3 - FY17	3.37	0.90	0.56	0.77	1.10	1.12	2.33	1.32	1.12	0.26	0.45
Q4 - FY17	2.71	0.80	0.53	0.74	1.01	1.02	2.19	0.99	0.90	0.28	0.56
Q1 - FY18	2.49	0.85	0.57	0.72	0.90	0.79	2.20	1.14	0.88	0.27	1.15
Q2 - FY18	1.56	0.65	0.40	0.55	1.05	0.91	1.50	0.62	0.65	0.23	0.70
Q3 - FY18	2.51	1.03	0.57	0.93	1.18	0.92	2.34	1.15	0.98	0.28	0.56

Figure 7: Energy intensity analysis of mining company



	Gold plant A	Gold plant B	Retreatment facility
Q1 - FY17	0.14	0.11	0.03
Q2 - FY17	0.14	0.10	0.02
Q3 - FY17	0.14	0.13	0.02
Q4 - FY17	0.13	0.12	0.02
Q1 - FY18	0.13	0.04	0.03
Q2 - FY18	0.09	0.02	0.02
Q3 - FY18	0.14	0.03	0.02

Figure 8: Energy intensity analysis of mining company continue

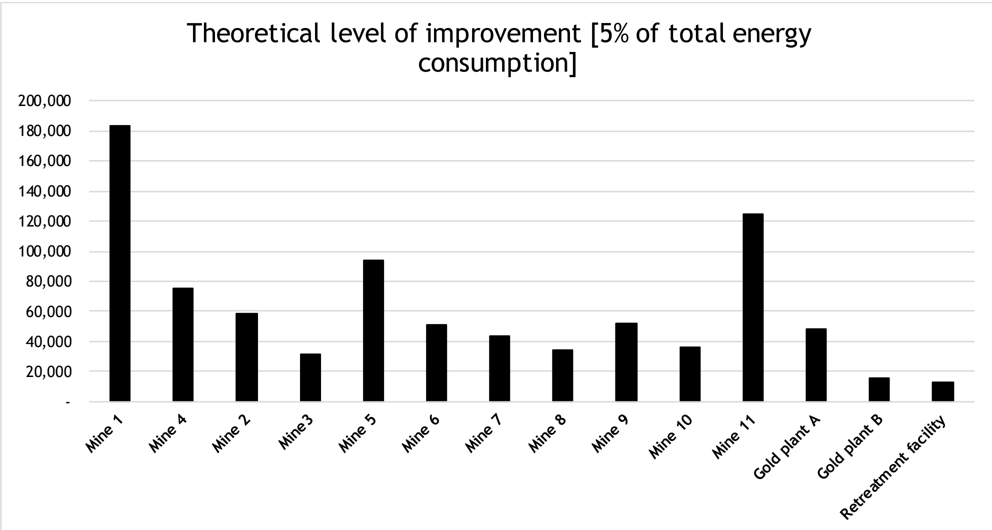
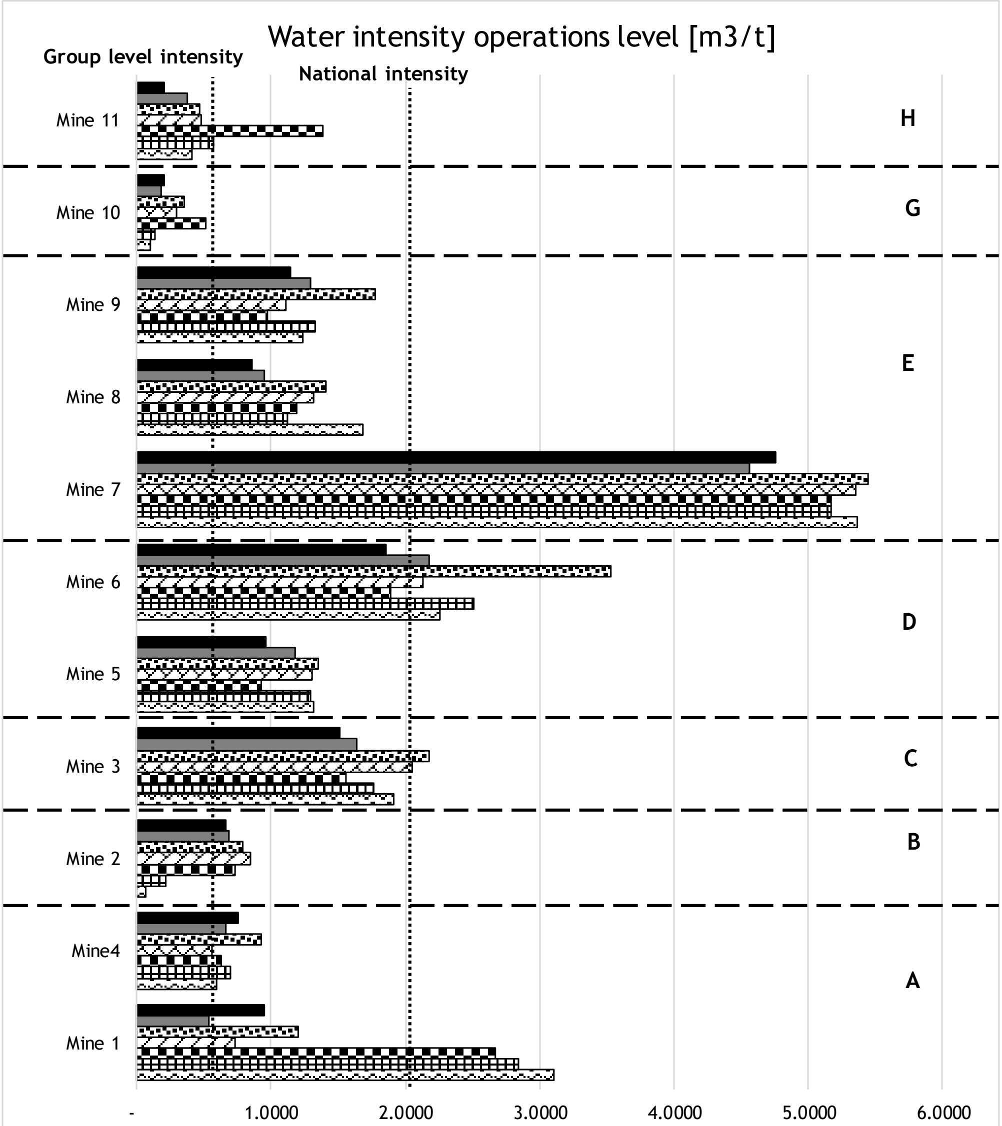
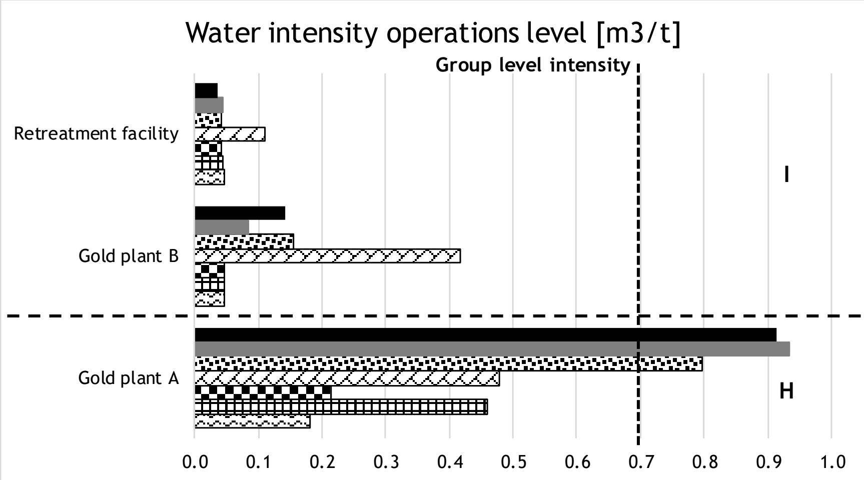


Figure 9: Theoretical potential for energy improvement



	Mine 1	Mine4	Mine 2	Mine 3	Mine 5	Mine 6	Mine 7	Mine 8	Mine 9	Mine 10	Mine 11
Q1 - FY17	0.9539	0.7545	0.6647	1.5071	0.9586	1.8575	4.7615	0.8539	1.1415	0.2062	0.2085
Q2 - FY17	0.5406	0.6602	0.6904	1.6355	1.1830	2.1779	4.5606	0.9515	1.2905	0.1830	0.3781
Q3 - FY17	1.1986	0.9304	0.7912	2.1739	1.3523	3.5305	5.4460	1.4076	1.7792	0.3515	0.4716
Q4 - FY17	0.7273	0.5616	0.8502	2.0476	1.3069	2.1302	5.3571	1.3177	1.1122	0.3005	0.4779
Q1 - FY18	2.6675	0.6318	0.7337	1.5576	0.9314	1.8963	5.1733	1.1965	0.9767	0.5122	1.3843
Q2 - FY18	2.8477	0.6980	0.2123	1.7699	1.2893	2.5146	5.1694	1.1246	1.3295	0.1413	0.5775
Q3 - FY18	3.1081	0.5967	0.0695	1.9102	1.3203	2.2585	5.3589	1.6852	1.2418	0.1059	0.4081

Figure 10: Water intensity analysis of mining company



	Gold plant A	Gold plant B	Retreatment facility
Q1 - FY17	0.9146	0.1419	0.0350
Q2 - FY17	0.9342	0.0843	0.0450
Q3 - FY17	0.7971	0.1542	0.0426
Q4 - FY17	0.4779	0.4183	0.1107
Q1 - FY18	0.2133	0.0469	0.0412
Q2 - FY18	0.4592	0.0457	0.0441
Q3 - FY18	0.1815	0.0468	0.0467

Figure 11: Water intensity analysis of mining company continue

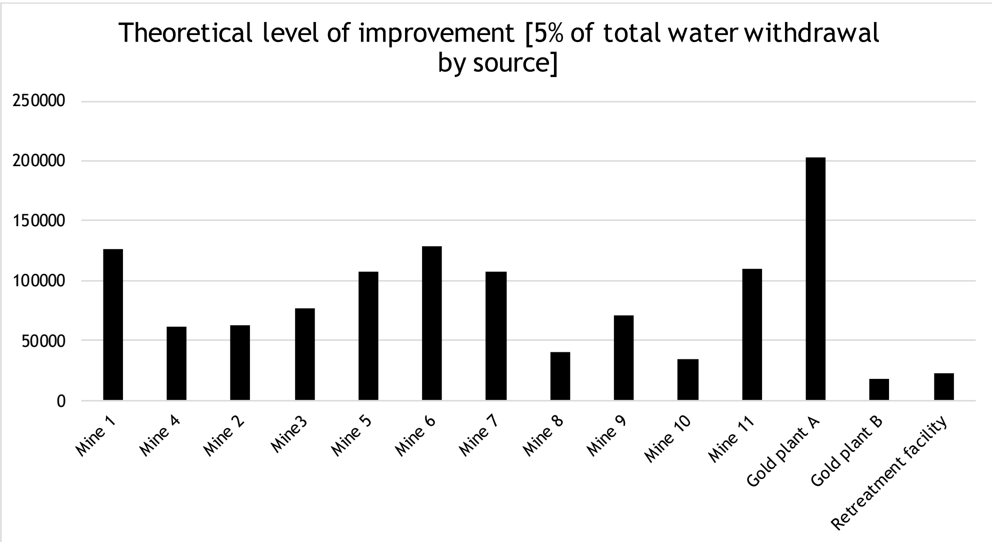


Figure 12: Theoretical potential for water improvement

5. VALIDATION AND OPPORTUNITIES IDENTIFIED

The key requirement of the report structure is to assist in identifying underperforming facilities and improve EPI improvement at these facilities. The report structure identified that Mine 1 is both inefficient and has the highest level of theoretical improvement for energy. For water, Mine 1 is under the top three worst performers and second with theoretical improvement. The mining company is investigating available opportunities in the company. Table 4 outline the key opportunities that have been identified in Mine 1. In some cases, these projects have been already implemented.

Table 4: Mine 1 key improvement opportunities identified

Environmental Indicator	Opportunities identified
Electricity	The evaporator pump at the fridge plant that supplies cold air to underground staff is not dynamically controlled resulting in overcooling.

Electricity	Compressed air is wasted through the underground refuge chambers.
Electricity	The installation of a larger bypass control value for each production level compressed air supply. Reduce compressed air wastage.
Water	22 Cooling cars at the working stations were removed without compromising ventilation requirements. A 50 l/s reduction in services water consumption was achieved.
Water	Leak repairs in the water network were identified.

The total impact the mine as made to date in the electricity cost avoidance at Mine 1 is over R 33 million. A quarterly energy intensity reduction due to these saving can be seen in
Figure 7. Further investigation into the water network was initiated by the mining company due to the identification of Mine 1's underperformance. The water network audits identified that the mine's single men's hostels were consuming above average amount of potable water. Figure 13 illustrates the comparison of Mine 1's consumption with a similar hostel in the mining group.

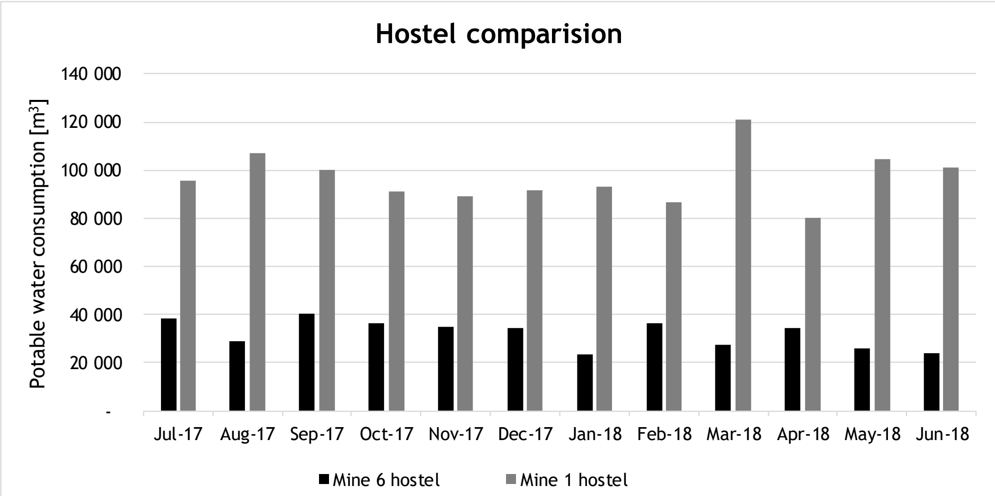


Figure 13: Mining hostel comparison

The occupancy of Mine 6 hostels is 1400 people whereas Mine 1 hostels only occupies 1104 people. This illustrates that there is large potential for water reduction savings at the Mine 1. Current investigations are underway to reduce consumption. The report structure methodology is therefore validated and can be used to identify underperforming facilities and EPI improvement can be achieved at these facilities.

6. CONCLUSION

The gold mining industries placed more focus on compliance, instead of focusing on optimisation opportunities identification. The ISO 14031:2013 standard can be used to evaluate a company's environmental performance. An EPE is a useful tool to identify key environmental performance indicators (EPIs).

Gold mines should first prioritise the improvement of energy and water consumption due to the operational costs and legislative risks. A web-based database should be used to centralise all the environmental data and increase the data confidence in terms of quality and integrity. An intensity analysis can be used to identify underperforming facilities. The mines with the same characteristics can also be compared to identify underperformance in terms of benchmarking.

The PDCA cycle of the ISO 14031:2013 was applied to the gold mining industry's reporting chain. It was found that the top management report is a key aspect of the reporting chain. This report can assist to stimulate improvement of EPIs water and energy consumption. The report should be presented in a manner that will assist top management in improved decision-making.

The methodology was applied to a gold mining company. It was found one of the mines, namely mine 1, was significant underperforming when compare with a similar mining operation and group level intensities. The cost avoidance in terms of electricity consumption is over R33 million to-date.

REFERENCES

[1] O. O. Ololade and H. J. Annegarn, "Contrasting community and corporate perceptions of sustainability:

A case study within the platinum mining region of South Africa," *Resour. Policy*, vol. 38, no. 4, pp. 568–576, 2013.
[2] M. Tost, V. Chandurkar, M. Hitch, P. Moser, and S. Feiel, "Is it time for a Global Mining Initiative 2.0?," *Geo-Resources Environ. Eng.*, vol. 2, no. 0, pp. 41–47, 2017.
[3] International Institute for Environment and Development, "Breaking New Ground: Mining, Minerals, and Sustainable Development," 2002.
[4] M. Tost, M. Hitch, V. Chandurkar, P. Moser, and S. Feiel, "The state of environmental sustainability considerations in mining," *J. Clean. Prod.*, vol. 182, pp. 969–977, 2018.
[5] International Organization for Standardization, "ISO 14031 Environmental management — Environmental performance evaluation — Guidelines." Switzerland, pp. 6–11, 2013.
[6] G. Jifa, "Data, information, knowledge, wisdom and meta-synthesis of wisdom-comment on wisdom global and wisdom cities," *Procedia Comput. Sci.*, vol. 17, pp. 713–719, 2013.
[7] P. Cooper, "Data, information, knowledge and wisdom," *Anaesth. Intensive Care Med.*, vol. 18, no. 1, pp. 55–56, 2017.
[8] Global Sustainability Standards Board, "Consolidated set of GRI sustainability reporting standards," 2016.
[9] T. M. Askham and H. M. Van der Poll, "Water sustainability of selected mining companies in South Africa," *Sustain.*, vol. 9, no. 6, pp. 1–16, 2017.
[10] Republic of South Africa, "National Water Act 36 of 1998," *Gov. Gaz.*, vol. 398, no. 19182, pp. 1–200, 1998.
[11] L. F. van der Zee, "Modelling of electricity cost risks and opportunities in the gold mining industry," North-West University, 2014.
[12] Republic of South Africa, "Draft carbon tax bill," Pretoria, South Africa, 2017.
[13] R. W. Wang and D. M. Strong, "Beyond accuracy: What data quality means to data consumers," *J. Manag. Inf. Syst.*, vol. 12, no. 4, p. 5, 1996.
[14] M. Zviran and C. Glezer, "Towards generating a data integrity standard," *Data Knowl. Eng.*, vol. 32, no. 3, pp. 291–313, 2000.
[15] M. Van Heerden, "Environmental data management for gold mines," North-West University, 2017.
[16] G. Shuttleworth, "Benchmarking of electricity networks: Practical problems with its use for regulation," *Util. Policy*, vol. 13, no. 4, pp. 310–317, 2005.
[17] D. F. Edvardsen and F. R. Førsund, "International benchmarking of electricity distribution utilities," *Resour. Energy Econ.*, vol. 25, no. 4, pp. 353–371, 2003.
[18] C. Cilliers, "Benchmarking electricity use of deep-level mines," North-West University, 2016.
[19] B. The and E. Consumption, *Benchmarking the Energy Consumption of C Anadian O Pen -P It M Ines*. .
[20] Canadian Industry Program for Energy Conservation, "Benchmarking the energy consumption of Canadian Underground Bulk Mines," Ottawa, 2005.
[21] J. I. G. Bredenkamp, "An integrated energy management strategy for the deep-level gold mining industry," North-West University, 2016.
[22] Department of Water and Sanitation, "Benchmarks for water conservation and water demand management (WC/WDM) in the mining sector," Pretoria, South Africa, 2016.

TOWARD THE DEVELOPMENT OF SUSTAINABLE BUSINESS MODELS FOR SOCIAL ENTERPRISES

R.M. Armstong[1*] & S. Grobbelaar[2]

[1]Department of Industrial Engineering
University of Stellenbosch, South Africa
17603609@sun.ac.za

[2]Department of Industrial Engineering and DST-NRF CoE in Scientometrics and Science,
Technology and Innovation Policy (sciSTIP)
Stellenbosch University, South Africa
ssgrobbelaar@sun.ac.za

ABSTRACT

With social depravation and inequities on the rise, social enterprise has been the response offered to combat this trend. These social enterprises are required to be autonomous, sustainable and meet the social needs that are not being fulfilled by government or the private sector. Research has been focussed on the definition of social enterprise but has not sufficiently examined the concepts of sustainability and business models within the social enterprise context. A conceptual framework is proposed that synthesizes sustainability, business model and social enterprise components.

[1]The author is enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding Author

1. INTRODUCTION AND PROBLEM STATEMENT

Social depravation and inequality is on the rise in South Africa [1]. Traditionally, government is held responsible to represent the interests of its citizens and alleviate social problems that would not be resolved otherwise [1]. It has however been evident that even governments in developed countries are not capable of resolving their social problems, much less governments in developing countries.

Teasdale [2] expands on four failures that explain the context within which Social Enterprises (SE's) originated. The failures are; (1) failure of state and market, which occurs when government is not able to counteract inequalities fostered by the market or when there is a lack of market presence by the private sector, (2) resource failure, the dependence of an organisation on its external environment for its resources, there is often a lack of control over the flow of these resources by the organisation (3) institutional theory failure, which occurs when the dominant practices of an industry are incorrectly accepted as best practice and finally (4) voluntary failure, which occurs as a result of the power imbalance between the donor and the organisation receiving the funding. Traditionally Non-Governmental Organisations (NGOs) serve the sector that SEs target. NGOs however do not always operate with a business model as they often do not capture value. Thus, NGOs are more dependent on donations than traditional enterprises. To overcome this dependence, other sources of income are desired [2]. The failures and dependencies identified illustrate that neither the government, NGOs or the private sector can comprehensively meet the social needs of society, or the subsequent gaps in the market. As can clearly be seen in South Africa, 60% of the national budget is spent on social services [3]. The budget deficit is already 24%, clearly the social need is greater than even government can cater for, which supports the case of state and market failure. From these statistics it is clear that in developing countries, SE's developed not because of institutional theory failure, but out of a need to provide for social needs.

With increased competition for resources, resources failure has highlighted the frailty of NGOs. NGOs limited control over their revenue streams cause NGOs to have variable certainty about their long-term liquidity and sustainability.

Something in between the public and private-sector was needed to fill the gap in the market. More specifically, an enterprise that could be sustainable, autonomous and address the social and market needs that are not being met by government, NGOs or the private sector. The resulting concept was that of the social enterprise [2]. For the purpose of this paper, SEs will be defined as: "Organisations seeking to achieve social goals through the application of commercial and business like activities" [4].

The definition and evolution of SEs has been well researched in literature [5] [6] [4]. Although sustainability is a vital component of SE policy, there has been little critical debate on how to achieve sustainability, and it is recognised as a subject that needs to be researched more rigorously [4].

Business models, in general, are a topic of study that have been neglected and rarely analysed [7]. Research has not uncovered many scalable and sustainable business models [8]. This is due to the lack of business models that are able to serve the SE market sustainably [6]. Recent research in business model innovation has also placed emphasis on sustainability issues [9]. Thus, it is required to research the nature of business models that are sustainable within the markets that SE's serve.

The preliminary investigation suggests that there are numerous examples of well-meaning SEs that aim to make a positive difference in society but are not able to make sustainable profits [10]. Business models that have been applied in the past required very high penetration rates and were not sustainable [10]. Scalable and sustainable business models have also not been sufficiently explored in literature [8]. SEs tend to struggle with hybridity tension that arises because of its dual mission. It is also evident that business models research is in its infancy in the SE field. Thus, there is a need to develop a conceptual framework, that will provide an understanding of the elements required of a business model, to ensure that SEs can be sustainable.

The research will be guided by the following research questions:

1. How can sustainability be defined for social enterprise?
2. What are the concepts that need to be accounted for in a business model that aims to serve the social enterprise sector?
3. What are the main tensions and issues that arise from the social and financial missions in SE's?
4. How can organisational sustainability theory and business models manage the tension between the social mission and financial sustainability within a social enterprise?
5. What are the factors that drive sustainability in social enterprises?
6. How can business models enable scaling of social enterprises?

2. RESEARCH METHODS AND CONCEPTUAL FRAMEWORK ANALYSIS

The topic of this study is multi-disciplinary in nature, with grounding in business management, economics, social sciences and engineering. This study will thus be qualitative in nature, reviewing material from various sources. Grounded theory, which is a commonly utilised qualitative research method [11], is capable of reviewing large amounts of multi-disciplinary data. Grounded theory does however not sufficiently provide a process to be followed when formulating a conceptual framework. Jabareen [12] formulated the conceptual framework analysis which provides a detailed process, based in grounded theory, that can be used to build a conceptual framework during qualitative studies. The conceptual frameworks that are subsequently developed are intended to provide an interpretive approach to society and provide understanding of the real world, rather than theoretical definitions [12].

The conceptual framework analysis defines the phases of formulating a conceptual framework very well, but it does lack 'tools' that are capable of achieving the objectives of the different phases. To solve this, systematic literature review tools are incorporated into the conceptual framework analysis.

Table 1 illustrates the synthesis of the conceptual framework analysis and the systematic literature review tools.

Table 1: Adapted conceptual framework analysis [12], [1], [11]

Conceptual Framework Analysis Phases	**SLR Phase**	**Systematic Literature Review Tools** [1], [11]	**Conceptual Framework Analysis Tools** [12]
Phase 1: Mapping the selected data sources	Searching for articles	• Keywords • Search strings • Databases	• Extensive review • Initial specialists interview • Comprehensive data collection
Phase 2: Extensive reading and categorizing of the selected data	Relevance appraisal	• Screening, inclusion and exclusion criteria • Quality appraisal (Four Criteria) aims and objectives stated, clear context, concept description, sufficient data for evidence • Open-coding	• Categorise data by discipline • Categorise by importance within the discipline • Atlas.ti line-by-line coding • Atlas.ti mapping of publishing dates and geographical location
Phase 3: Identifying and naming concepts	Extracting data	• Open coding	• Discover concepts • Read and re-read the coded data
Phase 4: Deconstructing and categorizing the concepts		• Axial coding - analysing context of the concepts	• Deconstruct each concept identify its main attributes, characteristics assumptions and role • Organise and categorise concepts by: features and ontological, epistemological and methodological role
Phase 5: Integrating concepts		• Axial coding - analyse which characteristics of the identified concepts overlap. Link those concepts.	• Integrate and group together concepts to form new categories
Future Research - Not Included in this paper			

Phase 6: *Synthesis, resysnthesis and making it all make sense*		• *Selective Coding - select core category, forms foundation of the framework*	• *Synthesize theoretical framework* • *Semi-structured interviews* • *Iterations*
Phase 7: *Validating the conceptual Framework*			• *Present at peer reviewed conferences* • *Framework ranking interviews, case study and case study interviews*
Phase 8: *Rethinking the conceptual framework*			• *Revise conceptual framework based on feedback, new insights and literature.*

3. CONCEPTUAL FRAMEWORK ANALYSIS

Jabareen [12] defines a conceptual framework as “a network or ‘plane’ of interlinked concepts that together provide a comprehensive understanding of a phenomenon or phenomena”. Conceptual frameworks are built on concepts alone, not variables. Concepts are groupings of similar data that together are given conceptual labels. These concepts are then related to each other by statements of relationships which result in new categories being defined. Categories are groupings of concepts that have been elaborated upon to describe the real world [11]. The framework is synthesized using the identified categories. This paper will document the first six phases of the conceptual framework analysis, up until the integration of the concepts into categories and an initial synthesis of a theoretical framework.

3.1 *Phase 1*: Mapping the selected Data sources

The goal of this phase was to comprehensively survey the topic of study and to collect relevant data. This was achieved by analysing the existing literature. Upon completion of Phase 1 it was desired to understand the data sources. This was achieved by first selecting the relevant key words for the study, shown in Figure 1.

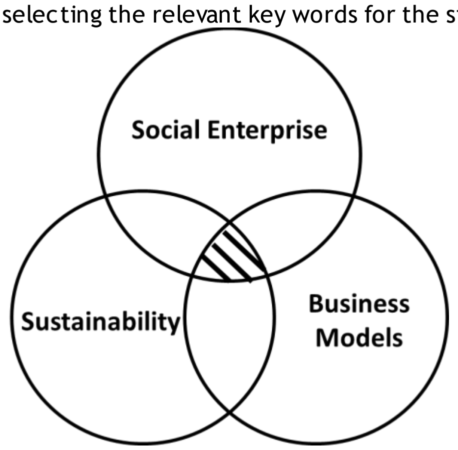


Figure 1: Key Word Venn Diagram

The keywords were combined to form the search terms. These terms were used to identify the literature that is relevant to the shaded area in Figure 1, where the three domains of the study overlap. The search terms were used to weed out papers that did not relate to the niche that the study is researching. The search terms are listed in Table 2. The online database utilised was Scopus, as Scopus is “the largest abstract and citation database of peer-reviewed literature” [13]. The initial search returned 376 articles. In addition to the articles found using the Scopus search, 14 grey literature articles were included, thus a total of 390 articles were identified.

Table 2: Search Strings and Scopus Results

Keywords	Search Strings	Scopus Results
Social Enterprise, Sustainability	("Social Enterprise") AND ("Sustainability")	214

Business Model, Sustainability	("Business Model") AND ("Organisational Sustainability")	18
Social enterprise, business model	("Social Enterprise") AND ("Business Model")	115
Social enterprise, business model, sustainability	("Social Enterprise") AND ("Business Model" OR "Business Model Innovation") AND ("Sustainability")	29

3.2 *Phase 2:* Extensive reading and categorizing of the selected data

The aim of this phase was to identify the articles that were relevant to this study and to subsequently read and categorise the data by coding.

3.2.1 Screening papers: inclusion and exclusion criteria

The titles, abstracts, and in some cases the full texts of the 390 articles identified were scrutinised to determine whether they were to be included in the study. If the paper met any of the exclusion criteria, the paper was rejected. The exclusion criteria were: (1) article did not focus on any of the keywords, (2) full text not available, (3) article not in English and (4) the paper did not meet the inclusion criteria.

For inclusion the paper needed to meet only one of the inclusion criteria; (1) definitions of relevant concepts provided (2) paper focuses on the relevant search terms (3) the paper synthesizes some of the search concepts and (4) the paper embarked on a case study in the relevant field.

3.2.2 Relevance appraisal

After the screening of the initial 376 articles using the exclusion and inclusion criteria, 30 relevant articles were identified. To further refine the search, a relevance appraisal was conducted on the remaining 30 articles. The articles were appraised using three metrics scored from 0-2. The metrics used were: (1) the article expressly deals with one of the three key search terms (2) the article discusses the relation between at least two of the three key search terms and (3) the article contains a validated definition or framework that relates directly to the key search terms. Only papers with a combined score of 5 or higher were included in the review. After the appraisal, 20 papers were identified to have a score of 5 or higher and were subsequently coded.

3.2.3 Categorisation of data

The data was categorised through open-coding, using the program 'Atlas.ti'. During coding four main themes, which are listed in Table 3, emerged from the data. Each of the themes were made up of several codes that were identified during the process of open-coding. Open-coding was used to identify concepts, which arose in the form of keywords, phrases and sentences.

Table 3: Themes and Codes that emerged during open coding

1. Research Aids	2. Social Enterprise	3. Business Model	4. Sustainability
• Gap in literature • Industry 4.0 • Research Lens • Research Methodology • Scaling ability • Tools and Frameworks	• Definition • Innovation • Challenges • Characteristics • Entrepreneur and staff • Sustainability	• Definition • Innovation • Business Model • Social Enterprise Business Models • Sustainability	• Definition • Challenges • Drivers • Financial

The code names were kept broad so that concepts were not pre-empted. This step can be viewed as 'housekeeping'.

3.2.4 Mapping of selected data

Research into SEs, business models and sustainability has been gaining popularity, this is supported by Figure 2, as the number of articles published is increasing per year. From the twenty articles selected, nine were published within the last two years. SE literature has been on the rise since 2006, but there has been a marked increase in the interest in its sustainability and business models [14].

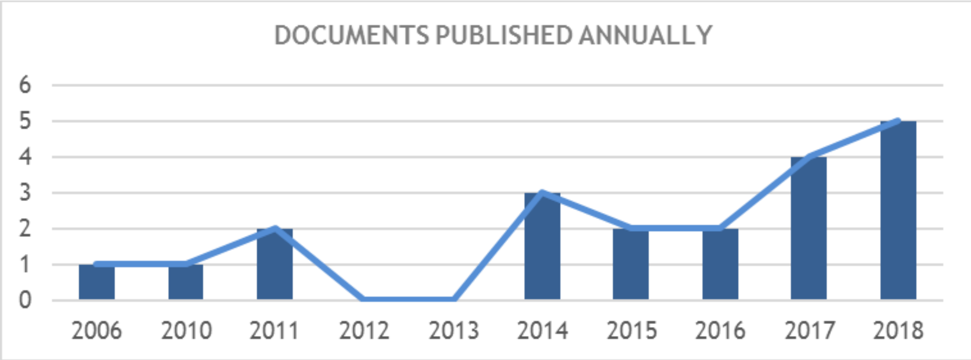


Figure 2: Articles published per year

It is interesting to note, from Figure 3, that the published papers are originating from developed and developing countries alike. Although 65% of articles originate from developed countries, the developing countries are contributing significantly to the literature, as in these countries SE's are filling the voids left by institutional failure [2].

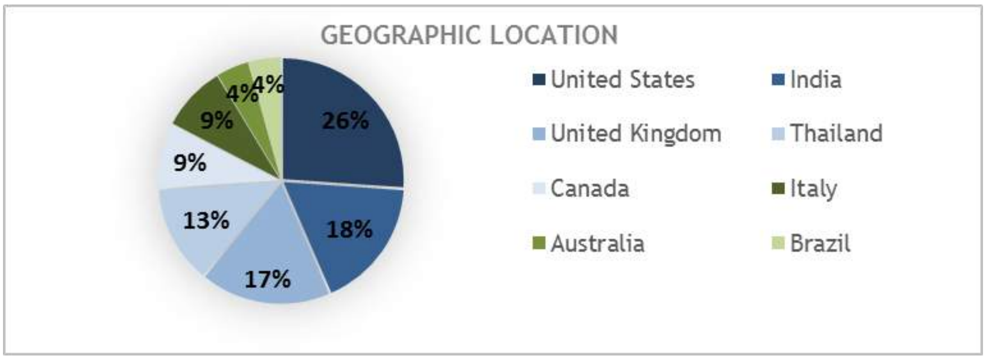


Figure 3: Geographic locations of papers published

Figure 4 supports the notion that social enterprise, business models and sustainability are multi-disciplinary topics of research. This confirms that the research methodology chosen, the conceptual framework analysis and systematic literature review, are appropriate for this study. The conceptual framework analysis in conjunction with the systematic literature review, are capable of handling large quantities of data, whilst also facilitating the formulation of conceptual framework based on the multi-disciplinary literature.

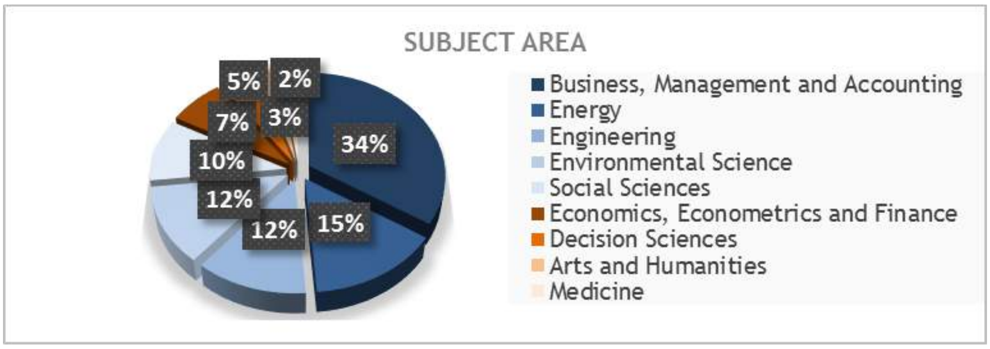


Figure 4: Subject areas of selected articles

3.3 *Phase 3:* Identifying and naming concepts

The goal of this phase was to allow concepts to emerge from the selected data. It was imperative for this phase that the data was not read with pre-conceived ideas about what the data would reveal, but rather for the prevalent concepts to be brought forward after reading and re-reading the selected data. This phase involved substantial data processing and 'ground work'.

The strategy for this phase was to identify concepts through the lens of the 3 main themes of the research; social enterprise, business models and sustainability. The results are indicated in Figure 5, Figure 6 and Figure 7. This phase is useful to identify concepts that previous authors have seen as important. It must be noted that the number of mentions that a concept has is not indicative of how important that concepts is. A concept with one mention could be just as important to the bigger picture as a concept with 10 mentions. The number of mentions gives an indication of how well grounded a concept is in literature.

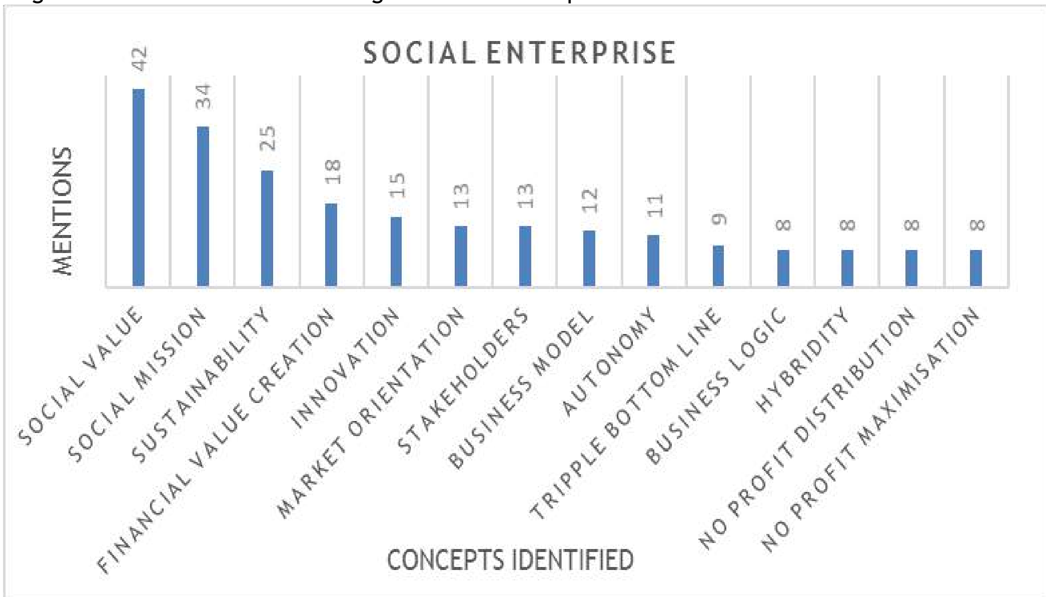


Figure 5: Social Enterprise Concepts Identified

For the SE theme, illustraed in Figure 5, it is interesting to note that sustainability and business models are frequently mentioned in the SE literature. Further, the author also identified that the SE is composed of 4 areas; community, financial logic, business logic and social logic. From this initial concept generation, it is evident that SEs are multi-facted and multi-diciplinary enterprises, with aspects of: economics, busness stretegy, business mangagement and the social sciences.

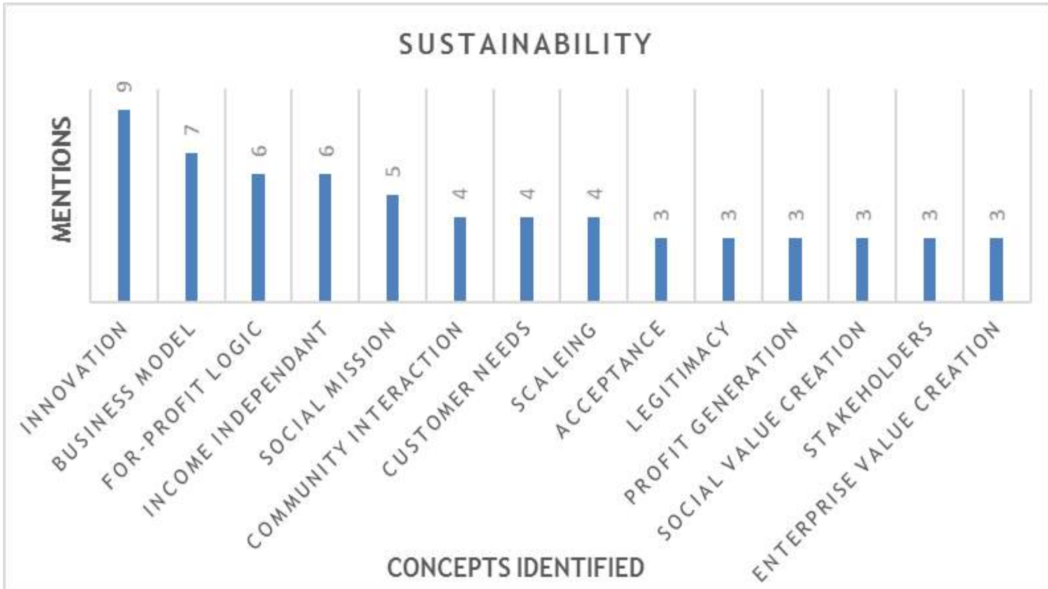


Figure 6: Sustainability Concepts Identified

Sustainability is interesting in that the concepts, as shown in Figure 6, that get the highest number of mentions in literature are financial in nature. Social mission only has the fifth most mentions in the articles reviewed. This is interesting and highlights the hybridity issue that SEs face: whether to pursue a financial mission, or social mission? It was also found that these concepts comprised the four areas of; community, financial logic, business logic and social logic. It is interesting to note how the SE and sustainability concepts correlate. Both focus on for-profit business logic, social mission, community engagement and social value creation.

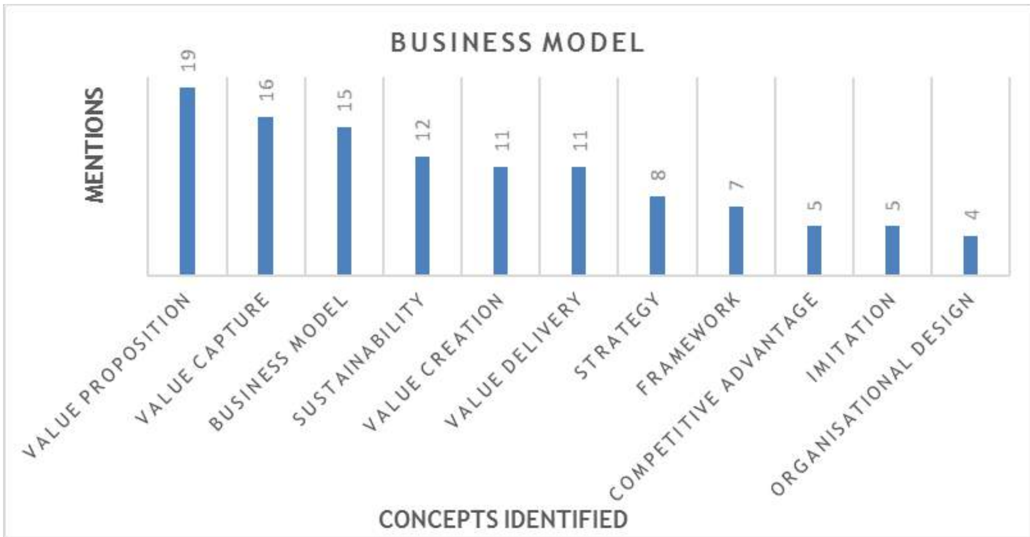


Figure 7: Business Model Concepts Identified

The initial identification of business model concepts, in Figure 7, gives an impression that business models are intrinsic to both SEs and organisational sustainability. An initial hypothesis is that business models are drivers of financial sustainability within SEs. The business model by itself will however not ensure sustainability of the SE, as the social mission also needs to be considered. The business model concepts identified in Figure 7 can be categorised in three groups: value, market and organisational design. These groups will be synthesized in Phase 5.

3.4 *Phase 4*: Deconstructing the concepts

In this phase, the goal was to dive deeper into the meaning of the concepts identified. It was desired to know the role, characteristics and assumptions of the selected concepts. Many concepts were identified in Phase 3 of the CFA. Before the concepts were researched in more depth, the identified concepts were reviewed to refine and synthesize the initial concepts identified. The concepts included in Phase 4 were selected based on: its importance to social enterprise, sustainability or business models, strength of relationship with surrounding concepts and how well grounded the concept was in literature. The concepts included are shown in Table 4.

Table 4: Deconstruction of concepts

Concept Name	Description	Reference
1. **Social Value**	Social value creation is the main goal of the SE. Social value is aimed at the community that the social mission is targeting, it includes: solutions to societal problems, social change, satisfaction of social needs, development of social goods and services, eliminating poverty, creation of jobs and ultimately, improving the quality of life of the community.	[15] [6]
2. **Social Mission**	It is the articulation of social value to be added in the community and is also expressed as social goals. The SE must have a clear, established and documented social mission. Social mission is prioritised in SE above economic drivers. The social mission is the driving force behind the SE's strategic direction. The social mission also affects the configuration of the enterprise.	[16] [6] [17] [8]
3. **Innovation**	Innovation leads to the integration of social mission, sustainability and solution design in SE's. It is a core element of delivering value and involves experimentation with new procedures, products and markets. It is also implemented in SE's to reduce the hybridity tension experienced from dual mission.	[8] [18] [15]
4. **Market orientation**	It is the orientation of the SE to meet its economic bottom-line. SE's incorporate business-like practices to ensure the financial continuity of the enterprise. This relieves the dependence on donations and ensures that the social mission can be met sustainably. It is also an articulation of how the organisation will meet the needs	[6] [19] [20] [8] [4]

	of its stakeholders. Market orientation is a driver for scaling an SE.	
5. **Stakeholders**	Stakeholders are the various parties impacted by the SE's activities. The market (and the public) is considered as stakeholders not customers. The views of stakeholders are incorporated into the objectives and values of the SE. The social and economic value generated by the SE, is aimed at its stakeholders. Thus, for an SE to be successful, it must satisfy its stakeholders.	[21] [6] [19] [20] [9]
6. **Autonomy**	SE's are characterised by being autonomous before the state, financially and in its decision making. Financial sustainability is imperative for autonomy and business models are used to bolster financial autonomy.	[16] [6] [18]
7. **Triple bottom line**	The social, economic and environmental goals that guide the SE. The needs of the people, profit and planet. SE's often employ the double bottom-line: the economic and social goals. Environmental goals are not a defining trait of SE's.	[6] [19]
8. **Hybridity tension**	The trade-off between social and economic missions cause internal tensions. Governed by double bottom line. SE's have a hybrid organisational setup, to ensure social mission is met, whilst generating sustainable economic returns. The SE organisational form also spans the public and private sector.	[8] [6] [19] [17]
9. **Scaling**	The ability of an SE to increase the reach of programs while maintaining financial sustainability. Seen as a central issue for sustainable business models, which has been identified as a key driver to achieving scale. Characteristics required for scaling: staffing, communication, alliance building, lobbying, earning generation, replication and stimulating market forces. Scaling can also be supported by increasing the range of value propositions.	[22] [23] [23] [8]
10. **Value Proposition**	Basic element of business model. The products and services that create value for customers and generate a profit for the enterprise. Value propositions are only valid for a specific customer segment. It represents the product-service system developed by a business. Its essential pillars are: the product/service, customer segments and customer relationships.	[23] [24] [14] [9]
11. **Value capture**	One of the minimum requirement for a business model. The two focusses of value capture are: how to generate revenue and how to keep costs low. Value capture is the mechanism used to secure a return from the value proposition.	[24] [14]
12. **Value Creation**	It refers to how processes are executed. It also details how resources are transformed into a value proposition. It relates to the organisational activities that form the value proposition. Value creation is also seen as venturing into new business areas and markets.	[24] [14] [9]
13. **Value Delivery**	Often linked with value creation. Value delivery is who the value proposition is targeted at and how it will be delivered to them. Customer segments are identified as well as distribution channels. It also deals with: resource acquisition, channel management, partner management and technology use.	[24] [14]

3.5 *Phase 5 & 6*: Integrating and synthesizing concepts

During this phase, concepts were integrated and synthesized to reduce the number of concepts and to indicate the relationship between concepts. The result is displayed in Figure 8. There are multiple ways for the concepts to be integrated and synthesized, Figure 8 merely represents the author's interpretation and insight.

4. CONCLUSION AND RECOMMENDATIONS

With social depravation and inequities on the rise, social enterprise has been the response offered to combat state and market failure. These SEs are required to be autonomous, sustainable and meet the social needs that are not being fulfilled by government or the private sector. Research has been focussed on definition of social enterprise but has not sufficiently examined the concepts of sustainability and business models within the SE context.

As research pertaining to sustainability and business models in the SE context is in its infancy, it was desired to derive a conceptual framework that could synthesize the key aspects of sustainability, business models and social enterprises. To achieve this, the conceptual framework analysis, proposed by Jabareen [12], was synthesized with systematic literature review tools. Phases 1-6 of the conceptual framework analysis, shown in Table 1, was completed and resulted in the framework presented in Figure 8.

The proposed framework identified the main components to be considered with regards to sustainable business models for social enterprises. For future research it is recommended that the underlying themes of; value capture, value creation, value delivery, social value and market orientation be researched in depth. Although the broad conceptualisation offers an understanding of the topic, there needs to be an in depth, 'nuts and bolts' analysis on each of these themes to discover what drives their success respectively.

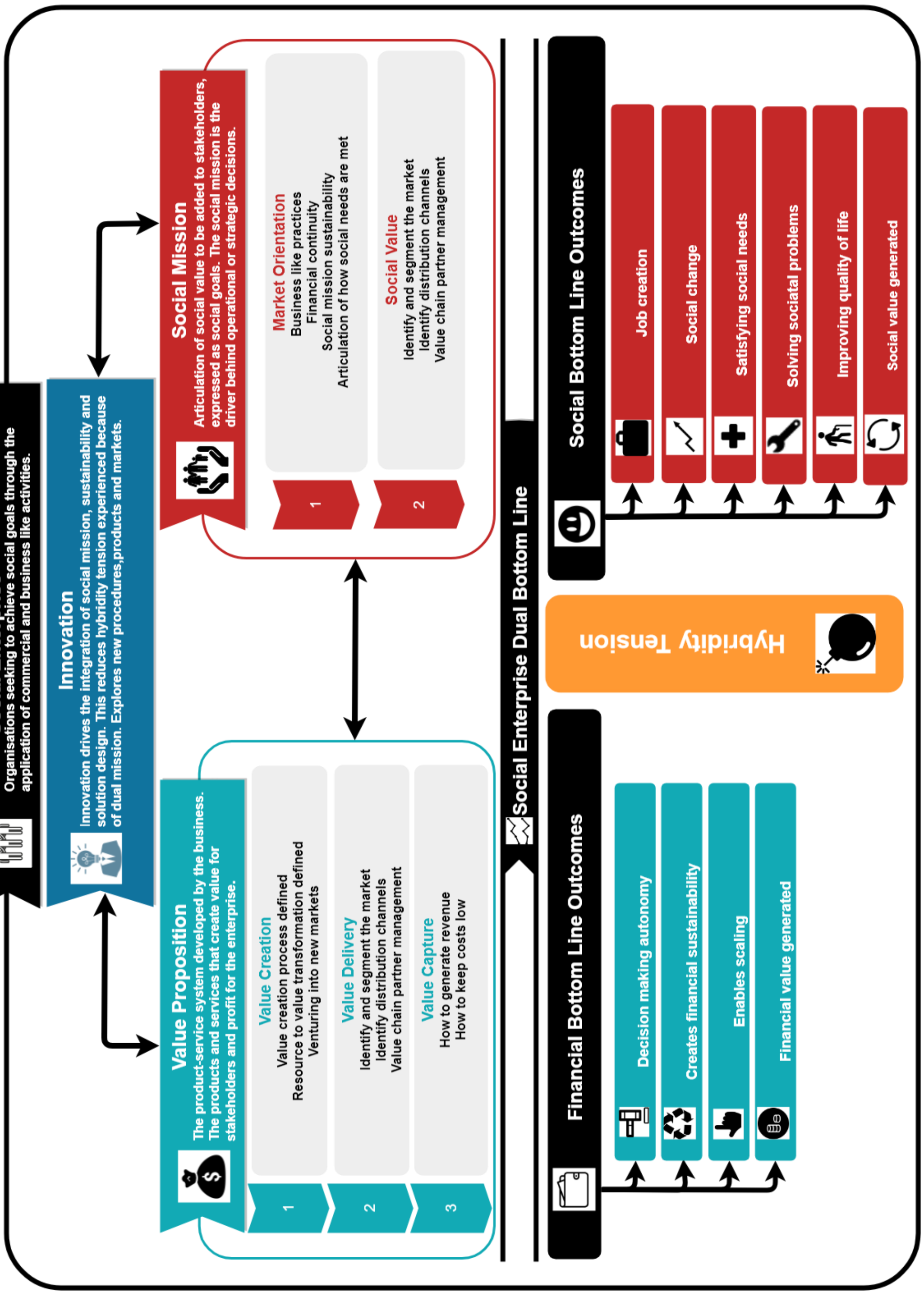


Figure 8: Synthesized framework of sustainability, business model and social enterprise concepts

5. REFERENCES

[1] **Ashraf, M., Razzaque, M., Siaw-Teng, L., Ray, P.,** and **Hasan, R.** 2018. Social business as an entrepreneurship model in emerging economy, *Management Decision*

[2] **Teasdale, S.** 2012. What's in a Name? Making Sense of Social Enterprise Discourses, *Public Policy and Administration*, 27(2), pp. 99–119.

[3] **National Treasury** 2018 Budget Highlights, 2018.

[4] **Granados, M. L., Hlupic, V., Coakes, E.,** and **Mohamed, S.** 2011. Social enterprise and social entrepreneurship research and theory, *Social Enterprise Journal*, 7(3), pp. 198–218.

[5] **de Souza João, I., Chiappetta Jabbour, C. J.,** and **Galina, S. V. R.** 2017. What Is a Social Enterprise? Revising Old Concepts and Interviewing Social Entrepreneurs, *Journal of Organisational Transformation and Social Change*, 14(2), pp. 127-147.

[6] **Sabella, A. R.** 2016. A strategic perspective of social enterprise sustainability, *Journal of General Management*, 41(4), pp. 71-89.

[7] **Teece, D. J.** 2010. Business models, business strategy and innovation, *Long Range Planning*, 43(2–3), pp. 172–194.
[8] **Goyal, S.**, **Sergi, B.**, and **Kapoor, A.** 2017. Emerging role of for-profit social enterprises at the base of the pyramid: the case of Selco, *Journal of Management Development*, 36(1), pp. 97–108.
[9] **Olofsson, S.**, **Hoveskog, M.**, and **Halila, F.** 2018. Journey and impact of business model innovation: The case of a social enterprise in the Scandinavian electricity retail market, *Journal of Cleaner Production*, 175, pp. 70–81.
[10] **Simanis, E.** 2012. Reality Check at the Bottom of the Pyramid, *Harvard Business Review*, , pp. 120–125.
[11] **Bryman, A.** and **Bell, E.** 2014. *Research methodology: Business and management contexts*, 1st Edition, *Oxford:* Oxford University Press.
[12] **Jabareen, Y.** 2009. Building a Conceptual Framework: Philosophy, Definitions, and Procedure, *International Journal of Qualitative Methods*, 8(4), pp. 49–62.
[13] **Scopus** Scopus, 2018. [Online]. Available: https://www.elsevier.com/solutions/scopus. [Accessed: 06-Jun-2018].
[14] **Davies, I. A.** and **Chambers, L.** 2018. Integrating hybridity and business model theory in sustainable entrepreneurship, *Journal of Cleaner Production*, 177, pp. 378–386.
[15] **Rahdari, A.**, **Sepasi, S.**, and **Moradi, M.** 2016. Achieving sustainability through Schumpeterian social entrepreneurship: The role of social enterprises, *Journal of Cleaner Production*, 137, pp. 347–360.
[16] **Margiono, A.**, **Zolin, R.**, and **Chang, A.** 2018. A typology of social venture business model configurations, *International Journal of Entrepreneurial Behavior & Research*, 24(3), pp. 626–650.
[17] **Doherty, B.**, **Haugh, H.**, and **Lyon, F.** 2014. Social enterprises as hybrid organizations: A review and research agenda, *International Journal of Management Reviews*, 16(4), pp. 417–436.
[18] **Goyal, S. and Sergi, B.** 2015. Social Entrepreneurship and Sustainability – Understanding the Context and Key Characteristics, *Journal of Security and Sustainability Issues*, 4(3), pp. 269–278.
[19] **Sengupta, S.**, **Sahay, A.**, and **Croce, F.** 2017. Conceptualizing social entrepreneurship in the context of emerging economies: an integrative review of past research from BRIICS, *International Entrepreneurship and Management Journal*, pp. 1–33.
[20] **Shuayto, N.** and **Miklovich, P. A.** 2014. Creating Organizational Sustainability in Social Enterprises: The Use of Evidence-based Positioning and Market Orientation, *Review of Bussiness & Finance Studies*, 5(2), pp. 109–115.
[21] **Darby, L.** and **Jenkins, H.** 2006. Applying sustainability indicators to the social enterprise business model, *International Journal of Social Economics*, 33(5/6), pp. 411–431.
[22] **Chutukuta, S.** 2016. Challenges to Scaling Inclusive Innovations : Four Case Studies from the Healthcare Sector in the Western Cape Province of South Africa.
[23] **Palomares-aguirre, I.**, **Barnett, M.**, **Layrisse, F.**, and **Husted, B. W.** 2018. Built to scale ? How sustainable business models can better serve the base of the pyramid, *Journal of Cleaner Production*, 172, pp. 4506–4513.
[24] **Carayannis, E. G.**, **Grigoroudis, E.**, **Sindakis, S.**, and **Walter, C.** 2014. Business Model Innovation as Antecedent of Sustainable Enterprise Excellence and Resilience, *Journal of the Knowledge Economy*, 5(3), pp. 440–463

MANAGING CUSTOMER EXPERIENCE USING DATA ANALYTICS IN A PARTNERING VENTURE

M. Roos[1*] & J. Bekker[2]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
roosmaryke@gmail.com

[2] Department of Industrial Engineering
Stellenbosch University, South Africa
jb2@sun.ac.za

ABSTRACT

Today's technologies are changing the physical, digital and biological worlds. This change impacts the economy and how industries operate. In the light of this, we ask how these changes can be incorporated into customer experience. To investigate this, a demonstrator was developed in which a customer's activities are simulated on a full-scale partnering platform while the business data are captured and analysed. The focus is on the domain of travel, in which customers use several modes of transportation while engaging with various collaborating enterprises. This paper will discuss the results and findings obtained, showing how customer experience can be managed and improved with the use of data analytics in a partnering venture.

[1] The author was enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

Technology has been around since the seventeenth century and has played a fundamental role in the economic growth, productivity and welfare of countries [1, 2]. The past three industrial revolutions have demonstrated these changes. The first industrial revolution introduced mechanical production, the second brought about mass production and the third enabled automation by using computers. Today we are in the fourth industrial revolution in which the interaction between technologies impacts the physical, digital and biological worlds. These technologies include nanotechnology, artificial intelligence, quantum computing, gene sequencing, *etc*.[3, 4]. But in this fourth industrial revolution, how does one manage and improve customer experience to increase revenue?

An example of how technology can be used to manage customer experience, is the Airbus A380 aircraft *first class suite* of Singapore Airlines. Some of its features include a seat that can rotate 360 degrees, a bed and a TV whose position can be changed depending on where you are positioned in the suite. The food is made according to the customer's preferences and the Airbus A380 has an entertainment system that keeps a record of the customer's choices [5, 6]. Another example of how technology can be used to manage a customer's experience is the *TripIt* digital system in which a customer sends his or her flight, car rental, hotel and restaurant details and the system then creates a master itinerary from it. The system alerts the customer on a real-time basis on the progress of the trip [7, 8]. Digital systems similar to this are *TripCase* and *WorldMate*. These systems focus on managing a customer's experience based on details provided by the customer, but it does not provide insights from it in order to improve a customer's experience.

Therefore, the aim of the study is to determine how data analytics can be used on a cross-functional business partnering platform in order to increase and manage a customer's experience. This study forms part of a mutual, ongoing research project done by the *Unit for System Modelling and Analysis* (USMA) in the Industrial Engineering Department of Stellenbosch University. For this study, a capability demonstrator will be developed, known as the *Trip Planner*. The Trip Planner will act as a *digital information and support system* in order to demonstrate how to manage and improve a customer's experience with the use of data analytics and business partnering. It will plan a trip for a customer based on a departure date, return date, destination and a budget entered by the customer, together with the customer's historical behaviour and preferences. Using this approach, the Trip Planner aim to deliver a 'superior' customer experience, because it will exceed the customer's expectation.

During the 28th annual SAIIE conference in 2017, the architecture for the Trip Planner demonstrator was presented [9] and the remainder of this paper will discuss how it is developed and the results and findings. The first section provides a concise summary of the theoretical concept investigated to get a better understanding of how the Trip Planner should be developed. The second section gives an overview of the model development for the Trip Planner demonstrator. The third section provides an overview of the simulator used in the Trip Planner demonstrator. The fourth section discusses the results and findings gained from it. The conclusion gives a summary of the key concepts discussed and how this study contributes to the management and improvement of customer experience, with the use of data analytics on a business partnering platform.

2. A LITERATURE BACKGROUND FOR THE TRIP PLANNER

The first step of this study was to conduct a literature study based on customer experience and its management, Big Data and its analytics, as well as business partnering on a cross-functional platform and how these three subjects integrate with each other. The paper in the SAIIE28 conference proceedings [9] provided an extensive overview of the literature study. This section will provide a concise summary of Customer Experience Management and Business Partnering on a cross-functional platform, as well as how machine learning contributes to the management of customer experience.

2.1 Customer experience management

To understand how a customer's experience can be improved and managed, it is important to know what Customer Experience is and to be familiar with the management approach required to address this.

Customer Experience is defined by Best et al. [10], Meyer & Shwager [11] and Richardson [12] as the "sum-total of experiences a customer has with a product and/or service. The experiences expand throughout all interactions a customer has with that particular product and/or service. That is from the first interaction until discontinued use". From this definition it is evident that it is more than just whether a customer is satisfied with a product or service. Customer Experience also takes place at various touchpoints during a customer's interaction with a product or service, whether it is during the buying, using and/or sharing phase. But how does one manage and improve it?

Various management approaches exist, such as Customer Experience Management, Customer Interaction Management, Customer Knowledge Management, Customer Relationship Management and Service Quality

Management. The two most used approaches in industry are Customer Experience Management (CEM) and Customer Relationship Management (CRM) [10, 11, 13, 14].

For this study CEM will be used. Firstly, because it is built on top of good practices and processes of CRM. Secondly, CEM is a proactive management approach whereas CRM is a reactive management approach [11]. One can define CEM as a strategic management approach in which an organisation transforms from a business-centric to a customer-centric organisation [10]. By following this approach, an organisation will have the ability to proactively respond to the drives and needs of a customer in order to manage and improve the Customer Experience with its product and/or services [11]. In other words, the organisation will change its skills, systems and processes to deliver a 'superior' Customer Experience. According to Schmitt [13], the focus will not only be on the customer's transactions but rather on building rich and lifelong relationships with customers.

Therefore, CEM is the appropriate management approach to use for this study, as the aim is to plan a trip for a customer in a proactive manner based on their needs and desires. These needs and desires will be learned from analysing historical customer data and behaviour.

2.2 Business partnering on a cross-functional platform

Unfortunately, a customer's historical and behavioural data are not freely available. Therefore, to gather all relevant data available about a customer, business partnerships are required. The organisations in these partnerships should be able to share the data amongst each other on a platform. Firstly, it is important to understand what business partnering is.

Business partnering is defined by Ellram & Hendrick [15] and Sheth & Parvatiyar [16] as a "relationship between two or more competing and/or non-competing organisations in which they share information amongst each other in order to share the reward or risk of such a relationship, whether it is for strategic or operational reasons". Burns et al. [17] point out that most business partnerships are used today for analytical and advisory purposes. Therefore, for organisations to enter into a business partnership, it is important to know who the parties involved are as well as the purpose of the partnership. Once the business partnerships have been established, the organisations can share data amongst each other. Secondly, it is important to know how the data will be shared.

An approach to this is to share data amongst parties across a platform, known as *business partnering on a cross-functional platform*. This term is synthesised by the authors, in which business partnering is combined with business platforms. In order to explain this concept, it is important to understand what a business platform is.

A business platform as defined by Truner [18] is a "platform which consists of a set of related systems, processes, people and data that deliver (or expose) a set of business services via a standard set of application program interfaces". Therefore, business platforms give an organisation the ability to transform its business processes by using a set of organisational principles and business capabilities. It is also a medium for an organisation to share and analyse its customer data.

It is also important to understand what a cross-functional platform is. An organisation that participates in the partnership, signs up to this platform for various functionalities. The functionalities include learning more about their customers, to build trusting relationships, to share data and to generate more revenue. The functionalities can be unique or the same for all the organisations involved. But it is important that all the organisations should share non-sensitive customer data.

Therefore, business partnering on a cross-functional platform will be a technological architecture that is information based. It is important to note that it is a technological architecture, since it will give the organisations the ability to share and exchange data in order to add value for the involved partners as well as the owner of such a platform. Figure 1 shows an example of how it works.

Let say that in *A-Friendly-City* there is an owner of such a platform. The first business partner who opts into the platform is *Mobile XYZ* that is the leading telecommunication business in *A-Friendly-City*. The second business partner is *Tom Fly Airlines* that only flies domestically. The third business partner is *Airline Bees* that flies internationally and domestically. Although *Tom Fly* and *Airline Bees* are competitors, they still share customer data and generate value by using each other's business opportunities. The fourth and last business partner is *Night B&B* that is the only bed and breakfast business in A-Friendly-City and they own three facilities.

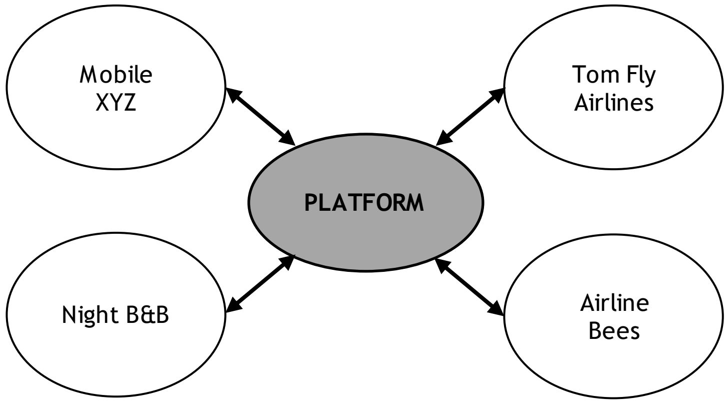


Figure 1: Business partnering on a cross-functional platform

All four of these partners can help each other with sharing relevant customer data. Mobile XYZ have access to all the browser and location information of its customers. By sharing this information, the other partners can learn how the customers travel in and around A-Friendly-City and what internet sites are accessed by the customers, in order to predict behavioural trends. Night B&B can share data about when the customers check in and out, so the airlines can better plan their flight schedules. The airlines can share information on when and how many customers are travelling in and out of A-Friendly-City and where these customers come from. More types of customer data can be shared on this platform in order for these organisations to create more value.

By looking at the example above, it is clear that business partnering on a cross-functional platform plays a vital role in the ability for the Trip Planner demonstrator to be a success. If partners do not opt into the Trip Planner's platform, it will not be able to improve and manage a customer's experience when they go on a trip.

2.3 Machine learning and artificial intelligence

Organisations have a lot of data available about a customer, but not all use it to learn more about their customers [19]. If all organisations applied analytical techniques on the customer datasets, they would have the ability to understand their customers better. By doing this, the organisation would be able to better manage and improve its customers' experiences and more revenue will be generated.

To do data analyses, Big Data Analytics is required. Machine Learning is used in the Big Data Analytics domain to perform data mining techniques on datasets in order to gain insights from it. As stated by SAS Insights [20], machine learning is a data analysis approach in which a computer can learn to perform specific tasks. It is made possible by learning from previous computations in order to gain insight from data and to produce decisions and results which are reliable and repeatable. The insights and results are determined by the application of complex mathematical calculations and algorithms.

But what is the link between machine learning and Artificial Intelligence (AI)? Marr [21] points out that machine learning forms a subset of AI. In simple terms, AI gives machines the ability to perform "smart" tasks and machine learning gives the machines the ability to learn how to do it. According to Hutson [22] "an important recent advance in AI has been machine learning, which shows up in technologies from spellchecking to self-driving cars and is often carried out by computer systems called neural networks".

A formal definition of AI as defined by Frankish & Ramsey [23] is that it "is a cross-disciplinary approach to understanding, modelling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices". In other words, AI aims is to understand and build intelligent agents or entities by using various approaches. AI also has the ability to not only to think and act rationally, but to also think and act humanely [24].

Therefore, for the Trip Planner demonstrator to be a success, machine learning techniques should be implemented in order to improve and manage each individual trip and to gain insight into the behavioural trends that exist. This will give the owner of such a system the ability to deliver a 'superior' customer experience in the travel domain and the participating organisations will be able to increase their revenue.

3. THE TRIP PLANNER DEMONSTRATOR

The focus of this study is to manage and improve a customer's experience in the domain of travel. It is based on a case study in which the purpose of the Trip Planner is demonstrated. The case study is about how a customer, Thandi, who lives in Cape Town, enters her trip requirement into the system. The trip requirement consists of a desired date of departure, a date of return, the destination and a budget. Thandi indicated that she wants to go to Durban next week Friday, return the following Monday and her budget is R5 000. The system will use this information to book her accommodation, transportation mode to get to Durban and the transportation mode used in Durban and Cape Town. After the booking is completed, the Trip Planner will notify Thandi of all the events that will occur. During the trip itself, alterations can be made to the trip, based on Thandi's feedback. The system will also notify Thandi about other events that might interest her based on her historical behaviour. For example, when she arrives at the airport, the system will inform her that the newest book by her favourite author is available in the bookstore.

To develop this demonstrator, first an architecture is required. The architecture was provided by Roos & Bekker [9], during the SAIIE28 conference. The model development is based on this architecture. As can be seen in Figure 2, the Trip Planner demonstrator consists of four parts, namely business partners, an information base, a simulator and a data analytics function. These four parts are interdependent of each other and they integrate with each other to perform as a digital information and support system.

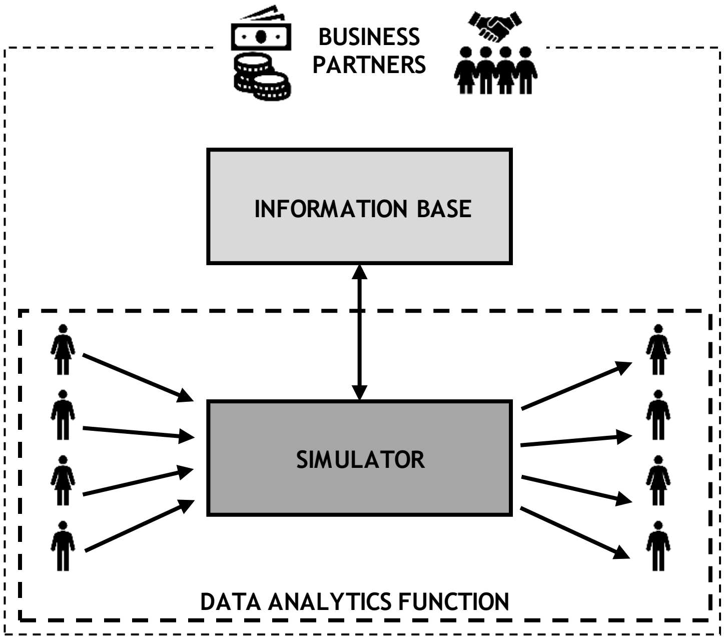


Figure 2: The Trip Planner demonstrator

The first part of the model is business partners. Without business partnerships, the Trip Planner demonstrator would not exist. The business partners are the backbone of the system, as they provide all the relevant customer data, but they are also the ones who will reap the benefits of this system.

The second part of the model is the information base. MS SQL Server is used as the database server for the Trip Planner information base. The data entities stored in the information base consist of the customers, customers' historical transactions, accommodation companies, long-distance transportation companies (airplanes and buses), short-distance transportation companies (hailing app taxis such as Uber, 'normal' taxis and car rental), customer preference for the type of accommodation, short-distance transport and long-distance transport, recorded customer experience at touchpoints, booked accommodation, long-distance transportation and short-distance transport, trip requirements and other data entities relevant to these data entities. All data used for the Trip Planner demonstrator have been extracted either from websites or fictional data have been simulated in Matlab based on statistics.

The third part of the model is the simulator, where all the magic happens. The simulator has been developed in Matlab. The first functionality of the simulator is that it will use data from the information base to plan a customer's trip, based on the trip requirement, customer preferences and customers' historical behaviour. The second functionality of the simulator is that it will manage the customer's trip while he or she is travelling. In other words, it will record the various touchpoints on the trip, make suggestions for other events that may be relevant to the customer and make alterations to the trip if required. The simulator will suggest unique trips for many customers. A customer can use the system repetitively and each time he or she will receive a unique trip that will manage their experience more effectively. The actions performed by the simulator is discussed in the next section.

The fourth part of the model is the data analytics function. Based on the previously completed trips, machine learning will be performed to gather insights. The machine learning tool that will be used for this is supervised learning as it makes future predictions by using a known dataset [25]. The two data mining techniques that are a subset of supervised learning are classification and regression techniques. Classification is a predictive data analysis technique in which associated class labels are used to divide data into predefined classes and regression is when relationships between dependent and one or more independent variables are explored in a given dataset [26, 27]. The data analytics function model will use the classification techniques in the simulator and to determine the results and findings of the Trip Planner.

4. THE SIMULATOR OF THE TRIP PLANNER DEMONSTRATOR

The simulator is the enabler of the processes in the Trip Planner demonstrator, as its prime functionality is to simulate unique trips for many customers. A unique trip will be simulated for a customer based on the trip requirements and the customer's historical behaviour and preferences. The simulator has been developed in the Matlab software. In order to develop this simulator, it is important to understand the different phases that it should undergo. In other words, what phases are required for a customer to go on an actual trip which is solely planned by the simulator based on the trip requirements in which the customer's experiences will be managed and improved.

In order to know what phases are required, it is important to know the processes required for the Trip Planner demonstrator. The processes can be divided into two processes as defined in the architecture, namely the *trip planning* process and *customer travelling* process. The trip planning process is the process of where a customer's trip is planned and booked. The customer travelling process is when the customer receives notifications on the upcoming events of the trip and the customer rates the experience at the various touchpoints. The phases required to enable the simulator to run these processes can be seen in the concept model of the simulator as presented in Figure 3.

The first two phases forms part of the tip planning process. In the first phase a customer accesses the system and enters their trip requirements in order for a trip to be booked and planned. The simulator will choose a random customer and it will then assign a start and end date, a destination that is in a different province from which the customer resides in and a budget that is depending on the duration of the trip as well as the destination of the trip.

The second phase of the simulator is the booking procedure. The booking procedure is done by applying rules on the information available about the customer's historical behaviour, preferences and loyalty programs where applicable. The simulator will choice the best fitted option for the booking of accommodation, long-distance transportation and short-distance transportation based on these rules.

The last three phases forms part of the customer travelling process. During the third phase, notifications are sent out to the customers as the events of the trip unfold. The events are all touchpoints on the customer journey, where the customer journey is the actual trip. In other words, the simulator will send a notification prior to an event to inform the customer with the necessary information. Events include upcoming flights, bookings for car rentals, taxi's arrival time, check-in and check-out times at the accommodation, special offers based on historical transactions, *etc*. The special offers will only occur during the time when a customer is at a long-distance transportation station or airport. It has only been limited to long-distance transportation to demonstrate to what extend a customer experience can be managed and improved.

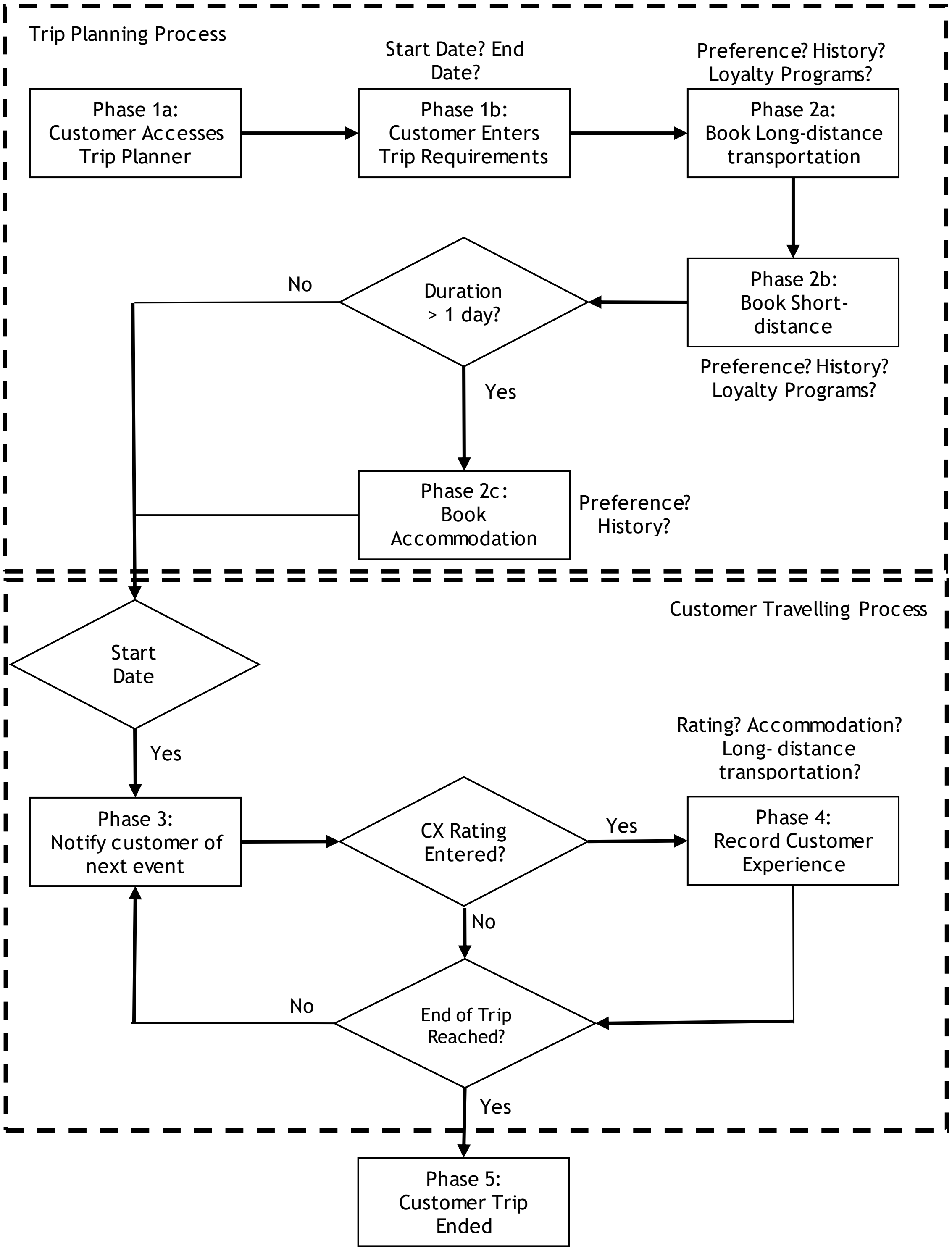


Figure 3: The concept model of the simulator

The fourth phase is when a customer rates the experience of a touchpoint, which can occur during or after the touchpoint. In order for the simulator to capture a customer experience, the customer has to physically rate their experience at these various touchpoints. But a customer will not necessarily rate their experience at every touchpoint in the trip and they will most probably only rate their experience if they had an above or below average experience. Therefore, two approaches can be used for the recording of the customer experience. Either an average experience can be recorded if the customer did not rate the experience or it will only record the rating entered by the customer considered that it will only be above or below average experience. For the purpose of this study, the latter approach will be used as the aim of the study is to show how this system can improve and manage a customer's experience. The customer experience rating is based on a truncated Poisson distribution (which is discrete and finite). The lambda parameter of the Poisson distribution is forced to vary by means of a truncated exponential distribution. This emulates unpredictable customer behaviour.

The fifth phase is when a customer has completed a trip. During this phase the rating of the touch points are analysed in order to improve the historical outcome of the rule-based application. It is also important to note that after the customer has rated a touchpoint, the only future events in a trip that can be changed is the taxi service of the short-distance transportation and special offers based on historical transactions. All other events remain fixed. But for the next trip of a customer, the accommodation, long-distance transportation or any type of short-distance transportation can be changed if necessary. This reflects to what levels a customer experience can be influenced.

The simulator is therefore the fundamental part of the Trip Planner demonstrator in order to show how the customer experience can be improved and managed in the travel domain. The simulator is a rule-based system in which the system became cleverer as history is built up. The outcome of the rule-based application is influenced as history is build.

5. RESULTS AND FINDINGS

For this article, a concise database is used to demonstrate the functionality of the Trip Planner. Only airplanes have been used for the *long-distance transport* and only customers living in Cape Town, Johannesburg, Durban and Bloemfontein have been included.

As has been highlighted in the previous section, in order to capture a customer experience, the customer has to physically rate their experience at various touchpoints. A customer will not necessarily rate an experience at every touchpoint during the trip and will usually only rate the experience if it was above or below average. Therefore, only the extreme points are analysed.

The results and findings obtained from the application of machine learning techniques, namely the weighted average and rule-based classification, can be one of two sets. The first is that it will show how a customer's experience has been improved and managed within a trip and/or how it has been improved from one trip to the next. The second represents the analysis that was done based on the customer experience rating. By doing analysis it will show which services most customers have been unhappy with and which services customers have rated higher. The latter results will indicate which companies need to improve their customer services and which ones have been doing well and it will give an overall view of how customer's experience journeys in the travel domain.

For the first set of results, the following two examples from the database will be discussed (shown in Table 1). The first case is for *Tendai Mbende*, who has travelled twice from Bloemfontein to Johannesburg. For the first trip, the system booked her flights with *KayLines* but she was highly dissatisfied and gave the experience a poor rating (1 out of 5). Based on this, the system booked her flights for the second trip on *Cheetah Airways* by using the rule-based analysis, which she gave an above average rating (4 out of 5). The second case is for *Brady Pieterse*, who travelled from Cape Town to Port Elizabeth. There he first used *XYTransport* but he was he was not happy with the experience and gave it a below average rating (2 out of 5). Therefore, for the next on-ground transportation (short-distance transport) he required, the system booked *Hop Around*. After he used Hop Around, he did not record the experience, which indicated that he was reasonably satisfied. Other similar examples can be drawn from the results. These examples represent the capability of the system to improve and manage a customer's experience during and after a trip by using the intelligence gained as history is build up.

Table 1: Example of trip improvements

Customer	Trip	Destination	Airline	Short Distance Transport	Accommodation
Tendai Mbende	1	Johannesburg	KayLines	Wheel on the hire	Angala Hotel de Boutique
	2	Johannesburg	Cheetah Airways	Wheel on the hire	Angala Hotel de Boutique
Brady Pieterse	1	Port Elizabeth	Brightest Airways	XYTransport Hop Around	None

For the second sets of results, various classification techniques have been applied on the touchpoints in order to determine which services need improvement by using the *weighted average* method. Analysis was done for the long-distance transportation, short-distance transportation, accommodation and customers. The output after applying the weighted average can be seen in Figure 4. The y-axis represents the Customer Experience Rating entered by the customer and the x-axis represents the company and customer ID number used in the database. The top and bottom-rated companies have also been identified and can be seen in Table 2.

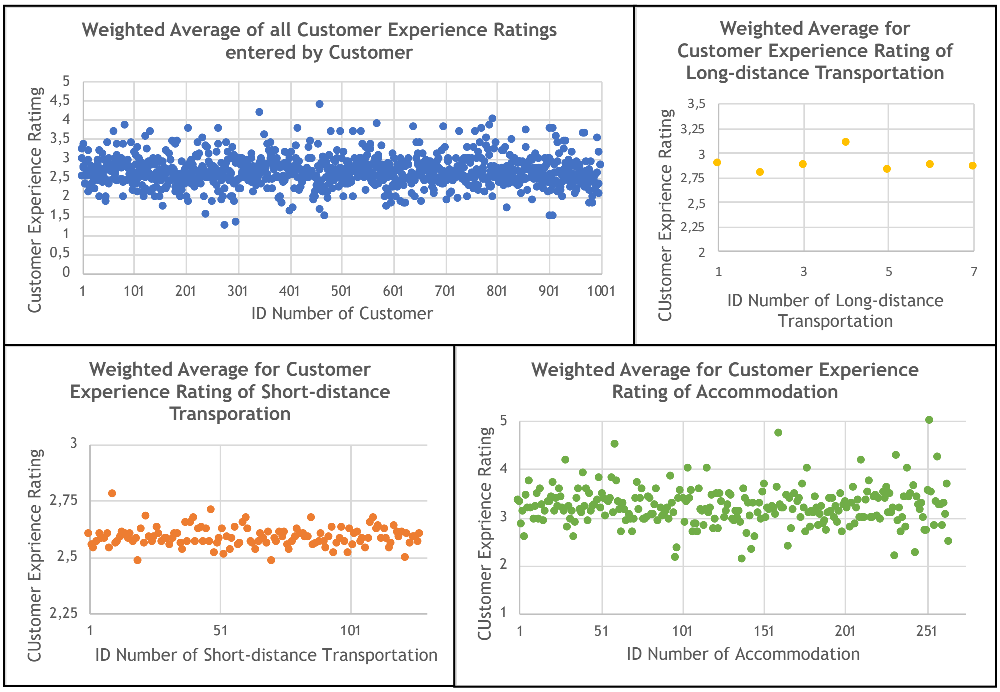


Figure 4: Weighted average plots

Table 2: Top and bottom-rated companies

Type of Service	Top companies	Bottom companies
Accommodation	1. Valdior de Vile (Durban)	1.Dock de Bridge Hotel & Suites (Cape Town)
	2. Lalaria de Sleep Plasa (Durban)	2. Di'o Hoek Country Hotel (Durban)
	3. Lofts de Bohemian (Cape Town)	3. Soonshowers Place (Pretoria)
Long-distance Transport	Yendiza SA	Central Airways
Short-distance Transport	1. Friendly Car Rental	1. The Dolphin Cab
	2. Mozzie and Jen Cabs	2. Willie's Taxi
	3. Friendly Haily (Sedan)	3. Ride and Pay (Sedan)

Based on these results, it is clear that the system can effectively book an entire trip for a customer and by doing so, it also has the ability to improve and manage a customer's experience. It can also highlight the top and bottom-rated companies, for these companies to either improve their services or to keep on doing what they are doing. Future work will include the expansion of the database. In other words, the type of long-distance transportation options offers airplanes and buses and will also include all areas in South Africa, not only the major cities.

6. CONCLUSION

This article provided insight into the work completed in the domain of travel in order to show how customer experience can be managed by using data analysis and business partnering on a cross-functional platform. This work contributes to the mutual ongoing project of the USMA research team in order to show how Big Data Analytics can be used in a business to create value.

Based on the results and findings obtained from the Trip Planner demonstrator, it is clear that the system can effectively manage and improve a customer's experience. It can also highlight top and bottom-rated companies. The Trip Planner demonstrator also shows how customer experience can be improved using modern technologies. For the implementation of such a system, an industrial engineer is an ideal candidate due to their ability to understand systems and the integration of it.

REFERENCES

[1] **Buchanan, R.A.** 2018. *History of Technology*, Encyclopaedia Britannica. [Online] Available: https://www.britannica.com/technology/history-of-technology. [Accessed: 21 May 2018].

[2] **Morrar, R.; Arman, H. & Mousa, S.** 2017. The Fourth Industrial Revolution (Industry 4.0): A Social Innovation Perspective, *Technology Innovation Management Review*, 7(11), pp. 12 - 20.

[3] **Liu, Y. & Xu, X.** 2017. Analysis, Industry 4.0 and Cloud Manufacturing: A Comparative Analysis, *Journal of Manufacturing Science and Engineering*, 139(3), pp. 034701-1 - 034701-8, 2017.

[4] **Schwab, K.** 2016. *The Fourth Industrial Revolution*, Switzerland: World Economic Forum.

[5] *New A380 Suites*. 2018. Singapore Airlines. [Online] Available: https://www.singaporeair.com/en_UK/gb/flying-withus/cabins/suites/new-a380-suites/. [Accessed: 21 May 2018].

[6] **The Points Guy**. 2018. *Review: Singapore's New A380 Suite, the World's New Best First Class*. [Online] Available: https://thepointsguy.com/2018/02/sq-new-a380-suite-review/. [Accessed: 21 May 2018].

[7] *What is TripIt - How it Works*. 2018. TripIt. [Online] Available: https://www.tripit.com/learnMore [Accessed: 21 May 2018].

[8] *Instantly Organize Travel Plans with TripIt*. 2017. YouTube. [Online] Available: https://www.youtube.com/watch?v=KRT2BBwHRIM. [Accessed: 21 May 2018].

[9] **Roos, M. & Bekker, J.** 2017. *An Architecture for Customer Experience in a Partnering Venture*, in SAIIE28 Conference Proceedings, South Africa, 2017.

[10] **Best, I.; Shamir, S.; Amalfitano, M.; Cotton, S.; Kawecki, A.; Chowdhury, S.; Mitra, S. & Sendel, R.** 2016. Customer Experience Management — Introduction and Fundamentals, *TM Forum*, GB962.

[11] **Meyer, C. & Schwager, A.** 2007. *Understanding Customer Experience*, Harvard Business Review. [Online] Available: https://hbr.org/2007/02/understanding-customer-experience. [Accessed: 23 May 2018].

[12] **Richardson, A.** 2010. *Understanding Customer Experience*, Harvard Business Review. [Online] Available: https://hbr.org/2010/10/understanding-customer-experie. [Accessed: 23 May 2018].

[13] **Schmitt, B.** 2003. *Customer experience management: A revolutionary approach to connecting with your customers*, New York: Wiley.

[14] **Walden, S.** 2017. *Customer Experience Management Rebooted: Are you an Experience brand or an efficiency brand*, London: Macmillan.

[15] **Ellram, L. & Hendrick, T.** 1995. Partnering Characteristics: A Dyadic Perspective, *Journal of Business Logistics*, 16(1), p. 41–64.

[16] **Sheth, J. & Parvatiyar, A.** 1992. Towards a Theory of Business Alliance Formation, *Scandinavian International Business Review*, 1(3), p. 71–87.

[17] **Burns, J.; Warren, L. & Oliveira, J.** 2014. Business partnering: Is it all that good?, *Controlling and Management Review*, 58, pp. 36 - 41.

[18] **Truner, A.** 2017. *Platforms: Design, deploy, diversify, repeat*, TM Forum Inform. [Online] Available: https://inform.tmforum.org/features-and-analysis/2017/05/plat forms-design-deploy-diversify-repeat/. [Accessed: 23 May 2018].

[19] **Hopkins, B.** 2016. *Think You Want to Be "Data-Driven"? Insight Is the New Data*, Forrester. [Online] Available: https://go.forrester.com/blogs/16-03-09-think_you_want_to_be_data_driven_insight_is_the_new_data/ [Accessed: 23 May 2018].

[20] *Machine Learning: What it is and why it matters*. n.d. SAS Insights. [Online]. Available: https://www.sas.com/en_za/insights/analytics/machine-learning.html. [Accessed: 23 May 2018].

[21] **Marr, B.** 2016. *What Is The Difference Between Artificial Intelligence And Machine Learning?*, Forbes, [Online]. Available: https://www.forbes.com/sites/bernardmarr/ 2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/ - 445045c72742 [Accessed: 23 May 2018].

[22] **Hutson, M.** 2017. AI Glossary: Artificial intelligence, in so many words, *Science*, 357(6346), p. 19.

[23] **Frankish, K. & Ramsey, W.M.** 2014. *The Cambridge Handbook of Artificial Intelligence*, Cambridge University Press.

[24] **Russell, S.J. & Norvig, P.** 2016. *Artificial Intelligence: A Modern Approach*, 3rd edition, Pearson.

[25] **Kotsiantis, S.B.** 2007. Supervised Machine Learning: A Review of Classification, *Informatica*, 31, pp. 249-268.

[26] **Ngai, E.; Hu, Y.; Wong, Y.; Chen, Y. & Sun, X.** 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature, *Decision Support Systems*, 50(3), pp. 559-569.

[27] **Han, J. & Kamber, M.** 2006. *Data Mining: Concepts and Techniques*, 2nd edition, Morgan Kaufmann Publishers, pp. 285-464.

A DECISION SUPPORT TOOL FOR QUANTIFYING THE RISK PROFILE OF SOUTH AFRICA'S PHARMACEUTICAL SUPPLY DISTRIBUTION NETWORK

C.W. Nolting[1]* & S.S. Grobbelaar[2] & J.H. van Vuuren[3]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
18186874@sun.ac.za

[2]Department of Industrial Engineering
Stellenbosch University AND DST-NRF Centre of Excellence in Scientometrics and Science, Technology and Innovation Policy (SciSTIP)
Stellenbosch University, South Africa
ssgrobbelaar@sun.ac.za

[3]Department of Industrial Engineering
Stellenbosch University, South Africa
vuuren@sun.ac.za

ABSTRACT

South Africa is facing a serious burden of disease, which is exacerbated by a dysfunctional public health care system. Blind, general estimates are often made in respect of crucial pharmaceutical inventory variables. This results in slow-moving pharmaceutical drugs being overstocked and fast-moving drugs being understocked at outlets and distribution centres. Consequently, large wastages are incurred in terms of drugs being discarded in bulk after passing their expiration date. As a result, patients are deprived of essential medicines due to enduring stock-outs. This paper aims to present the conceptual design of a decision support tool, which aids decision makers in determining pharmaceutical inventory variables that align with key objectives and keep the best interest of patients in mind.

[1] Enrolled for an MEng (Engineering Management) degree in the Department of Industrial Engineering, Stellenbosch University
[2] Associate Professor at the Department of Industrial Engineering, Stellenbosch University.
[3] Professor of Operations Research at the Department of Industrial Engineering, Stellenbosch University
*Corresponding author

1. INTRODUCTION

1.1 Background

For decades South Africa has been burdened by epidemics such as the human immunodeficiency virus (HIV/AIDS), tuberculosis (TB) and malaria. According to WHO, HIV/AIDS was the leading cause of death of South Africans, killing 202 100 people in 2012 [1]. Not only is the country haunted by a high frequency of communicable diseases, it also faces staggering numbers of non-communicable diseases, perinatal and maternal mortality, and violence-related injuries. South Africa therefore faces a quadruple burden of disease [2]. Patients receiving chronic medication related to communicable or non-communicable diseases often need to adhere strictly to their prescribed regimen in order to prevent drug resistance [3]. For this reason it is crucial that the right medicine is available to affected South Africans at all times.

It is not common knowledge that a large portion of primary health care facilities in the public health sector experience stock shortages of essential medicines on a regular basis. The Stock Outs National Survey, published in 2015, reported that on the day of contact approximately one in five facilities was affected by stock-outs of antiretroviral (ARV) and TB medicines, and one in ten experienced vaccine stock-outs [4]. Furthermore, 70% of stock-outs in South Africa lasted more than a month, indicating poor provincial supply management and a lack of emergency protocols. For this reason, initiatives such as the Stop Stockouts Project (SSP) exist. This project was established as a result of the Mthatha depot crisis in the Eastern Cape in 2013, where poor management led to the depot staff going on an unprotected strike, which subsequently led to stock shortages at the 300 medical facilities serviced by the depot [5].

The SSP monitors the availability of all essential primary health care medicines and children's vaccines by conducting annual telephonic surveys and logging reports through its SSP hotline. All gathered data are added to the case-management database, which is used to track reported national stock-outs up to a sub-district level. The SSP makes its gathered information available to the National Department of Health (NDoH), so that the department can attend to inadequate services experienced by citizens [4]. Regrettably, the annual survey has been discontinued, but the SSP hotline is still available and reports are still logged on *stockouts.org*.

In order to address the problem of stock-outs in the country, the NDoH commissioned multiple companies to develop solutions that can be implemented nationally. The Centralised Chronic Medicines Dispensing and Distribution (CCMDD) programme was introduced and implemented in 2014. It was developed in order to provide improved access to patients in need of chronic medication, as well as to decrease patient congestion at public health care facilities. By the end of 2014, the CCMDD covered prescriptions for 183 989 patients [6]. In 2015, electronic stock management solutions (ESMS) were implemented in 39 hospitals in order to monitor the availability of medicines and to effectively enhance demand planning [6]. In 2016, a proactive drug monitoring system, called Stock Visibility Solution (SVS), was introduced and implemented in 1 849 clinics across South Africa [7]. SVS is a platform developed by Mezzanine Ware in partnership with Vodacom to track stock levels at clinics in remote areas. SVS makes use of the capability of smartphones to scan barcodes and utilises Vodacom's vast mobile network to collect data in a central database. Trained medical staff are required to log their available stock once a week by scanning the barcode on a specific medicine and specifying the amount available of a particular drug, as well as its expiry date. The information is stored on a cell phone, until it can be uploaded to the central database via mobile network. The resulting database is a real-time status of the available stock of essential drugs across the network of participating clinics. In this manner clinics can be identified that are low on stock and need to be assigned a higher priority for resupply, but also to single out clinics that continually run out of stock and require further investigation by the NDoH's supply chain managers [8].

By August 2017, the CCMDD programme covered the prescriptions of 1 252 000 patients in need of chronic medicine, the ESMS was implemented at 123 hospitals, and SVS was implemented at 3 121 clinics. Additionally the Minister of Health, Dr Aaron Motsoaledi, stated that through the implementation of SVS the availability of ARV and TB medication has increased from 69.5% to 92.5% and from 65.7% to 88.5%, respectively [8].

The above-mentioned solutions address the stock management problem on an end-user outlet level and facilitate an important step towards end-to-end visibility in the South African pharmaceutical supply chain. These technologies, however, are not the solution to South Africa's primary health care supply problem, but are rather tools that can be utilised in the pursuit towards a functional and efficient demand-driven pharmaceutical supply chain.

1.2 Problem Statement

Primary health care workers in the South African public health care sector are overburdened and sacrifice a lot of time to capture stock data so that managers in the various levels of the NDoH have an overview of how specific facilities are performing. While solutions such as SVS have greatly simplified the stock-taking procedure and guarantee reliable data, the heavy burden of having to report on a regular basis and justify minor wastages persists. Having stock and demand visibility on a facility level is a very important step towards achieving an effective supply chain, but should not be treated as the solution − it is rather the beginning of a solution. The next step should be to establish inventory visibility between supply-chain nodes to allow for effective planning at all levels of the pharmaceutical supply chain and to move towards a preventative strategy rather than a reactive strategy.

This study addresses the problem of ineffective and inefficient inventory management across the South African pharmaceutical supply chain and proposes a possible solution to said problem by the conceptual design of a technology platform aimed at lending decision support to key role players of the pharmaceutical supply chain by providing them with drug priority rankings for each health district.

1.3 Scope and Objectives

To address the problem outlined in Section 1.2, the following objectives are pursued as part of this ongoing research for which preliminary results are presented in this paper:

I. *Review* the academic literature relevant to this project;
II. *Identify* the system user requirements and subsequent system specifications from the literature and related documents.
III. *Propose* a conceptual framework of an ideal modular decision support platform that encapsulates data sourcing, analysis and user interaction, assuming that perfect data are available;
IV. *Identify* possible sources of data that can be used to predict demand patterns of specific drugs, extract performance indicators, develop risk profiles, identify supply-chain bottlenecks, and build user profiles for various role players in the South African pharmaceutical supply chain;
V. *Outline* the planned future work.

1.4 Research Methodology

The *systems development life cycle* (SDLC) is used as a guideline throughout this research and the design of the system is conducted according to a top-down approach, starting at the systems level and delving deeper into the various sub-systems. More specifically, an adaptation of the waterfall model of software development will form the roadmap of this ongoing research, as depicted in Figure 1.

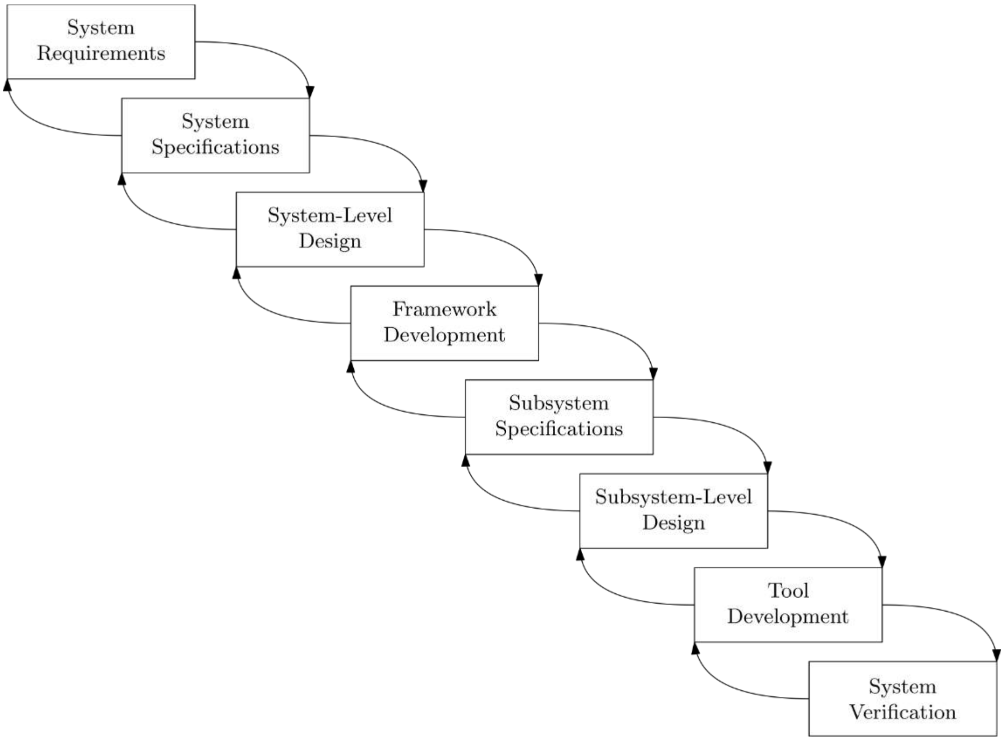


Figure 1: An adaptation of the waterfall methodology

This paper will only span the process up to and including an initial framework development.

2. LITERATURE REVIEW

This section contains a brief overview of the literature relevant to this project.

2.1 Supply Chain Concepts

The concepts reviewed in this section include *demand-driven supply-chain management* (DDSCM), *supply chain visibility* (SCV), the bullwhip- and ripple-effect, the *visibility and analytics network* (VAN), and *supply chain risk management* (SCRM).

2.1.1 Demand-Driven Supply Chain Management (DDSCM)

A demand-driven supply chain, otherwise known as an agile supply chain, responds to the actual product consumption instead of attempting to anticipate demand in advance through general, standardised heuristic forecasting techniques. This concept is most often applied to industries that cater to the distribution of a variety of different products with different properties where the demand is volatile, such as the retail and health care industries.

Capturing actual consumption data through modern visibility technologies allows decision makers to visualise the current area and quantity of demand of specific products, and also allows for a responsive management strategy. To implement a demand-driven supply chain requires *demand-driven supply-chain management* (DDSCM). A DDSCM strategy allows for timely and accurate order fulfilment and ensures customer satisfaction by providing the right product at the right time in the right quantity. This is achieved by focusing primarily on real-time information sharing, advanced inventory-management techniques, lead time reduction, and stakeholder collaboration [9].

2.1.2 Supply Chain Visibility (SCV)

Supply Chain Visibility (SCV) is defined by Goh *et al.* [10] as follows: *"SCV is the capability of a supply chain player to have access to or to provide the required timely information/knowledge about the entities involved in the supply chain from/to relevant supply chain partners for better decision support."*

SCV can be further characterised by having visibility of the process status, the current amount of inventory, the demand pattern, and visibility of exceptions and events in the supply chain, as visualised in Figure 2.

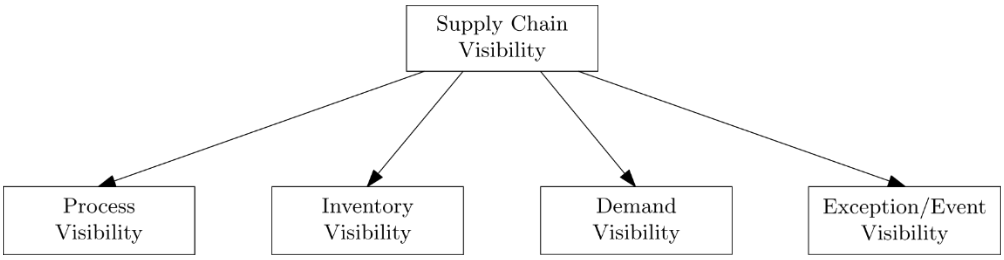


Figure 2: The concept of supply chain visibility adapted from Goh *et al.* [10]

2.1.3 Bullwhip-effect and Ripple-effect

The term bullwhip-effect refers to the phenomenon of increasing variability in order-frequency and order-quantity along a supply chain, due to a misinterpreted perception of demand-variability at the end-user level. The bullwhip-effect is often associated with high-frequency and low-impact disruptions, such as lead time and demand fluctuations, that can be rectified and balanced out within a short amount of time [11].

The ripple-effect, on the other hand, is the phenomenon of disruption propagation in a supply chain and its impact on the output performance [12]. The ripple-effect applies to low-frequency, high-impact disruption events, that might have long-lasting consequences for either the whole supply chain or dependants of the supply chain [13]. The disruption events can be caused by natural disasters, political conflicts, terrorism, and other force majeure events. Figure 3 contains a visualisation of the risk associated with the bullwhip-effect and the ripple-effect respectively, in the form of a risk matrix.

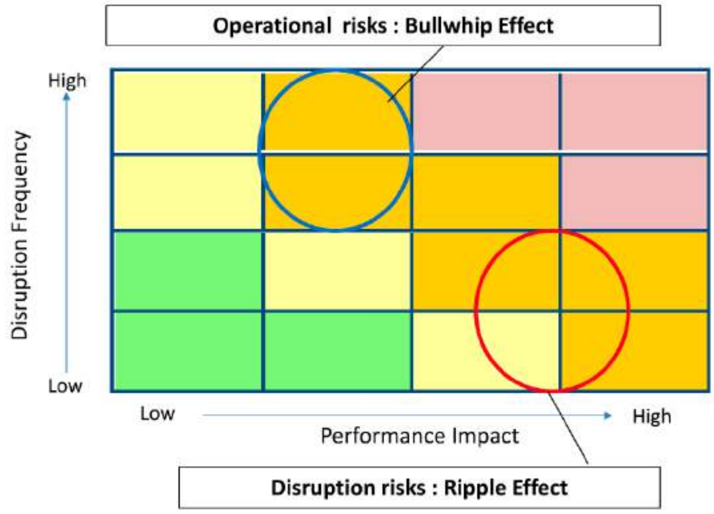


Figure 3: The bullwhip-effect and the ripple-effect on a disruption frequency versus performance impact matrix from Ivanov [13]

2.1.4 Visibility and Analytics Network (VAN)

To improve the availability of medicine in the public sector in Sub-Saharan African countries, the *Visibility and Analytics Network* (VAN) framework was developed in 2015 with funding from the Bill & Melinda Gates Foundation. Theoretically, the VAN is applied to a supply chain and, if implemented correctly, continuously improves the performance of the supply chain by capturing, aggregating, and analysing data and acting on gained insights.

For the VAN to work effectively, it is required that there be end-to-end visibility across the supply chain, meaning that reliable data are available at every node across the supply chain. The various data sets need to be combined and analysed, such that meaningful insights can be gained [14].

Currently the South African medical supply chain functions on an "uninformed pull" system, in which the medical staff are required to place orders based on simple heuristic calculations. This puts considerable unnecessary stress on the medical staff who are often not trained in stock management and can lead to stock-outs and wastage. To disburden the medical staff and allow them to focus on their duties as health care practitioners, the South African medical supply chain needs to transition to an "informed push" system and ultimately to an "informed pull" system [15].

2.1.5 Supply Chain Risk Management

The concept of risk is defined by Schlegel and Trent as: *"The probability of realising an unintended or unwanted consequence that leads to an undesirable outcome such as loss, injury, harm, or missed opportunity."* [16]

Risk is a subject of perspective and context. Risk is often viewed as something to be avoided completely, modelled as the concept of risk aversion in the decision-making process. Risk may, however, also be viewed as an opportunity that needs to be seized, while accepting the possible consequences if it means that there is a chance of achieving something greater.
In the context of the pharmaceutical supply chain, there are, of course, consequences that should not be gambled with, such as the possibility of losing lives, but if the consequence is a non-permanent one, decision makers might be inclined to take some measure of risk.

In his book on strategic risk taking, Damodaran [17] outlines three important distinctions in various definitions of risk: risk versus probability, risk versus threat, and all outcomes versus negative outcomes. It is emphasized that the more comprehensive definitions of risk address both the probability of an event occurring as well as its consequences.

Furthermore, an emphasis is made on low probability, high negative impact events. These events are commonly referred to as threats if they cannot be associated with a probability, due to their unforeseeable nature. Of course, some geographical areas are more prone to certain threats and tell-tale signs can often be discerned prior to the event, but these instances call for fast and efficient emergency response and often cannot be avoided.

Lastly, some definitions consider only the negative impacts of risk and aim to minimise the probability and contain the outcome, while other definitions consider all variability as risk and aim to exploit the positive outcomes.

On this subject Damodaran points out the relevance of the Chinese symbol for risk, which is a combination of the symbols of danger and opportunity. This represents the definition of risk and accurately captures the two opposing sides of risk [17].

An important concept to highlight with regards to risk, is that risk can neither be controlled nor avoided, but it can be managed to some extent in search of the best possible outcome.

Schlegel and Trent define *supply-chain risk management* (SCRM) as the minimisation of risk to a system, by ensuring that all components are obtained from trusted, identifiable, and top-grade sources [16].

Another definition by Wieland and Wallenburg [18], describes SCRM as the management of daily and exceptional risks along the supply chain through the implementation of strategies that ensure the continuous assessment of risk with the objective of vulnerability reduction and ensured continuity.

The most popular method to assess risk is the one of a risk matrix scoring method [19]. Here the likelihood of an event occurring is weighted against the impact of the event. The most basic form of this method adopts a high, medium and low scale for both variables, as can be seen in Figure 4.

IMPACT				
	High	Medium	High	High
	Medium	Low	Medium	High
	Low	Low	Low	Medium
		Low	Medium	High
			LIKELIHOOD	

Figure 4: Basic risk matrix

While this concept is valuable, it does not include the perception of risk, the willingness with respect to exposure to risk, and the uncertainties that come with making risky decisions [19]. When signing an agreement with an investment manager, clients often have to fill out a risk profile questionnaire. This includes questions that assess the amount of risk clients would be willing to subject themselves to for an increased chance at a higher return on investment. Their resilience and level of tolerating losses in particular are central to these types of questions. Oxford Risk is a company that specialises in improving financial decisions through the application of concepts from behavioural science. The company has developed a risk tolerance questionnaire that informs investors of their level of risk tolerance and advises them on the type of investments they should make, based on their behavioural risk profile [20]. This concept aligns well with decision making in a supply chain context, except that a risk profile needs to be developed for the organisation as a whole, its various sub-components and the specific decision makers.

A relevant example of risk tolerance in a medical supply chain context would be the risk associated with vaccine inventory. The extremes of this example would be a total stock-out on the one end of the spectrum and a complete over-supply on the other. A complete stock-out would mean that a newly born infant would be deprived of their first immunisation against polio and tuberculosis which might put them at great risk of potential infection. Furthermore, if a mass immunisation programme runs out of stock before the whole population has received its immunisation booster, it might run the risk of the population not being protected against preventable diseases [21].

On the other hand, over-supply could lead to wastage in terms of vaccines passing their expiration dates or spoilage may occur due to incorrect storage, resulting in possible financial losses [22]. Decision makers have to decide on which side of the spectrum they should operate. In the case of vaccinations, the risk of overstocking may lead to a potential financial loss, which is deemed a less-serious outcome than the alternative. The adoption of a risk aversion strategy should, however, not invoke thoughtless caution, which is where supply chain visibility is a very important concept.

Tolerance with respect to risk is a critical part of assessing risk, but another concept that many risk assessment techniques omit, is the use of time-based risk dimensions. The impact of an event might not be evident immediately after it occurs; it might take a while for the negative or positive effects of the event to unfold. On the other hand, the impact might be felt immediately, but might take a while to subside [23].

An example of a 3-dimensional risk matrix can be seen in Figure 5. This concept includes the conventional probability versus severity matrix, but expands into the third dimension of expected time until impact [24].

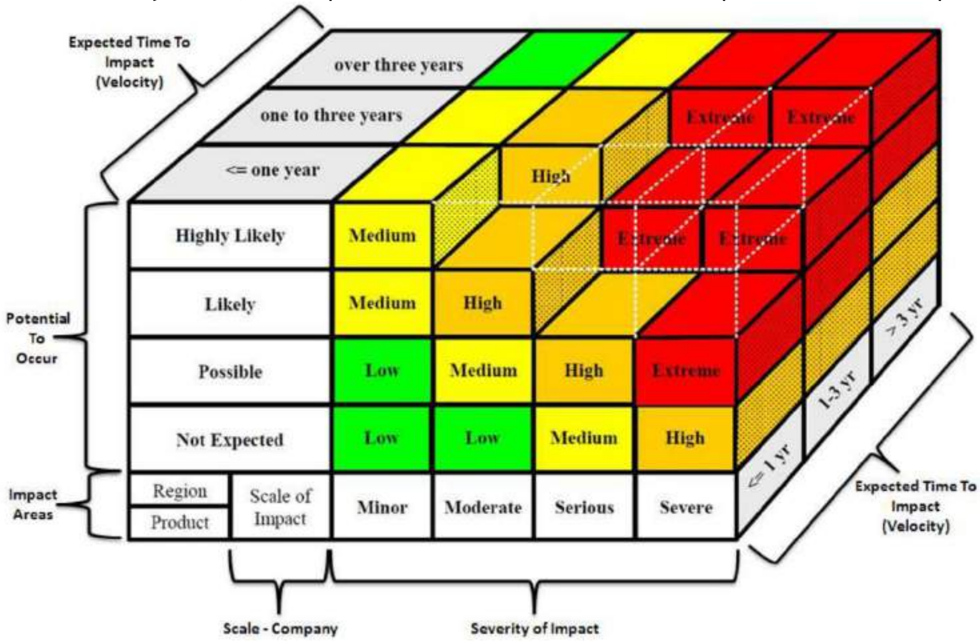


Figure 5: An example of a 3D risk matrix [24]

The crisis at the Mthatha depot, mentioned in Section 1.1, is a very relevant example of the time dimension of risk. It was not predicted that an unprotected strike would lead to 300 medical facilities running out of stock and it was not predicted how long it would take to rectify the situation.

To build on the concept of perception of risk and 3-dimensional risk, one may consider the concept of risk vectors. Hubbard [25] states that one cannot assume a person to be risk-neutral, and thus should not present them with a single value of risk. Instead he suggests leaving the quantity of risk in its separate components in a tabular format and using the entire table as the risk vector quantification. In this way all the outcomes and their associated probabilities for one event can be represented in a tabular format, such that decision makers can see the full picture and decide for themselves whether they are comfortable with the amount of risk to which they are subjecting themselves. A visual representation of a risk vector quantification in a 3-dimensional risk space can be seen in Figure 6.

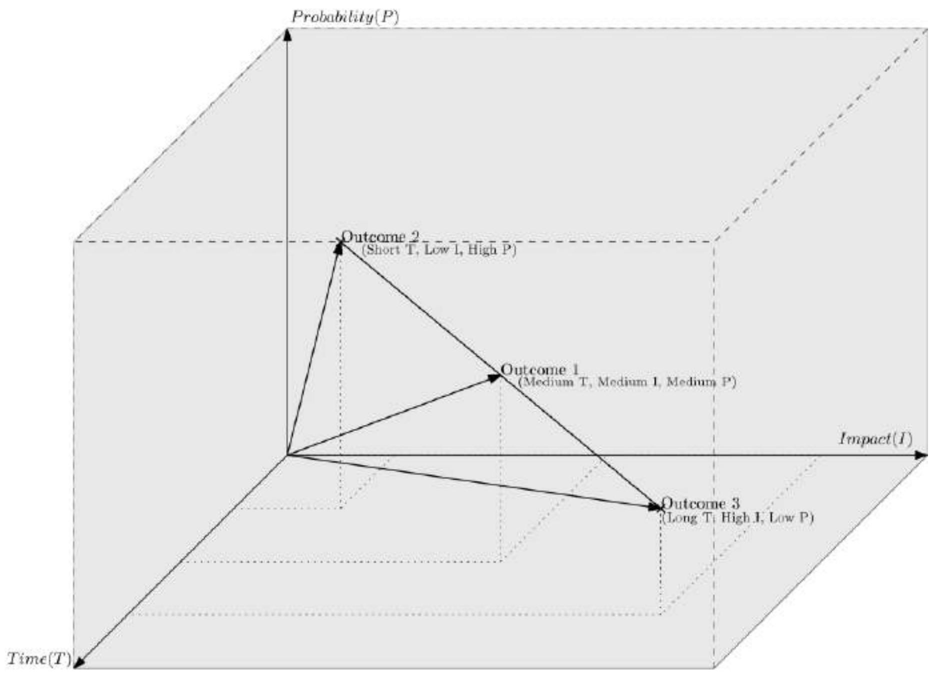


Figure 6: Visual representation of a risk vector quantification

The idea is to provide a decision maker with a direct comparison of all the possible outcomes of an event. This can be achieved in the form of a table, with measured numerical values for probability and impact. The table may then be represented in a 3-dimensional or 2-dimensional space.

2.2 Decision Support Systems

Power *et al.* [26] define decision support systems (DSS) as follows:
"DSS is a general term for any computer application that enhances a person or group's ability to make decisions. In addition, DSSs refer to an academic field of research that involves designing and studying DSSs in their context of use. In general, DSSs are a class of computerized information system that supports decision-making activities."

This definition is a very broad one, but it encapsulates the general purpose of any DSS. There is an abundance of different decision support systems of various types and purposes, and each one is designed to support decisions to a specific kind of problem. DDSs can be stand-alone applications or can form a functional part of an information system. In general, a DSS is used as an inclusive term for various types of information systems [27].

Decision making is a complex concept in the sense that there is no single right or wrong answer to most practical problems. Although there might be an answer that is more applicable than another, there is no guarantee that it is right [28]. It is mostly a matter of perspective or priority – a certain course of action may be a good answer from an economical point of view, but not from an ethical one or *vice versa*. In such instances a DSS only provides indicative outputs instead of definitive ones.

Currently the field of DSS research is dominated by buzz phrases such as Big Data Analytics, Business Intelligence, Internet of Things, Enterprise Information Systems, Cloud Computing, and others [29–31]. What all these concepts have in common is the exceptionally fast growth of internet connectivity, especially through mobile networks, and the abundance of available data. New methods of data creation are being readily implemented. Machine learning techniques are being implemented to uncover previously undiscovered avenues of analysis and insight generation. The age of digital information capture and analysis has made it possible to provide visibility of human behaviour in a real-time manner, where experts previously had to provide broad forecasts and make general assumptions about population groups [32, 33].

Modern day DSSs in conjunction with machine learning techniques have made it possible to provide decision makers with information that is custom-tailored to their preferences and the problems they are trying to solve. Furthermore, the extensively competitive research and development with respect to online application hosting has made it possible to develop web-based DSSs. The design of a Web-DSS is based on a three-tier web architecture: A webpage consisting of a *graphical user interface* (GUI) and a web-server as a front-end, an application server as middleware, and a relational database as the back-end [34].

The concept of web-based DSSs facilitates collaboration by many stakeholders and decision makers in different geographic locations, by providing them with universally relevant information but also with customised information based on their particular roles and what kind of information they would like to see.

3. SYSTEM DESIGN

This section outlines the process followed in the design of the DSS framework. First the user requirements are collected, followed by the identification of relevant data sources. The specifications for a DSS are deduced from the user requirements and a framework is drafted as a starting point for the detailed design and implementation of the system. Lastly, the different dimensions of risk are outlined in the context of the pharmaceutical supply chain.

3.1 System User Requirements

Before the design of a system is attempted, the requirements set out by the end user should be identified. For this project, the requirements were sourced from a collection of documents related to the management of South Africa’s pharmaceutical supply chain, as shown in Table 1.

As this project aligns with the concept of DDSCM, it is fitting to design the system with the key success factors of DDSCM in mind. Bvuchete [35] summarised the key concepts that ensure the success of DDSCM in practise and categorised these concepts by visibility, technology, collaboration, change management, distribution management, and performance management. The concept categories best suited for supporting the identified user requirements have been included in the user requirement table.

Table 1: User Requirements

UR_ID	*User Requirement*	*DDSCM Concept Category* [35]	*Source*
1	Identify Distribution Challenges and Bottlenecks	Visibility	[36]
2	View the current state of stock availability	Visibility	[36]
3	Investigate accusations of underperformance	Performance Management	[37]
4	Simulate outcome of different scenarios	Technology; Change Management	[38]
5	Plan for different seasons and demand spikes	Distribution Management	[36]
6	Share information	Visibility	[38]
7	View concise reports	Visibility; Performance Management	[36]
8	A decision recommendation capability within a what-if scenario context.	Technology; Change Management; Collaboration	[38]
9	Recommended maximum and minimum stock levels	Distribution Management	[36]
10	Recommended stock reporting schedules	Performance Management	[36]
11	Pre-emptive stock-out predictions	Performance Management; Distribution Planning; Visibility	[36]
12	User-friendly graphical user interface	Technology	[36], [38]
13	Usable by people with narrow operations research knowledge	Collaboration; Change Management; Distribution Management; Performance Management	General
14	Various visualisation capabilities	Visibility; Technology	[36], [38]
15	Current system compatibility	Technology	General
16	Nurture collaboration	Collaboration	[39]
17	Can be upgraded and easily maintained	Technology	General
18	Multiple different user profiles	Collaboration	[36]
19	Restricted access through internet	Technology	[36]
20	National stock level performance	Performance Management	[36]
21	Identify high and low priority drugs	Visibility;	[38]

		Distribution management; Performance Management	

3.2 South African Data Sources

The identification of possible quantitative data sources is a crucial part of designing a DSS. Apart from general, public statistics on international websites such as those of WHO and UNAids, the most relevant data sources have been summarised in Table 2.

Table 2: Relevant South African data sources

DS_ID	*Name*	*Status*	*Data Type*	*Description*
1	**Master Procurement Catalogue**	**Online, Public**	**Excel and Report**	A list of all essential medicines and medical supplies for South Africa. Includes names, lead times, manufacturers, barcodes, and therapeutic class.
2	**StopStockouts**	**Online, Public**	**Report and online Hotspot Map**	An organisation that allows citizens to report a drug stock-out. All reports are logged in a database and compiled in an annual report, and are shown on their hotspot map on stockouts.org
3	**District Barometer**	**Online, Public**	**Excel and Report**	An annual report published by the Health Systems Trust, outlining the annual health statistics of each health district. The data presented in the report are also available in an electronic spreadsheet format.
4	**DHIS2**	**Online, Restricted Access**	**Database**	The District Health Information System is an open source, online, cloud-based health data repository. It features health demographics, district level headcounts, and other health-related statistics, and is run by the NDoH.
5	**SVS**	**Online, Restricted Access**	**Database**	Stock Visibility Solution is a database of logged clinic inventory levels.
6	**SVS User Stories**	**Online, Public**	**Report**	Mezzanine Ware interviewed employees of the NDHoH and determined the features they'd like to see on a user dashboard. This document summarises the user hierarchy as well as a portion of the user profiles.
7	**etr.net**	**Online, Registered Access**	**Database**	An online register for TB patients.

Of these data sources the stock visibility solution, the master procurement plan, and the district health barometer seem to have promising potential to be employed in the context of for a quantitative analysis of the South African pharmaceutical supply chain. The stock visibility solution is a structured database, which is essential for quantitative analysis. The master procurement plan gives an indication of drug prices and theoretical lead times. The district health barometer is a collection of the most relevant information from the DHIS2 database and gives an indication of annual health statistics per health district.

3.3 Technical Specifications

From the identified user requirements, as well as their supporting DDSCM concepts, preliminary technical system specifications have been deduced, as shown in Table 3.

Table 3: System specifications

SS_ID	*System Specification*	*Reference User-requirement No.*	*Tools and Techniques*
1	**Global, relational database**	**2, 5,6,7**	SQL database
2	**Broad user proficiency compatibility**	**12,13**	Basic and advanced settings
3	**Web-based system**	**19**	Web-API development

4	**Modular system**	**17**	Component Based Software Engineering
5	**Multi-module model base**	**8, 14, 15, 17, 18**	Application server
6	**Current and projected inventory control risk profiles**	**5, 11, 20, 21**	Based on Seasonal district health profiles, Stock event severity profiles with probabilities, Strategy simulation outputs, Calculated inventory control variables
7	**Supply disruption reporting module**	**5, 11**	Social media scanning and app
8	**Seasonal district health profiles**	**5**	Time dependent clustering
9	**Strategy simulation module**	**4**	Agent-based model
10	**Inventory control variable calculation**	**9, 10**	Inventory management
11	**Compatibility with current NDoH information systems**	**15, 16**	Open source DHIS2 development kit or Tableau
12	**Stock-event severity profiles**	**5**	Bayes-Network, decision trees
13	**Geographical visualisation**	**1, 3, 12, 14, 20**	Google Maps API, QGIS, Tableau
14	**Statistical visualisation**	**1, 3, 14, 20**	Boxplots, cluster plot, distribution charts,
15	**Data integrity confirmation module**	**15**	Correct file type, structure and statistical feasibility

These specifications are a basic collection of tools and concepts that will be used in the design and/or will form part of the development of this system. The project is still ongoing and more specific specifications of the sub-systems will be developed as part of the future work.

3.4 System-level Functional Decomposition

The basic functionality of the DSS was based on collected requirements and specifications and is envisioned to conform to the structure in Figure 7.

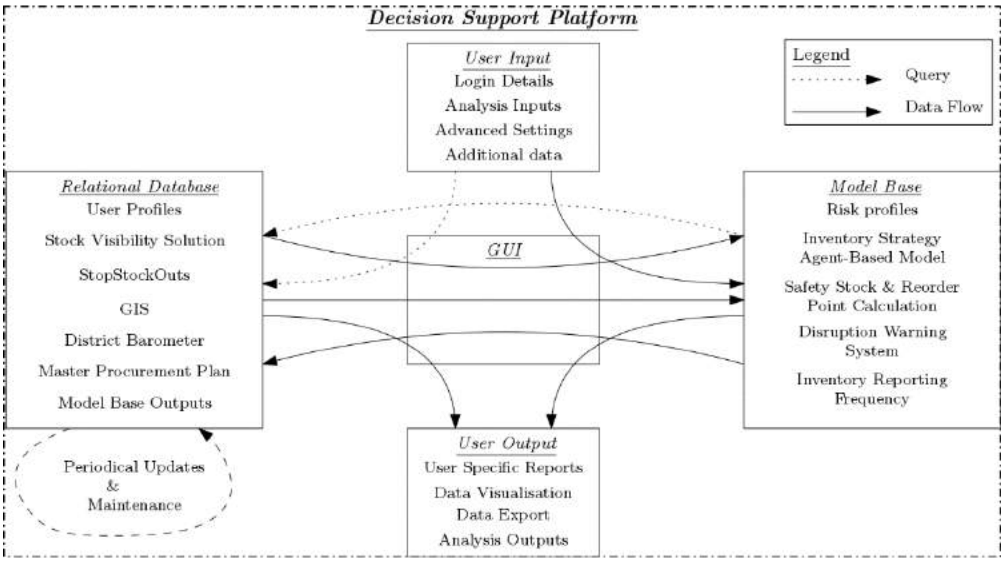


Figure 7: Decision Support Platform Framework

The idea is that the user provides his/her login details through the GUI. The user details are queried with the relational database and the user's profile is retrieved by the GUI. The default user settings are sent to the model base, and these settings are used as an input for the relevant modules. The model base queries the database and receives the data necessary for analysis. The data are sent through the GUI to the model base, where an analysis is carried out. The necessary outputs are then sent to the GUI, where the aggregated information is displayed to the user. This forms the baseline of operation, which occurs automatically upon login. From this point the user can choose to perform more in-depth analyses, and to view more detailed information.

3.5 Dimensions of risk in pharmaceutical supply network

This section aims to expand on the concept of dimensions of risk as defined in Section 2.1.5, and to put it into context for stock-management risk.

The two main risks in a stock management environment are a stock-out, and an over-supply of stock. If one were to attempt to quantify the risk of a stock-out, it would be of no help to provide a decision maker with a single risk value, as this does not provide enough information to manage the risk effectively. It would be more useful to provide all the possible outcomes that fall under the category of a stock-out. A very basic example of this concept can be seen in Table 4. In this example the length of a stock-out is weighed against its impact, and the probability of it occurring.

Table 4: Basic example of the risk of a stock-out as a vector quantity

Event	*Probability*	*Impact*	*Time*
Trivial stock-out	High	Low	Short
Minor stock-out	Medium	Medium	Medium
Enduring stock-out	Low	High	Long

Another example of a more specific stock-out event can be seen in Table 5. In this example the probability of a pharmaceutical needed by a patient is weighed against the impact of its unavailability to the patient, and the timeframe until the impact is felt by the patient.

Table 5: Basic example of the risk of unavailable pharmaceuticals to patients, as a vector quantity

Event	*Probability*	*Impact*	*Time*
Analgesic	High	Low	Long - Not at all
HIV/AIDS medicine	Medium	Medium	Short - Medium
Snake bite anti-venom	Low	High	Immediate – Short

These two cases serve merely as examples of how risk quantification can be attempted in a pharmaceutical supply context. There are many avenues that can be explored, and these examples do not yet include the dimension of hierarchy (*i.e.* if a hospital pharmacy runs out of stock or is shut down, all the dependent clinics will be affected as well).

4. FUTURE WORK AND POSSIBLE IMPACT

The next phase of this project is to develop a concept demonstrator of the system proposed above. It is planned to use Tableau as a front-end user-interface. Tableau is a data visualisation software suite that has already been implemented in some sectors of the department of health.

A back-end application will be built, which can successfully aggregate structured data contained in a relational database, to quantify the risk associated with the state of current drug inventory levels in a district. The application will be developed within the programming language R. The concept demonstrator will aim to form a base structure for other decision support tools.

The decision support tool developed for this specific project will aim to rank drugs according to their priority levels, which will be based on their specific risk profiles. Users can specify which dimension of risk they'd like to see, be it the risk of over-stocking or of a complete stock-out. Furthermore, the user can specify whether they know of any supply disruption event that has occurred, so that it will be included in the analysis. It can be specified where such an event has occurred, and it will be determined which facilities will be affected by the event. This will allow for effective emergency and pre-emptive planning and could possibly decrease the effect of enduring stock-outs and supply disruptions.

The plan is to use machine learning techniques, such as Naïve-Bayes Classification and Kernel-Support Vector Machines, to determine the probability of a stock-out occurring within a sub-district, based on historical data. If this is done successfully, the use of more sophisticated machine learning techniques, such as artificial neural nets can be considered.

The quantification of risk of a supply disruption, such as a stock-out, could have a significant impact on the future planning of decision makers along the pharmaceutical supply chain in South Africa.

5. CONCLUSION

In this paper, the problem of an inefficient and dysfunctional pharmaceutical supply chain was addressed by proposing a preliminary conceptual top-level design of a decision support platform aimed at helping decision makers evaluate the current state of the supply chain and what sectors and products require more attention from a management perspective.

We presented a summary of relevant concepts in the literature that are considered in the system design. These include demand-driven supply chain management, supply chain visibility, the ripple-effect, supply chain risk management, and DSSs.

Following this, a list of user requirements was compiled from different sources and summarised in Table 1. User requirements were also confirmed to be in line with the concept of DDSCM, by indicating their respective categories of key concepts of success for DDSCM, as proposed by Bvuchete [9].

Some of the more prominent data sources of the South African pharmaceutical supply chain that may be used in a DSS were showcased in Table 2. The user requirements from Table 1 were used to compile a collection of basic technical specifications that the proposed system should meet to satisfy the user requirements. These were summarised in Table 3.

The user requirements and system specification were used to draft a simple top-level decision support platform framework that can be used as a starting point in the design of the system and all its sub-systems.

This was followed by bringing the concept of risk dimensions and risk vector quantities into the context of pharmaceutical stock management.

Lastly, planned future work was outlined, as well as the impact that a supply chain risk quantification decision support tool could have on the pharmaceutical supply chain.

The Stock Visibility Solution developed by Mezzanine Ware is a valuable and crucial first step towards a functional and efficient medicine supply chain.

REFERENCES

[1] **World Health Organization** 2015. South Africa: WHO statistical profile. [Online]. [Accessed 20 Jun 2018]. Available from http://www.who.int/gho/countries/zaf.pdf?ua=1.

[2] **Mayosi, B.M., Flisher, A.J., Lalloo, U.G., Sitas, F., Tollman, S.M., & Bradshaw, D.** 2009. The burden of non-communicable diseases in South Africa, *The Lancet*, 374(9693), pp 934–947.

[3] **Byarugaba, D.K.** 2004. Antimicrobial resistance in developing countries and responsible risk factors, *International Journal of Antimicrobial Agents*, 24(2), pp 105–110.

[4] **Govender, I.** 2016. 2015 Stock Outs National Survey: Third Annual Report – South Africa. [Online]. [Accessed 20 Jun 2018]. Available from http://stockouts.org/Downloads/2015_stock_outs_national_survey.pdf.

[5] **Medecins Sans Frontieres, Rural Health Advocacy Project, Section 27, & Treatment Action Campaign** 2013. Emergency Intervention at Mthatha Depot: The hidden cost of inaction. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/mthathaemergency.

[6] **National Department of Health** 2015. National Department of Health: Annual Report 2014/15. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/ndohannualreport2015.

[7] **National Department of Health** 2016. National Department of Health: Annual report 2015/16. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/ndohannualreport2016.

[8] **National Department of Health** 2017. National Department of Health: Annual Report 2016/2017. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/ndohannualreport2017.

[9] **Bvuchete, M., Grobbelaar, S.S., & Van Eeden, J.** 2018. A Maturity Model for Demand Driven Supply Chain Management. (Unpublished) MEng Thesis. Stellenbosch University.

[10] **Goh, M., De Souza, R., Zhang, A.N., He, W., & Tan, P.S.** 2009. Supply chain visibility: A decision making perspective, *2009 4th IEEE Conference on Industrial Electronics and Applications, ICIEA 2009*, pp 2546–2551.

[11] **Syntetos, A.A., Babai, Z., Boylan, J.E., Kolassa, S., & Nikolopoulos, K.** 2016. Supply chain forecasting: Theory, practice, their gap and the future, *European Journal of Operational Research*, 252(1), pp 1–26.

[12] **Ivanov, D., Sokolov, B., & Dolgui, A.** 2014. The Ripple effect in supply chains: Trade-off "efficiency-flexibility- resilience" in disruption management, *International Journal of Production Research*, 52(7), pp 2154–2172.

[13] **Ivanov, D.** 2017. Simulation-based ripple effect modelling in the supply chain, *International Journal of Production Research*, 55(7), pp 2083–2101.

[14] **VillageReach** 2015. Enhanced visibility, analytics and improvement for the Mozambique immunization supply chain.

[15] **Llewellyn, E.** 2016. National Strategy for Improved Availability and Access: Visibility and Analytics Network.

[16] **Schlegel, G.L., & Trent, R.J.** 2015. *Supply Chain Risk Management: An Emerging Discipline*, 1st Edition.

[17] **Damodaran, A.** 2008. *Chapter 1 - What is Risk?*. Strategic Risk Taking: A Framework for Risk Management, 1st Edition. Wharton School Publishing, pp 1–13.

[18] **Wieland, A., & Marcus Wallenburg, C.** 2012. Dealing with supply chain risks: Linking risk management practises and strategies to performance, *International Journal of Physical Distribution & Logistics Management*, 42(10), pp 887–905.

[19] **Hubbard, D.W.** 2009. *Chapter 7 - Worse Than Useless: The Most Popular Risk Assessment Method and Why It Doesn't Work*. The Failure of Risk Management: Why it's Broken and How to Fix it. John Wiley & Sons, Ltd, NJ.

[20] **Oxford Risk Ltd.** Risk Profiling. [Online]. [Accessed 15 Jun 2018]. Available from https://www.oxfordrisk.com/our-methods/risk-profiling.

[21] **Lydon, P., Schreiber, B., Gasca, A., Dumolard, L., Urfer, D., & Senouci, K.** 2017. Vaccine stockouts around the world: Are essential vaccines always available when needed?, *Vaccine*, 35(17), pp 2121–2126.

[22] **Weir, E.** 2004. Preventing cold chain failure: vaccine storage and handling, *Canadian Medical Association Journal*, 171(9), pp 1050–1050.

[23] **Bernstein, P.L.** 2007. The Time Dimension of Risk, *Journal of Portfolio Management*, 33(4), pp 1.

[24] **McCormick, M.** 2011. 3D Risk Model Concept. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/3Driskmatrix.

[25] **Hubbard, D.W.** 2009. *Chapter 5 - An Ivory Tower of Babel: Fixing the Confusion about Risk*. The Failure of Risk Management: Why it's Broken and How to Fix it. John Wiley & Sons, Ltd, NJ.

[26] **Power, D.J., Sharda, R., & Burstein, F.** 2015. *Decision Support Systems*. Wiley Encyclopedia of Management. John Wiley & Sons, Ltd, Chichester, pp 1–4.

[27] **Power, D.J.** 1997. What is a DSS? [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/whatisadss.

[28] **Belton, V., & Stewart, T.** 2002. *Multiple Criteria Decision Analysis: An Integrated Approach*. Springer Science & Business Media.

[29] **Zhong, R.Y., Newman, S.T., Huang, G.Q., & Lan, S.** 2016. Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives, *Computers and Industrial Engineering*, 101, pp 572–591.

[30] **Bohlouli, M., Merges, F., & Fathi, M.** 2014. Knowledge integration of distributed enterprises using cloud based big data analyticsIEEE International Conference on Electro Information Technology. IEEE, pp 612–617.

[31] **Duan, L., Street, W.N., & Xu, E.** 2011. Healthcare information systems: Data mining methods in the creation of a clinical recommender system, *Enterprise Information Systems*, 5(2), pp 169–181.

[32] **Gunter, B.** 2016. *The Psychology of Consumer Profiling in a Digital Age*, 1st Edition.

[33] **Sumner, C., Byers, A., Boochever, R., & Park, G.J.** 2012. Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets 11th International Conference on Machine Learning and Applications. IEEE, pp 386–393.

[34] **Power, D.J.** 2000. Web-Based and Model-Driven Decision Support Systems: Concepts and Issues, *Americas Conference on Information Systems*, pp 352–355.

[35] **Bvuchete, M., Grobbelaar, S.S., & Van Eeden, J.** 2019. Best practises for demand driven supply chain management in public healthcare sector: A systematic literature review. South African Journal of Industrial Engineering (Accepted to appear).

[36] **Rademeyer, A., & de Vos, J.** 2014. Statement of Work for Stock Visibility Solution.

[37] **Motsoaledi, A.** 2018. Press Conference on Status of Healthcare in the Country. [Online]. [Accessed 20 Jun 2018]. Available from https://tinyurl.com/speech2018motsoaledi.

[38] **Grobbelaar, S.S., & Van Eeden, J.** 2018. Improving supply network design and management through a demand driven supply chain technology demonstrator. (Unpublished) Project Proposal.

[39] **Candea, C., Candea, G., & Filip, F.G.** 2012. iDecisionSupport — A web-based framework for decision support systemsProceedings of the 14th IFAC Symposium on Infiormation Control Problems in Manufacturing. IFAC, Bucharest, pp 1117–1122.

DEVELOPMENT OF A DEMONSTRATOR OF BIG DATA ANALYTICS

Rhett Desmond Butler[1]* & Prof. James Bekker[2]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
17647657@sun.ac.za

[2]Department of Industrial Engineering
Stellenbosch University, South Africa
jb2@sun.ac.za

ABSTRACT

Big Data Analytics is now not only being applied in the fields of science and business, but in healthcare and economic development, by organisations such as the United Nations. The research presented in this article provides a demonstration of developing a Big Data Analytics Demonstrator by integrating selected hardware and software. The components of such an analytics tool are presented, as well as the analysis of results of test data sets. Experience gained when setting up a proprietary data analytics suite is shared, and practical recommendations are made.

The goal of this demonstrator is to illustrate that a system could be built to provide meaningful insights into a given dataset, by making use of free-to-use software, commodity hardware and leveraging machine learning to mine the data for these insights.

---

[1] The author was enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION

Big Data Analytics is gaining widespread adoption due to the inherent properties that define Big Data, which provide the means to analyse large volumes of data in rapid fashion [1]. These properties are the four V's that define Big Data: *Volume, Variety, Veracity* and *Velocity*. These definitions were discussed in detail in [2] on which this article expands. The article by [2] investigated what Big Data Analytics (BDA) is, demonstrating the workings of the tools commonly associated with Big Data Analytics, namely, *MapReduce* and *Hadoop,* and provided an architecture which was used to develop a Big Data Analytics Demonstrator (BDAD).

This article expands on the research conducted by [2], by taking the idea and principles, and using it to aid the building and development of the BDAD outlined in this paper, using the open-source software Hadoop and Apache Spark. This paper therefore not only outlines what is Big Data, but also provides a demonstration of the technologies along with practical considerations when working with such a system. The Apache Spark software, which was developed in 2009 by [3] at UC Berkley, was built due to the need to analyse large volumes of data but with greater speed than what *MapReduce* offered. The method by which Apache Spark analyses data is the focus of the literature review in the following section, also providing reasons as to why this Analytics software was more applicable to the BDAD, and the ultimate goal for this project to better understand BDA.

The goal of the demonstrator developed in this project, was to illustrate how such a system could be built making use of available commodity hardware and open-source software to aid in decision-making. This is because documentation is scattered or lacking, which provides a concise picture of Big Data and all the components, associated with it. The demonstrator would do this by analysing given structured-datasets in a timely and accurate manner, using Big Data technologies and machine learning tools and techniques available. The results are then compared to those of standard analytics systems hardware and software which do not use Big Data technologies. Such technologies offer the means to analyse engineering-related data, from production quantities, worker productivity and the effect of different factory layouts on output. For this reason, an investigation and research was conducted into Big Data Analytics. BDA is a tool which aids the data analysis and decision-making of industries involved in the fourth industrial revolution.

The article is organised to first provide background around the Spark Analytics software used in this project in order to conduct the Big Data Analytics. Thereafter in section 3, the methods followed in order to develop the BDAD are outlined, discussing the practical challenges experienced during this phase of the project in order to develop the Big Data Storage and Analytics components required for such a system. Finally, in section 4, the demonstrator's capabilities are validated against a standard analytics system which does not make use of Big Data Analytics technologies. The validation is executed by comparing the time and accuracy taken to analyse ever larger datasets and making use of different machine learning (ML) tools and techniques as outlined in [2]. The validation also serves to demonstrate not only the Demonstrator's capabilities, but also the capabilities of BDA and why it should be implemented in more industries.

2. LITERATURE REVIEW

The literature review provides background to the workings of the Apache Spark Analytics software used in this project. The method by which Apache Spark analyses data is firstly discussed; then the different properties such as APIs (*Application Programming Interface)* and libraries of Apache Spark are outlined.

2.1 Apache Spark

To understand why Apache Spark (Spark) was developed, the *MapReduce* framework needs to be understood. As from [2], a *JobTracker* which is located on the master node, is used in *MapReduce* to manage the processing of a query (locate stored files, execute the jobs and store metadata of the processing). The *TaskTrackers*, which are located on the slave nodes, conduct the processing of a query that was submitted to the *JobTracker*. The query and the processing thereof is mapped to various partitions located on slave nodes which contain the source file. In order to manage this process, *MapReduce* makes use of *key-value* pairs (*k1,v1)*, where similar keys are joined, sorted and filtered in the *Map-phase*. The final step is to consolidate the results back from all the slave nodes to the master node via the shared *key-value* pairs, to where the *JobTracker* is located in the *Reduce-phase*, [2] and [4]. This process creates a lot of overhead, which increases the time per analysis [5] as each process is read and written to disk.

Because of this lengthy process, Spark was created. Spark makes use of in-memory processes in order to address this shortcoming of *MapReduce*. Spark uses a *Resilient* Distributed Dataset (RDD) for its processing. An RDD is a collection of data packets stored and distributed across a cluster in memory. This RDD, when consolidated, represents the data currently manipulated by an analyst. RDDs are immutable (cannot be altered, only a new copy can be created) which ensures fault-tolerance throughout a cluster by storing the metadata of the RDD copies. RDDs can be created in four ways [5], the first being loading a file from a shared file system such as HDFS (Hadoop Distributed File System) as discussed in [2], or loading an array generated previously within specifically the Scala programming language environment. Thirdly, by transforming an already existing RDD, or lastly, by persisting (caching) the RDD in memory or to disk (storing in a different file system).

The process to create and analyse data using an RDD within a cluster is given in Figure 1. Similar to the process used in *HDFS* and *MapReduce*, Spark has a *JobTracker* located on the Spark Driver. Located on Worker nodes or Slave Nodes are the *Executors*, these have *TaskTrackers* which execute processing of the RDD, sent to them from a *JobTracker*. As shown in the figure, there are many *TaskTrackers* conducting processing and sending the results back to the *JobTracker*, all of the processing occurring in memory [6] and [5]. For this reason, Spark is considered much faster than *MapReduce*; up to 100 times faster [3].

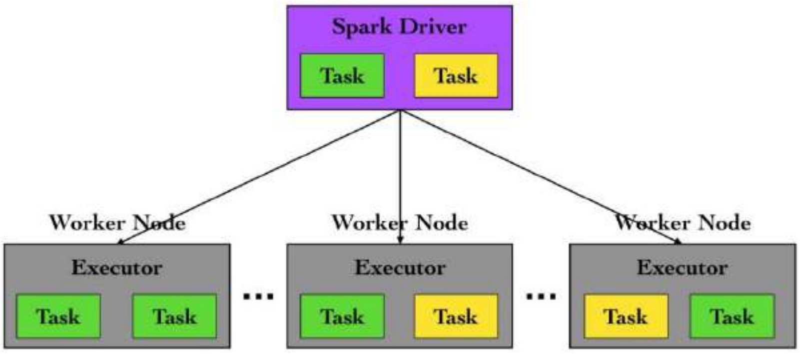


Figure 1: The process by which Spark submits jobs to a cluster of multiple nodes [3], making use of *JobTrackers* and *TaskTrackers* to execute and manage RDDs

Even though Spark is faster at conducting an analysis, it is not suited to all applications. As discussed in [7], *MapReduce* can be better suited when conducting linear large-scale processing that can be considered trivial, or when no immediate results are required, and therefore the rate of processing is of less importance.

Spark, at the time of writing, provides multiple APIs for different popular programming languages; these include R, Scala, Java and Python [3]. This allows pre-built programs to be easily adapted in the respective languages to make use of the Spark technology, increasing the user base. Spark also contains a pre-built library of ML tools (Supervised, Unsupervised ML *etc.*) and techniques (Linear Regression, k-means clustering *etc.*) for each of these languages. These additions negate the need for third-party ML applications and language packages that have to be added.

In the next section, the steps followed and practical challenges experienced to develop the BDAD cluster of computers are outlined. This includes connecting the hardware together over a network, and configuring the software, namely the Hadoop and Spark software used as the Big Data storage and analytics solutions.

3. BUILDING A BIG DATA ANALYTICS DEMONSTRATOR

There are multiple avenues available for an analyst to gain access to a Big Data analytics suite. These range from using pre-built solutions offered by companies such as Cloudera and Hortonworks [6] and [8]. These suites come with the relevant Big Data storage and analytics environments built-in and just have to be connected to an existing number of computing nodes. Another solution is to make use of cloud computing services offered by companies such as Amazon Web Services, which offer computing nodes on a wide variety of performance scales and prices (pay-as-you-use basis). These nodes are free to be configured however desired or have to be pre-built in a similar fashion. This makes it quick and easy to create a Big Data Analytics cluster in order to start with an analysis, without the need to configure the system and all its related components [9].

The goal of this project was to provide a demonstration of Big Data by making use of commodity hardware, and therefore a solution had to be configured which demonstrated how to make use of a combination of spare computing nodes and open-source software. For this reason, a pre-built system would not offer the means to learn how Big Data technologies work, and how they can be customised or expanded. The pre-built solutions are also better suited to companies and individuals looking to contract their analytics out and not develop in-house solutions. By building a demonstrator, the fundamental components could be demonstrated, making use of commodity hardware joined together to provide insights at a larger scale. By building this BDAD, a better understanding was formulated as to what Big Data Analytics entails. This included using technology to increase the rate at which an analysis of large datasets could be conducted, which is not possible using standard analytics software.

For this project, the BDAD made use of three computing nodes of which one was also the master node. The master node was responsible for process management and metadata storage and from which the queries were submitted to the cluster, the slave nodes being responsible for the computations of the query. The subsequent discussion is divided into hardware and software sections. In the hardware section, the computing nodes' hardware specifications are provided, as well as how the computing nodes were connected together on the network to form the BDAD cluster. Thereafter, in the software section, the operating system (OS) choice is discussed, outlining why the Big Data Analytics storage solution (HDFS) and analytics solution (Spark) were chosen, along with the process to configure these software applications together in a cluster. Finally, a list of practical challenges experienced during the development of the BDAD is provided, highlighting what to consider if such a system were to be replicated.

3.1 Hardware used in the Big Data Analytics Demonstrator

Figure 2 provides a visual representation of the configuration used by the BDAD. The term *NameNode* and *DataNode* will be used in this article interchangeably with master node and slave node.

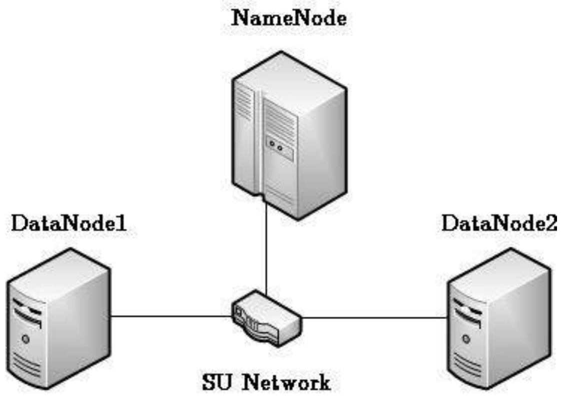


Figure 2: A representation of the three computing nodes used in the Big Data Analytics Demonstrator, connected together over the university network

The hardware used was chosen as it aligned with the scope of the project, to make use of commodity hardware. The specification for the three computing nodes is provided in Table 1. It is not suggested that use of this hardware configuration be made when developing a similar Big Data Analytics system in future, as due to technological advances over time, this hardware selection will be inefficient. This configuration was merely used at the time of this project as it aligned with the scope of this project. The information regarding the hardware specifications listed in Table 1 was from [10] and [11].

Table 1: The hardware specifications of the three computing nodes used in the Big Data Analytics Demonstrator

Hardware	Master Node (*NameNode*)	Slave Nodes (*DataNode 1 & 2*)
CPU	Intel 4th Generation Core i7-4770 with a clock speed of 3.4GHz	Intel Xeon 5500 series processors W3550 with a clock speed of 3.07GHz
RAM	Non-ECC dual channel 1600MHz DDR3 SDRAM, with 24 GB of memory	ECC 1333MHz DDR3 memory configured to 12 GB.
GPU	Integrated Intel HD Graphics 4600	Support for 2 PCI Express x16 Gen 2 graphics cards up to 150 W, NVIDIA Quadro 4000
Storage	One 500 GB disk drive (ST500DM0 02-1BD142 SCSI Disk Device) and one solid state drive of 256 GB (Samsung SSD SM841N 2.5 SCSI Disk Device)	One Disk Drive with 500 GB of storage (ST3500418AS ATA Device)
Networking	Integrated Intel I217LM Ethernet LAN 10/100/1000; supports optional PCIe 10/100/1000 network card	Integrated Broadcom 5761 Ethernet controller with Remote Wake UP and PXE support

The next step after acquiring the hardware was connecting the three nodes together over the Stellenbosch University network. The steps outlined were specific to the given OS used and may differ depending on the OS; however, the fundamentals remain the same. The next section outlines the steps followed as well as what to consider during configuration.

3.2 Connecting the computing nodes together over a network

The first step in the development of the cluster was to create unique identifiers for each node, namely *hostnames* that were used in place of the IP (Internet Protocol) address of a computer, to identify each on the network. Thereafter, creating a common user account, which has all the necessary permissions that is separate from the system accounts of the nodes, for consistency and good practice (should problems arise). Firstly, when configuring the cluster, a connection had to be established between the nodes, to allow for seamless data transfer, whilst being secure. For this project, the nodes' hostnames were:

- For the master node: *master*
- For the two slave nodes: *datanode1* and *datanode2*

Using these hostnames and IP addresses, the different nodes were located on the network. After establishing a linkage between the nodes, a secure connection, which would allow for data transfer without requiring passwords, was required. Using SSH (Secure Shell) technology, an encryption key was created and distributed between the nodes and the connection was tested. To test the connection, a *ssh accountname@hostname* (account name and hostnames created) command could be run to access another node securely and without a password, thereby creating the fundamental connections of the BDAD.

3.3 Software used for the Big Data Analytics Demonstrator

As mentioned in section 3.2, the OS used for the BDAD was Linux, specifically the CENTOS 7 distribution. The OS was chosen after experimenting with pre-built Big Data Analytics suite solutions by Cloudera and Hortonworks which both use Linux CENTOS OS, as well as after reading articles and blog posts by persons who have developed Big Data Analytics systems. Other OSs can also be used, as this article does not prescribe the exclusive use of Linux, but rather Linux was chosen as it was easily configurable with the Big Data technologies used for this project (Hadoop and Spark), and it is available as open-source. For the programming, the Python programming language was used, as it is easily interpreted and widely used by the scientific community, while being offered on Spark as an API, namely *PySpark*.

Following on from the literature review, the Spark analytics software was installed on the BDAD. The steps by which to configure the system were a combination of steps from [12], [13] and [14] providing the means to install the Spark software version 2.2.1, on the Big Data cluster. After this, the Spark environment was configured and the web dashboards to monitor spark jobs could be called by typing the following in the web browser:

- *master:8080* and,

In addition, when a query was executed:

- *master:4040*

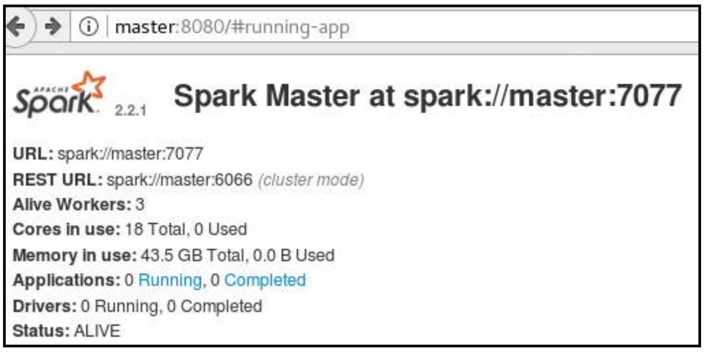


Figure 3: The Spark dashboard provided (master:8080), indicating the number of nodes (shown as workers) and cluster configuration along with the URL for calling the cluster

Using these dashboards, the system was inspected to see if the desired configuration was successful. For this project, a total of 43.5GB of RAM was configured (10.6GB per slave node and 22.3GB from the master node) per node and six of the eight CPU cores were configured for the Spark system per node (a total of 18 cores). This was validated by the output given in the master:8080 URL (*Uniform Resource Locator)* in Figure 3. To access the Spark cluster, the *spark://master7077* command is used at the start of a query.

The final software component used in this Big Data cluster was the HDFS storage solution from Hadoop. For this project, the Hadoop HDFS version 2.8.3 was used. The steps on how this distributed file storage system was configured were a combination of steps from [15], [16], [17] and [18]. The Hadoop program was then started, and then using the master:50070 command to view the Hadoop dashboard, the successful configuration was confirmed.

For this project, due to the limited number of storage and computational nodes, the master node was also used as a computing and storage node (not just to manage the processes typically associated with the master nodes). For this reason, the number of nodes indicated in Figure 4 (*Live Nodes*) which offer storage, was three and not two. The same applied for the analytics system, and why 18 cores were configured and not 12.

Configured Capacity:	796.99 GB
DFS Used:	1.42 GB (0.18%)
Non DFS Used:	24.62 GB
DFS Remaining:	770.95 GB (96.73%)
Block Pool Used:	1.42 GB (0.18%)
DataNodes usages% (Min/Median/Max/stdDev):	0.00% / 0.17% / 0.25% / 0.11%
Live Nodes	3 (Decommissioned: 0)

Figure 4: The summary provided by Hadoop, indicating the complete system storage configuration and the number of nodes used

Information provided by the Hadoop dashboards allow the analyst to monitor the files stored by the system, along with the number of data packets distributed across the system. As discussed previously in [2], the Hadoop system divides a file into data packets and stores and distributes these packets of information across the cluster to provide a fault-tolerant and scalable storage solution. The other dashboards to inspect the *DataNodes* were accessible by using the *hostname*:50075 (*hostname* of the node) command.

Other storage options that are open-source were also available for use, such as the MongoDB, Cassandra, Hive *etc*. However, it was decided to use the HDFS as it was a simple system requiring that data be stored in a distributed manner, and it was able to demonstrate the capabilities of distributed computing. The other storage solutions offer capabilities not required for this system, *e.g.* Cassandra offers high speed access suited to near-real-time analysis, or MongDB offers a NoSQL database, and they were therefore not chosen.

An advantage of using such Big Data technologies is that adding another computing node to the network requires that the current environment be stopped, and the new node's details and software files (Hadoop and Spark) then simply have to be added and installed. The system then requires a restart after which the new node is then added to the cluster. This simple process allows more nodes to be added in parallel, allowing for scalability in storage and performance as desired. This means the computing power can be increased should the query complexity increase and the need for more computing power be required.

3.4 Practical challenges faced during the Demonstrator Development

The following is a list of the various practical challenges experienced during the development of the BDAD. The list provides a guide as to what should be considered during the development in order to reduce set-up time and improve the ease of development.

- During the period when the clusters of computers were being connected to a network, it was important to ensure that the global network IP address was selected, and not the nodes' internal IP (typically a 192.168 or 10.0 number), as the nodes were not then visible to each other and could not share data.
- The firewall had to be altered in order to allow the connection on the specific port through, as at first a connection was established, but information could not flow bi-directionally, only from the master node to the slave node and not in the opposite direction.
- Ensuring the SSH key was correctly stored and visible to create the connection was a common issue which prevented the system from allowing data transfer to occur.
- A practical challenge experienced during the Spark installation was configuring the different *DataNodes* to the cluster, as Spark would not connect these to the cluster due to incorrect read and write permissions (using Linux commands *chown* and *chmod* to give these permissions) given to the files where Spark locates the *DataNodes* and *NameNodes*.
- Similar to Spark, for Hadoop, the folders and their locations where the Big Data programs were located had to be given the correct permissions, as previously mentioned by using the *chmod* and *chown* commands. Hadoop during set-up would not otherwise be able to create the desired storage directories for the cluster.
- It must be ensured that all the Java packages were installed and be up to date, while the correct location of Java program files must be directed to Hadoop; it makes use of Java-related technologies to conduct its processes.
- A considerable amount of time was spent during the development phase to ensure correct configuration files (.xml files which provided settings, paths *etc.*) for the Hadoop environment were selected, as the example files found online are system specific and each file needed to be altered to suit the BDAD of this project.
- Before the Hadoop software was started, it was also important to use correct configuration files, specifying the number of times a file is replicated across the cluster. This has to be set to only the number of *DataNodes* configured, along with ensuring an appropriate block-size (size the file is broken up into) is used (*e.g.* if working with very large files opting for 128 Mb blocks instead of 64 Mb). By not doing this, duplicate files were stored on a node, which caused file corruption and then destroyed the HDFS storage system, preventing data from being read from the database.

In section 0, validation of the BDAD is presented. This was done by analysing ever-larger samples of a dataset stored in Hadoop, by applying different ML techniques under each tool. For each ML tool, the appropriate dataset was used during the analysis. The results of the Big Data system were then compared to that of a *standard* system, to compare the time taken to conduct an analysis and its accuracy. The validation serves to demonstrate the BDAD capabilities along with the benefits of using BDA.

The *standard* system, used to compare the Big Data Analytics Demonstrator to, was created by taking the master node, but not coupling it to a cluster. This *standard* system made use of an open-source ML library S*cikit-learn,* available in Python, in order to conduct the same analysis as that on Spark, but not in a parallel-distributed manner.

4. VALIDATING THE DEMONSTRATOR

The validation of the BDAD is divided into three sections; each taking the commonly used ML tools (Classification, Regression and Clustering*)* for each system under Supervised and Unsupervised ML. For each tool, a dataset containing data types that are appropriate for the tool was used, which are publicly available and found online. The datasets used in each tool were then divided into different sizes in an ascending order (1 000 rows, 500 000 rows *etc*.), taking random samples (datapoints) from the total dataset to fill each sample size with the specified number of data points. This sample size, of e.g. 1000 rows of randomly selected datapoints from the complete dataset, was then divided randomly into training and testing sets. The training set consisted of 70% of the samples, and 30% for the testing set. These training and testing sets from the specific sample size was then used by the ML algorithm, and the results were recorded. This process was repeated 100 times for each sample size. The time and accuracy for each of the 100 samples was recorded and plotted, in order to identify and compare the results from the Big Data system with that of the Standard system (which uses the open-source Python ML library *Scikit-learn*), with regards to time and accuracy. Each sample

An example of this would be for Linear Regression. By taking dataset $\boldsymbol{X}$, which is the total dataset and randomly dividing the dataset into smaller datasets $\mathbf{x}_{(i;j)}$, where *i* is the sample *number* $i \in \{1, 2, \ldots, n\}$ and $n = 100$ samples, for dataset sample size $j \in \{1\,000;\ 5\,000;\ 50\,000;\ 500\,000;\ 1\,000\,000;\ 5\,000\,000;\ 10\,000\,000;\ 20\,000\,000\}$. For each of these 100 samples, the results were then plotted (time *vs*. accuracy), e.g. $\mathbf{x}_{(38;1\,000\,000)}$ would be the 38th sample of sample size 1 000 000 rows. For each sample, randomly selected datapoints were taken until the 1 000 000 rows for the sample were filled. This sample was then randomly divided into training (700 000) and testing sets (300 000) respectively, followed by analysis, and the time and accuracy for this 38th analysis being recorded.

To ensure for a controlled experiment on both systems, only the time and accuracy of each analysis was recorded, and not the time required performing the pre-processing. Along with this, only the necessary steps required during the ML analysis were added, such that the ML algorithm could perform the analysis. Any parameter to *e.g.* smoothen, reduce over-fitting or improve the accuracy of each individual analysis was not added, but simply running the data through the algorithm and recording its performance, to ensure comparability between the two systems ML algorithms.

The goal of this validation was to illustrate the benefits of using Big Data technology when working with larger datasets, the BDAD's ability to analyse larger datasets than those of the standard system; when it is required to obtain results in a timely and accurate manner, and also to validate the BDAD and demonstrate Big Data Analytics.

4.1 Results from the classification analysis

The different Classification algorithms used to test the BDAD were chosen as they are commonly used and require different data types, which would allow the system capabilities to be tested and compared to the standard system. The datasets used were chosen to ensure the data types of the variables/features were labelled and contained categorical features which would allow for classification. An example where such classification algorithms could be used in engineering is to classify faulty or non-faulty products after production, given a set of features that describe what constitutes a faulty item.

4.1.1 Decision Trees

Due to its popularity, the first classification algorithm namely Decision Trees (DT) was considered. The dataset used to analyse both systems was a freely available NYC taxi dataset. The dataset had 28 million rows of taxi trips made for the first three months of 2017. The dataset contained 17 features, of which the *passenger_count* (number of people in the taxi for the specific trip), *trip_distance* (in miles), *Ratecode_ID* (what rating the ride was, dependent on trip surcharges *etc*.), *payment_type* (form of payment made) were used to predict the cost of the trip to the customer. The feature used to predict the cost was *fare_amount* (cost per trip in USD), the trip cost was divided into 15 cost ranges (from [0-2], [2-4],…,[400-450] *etc*.), in order for the DT ML algorithm to classify which cost range a customer could fall into given the input features. The dataset after pre-processing had 21 million usable rows from which to make predictions, therefore the samples were divided into sample sizes 1 000, 5 000, 50 000, 500 0000, 1000 000, 5000 000, 10 000 000 and 20 000 000. This was to get a broad idea of both systems' performance capabilities with different sample sizes. One hundred randomly selected samples were chosen for each sample size and a prediction was made whether a person would fall into one of the respective ranges, given the set of input parameters.

The results from the BDAD and standard systems are as follows. Figure 5 shows the mean results from the analysis on the standard system. As can be seen, as the sample sizes increased the mean accuracy of predictions increased. The BDAD was able to conduct the analysis in a reasonable time, but as can be noted in Figure 5, the time taken to analyse the sample size of 20 million was considerably longer at around 580 seconds per sample analysis.

Upon further investigation of the standard deviation shown in **Table 2**, the general trend indicates a small standard deviation in accuracy, therefore predictions did not differ by much from the mean for the BDAD. The *Standard* system was able to provide results that were generally lower in accuracy and with a larger standard deviation, shown in **Figure 6** and **Table 3** respectively. The standard system was, however, able to complete the analysis in less time than the BDAD. In addition, the standard system was not able to analyse the final sample size of 20 million, outputting a memory error.

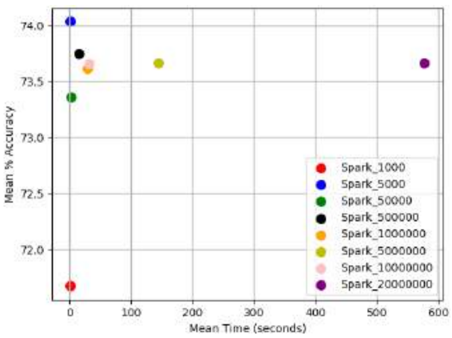


Figure 5: The mean results for the BDAD using the DT. As noted, the larger the sample size, the more accurate the prediction accuracy

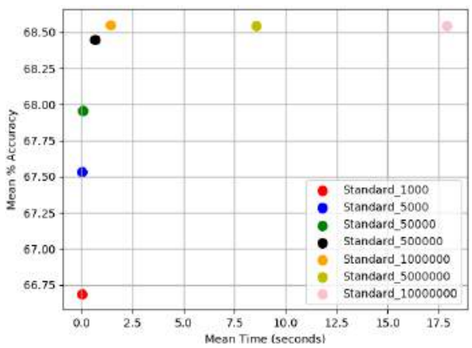


Figure 6: The mean results from all sample sizes on the Standard System. As noted, the larger the sample size, the more accurate the prediction accuracy

Table 2: Table of the standard deviations of the prediction accuracies for the BDAD for the DT

Sample Size	Standard Deviation
1000	0.0292
5000	0.0114
50000	0.0040
500000	0.0012
1000000	0.0034
5000000	0.0006
10000000	0.0006
20000000	0.0008

Table 3: Table of the standard deviation of the prediction accuracy for the DT on the Standard System

Sample Size	Standard Deviation
1000	2.8762
5000	1.4801
50000	0.7871
500000	0.3884
1000000	0.1420
5000000	0.0416
10000000	0.0257

4.1.2 Naïve Bayes

The same dataset that was used in demonstrating the *Decision Trees* ML algorithm was used for the Naïve Bayes ML algorithm. As with the NYC taxi dataset, the samples were divided randomly as before, with the same sample sizes for the 100 samples per sample size.

The results from the analysis on the BDAD and Standard system are shown in Figure 7 and Figure 8. From the figures, the mean accuracies for the predicted values of *fare_amount* decreased for the BDAD, and for certain sample sizes for the standard system. Spark predicted accuracy was lower than that on larger datasets, but critically, was able to analyse the larger datasets (the standard system was unable to analyse the 20 million-row dataset), compared with the standard system. The standard deviations were as before lower on the BDAD in Table 4 compared to the standard system shown in **Table 5**, which indicates that even though the accuracy was lower; the values remained close to the mean value. The reason for the lower prediction accuracy could be due to the dataset not being optimal for the algorithm on Spark; Naïve Bayes is typically applied to text classification.

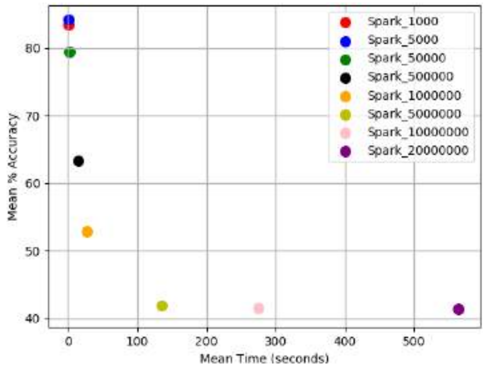


Figure 7: The mean accuracy results from the Naive Bayes analysis on the BDAD

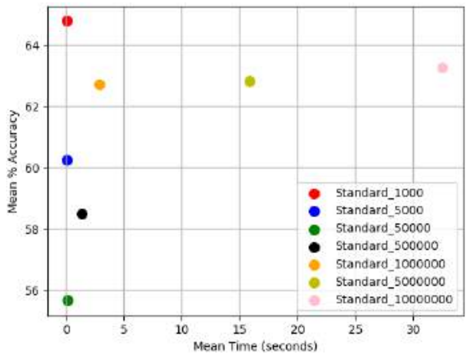


Figure 8: The mean accuracy results from the Naive Bayes analysis on the Standard system

Table 4: Table of the standard deviations of the prediction accuracies for the BDAD

Sample Sizes	Standard Deviation
1000	0.0260
5000	0.0186
50000	0.0758
500000	0.1954
1000000	0.1802
5000000	0.0009
10000000	0.0007
20000000	0.0008

Table 5: Table of the standard deviation of the prediction accuracy for the Standard System

Sample Size	Standard Deviation
1000	4.8724
5000	4.9056
50000	5.0318
500000	7.9217
1000000	4.8256
5000000	2.8657
10000000	2.3236

4.1.3 Logistic Regression

The dataset used in order to test and demonstrate the BDAD to the standard system was a freely available dataset, 23 features of which were used to indicate whether a woman would be likely to have a child which is born live, or *still/have an abortion*. The dataset had 7 259 473 rows. The dataset was a collection of results from questionnaires from a study conducted in different states in India. The dataset was modified so that if it was determined the woman had a birth which was *live*, a value of one was assigned, and zero for *still/abortion*. Logistic regression requires binary classification given a set of multiple features; therefore, the dataset used in *Decision Trees* was not applicable. The sample sizes were divided 1 000, 5 000, 50 000, 500 000, 1 000 000, 2 000 000, 4 000 000, and 6 000 000 accordingly, and as before, 100 samples per sample size were taken and analysed by the BDAD and standard system.

The results from the analysis on the BDAD and standard system show high accuracies were achieved as shown in **Figure 9** for the BDAD, with the standard system performing better with smaller dataset sizes as expected (the BDAD is designed for large datasets) shown in **Figure 10**. The standard system could only analyse sample sizes of up to 2 million rows before encountering limitations due to the hardware. The standard deviation also shows better performance with smaller datasets in **Table 6** compared with **Table 7** . From the results, it was shown that the standard system provides accurate results, but it could not analyse larger datasets, which was the purpose of the study, to demonstrate the limitations of non-Big Data software compared to the BDAD developed by the author.

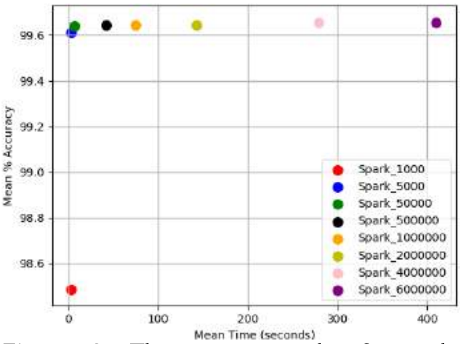


Figure 9: The mean results from the Logistic Regression Analysis on the BDAD

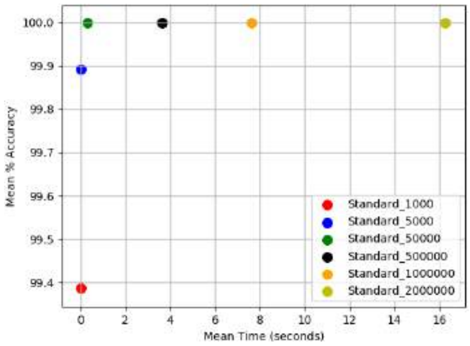


Figure 10: The mean results from the Logistic Regression Analysis on the standard system

Table 6: Standard deviations of the prediction accuracy of Logistic Regression on the BDAD

Sample Size	Standard Deviation
1000	0.0123
5000	0.0012
50000	0.0004
500000	0.0001
1000000	0.0001
2000000	0.0001
4000000	0.0000
6000000	0.0000

Table 7: Standard deviations of the prediction accuracy of Logistic Regression on the Standard System

Sample Size	Standard Deviation
1000	0.0034
5000	0.0006
50000	0.0000
500000	0.0000
1000000	0.0000
2000000	0.0000

4.1.4 Support Vector Machines (SVM)

The same dataset, sample, and sample size divisions used in the analysis of *Logistic Regression* for the BDAD and standard system, were used in the *SVM* analysis. SVM and Logistic Regression are typically being applied in text classification and image recognition, and therefore the datasets used to demonstrate the BDAD were chosen due to the size to demonstrate use cases of BDA. Therefore, these datasets are not optimal, but due to the difficulty to locate large enough datasets, these datasets were chosen.

The results in

Figure 12 show that for small sample sizes, the standard system outperforms the BDAD, but as sample sizes increased, the standard system was unable to perform a SVM analysis and only up to 500 000 rows could be analysed. This demonstrates the ability of the BDAD to make predictions on large datasets as all sample sizes could be analysed shown in Figure 11 with good accuracy compared to the standard system. The standard deviation values of the standard system in **Table 9** indicate the better performance at smaller sample sizes compared to that of the BDAD in **Table 8**.

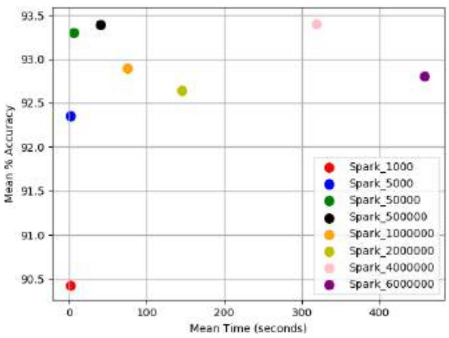


Figure 11: The mean results from the SVM Analysis on the BDAD

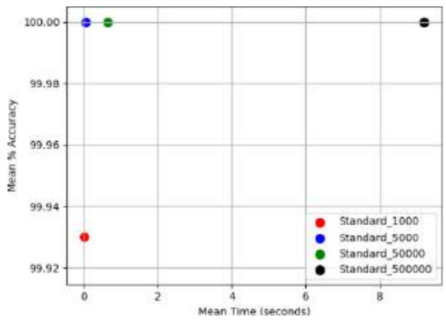


Figure 12: The mean results from the SVM Analysis on the standard system

Table 8: Standard deviations of the prediction accuracy of SVM on the BDAD

Sample Size	Standard Deviation
1000	0.0208
5000	0.0115
50000	0.0203
500000	0.0189
1000000	0.0140
2000000	0.0099
4000000	0.0189
6000000	0.0000

Table 9: Standard deviations of the prediction accuracy of SVM on the Standard System

Sample Size	Standard Deviation
1000	0.0024
5000	0
50000	0
500000	0

4.2 Results from the Regression analysis

To demonstrate the BDAD capabilities on regression analysis, a Linear Regression analysis was conducted with both systems. The datasets used for this were chosen such that the data contained data types that were labelled and numerical. A use case example in engineering could be to use regression analysis to predict production output over time producing a number of products on a production line.

4.2.1 Linear Regression

The analysis of the BDAD capabilities in conducting a Linear Regression analysis was tested using the same NYC taxi dataset as for the Decision Tree analysis. The same sample sizes and number of samples per sample size were used. For this analysis, however, the relationships between the *trip_time*, *trip_distance* and *fare_amount* were analysed and a Linear Regression analysis applied to predict how much a person could expect to pay for a taxi trip given the time and distance they expected to travel from the current position. This is because NYC taxis take into account time and distance when calculating the cost for a fare (time used when taxi is still, distance for when the taxi is moving).

The results from the analysis showed higher predicted values for the standard system in Figure 14 than the BDAD in Figure 13, with the standard system having slightly higher standard deviation values in Table 11, compared to the BDAD in Table 10. This means the predicted values vary more around the mean for the standard system. However, it can be noted that the standard system was not able to analyse the data for a sample size of 20 million, whereas the BDAD was able to provide results with small standard deviations. This analysis shows that the BDAD is able to analyse a larger dataset than the standard system, whilst providing accurate results in a timely manner.

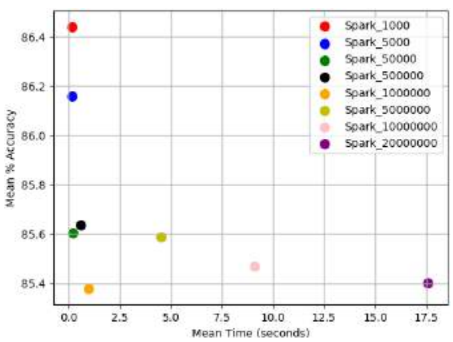


Figure 13: The mean results from the Linear Regression Analysis on the BDAD

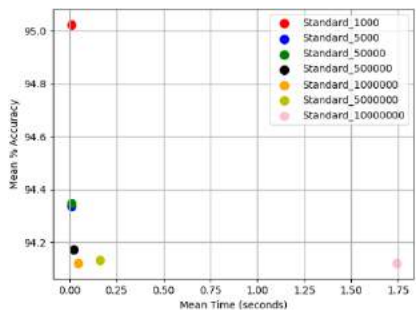


Figure 14: The mean results from the Linear Regression Analysis on the standard system

Table 10: The mean results from the Linear Regression Analysis on the standard system

Sample Size	Standard Deviation
1000	0.0800
5000	0.0358
50000	0.0168
500000	0.0055
1000000	0.0059
5000000	0.0018
10000000	0.0023
20000000	0.0011

Table 11: Standard deviations of the prediction accuracy of Linear Regression on the Standard System

Sample Size	Standard Deviation
1000	0.0472
5000	0.0320
50000	0.0166
500000	0.0055
1000000	0.0053
5000000	0.0023
10000000	0.0019

4.3 Results from the Cluster analysis

The k-means clustering ML algorithm was used to demonstrate the capabilities of the BDAD. This algorithm requires data to be unlabelled and numerical. Because the data is unlabelled, no predictions are made, but the algorithm investigates the similarities between the data points, grouping common data points together in clusters given their similarities with one another. An example would be to perform a cluster analysis of retail customers given a set of features, grouping similar customers together.

4.3.1 k-means

The dataset used to analyse the clustering capabilities of the BDAD was a freely available dataset of radiation readings taken from across the world totalling 80 million readings. The reason for using this dataset was due to the volume of data it contained, which could be clustered, namely the geographical positions of all the readings taken. Due to the unsupervised nature of clustering, the analysis of the clusters was done by making use of elbow plots, which required visual inspection in order to determine the best number of clusters after the clustering algorithm had been run.

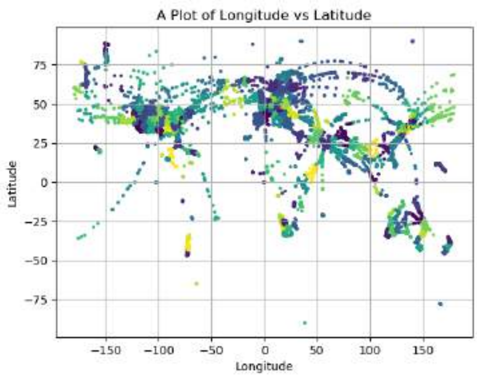


Figure 15: 195 Clusters formed after k-means on the BDAD. Only displaying a sample of the 80 million points

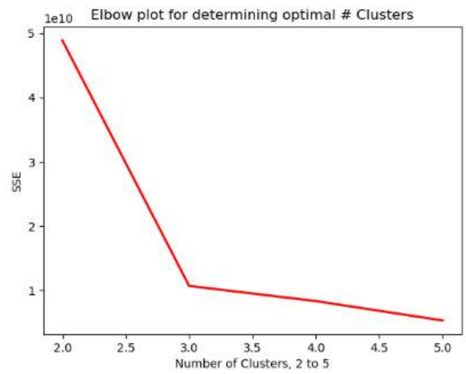


Figure 16: The elbow plot showing the SSE, indicating best number of clusters to be three on the BDAD. The plot is for illustration purposes, so as to identify the elbow in the plot

The results from performing k-means on the BDAD are shown in Figure 15 and Figure 16. These are the preliminary results indicating all 195 clusters, which represent the current 195 countries worldwide. The reason for choosing this number is to conduct an investigation into the radiation levels per country (taking into consideration also that some readings were taken by planes and ships, as shown in the figure, which clearly should not belong to a 'countries cluster' and is a contextual error of the clusters). When performing k-means in order to determine the best number of clusters used by the *elbow* method, these methods use the SSE (Sum Square Error) to determine the error in the distance between the datapoints and cluster centroid. The resulting *elbow plot* given in Figure 16 shows the optimal number of clusters to be three. In the context of the data provided, given that the analysis is correct and three clusters were optimal, the three clusters were deemed not to provide information of any further use. Figure 15 shows the radiation clustered roughly according to country, providing an analyst with information around radiation levels per country (making the results more usable).

The standard system was unable to perform *k-means* clustering on the dataset and was therefore unable to provide any meaningful insights. This example also demonstrated the superiority of the BDAD over the standard system in analysing larger volumes of data.

This dataset showed that the analysis undergone needs to be context specific, and that the acquired results need to be questioned for their applicability to the problem or question being posed.

5. FUTURE WORK

For the BDAD, proposed future work is to deploy more ML algorithms on the BDAD cluster in order to test the system capabilities. Another option could be to create a distributed non-relational database using technologies such as MongoDB. Using this, a datastore can be created, which offers the means to store data, not only in a distributed manner but also to access relevant data directly. This is because currently the database used for this project stores and distributes a file containing the relational information, which has to first be pre-processed to access the relevant fields. Another option is to test and use fast access databases such as Cassandra, in order to test the system using near real-time data and conduct near real-time data analysis, instead of batch analysis as was done for this project.

6. CONCLUSION

The BDAD developed for this project and as outlined in this article, demonstrates the use case for Big Data technologies; in order to conduct Big Data Analytics, typically unavailable or misunderstood by many. This article serves to illustrate to industrial engineers how such a BDA system could be developed using commodity hardware and software. The various analyses demonstrated the capabilities of the system compared to a *standard* system, which makes use of non-Big Data technologies in order to analyse data, using machine learning.
The results indicate the ability of the BDAD to provide accurate predictions with larger datasets than the standard system, thus demonstrating the capability of the BDAD and Big Data Analytics.

Such a system is of benefit to engineering, as more data is generated than previously and requires deeper analysis in less time. Deploying BDA and machine learning to aid analysis in fields where industrial engineering practices are required can benefit engineering as a whole. The information could then aid the decision-making processes of industrial engineers. This project and the BDAD is seen as a tool which can be used by systems designed with the fourth industrial revolution in mind, by providing BDA to discover deeper insights to aid decision making as well as automate and manage the processes of these new systems.

REFERENCES

[1] **P. C. Zikopoulos, D. DeRoos, K. Parasuraman, T. Deutsch, D. Corrigan, and J. Giles**, *Harness the Power of Big Data*. 2012.
[2] **R. D. Butler and J. Bekker**, "Development of a Big Data Analysis Demonstrator," in *SAIIE 28th Annual Conference*, 2017, pp. 1–13.
[3] **A. Spark**, "Apache Spark," *Apache Spark*, 2018. [Online]. Available: http://spark.apache.org/. [Accessed: 01-Feb-2018].
[4] T. White, *Hadoop: The Definitive Guide*, 3rd ed. Sebastopol, CA: O'Reilly, 2012.
[5] **M. Zaharia**, M. Chowdhury, M. J. Franklin, S. Shenker, and I. Stoica, "Spark : Cluster Computing with Working Sets," *HotCloud'10 Proc. 2nd USENIX Conf. Hot Top. cloud Comput.*, p. 10, 2010.
[6] **Cloudera**, *Spark Guide*, 1st ed. 1001 Page Mill Road, Bldg 3 Palo Alto, CA 94304: Cloudera, 2017.
[7] **A. Bekker**, "Spark vs. Hadoop MapReduce: Which big data framework to choose," *ScienceSoft*, 2017. [Online]. Available: https://www.scnsoft.com/blog/spark-vs-hadoop-mapreduce. [Accessed: 01-Mar-2018].
[8] **Hortonworks**, "Hortonworks Data Platform," 2016.
[9] Amazon, "Amazon Web Services," *Amazon*, 2018. [Online]. Available: https://aws.amazon.com/. [Accessed: 10-Jun-2018].
[10] **Dell**, "Dell Optiplex 9020 Specifications." Dell Computers, p. 3, 2018.
[11] Dell, "Dell Precision T3500 Specifications." Dell Computers, p. 2, 2018.
[12] **DataDotz**, "Spark Master/Slave installation in Multi Node," *Chennai Hadoop User Group*, 2018. [Online]. Available: http://chennaihug.org/knowledgebase/spark-master-and-slaves-multi-node-installation/. [Accessed: 02-Mar-2018].
[13] **S. Mitra**, "How to Setup Hadoop Multi Node Cluster - Step By Step," *DWBI*, 2016. [Online]. Available: https://dwbi.org/etl/bigdata/201-install-spark-in-hadoop-cluster. [Accessed: 02-Mar-2018].
[14] **DeZyre**, "Step-by-Step Apache Spark Installation Tutorial," *DeZyre*, 2018. [Online]. Available: https://www.dezyre.com/apache-spark-tutorial/apache-spark-installation-tutorial. [Accessed: 02-Mar-2018].
[15] **Y. Darji**, "Easily Setup Multi-Node Hadoop Cluster in YARN Mode on CentOS/Ubuntu," *Medium*, 2017.

[Online]. Available: https://medium.com/@yogeshdarji99/installing-multinode-hadoop-cluster-in-yarn-mode-86741c3df45b. [Accessed: 27-Feb-2018].
[16] **Fibrevillage**, "No Title," *Fibrevillage*, 2018. [Online]. Available: http://fibrevillage.com/storage/617-hadoop-2-7-cluster-installation-and-configuration-on-rhel7-centos7. [Accessed: 27-Feb-2018].
[17] **DWBI**, "How to Setup Hadoop Multi Node Cluster - Step By Step," *DWBI*, 2016. [Online]. Available: https://dwbi.org/etl/bigdata/183-setup-hadoop-cluster. [Accessed: 27-Feb-2018].
[18] **K. Rahul**, "How to Set Up Hadoop Multi-Node Cluster on CentOS 7/6," *tecadmin*, 2013. [Online]. Available: https://tecadmin.net/set-up-hadoop-multi-node-cluster-on-centos-redhat/. [Accessed: 27-Feb-2018].

ANALYSIS OF DEMAND FORECASTING STRATEGIES AND PRACTICES FOR EMPTY RAIL WAGONS: A SOUTH AFRICAN CASE STUDY.

R.P. Mutyavavire[1*] & M.M. Mazibuko[2]

[1]Department of Industrial Engineering
University of Johannesburg, South Africa
rmutyavavire@uj.ac.za

[2] Department of Industrial Engineering
University of Johannesburg, South Africa
meshaelmazibuko@gmail.com

ABSTRACT

Optimized demand management strategies and associated forecasting approaches make critically important contributions towards enhanced customer demand satisfaction and overall business sustainability.

In this paper we interrogate current empty rail wagon demand forecasting practices in the Southern African rail industry and present inadequacies noted. We further analyze case company time series demand data for empty wagons over a period of four years. A quantitative forecasting model based on the Holts-Winters method is developed and tested against real life data. Good to reasonable forecast results are recorded for forecast periods up to twelve months with the best forecasts recorded for the six to twelve months' forecast period.

[2] The author was enrolled for an M Eng (Industrial) degree in the Department of Industrial Engineering, University of Johannesburg, South Africa
*Corresponding author

1. INTRODUCTION

Freight transportation systems face challenges of optimally managing demand and supply of empty Loading units. In rail freight transportation, loading units exist in the form of wagons. Optimal allocation, customer satisfaction and efficient utilization are critically important in Empty wagon distribution (EWD) management [1], [2].

Rail transportation systems are characterized by general inadequate supply of empty wagons [2]. As the demand for empty wagon distribution grows, the need for optimized demand management practices and efficient infrastructure and resource utilization increases [3].

Forecasting is critically important in planning and demand management in general [4]. Inaccurate demand prediction adversely impacts demand management [5]. Forecasting, is an integral part of effective demand management and plays a fundamental role in predicting future customer demands for empty wagons.

A review of the case company current-state-of-practice revealed:

- High dependency on customer generated forecasts.
- The absence of scientific demand forecasting by the rail operator.
- Customer demands are uncertain and subject to change at short notice.

The absence of good forecasting practices adversely impacts planning processes as well as operational efficiency. In this paper we develop a forecasting model based on the Holt-Winters methodology and test its suitability and accuracy in the freight rail industry with specific emphasis on empty wagon demand forecasting.

2. Literature Review

The balance between demand and supply plays a critical role in the operational viability of most companies [5]. Customer demand is a function of economic trends. Demand is characteristically stochastic. The fundamental objective of Production planning is to satisfy the uncertain demand.. Demand and supply patterns are frequently out of phase. The challenge of synchronizing the two variables in pursuit of sustained demand satisfaction is real.

Demand management (DM) practice seeks sustainable customer demand satisfaction. DM integrates forecasting, planning and scheduling in a supply chain in order to achieve this objective over a specific period [6]. DM practice seeks a sustainable balance between the customers' requirements and the capabilities of the supply chain [7]. Demand management enables organizations to forecast customer demands. Good demand management is characterized by a company's ability to proactively anticipate customer demand and be reactive to unanticipated customer demand [7].

In the case of empty wagon distribution, the application of demand management practices will provide valuable insight into past supply and demand patterns of empty wagons within the rail network. Ultimately such information can be shared within the organization to determine the future demand requirements for empty wagons through qualitative and quantitative methods and assist in allocating the required resources so that customer demands are met. This information helps ensure that adequate capacity and resources are available to meet forecasted demand [6].

Forecasting is essential in planning and decision-making processes, particularly in a dynamic operational environment. It is important to focus on the management of the forecasting process with respect to other crucial DM issues [8, 9, 10].

The forecasting model should be designed to suit a given data set [11]. Chase [11] proposes a nine step evaluation and selection criteria as a general guideline for the most appropriate forecasting approach for a given data set. Good forecasting practice recommends that all departments in a company should share a single database [9]. Good forecasting management facilitates understanding of market dynamics and reduces the uncertainty of future events [10]. The forecasting process should be alive to environmental factors and requires periodic revision, maintenance and modification [12].

Time-series methods attempt to predict the future as a function of historical data [13]. Time-series methods identify historical demand patterns and assume future demand will mimic the patterns of the past [11]. It is important to have an insightful understanding of the causal factors influencing the patterns when applying time-series methods. Hanna *et al* [13] identify four time-series patterns, namely: the *trend pattern*, *seasonality pattern*, *cyclic pattern* and *random variation*. Mentzer *et al* [14] describes random variations as noise. Good forecasting methods provide for majority of these patterns. The Holtz-Winters (HW) method accommodates the level, trend and seasonality patterns. It is acknowledged as a top performing method in literature. The HW method forms the basis of our model.

3. RESEARCH METHODOLOGY

The deductive research approach was adopted founded upon Experimental quantitative research design. A mixed method research strategy was employed. Research data constitutes secondary data extracted from the case company central database. Explanatory information was generated from targeted interviews.

Time series data, covering a period of four years (2012-2016) was analyzed in this study. A minimum of three years panel data is necessary to accommodate possible seasonal patterns in the data set [11]. The data constitutes empty wagon demand on selected nodes of the case rail operator network.

A forecasting model developed from the HW method was developed. The model performance and forecast accuracy was tested on both in-sample and out-of-sample data. The model evaluation was separated into two sets, namely in-sample and out-of-sample. The in-sample data was for 36 months and the out-of-sample data was for 12 months. A 36-month period (April 2012-March 2015) data was used as the in-sample data set. This was done to test how well the data fits the additive and multiplicative versions of the Holt-Winters method. The model performance on in-sample data was evaluated. The Ljung-Box (18) statistic was applied to evaluate randomness of the residual errors of the model. The SPSS 23 tool was used for computations.

The model was further evaluated over a 12 month out-of-sample period. The forecasts results were then compared against actual recorded demand data over the same period. The 12 month out-of-sample period was divided into four quarters. The performance of the model in each quarter was independently evaluated.

4. The Model

Model component functions are briefly characterized below.

Simple Exponential Smoothing
Simple exponential smoothing is applied if a time series fluctuates about a base level where there are no trend and seasonal components [18]. The formulation of forecasts in simple exponential smoothing is based on the combination of past demand weighted by a smoothing parameter known as alpha (α). Alpha ranges from a scale of 0 to 1. Forecasting accuracy is dependent on the value of the smoothing parameter selected. The simple exponential smoothing model to forecast the next period demand is given by the equation:

$$F_t = (1-\alpha)F_{t-1} + \alpha x_t \quad (1)$$

Where:
F_t = new smoothed forecast (new forecast value)
F_{t-1} = previous smoothed value at time t
x_t = most recent actual value at time t
α = smoothing parameter ($0< \alpha < 1$)

Holts method
The Holt's method is one of the methods used when the data set exhibits a trend in a time-series. For the purpose of exponential smoothing a trend can be thought of as a series of changes in the level signifying the level either stepping up or stepping down with each successive period [16]. The Holt's method is similar to simple exponential smoothing except that it contains two components - *level* and *trend* which must be updated each period. According to [11,15,17], the Holt's method can be expressed by the following equations:

$$L_t = \alpha x_t + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (2)$$
$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (3)$$
$$F_{t,k} = L_t + kT_t \quad (4)$$

Where:
x_t = most recent actual value at time t
L_t = an estimate of the base level of the series at time t
T_t = an estimate of the trend data series at time t
L_{t-1} = previous estimate of base level at time t
T_{t-1}= previous estimate of trend at time t
α = smoothing parameter ($0< \alpha < 1$)
β = smoothing parameter ($0< \beta < 1$)
$F_{t,k}$= forecast for x_{t+k} at the end of period t to forecast for k periods ahead
k= the number of periods into the future to forecast

Holt-Winters method
This method is founded upon the Holt's method allows time-series data to be modeled by a local mean the level, a local trend and a local seasonal factor which are all updated by exponential smoothing [18]. The HW method applies three smoothing parameters α, β, and γ to represent level, trend and seasonality respectively.

Albright et al [19] notes that seasonal models can be classified under two categories – additive and multiplicative versions. The multiplicative version is appropriate for use in a time-series when the amplitude of the seasonal pattern is proportional to the level of the series. The additive version is appropriate for use in a time-series when the amplitude of the seasonal pattern is independent of the average level of the series [20,21].

The level, trend and seasonal fluctuations equations are updated as follows for the multiplicative version [15]:

$$L_t = \alpha\frac{x_t}{S_{t-C}} + (1-\alpha)(L_{t-1} + T_{t-1}) \quad (5)$$

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \quad (6)$$

$$S_t = \gamma\frac{x_t}{L_t} + (1-\gamma)T_{t-C} \quad (7)$$

Where:
L_t = an estimate of the smoothed level of the series at time t
T_t = an estimate of the smoothed trend of the series at time t
S_t = an estimate of the smoothed seasonal adjustment of the series at time t
x_t = most recent actual value at time t
C = is the number of periods in the length of the seasonal pattern (c =12 (months) and c =4 (quarterly))
α = smoothing parameter updating the level (0< α < 1)
β = smoothing parameter updating the trend (0< β < 1)
γ = smoothing parameter updating the seasonality (0< γ < 1)

Forecast Error
Model accuracy (forecasting error) is evaluated using the following criteria: the Mean absolute deviation (MAE), Mean squared error (MSE) and Mean absolute percentage error (MAPE).

The lower the value of MSE, MAE and MAPE the higher the forecast accuracy. An interpretation judgment criterion for the MAPE as shown in Table 1.

Table 1: Forecast accuracy judgment criteria of MAPE [22].

MAPE (%)	Interpretation
< 10	Highly Accurate Forecasting
10 - 20	Good Forecasting
20-50	Reasonable Forecasting
> 50	Inaccurate Forecasting

MSE and MAE are standard measures for model forecast accuracy, more relevant when comparing forecasting methods. Generally the lower the value the higher the forecast accuracy of the model.

5. RESULTS

The model fit statistics shown in Table 2 indicate the results obtained for each destination area when fitting the additive Holt-Winters model. Forecast results are presented as follows:
Bloemfontein
The proposed model performs better than the baseline model exhibiting a positive stationary R-squared value of 0.794. The model's RMSE, MAPE and MAE were found to be 1407.812, 15.594 and 1114.785 respectively at the model fitting stage. The model's MAPE is 15.594% for the in-sample data, thus the model generates good forecasts based on the fitted data. The Ljung- Box Q(18) statistics show that the significance output value is 0.021 which is less than the p-value of 0.05, thus showing that the model is significant and that the residual errors obtained from the time series are not random.

Botswana
The proposed model exhibits superior performance relative to the baseline model, a positive stationary R-squared value of 0.811. The model's RMSE, MAPE and MAE were found to be 1264.471, 10.835 and 979.084 respectively at the model fitting stage. The model's MAPE is 10.835% for the in-sample data indicating that the model generates good forecasts. The Ljung- Box Q(18) statistics show that the significance output value is 0.135 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Nelspruit
Applied on Nelspruit data, the model performed better than the baseline model exhibiting a positive stationary R-squared value of 0.824. The model's RMSE, MAPE and MAE were found to be 908.507, 9.996 and 622.453 respectively at the model fitting stage. The model's MAPE is 9.996% for the in-sample data indicating that the model generates forecasts that are highly accurate. The Ljung- Box Q(18) statistics show that the significance output value is 0.777 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Polokwane
The new model performs better than the baseline model exhibiting a positive stationary R-squared value of 0.635. The model's RMSE, MAPE and MAE were found to be 1253.520, 23.279 and 974.576 respectively at the model fitting stage. The model's MAPE is 23.279% for the in-sample data, thus the model generates reasonable forecasts. The Ljung- Box Q(18) statistics show that the significance output value is 0.117 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Table 2: Model Statistics for Additive Holt-Winters method for all destination areas

Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAPE	MAE	Statistics	DF	Sig.	
BLOEMFONTEIN-Model_1	0	.794	1407.812	15.594	1114.785	28.136	15	.021	0
BOTSWANA-Model_2	0	.811	1264.471	10.835	979.084	21.055	15	.135	0
NELSPRUIT-Model_3	0	.824	908.507	9.996	622.453	10.643	15	.777	0
SWAZILAND-Model_4	0	.856	738.291	15.634	542.977	16.293	15	.363	0
POLOKWANE-Model_5	0	.635	1253.520	23.279	974.576	21.645	15	.117	0

The Multiplicative Holt-Winters method

The model fit statistics in Table 3 shows the results obtained for each node when applying the multiplicative Holt-Winters method. The results are as follows:

Bloemfontein
Forecast results are better than the baseline model exhibiting a positive stationary R-squared value of 0.561. The model's RMSE, MAPE and MAE were found to be 1576.616, 18.561 and 1213.622 respectively at the model fitting stage. The model's MAPE is 18.561% for the in-sample data, thus the model generates good forecasts. The Ljung- Box Q(18) statistics show that the significance output value is 0.013 which is lesser than the p-value of 0.05, thus showing that the model is significant and that the residual errors obtained from the time series are not random.

Nelspruit
The model yields better forecasts relative to the baseline model exhibiting a positive stationary R-squared value of 0.628. The model's RMSE, MAPE and MAE were found to be 1059.997, 10.577 and 676.855 respectively at the model fitting stage. The model's MAPE is average of 10.577% for the in-sample data indicating that the model generates good forecasts. The Ljung- Box Q(18) statistics show that the significance output value is 0.873 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Swaziland
Improved forecast results are recorded relative to the baseline model exhibiting a positive stationary R-squared value of 0.621. The model's RMSE, MAPE and MAE were found to be 909.876, 18.468 and 661.654 respectively at the model fitting stage. The model's MAPE is 18.468% for the in-sample data indicating that the model generates good forecasts. The Ljung- Box Q(18) statistics show that the significance output value is 0.737 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Polokwane
The proposed model provides better forecasts relative to the baseline model exhibiting a positive stationary R-squared value of 0.353. The model's RMSE, MAPE and MAE were found to be 1469.983, 28.282 and 1100.464 respectively at the model fitting stage. The model's MAPE is 28.282% for the in-sample data, thus the model seems to generate forecasts that are reasonably accepted. The Ljung- Box Q(18) statistics show that the significance output value is 0.280 which is greater than the p-value of 0.05, thus showing that the model is insignificant and that the residual errors obtained from the time series are random.

Table 3: Model Statistics for Multiplicative Holt-Winters method for all destination areas

Model Statistics

Model	Number of Predictors	Model Fit statistics				Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	RMSE	MAPE	MAE	Statistics	DF	Sig.	
BLOEMFONTEIN-Model_1	0	.561	1576.616	18.561	1213.622	29.630	15	.013	0
BOTSWANA-Model_2	0	.668	1462.522	11.940	1054.233	13.117	15	.593	0
NELSPRUIT-Model_3	0	.628	1059.997	10.577	676.885	9.084	15	.873	0
SWAZILAND-Model_4	0	.621	909.876	18.468	661.654	11.224	15	.737	0
POLOKWANE-Model_5	0	.353	1469.983	28.282	1100.464	17.669	15	.280	0

The additive Holt-Winters model had the lower values of the RMSE, MAPE and MAE for all the nodes as shown in Table 4. The additive model is comparatively more accurate. Consequently we select the addictive model as the model of choice for out-of-sample forecasts.

Table 4: Comparison of Additive and Multiplicative HW methods

Models	Additive Holt-Winter's			Multiplicative Holt-Winter's		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE
Bloemfontein - Model 1	1407.812	15.594%	1114.785	1576.616	18.561%	1213.622
Botswana -Model 2	1264.471	10.835%	979.084	1462.522	11.940%	1054.233
Nelspruit - Model 3	908.507	9.996%	622.453	1059.997	10.577%	676.885
Swaziland - Model 4	738.291	15.634%	549.977	909.876	18.468%	661.654
Polokwane - Model 4	1253.52	23.279%	974.576	1469.983	28.282%	1100.464

1.1.1 Out-of-sample Forecasting

Out-of-sample forecasts applying the Additive HW method for a time period of 12 months, divided into four seasons, were generated by the model for each node. The respective forecast errors were evaluated. The results are as follows.

Bloemfontein

Table 5 represents the selected smoothing parameters of alpha, gamma and delta respectively which were used to estimate the level, trend and seasonal components.

Table 5: Smoothing parameters selected

Model	Smoothing parameters estimate	
BLOEMFONTEIN_Model 1	Alpha (Level)	0.098
Additive Holt-Winter's	Gamma (Trend)	2.996E-07
	Delta (Season)	3.162E-05

Figure 1 below illustrates the out-of-sample forecasts are generally within boundary limits. The month of November recorded inaccurate forecasts.

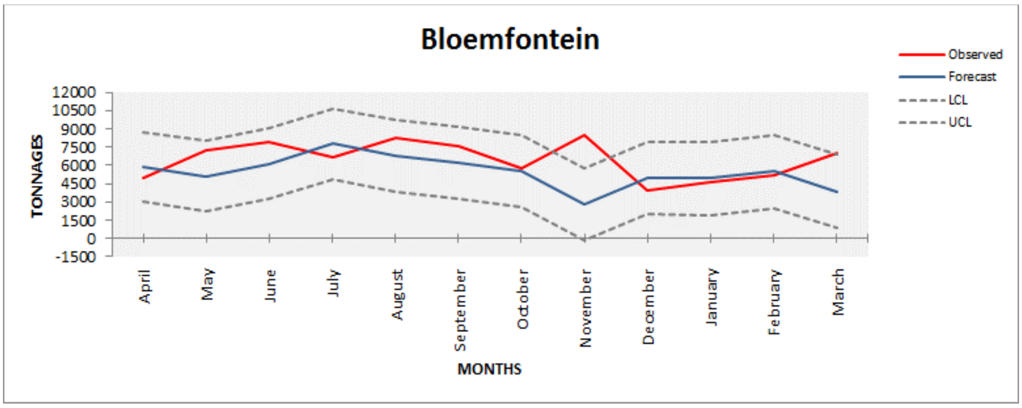


Figure 1: Time series of test data set and out-of-sample forecasts (April 2015-March 2016)

Botswana

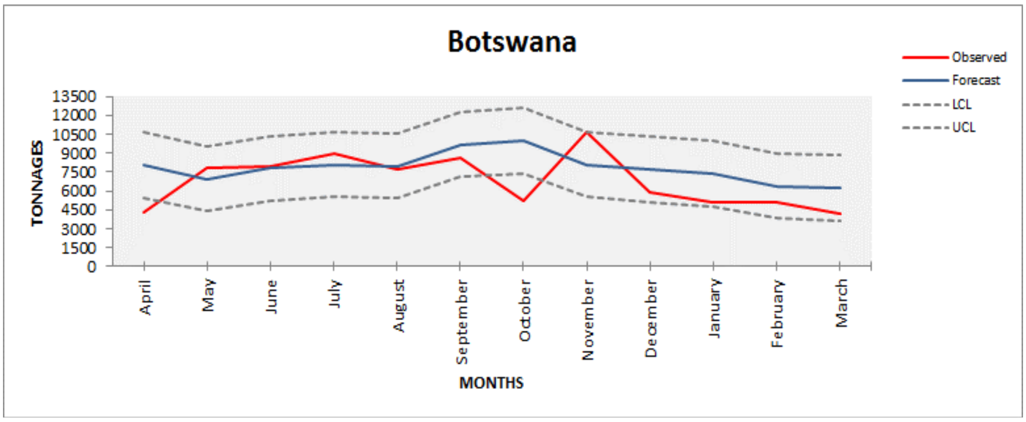


Figure 2: Time series data set and out-of-sample forecasts (April 2015-March 2016)

Satisfactory forecasts were recorded with the exception of October estimate.

The model yields the most accurate forecast within the six-month time horizon as illustrated in table 6 below. The MAPE value is 21%.

Table 6: Model accuracy in different time periods

	3 Months period	6 Months period	9 Months period	12 Months period
RMSE	2236.576	1680.206	2347.571	2240.410
MAPE	33.724%	21.165%	30.386%	32.446%
MAE	1603.173	1172.1475	1798.918	1808.464

Nelspruit

The smoothing parameter values of alpha, gamma and delta respectively are shown in Table 7 below.

Table 7: Selected smoothing parameters

Model	Smoothing parameters estimate	
NELSPRUIT_Model 1	Alpha (Level)	0.095
Additive Holt-Winter's	Gamma (Trend)	2.795E-07
	Delta (Season)	6.549E-05

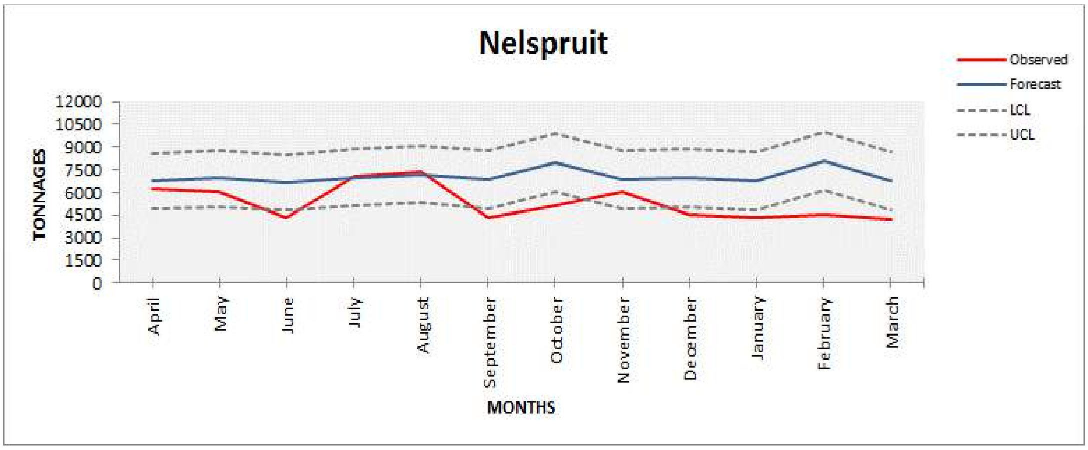


Figure 3: Time series data and out-of-sample forecasts (April 2015-March 2016)

Reasonably good forecast results were recorded during the year for Nelspruit. Forecasts fall within the UCL and LCL. Table 8 shows the best forecast horizon is within six months.

Table 8: Model accuracy in different time periods

	3 Months period	6 Months period	9 Months period	12 Months period
RMSE	1500.303	1484.974	1742.580	2080.400
MAPE	26.526%	23.601%	28.971%	37.797%
MAE	1286.982	1104.353	1398.704	1754.015

Swaziland

The model yields good forecast results over the twelve-month period as illustrated in figure 4 below.

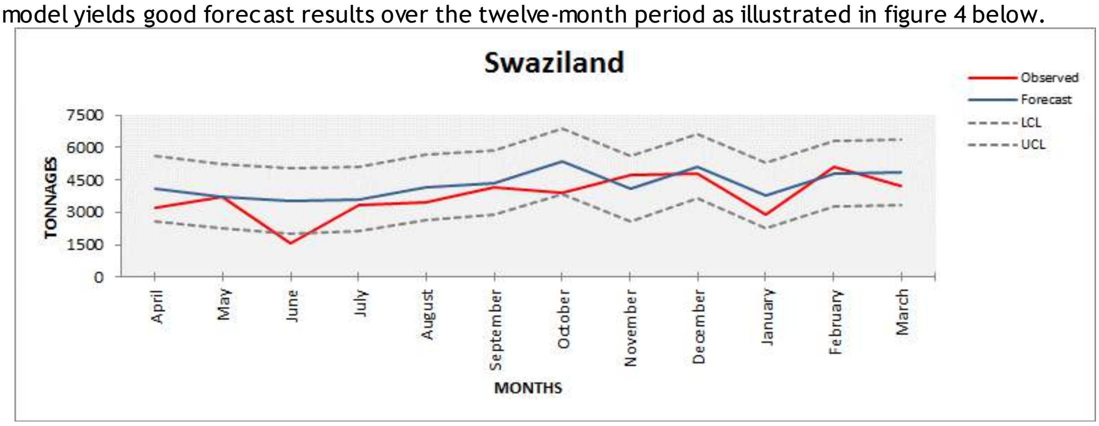


Figure 4: Time series data set and out-of-sample forecasts (April 2015-March 2016)

Table 9 below illustrates the forecast results. Most accurate results are obtained between nine and twelve months' time horizons.

Table 9: Model accuracy over different time periods

	3 Months period	6 Months period	9 Months period	12 Months period
RMSE	1231.492	923.988	928.052	867.827
MAPE	50.109%	30.507%	26.701%	24.341%
MAE	950.956	668.0748	713.341	686.804

Polokwane

Table 10 represents the smoothing parameters of alpha, gamma and delta respectively which were used to estimate the level, trend and seasonal components.

Table 10: Selected smoothing parameters

Model	Smoothing parameters estimate	
POLOKWANE_Model 1	Alpha (Level)	0.4
Additive Holt-Winter's	Gamma (Trend)	3.48E-05
	Delta (Season)	0.000

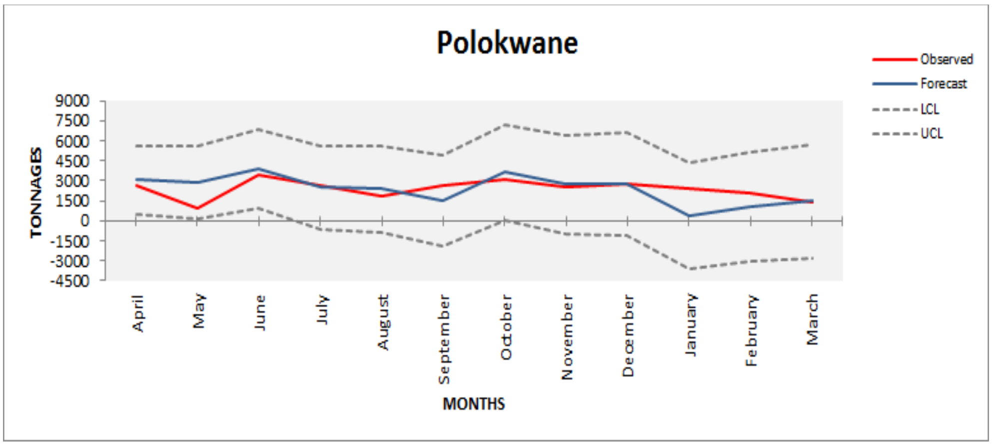


Figure 5: Time series data set and out-of-sample forecasts (April 2015-March 2016)

Good forecast results were obtained for the entire forecast period for this node.

Table 11 below reflects the forecast results are best obtained between nine and twelve months' time horizons.

Table 11: Model accuracy over different time periods

	3 Months period	6 Months period	9 Months period	12 Months period
RMSE	1171.712	967.594	814.890	2095.936
MAPE	80.649%	52.881%	38.396%	40.762%
MAE	952.172	765.3165	601.241	722.914

6. CONCLUSION

The model generally gives good forecast results for in-sample data. The additive HW method however yielded better forecasts relative to the multiplicative method. The additive HW model performance on out-out-sample data ranged from good to poor depending on forecast horizon. The six to twelve month time horizon recorded more accurate forecasts.

The forecasting model presented would assist the rail operator improve forecasting, planning activities and associated demand management processes.

7. REFERENCES

1. Haghani, A.E. 1986, *A combined model of train routing, makeup, and empty car distribution*, North Western University.
2. Joborn, M., Crainic, T.G., Gendreau, M., Holmberg, K. & Lundgren, J.T. 2004, "Economies of scale in empty freight car distribution in scheduled railways", *Transportation Science*, vol. 38, no. 2, pp. 121-134.
3. Tatineni, V.C. & Demetsky, M.J. 2005, *Supply chain models for freight transportation planning*.
4. Makridakis, S. 1996, "Forecasting: its role and value for planning and strategy", *International Journal of Forecasting*, vol. 12, no. 4, pp. 513-537.
5. Klassen, K.J. & Rohleder, T.R. 2001, "Combining operations and marketing to manage capacity and demand in services", *Service Industries Journal*, vol. 21, no. 2, pp. 1-30.
6. Martin, J.W. 2008, *Operational Excellence : using lean six sigma to translate customer value through global supply chains*, Auerbach Publications, New York.

7. Croxton, K.L., Lambert, D.M., García-Dastugue, S.J. & Rogers, D.S. 2002, "The demand management process", *The International Journal of Logistics Management,* vol. 13, no. 2, pp. 51-66.
8. Armstrong, J.S. 2001, *Principles of forecasting: a handbook for researchers and practitioners,* Springer Science & Business Media.
9. Moon, M.A., Mentzer, J.T. & Smith, C.D. 2003, "Conducting a sales forecasting audit", *International Journal of Forecasting,* vol. 19, no. 1, pp. 5-25.
10. Danese, P. & Kalchschmidt, M. 2011, "The role of the forecasting process in improving forecast accuracy and operational performance", *International Journal of Production Economics,* vol. 131, no. 1, pp. 204-214.
11. Chase, C.W. 2013, *Wiley and SAS Business Series: Demand-Driven Forecasting : A Structured Approach to Forecasting (2),* Wiley, Somerset, USA.
12. Herbig, P.A., Milewicz, J. & Golden, J.E. 1993, "The do's and don'ts of sales forecasting", *Industrial Marketing Management,* vol. 22, no. 1, pp. 49-57.
13. Hanna, M.E., Render, B. & Stairs, R. 2009, "Quantitative Analysis for Management",.
14. Mentzer, J.T., Moon, M.A., Estampe, D. & Margolis, G. 2007, "Demand management", *Handbook of global supply chain management,* , pp. 65-85.
15. Winston, W.L. and Goldberg, J.B., 2004. Operations research: applications and algorithms (Vol. 3). Belmont: Thomson Brooks/Cole.
16. Mentzer, J.T. & Moon, M.A. 2004, Sales forecasting management: a demand management approach, Sage Publications.
17. Mentzer, J.T. & Moon, M.A. 2005, Sales forecasting management : a demand management approach, 2nd edn, Sage Publications, Thousand Oaks, California.
18. Chatfield, C. & Yar, M. 1988, "Holt-Winters forecasting: some practical issues", *The Statistician,* pp. 129-140.
19. Albright, C.S., Winston, W.L. & Zappe, C.J. 2009, *Data analysis & decision making with Microsoft Excel,* 3rd Revised edn, Thomson South-Western, Australia.
20. Makridakis, S. & Wheelwright, S.C. 1989, *Forecasting methods for management,* 5th edn, Wiley, New York.
21. Kalekar, P.S. 2004, "Time series forecasting using holt-winters exponential smoothing", *Kanwal Rekhi School of Information Technology,* vol. 4329008, pp. 1-13.
22. Pan, R. 2010, "Holt–Winters Exponential Smoothing", *Wiley Encyclopedia of Operations Research and Management Science.*
23. Lewis, C.D. 1982, *Industrial and business forecasting methods: A radical guide to exponential smoothing and curve fitting,* Butterworth Scientific, London, Boston.

A METHODOLOGY TO ASSIST PROJECT-BASED ORGANISATIONS TO UTILISE AND DEVELOP PROJECT MANAGEMENT CAPABILITY AS A STRATEGIC ASSET TO INCREASE COMPETITIVE ADVANTAGE

C.J. Nel[1*] & C.S.L. Schutte[2]

[1]Department of Industrial Engineering
University of Stellenbosch, South Africa
coenraadnel@gmail.com

[2]Department of Industrial Engineering
University of Stellenbosch, South Africa
corne@sun.ac.za

ABSTRACT

Project Management, in business, plays both a functional and a strategic role in the sustainability of the organisation. The former, and its role towards the success of the organisation, has been researched intensively. The strategic role of Project Management towards the long-term sustainable success of an organisation is a new research topic with limited existing work done. A need therefore exists in project-based organisations to utilise their Project Management capabilities, not only towards functionally delivering successful projects but to apply it as a strategic asset to increase long-term organisational competitive advantage. The aim of this study, and goal of this paper, is to research the importance of Project Management as a strategic asset and to develop a model which can be used to evaluate the strength of Project Management as a strategic asset. This aim will be reached by conducting an integrated literature study on the knowledge areas of Project-Based Organisations, Project Management Capability- and Maturity Models and Strategic Management.

[1] The author was enrolled for a PhD in the Department of Industrial Engineering, University of Stellenbosch
*Corresponding author

1. INTRODUCTION

Project-based organisations can be defined as organisations with organisational structures designed around projects, and the delivery of projects on an on-going basis [1]. The business processes of project-based organisations are designed and structured around projects and the support of projects on an ongoing basis. Projects within these types of organisations are mostly long in duration, high capital value and complex. Furthermore, these organisations are cross-functional and generate income through designing, constructing and selling projects and services that are not of a standardised nature. In other words, the products and services offered by these organizations are unique to the client's requirements and budget.

Project Management in non-project-based organisations are in place to support the organisation's improvement strategy and to support its core value chain [2]. It can thus be said that *Project Management* in these organisations are put in place either to design and improve the existing business offering or to design and improve processes in which the existing product or serves are produced and offered to the client. On the other hand, in project-based organisations, Project Management forms part of the core business processes. For these organisations, the long-term sustainable success of these businesses relies on the organisation's *Project Management Capability*. As a result, it is important to understand the strategic importance of *Project Management* in these organisations, as well as to be able to understand how to measure the strategic strength, or aligned, as well as how competitive the organisation's project management capability is within these organisations. A need, therefore, exists for a method, framework or model to measure these aspects, and in the process supply a solution for these organisations to improve.

2. DESIGN REQUIREMENTS AND SUB-MODELS

The research took a *reverse design* approach whereby the research problem statement was taken as a point of departure, and through a combination of *systems engineering* and *iterative design*, developed the detail for a proposed model which could solve the above mentioned problem. The research problem is defined as follow:

> *"Project-based Organisations have the need to assess the alignment of their Project Management Capability as a strategic business asset, in order to ensure sustainable competitive advantage within the industry."*

By applying the *reverse design* approach and through the processes of *systems engineering problem solving* the research problem statement were further developed into nine model design requirements, which were then sub-developed into Output Requirements (OR), Input Requirements (IR) and Process Requirements (PR). These nine sets design requirements, shown in the knowledge areas they reside in, are shown in Table 1 below.

Table 1: Model Design Requirements

	DESIGN REQUIREMENTS
I.	PROJECT BASED ORGANISATION
OR1	The model should be able to produce a quantitative score, or status, on whether, and how strongly the organisation can be categorised as a PBO.
PR1	The model should be able to assess to what extent the organisation can be categorised as a project-based organisation.
IR1	The model should be able to accept information which can be utilised to do an assessment of the organisation's PBO status.
II.	PROJECT MANAGEMENT MATURITY
OR2	The model should be able to produce a quantitative score, or status, on the maturity level of the organisation's PMC.
PR2	The model should be able to assess the organisation's PMM.
IR2	The model should be able to accept information which can be utilised to assess the organisation's PMM.
OR3	The model should be able to produce specific information on how the organisation's PMM can be improved.
PR3	The model should be able to calculate or highlight areas of improvement to improve the PMM level.
IR3	The model should be able to accept information which can be utilised to highlight areas of improvement to improve the level of PMM.
III.	STRATEGIC ASSET

OR4	The model should be able to produce a quantitative score, or status, on how closely the organisation's PMC is aligned towards being a strategic asset.
PR4	The model should be able to assess the organisation's PMC towards being a strategic asset.
IR4	The model should be able to accept information which can be utilised to assess the organisation's PMC towards being a strategic asset.
OR5	The model should be able to produce specific information on how the organisation's PMC can be improved to be more aligned towards being a strategic asset for the organisation.
PR5	The model should be able to calculate or highlight areas of improvement within the PMC to improve the alignment towards being a strategic asset for the organisation.
IR5	The model should be able to accept information which can be utilised to improve the alignment towards being a strategic asset for the organisation.
IV.	SUSTAINABLE COMPETITIVE ADVANTAGE
OR6	The model should be able to produce a quantitative score, or status, of how competitive the organisation's PMC is within the industry it operates in.
PR6	The model should be able to assess how competitive the organisation's PMC is within the industry it operates in.
IR6	The model should be able to accept information which can be utilised to assess how competitive the organisation's PMC is within the industry it operates in.
OR7	The model should be able to produce specific information on how the organisation's PMC can be improved to be more competitive in the industry it operates in.
PR7	The model should be able to calculate or highlight areas of improvement to improve the competitiveness of its PMC within the industry it operates in.
IR7	The model should be able to accept information which can be utilised to improve the competitive advantage of its PMC.
OR8	The model should be able to produce a quantitative score, or status, on how sustainable its PMC competitiveness is within the industry it operates in.
PR8	The model should be able to assess how sustainable its PMC competitiveness is within the industry it operates in.
IR8	The model should be able to accept information which can be utilised to how sustainable its PMC competitiveness is within the industry it operates in.
OR9	The model should be able to produce specific information on how to improve the sustainability of the organisation's PMC competitive advantage.
PR9	The model should be able to calculate or highlight areas of improvement to improve the competitive advantage sustainability of its PMC.
IR9	The model should be able to accept information which can be utilised to improve the competitive advantage sustainability of its PMC.

The framework for the required model was then developed by proposing five *sub-models*, each addressing a set of design requirements as mentioned above and listed below.

- Sub-Model 1 – Project Based Assessment
- Sub-Model 2 – Project Management Maturity Assessment
- Sub-Model 3 – Strategic Asset Alignment Assessment
- Sub-Model 4 – Competitive Advantage Assessment
- Sub-Model 5 – Competitive Advantage Sustainability Assessment

The detail for the five *sub-models* was developed by applying in-depth literature studies on the knowledge areas presented by the *sub-models* and *design requirements* listed above. From the main study, the knowledge areas are *Project Management & Project-Based Organisations*; *Project Management Capability & Maturity*; and *Strategic Management, Strategic Assets & Competitive Advantage*.

2.1 Project Management & Project-based Organisations

The first set of design requirements focuses on defining and categorising Project-based Organisations (OR1, PR1, IR1). Traditionally the aim of *Project Management*, within organisations, was to support and execute projects successfully. This fact is changing as the result of the high demand on organisations to be more efficient and more competitive [3]. Bollinger and Smith [3] further state that organisations can no longer rely on traditional ways of Project Management, but should position themselves in a way that *Project Management* takes on a more strategic role in the organisation to assure sustainable success. This approach is even more imperative in organisations in which Project Management plays a strategic role towards the success of the organisation, i.e. *Project-based Organisations*. *Project-based Organisations* can be defined as organisations with organisational structures designed around projects, and the delivery of projects on an on-going basis [1]. By applying the studies of Archibald and Peppard [4], Kerzner [5] and Morris [2], the criteria for *Project-Based Organisations* were identified. These criteria were further developed into *Project-Based Organisations Assessment Statements*, which is listed below.

1. Your organisation's primary business is made up of delivering successful projects in an ongoing and sustainable manner.
2. Your organisation's growth strategies are positioned around the size, type, location and nature of the projects it takes on.
3. When comparing resource allocation between projects and operations, most of your organisation's resources are allocated towards the execution of successful projects, on an ongoing basis.
4. Project management is part of your organisation's core business processes.
5. In the development of your organisation's project, the following generic sequence is followed rigidly: *idea, outline, concept and strategy, and close-out*.
6. Your organisation has a value chain set up and designed to support the construction and selling of non-standardise products, goods and services.

These six *Project-Based Organisations Assessment Statements* were then applied into *sub-model 1,* to assess whether the organisation can be categorised as a *Project Based Organisations* and how strong it can be categorised as a *Project Based Organisations*. *[6] [7] [8] [9] [10] [11] [12] [13] [14] 15] [16]

2.2 Project Management Capability and Maturity Levels

The second set of design requirements has to do with measuring the organisation's *Project Management Capability*, and how it can be improved (OR2, PR2, IR2, OR3, PR3, IR3). It can be said that *Organisational Capability* is the organisation's ability to function as it should, or to do what it is there to do. Kelchner [17] take this further by stating that organisational capability is the organisation's ability to manage its resources effectively and to utilise this to gain a further *competitive advantage* in the industry. Scott [18] supports this theory by stating that *Business Capabilities* describe the organisation's unique and collective ability which must be applied to reach a specific desirable outcome. Peppard and Ward [19] state that organisational capability refers to how the organisation is applying their competencies, or capabilities, strategically. Therefore, if it can be said that the organisation is categorised as a *Project-based Organisation*, its strategic capabilities should be focused around the function of *Project Management*. A way to measure the state, or maturity, of the organisation's *Project Management Capability*, is through describing and quantifying the organisation through its *Project Management Maturity*.

To solve the bigger research problem and to link to what was said above, a need therefore exists for a way to measure the organisation's *Project Management Capability*, i.e. *Project Management Maturity*. To accomplish this, a critical analysis on existing *Project Management Maturity Models* were done. From this analysis the following criteria were developed which are to be taken into account when choosing, or developing, a *Project Management Maturity Model* for the proposed model:

- It should be fully customisable, from one organisation to another;
- It should be able to align *Project Management Capabilities* with organisational strategic objectives;
- It should be able to measure predefined *Key Performance Indicators*, aligned with the Project Management knowledge areas;
- It should be able to measure *Project Management Capabilities* in the context of the Project Management domains, i.e. *Projects, Programs* and *Portfolios*;
- It should be able to incorporate the iterative improvement framework of *plan-do-check-act*;
- It should be an evidence-based model;
- It should be simple and easy to use;
- It should supply output in a single value of assessment plus highlight areas of improvement towards next assessment; and
- It should be able to form part of a bigger assessment model.

After an assessment of existing *Project Management Maturity Models* in the context of the criteria mentioned above, it was found that none of the existing models satisfy these criteria. The criteria were then applied to *Project-Management-Maturity-Model* theory to come up with a new *Project Management Maturity Model*. This new model was applied to form *sub-model 2, as mentioned above,* of the proposed model. *[20] [21] [22] [23] [24] [25] [26] [27] [28] [29] [30]

2.3 Strategic Management, Strategic Assets and Competitive Advantage.

The third set of design requirements (OR4, PR4, IR4, OR5, PR5, IR5) has to do with how strongly the organisation's *Project Management Capability* is aligned into being a strategic asset for the organisation, and how this *strategic asset alignment* can be improved. Amit and Schoemaker [31] states that a *strategic asset* is an organisational asset which are, *scarce, hard to trade, hard to imitate* and which can add *value to the strategic goals* of the organisation. Bollinger and Smith [32] take this further by stating that these types of assets are a key factor in the sustainable success and competitive advantage of any organisation. By taking the above mentioned criteria,

[31], into account, and by applying it to the *Project Management Capabilities*, identified in *Project Management Maturity Model*, a *Strategic Asset Alignment* assessment is proposed, *sub-model 3*.

The fourth set of design requirements (OR6, PR6, IR6, OR7, PR7, IR7, OR8, PR8, IR8, OR9, PR9, IR9) has to do with how competitive the organisation is, as a result of its *Project Management Capability* in the industry it operates in. In this regard, Chakraborty [10] said that an organisation can gain competitive advantage through various ways, which might include some of the following: The organisation's activities (to add value, see below for Porter's perspective on Competitive Advantage); Organisational culture; Processes; Structure; and Innovation. *Strategic Positioning*, on the other hand, is defined by Porter [33] as the process in which organisations attempts to achieve a *sustainable competitive advantage*. This *Strategic Positioning* is achieved through keeping to itself the distinctive activities adding value to the organisation and putting them apart from other industry role players. In other words, it means to perform different activities, or performing similar activities in different ways than competitive industry role players. It is therefore not only important to achieve high levels competitiveness in the industry you operate in, but also to have systems and processes in place to assure *sustainability* in that competitiveness. Porter [33] therefore states that for the organisation to be categorised as competitive in its industry, it needs to either perform different activities than industry competitors, or perform similar activities different than industry competitors. This notion was then applied to the proposed model, to form *sub-model 4*, in which an assessment is proposed whereby Porter's requirement for competitiveness is applied to the *Project Management Capabilities* identified earlier.

In his study, Porter [33] further mentions that a direct correlation exists between the organisation's ability to apply innovation and how sustainable it is in its competitiveness. Because one of the *Key Performance Indexes*, identified as part of the *Project Management Maturity Model* (sub-model 2), under which the *Project Management Capabilities* are categorised in, is innovation, it was proposed that this score be applied to the model as *sub-model 5*. This scoring is done to give an indication, or probability, of how sustainable the organisation's sustainable competitiveness is. *[34] [35] [36] [37] [38] [39] [40] [41] [42] [43] [44] [45]

Figure 1 shows visually what was described above, in the context of the proposed model framework.

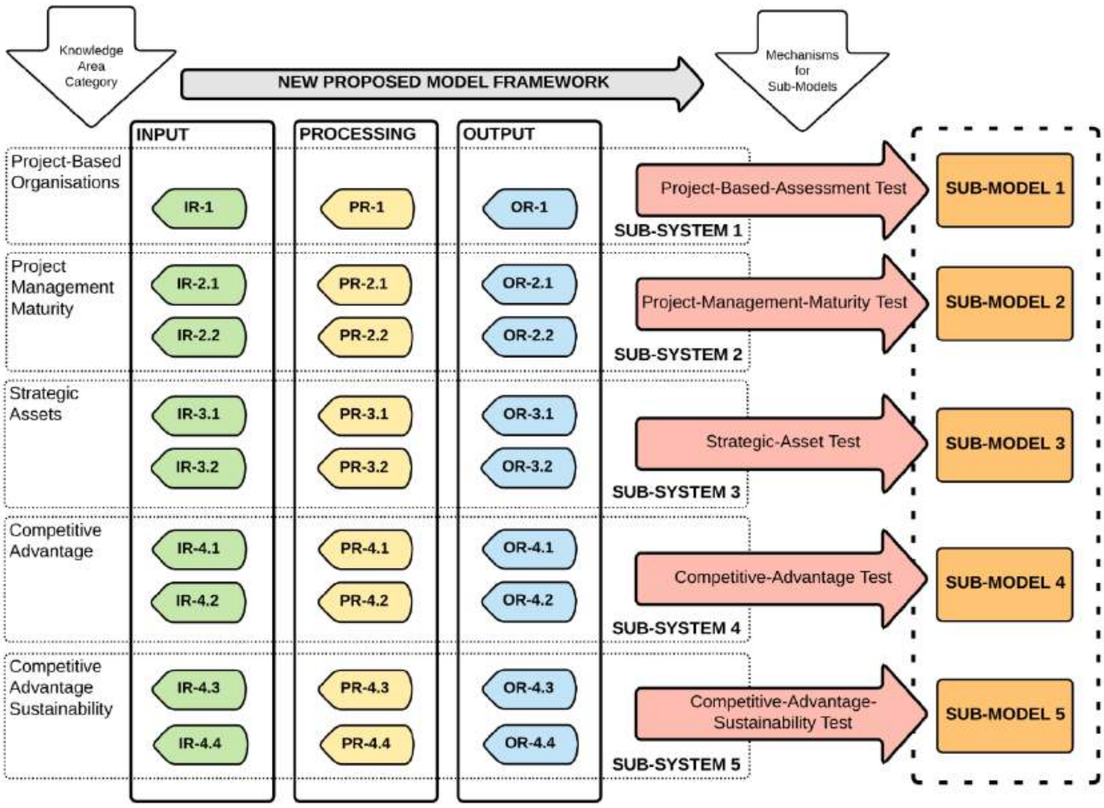


Figure 1: Proposed Model Framework

3. INITIAL MODEL DESIGN

Knowing the detail of how the design requirements will be met, through the various *sub-models* one to five, and by applying the framework, shown in Figure 1, the detail of the model was designed by taking on a break-down approach, shown in Figure 2.

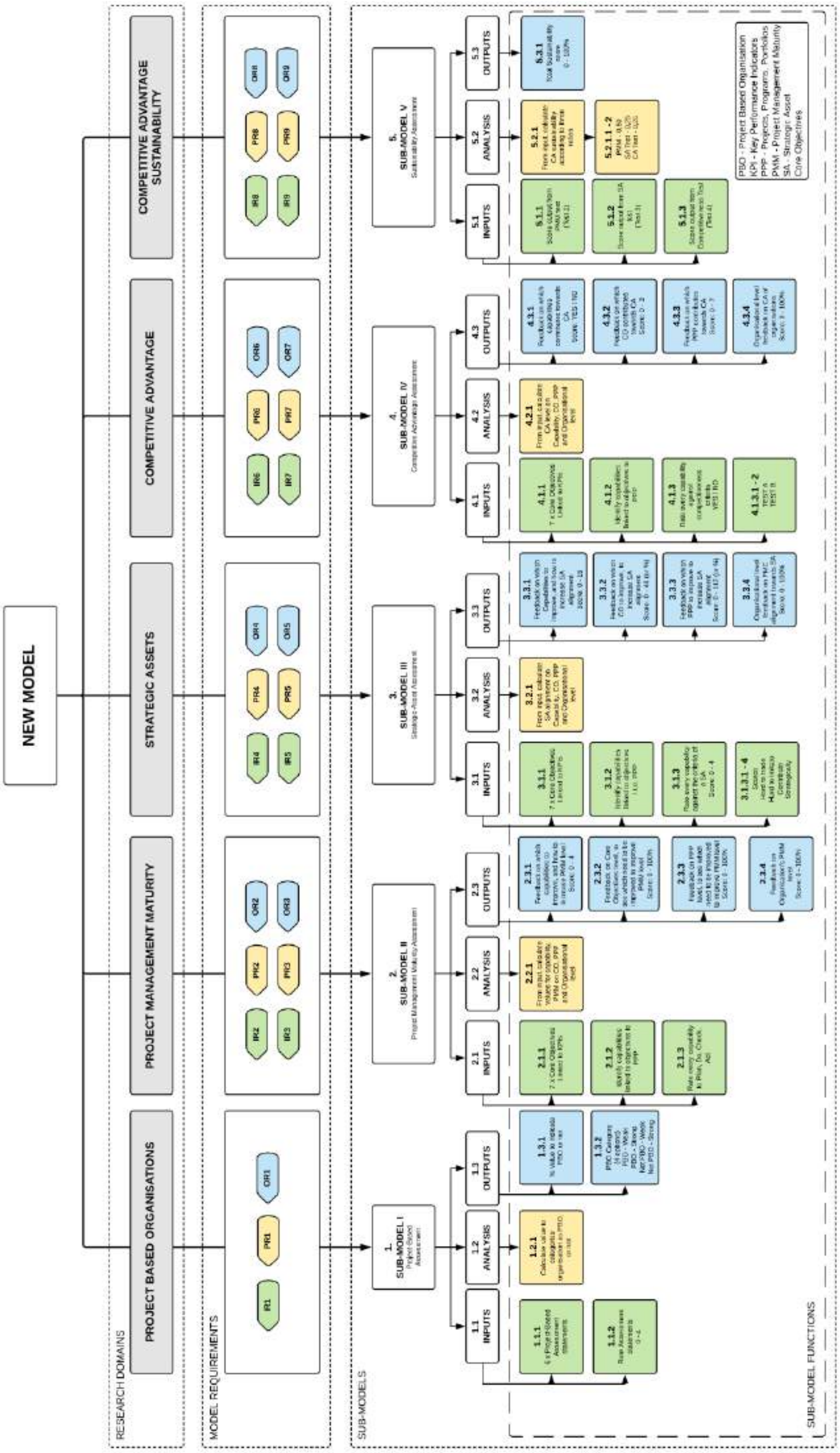


Figure 2: New Model Component Layout

The underlining detail of the subcomponents mentioned above were then sequenced into a work flow which forms the inner working of the proposed initial model (Rev_00). See Figure 3.

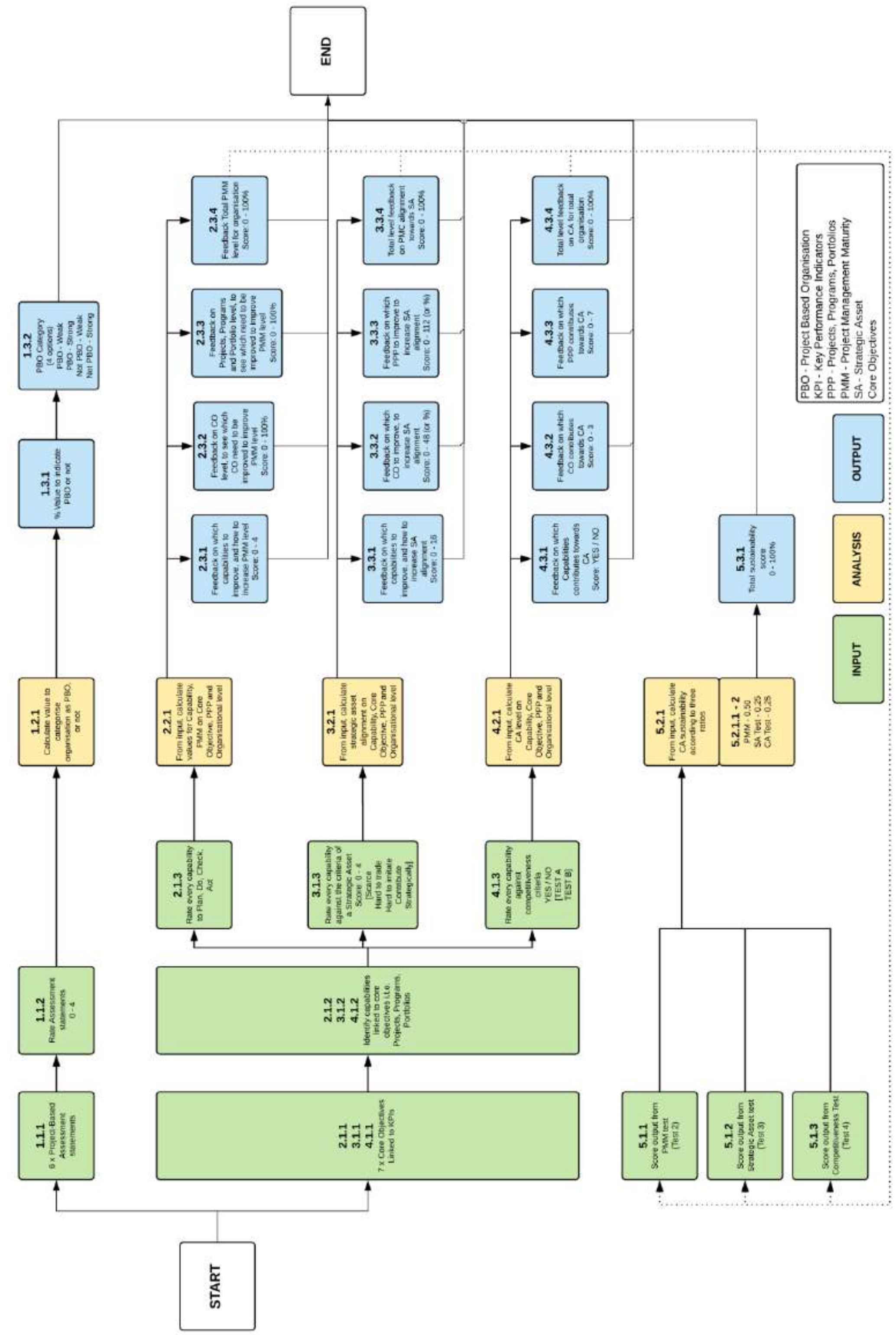


Figure 3: New Model Workflow

From what is shown above, it can be seen that for the model to function, the following input values are required:

1. Rate the organisation according to the *Project Based Assessment Statements*.
2. Identify seven core objectives linked to the *Key Performance Indexes* and *Project Management Knowledge Areas*, as part of the *Project Management Maturity Model*.
3. Identify the *Project Management Capabilities*, linked to each core objective in the context of the Project Management domains, i.e. *Projects, Programs* and *Portfolios*.

4. Rate every *Project Management Capability* in terms of the framework of *Plan, Do, Check, Act*, as part of the *Management Maturity Model*.
5. Rate every *Project Management Capability* against the criteria of being a strategic asset, i.e. *scarce, hard to trade, hard to imitate and adding value strategically*.
6. Rate every *Project Management Capability* against the criteria of sustainable competitive advantage, i.e. *perform different activities, or performing similar activities in different ways than industry competitors*.

By applying these inputs to the proposed model, the following outputs are generated:

1. ***Project Based Organisation Score*** - Quantitative value on how strong the organisation can be categorised as a *Project Based Organisation*. (0 – 100%).
2. ***Project Management Maturity Score*** - Organisation wide-, Domain Level-, Core Objectives-, and Capability Level- quantitative feedback on the *Project Management Maturity* level. (0 – 100%)
3. ***Strategic Asset Alignment Score*** - Organisation wide-, Domain Level-, Core Objectives-, and Capability Level- quantitative feedback on the Strategic Asset Alignment. (0 – 100%)
4. ***Competitive Advantage Score*** - Organisation wide-, Domain Level-, Core Objectives-, and Capability Level- quantitative feedback on how competitive the organisation is in the industry it operates in. (0 – 100%)
5. ***Competitive Advantage Sustainability Score*** - Organisation wide quantitative feedback on the *Sustainability* of the *Competitive Advantage*. (0 – 100%)

4. INITIAL MODEL VERIFICATION

After the initial design of the proposed model, the model was shown and explained to academic experts. The reason for this is twofold: firstly to gain constructive and informed feedback on the construction, usability and academic integrity of the model, and secondly to identify areas of improvement.

For this task five academic experts were chosen based on the following criteria: They have to be established academic professionals, who are able to add value by supplying constructive feedback in the knowledge areas of *Project Management* and *Strategic Management* and should also have prior knowledge and (or) experience in leading research or doing research themselves with regards to *new model*- or *framework development* in the academic environment. The interviews were held as semi structured interviews. The credentials of the academic experts taking part in this study is shown in Table 2.

Table 2: Academic Experts Credentials

	Academic Expert	Credentials
1.	**Dr. John Morrison**	PhD, Engineering, Project Management
2.	**Prof Herman Steyn**	PhD and Professor in Project Management
3.	**Dr Dirk Le Roux**	PhD Project Management, Information Technologies, IT strategy development, Project Portfolio Management, Programme Management
4.	**Dr Gerhard Ungerer**	PhD Industrial Engineering, Digital Enterprise Strategy, Digital Enterprise Strategist, Management Consultant
5.	**Prof Carl Marnewick**	PhD and Professor in Project Management

From the interview outputs, is was clear that a need for such a model does indeed exists as well as the fact that the model was based and developed on sound academic principles. From the interviews, the following updates to the model were proposed:

- **Proposed update 1**: As part of the output of the model, supply a high level (aggregated) feedback, giving a total combined model score for the assessment done.
- **Proposed update 2**: Give maturity feedback in terms of a five point scale, in line with other Project Management methodologies and maturity models.
- **Proposed update 3**: On the first iteration of the model, the total values were equally divided into the outputs of the three *Project Management Domains*, (Projects, Programs and Portfolios). As a result of organisational priorities and developments differences, it was proposed that the organisation under assessment gets to weight the relative importance of the domains to normalise feedback given. It was therefore decided to give output of the model in two sets, unweighted feedback (for external benchmarking) and weighted feedback (for internal benchmarking).

- **Proposed update 4**: It was further proposed that summary level feedback is also given summarising the output for the model in graphs, highlighting the areas of improvements.

These updates were then incorporated into the second iteration of the proposed model (model Rev_01).

5. CASE STUDY VALIDATION

One of the biggest advantages of applying case study validation it that a variety of sources can be applied to test a new framework or methodology. With the new model now develop, it is proposed that the model be tested by means of a case study validation process. Thomas [46] defines a *case study* follow:

> *"Case study is an in-depth exploration from multiple perspectives of the complexity and uniqueness of a particular project, policy, institution, program or system in a "real life" context."*

In the same paper, George and Bennet [47] notes that six types of case studies exist. Each with its own purpose within the research process.

- *Atheoretical / configurative idiographic*: Case studies supporting studies or research not contributing to theory;
- *Disciplined configurative*: Case studies applying theories to explain the case or phenomenon;
- *Heuristic*: Case studies applied to identify new causal paths;
- *Theory testing*: Case studies being applied to test the validity and scope conditions of single or competing theories;
- *Plausibility probes*: Case studies being applied to do preliminary studies in the process of determining whether further studies are warranted; and
- *Building block*: Case studies being applied to identify common patterns or investigate whether a new model or framework is fulfilling its purpose.

After the second iteration of the model was finalised, the model was validated by testing it on real life organisations. The purpose of the validation was therefore to test whether the new proposed model is satisfying its purpose, *building block case study* validation was therefore applied for this purpose. For the validation process, four companies were chosen to take part on the basis of the following criteria:

- The companies had to be local and accessible to conduct both the assessment and the questionnaire after the assessment;
- The companies had to be small to medium size organisation, with the researcher having direct access to management and/or someone with insight into strategic management process or the company;
- The companies had to operate in the project environment;
- The companies had to be from various industries; and
- The companies had to be established, in both business and product offering, in the industries they operate in.

By applying these criteria, four companies were identified, as shown in Table 3 below.

Table 3: Case Study Organisation Details

	COMPANY	CONTACT PERSON	ROLE	INDUSTRY	TURNOVER /YEAR	BASED
1.	Company November	Mr. CB	Project Director	Industrial Manufacture, Marine	R 600 Million	Cape Town
2.	Company Whiskey	Mr. WVN	Owner & Managing Director	Construction	R 25 Million	George
3.	Company Mike	Mr. JVW	Owner & CEO	Industrial Manufacture	R 120 Million	Cape Town
4.	Company Betha	Mr. FD	CEO	Control Systems, Mining	R 100 Million	Stellenbosch

The model was applied to the four companies mentioned above, after which the output of the assessments was discussed with the organisations. Due to the length limitations to this paper, only the organisational level output of the model (as applied to the four companies) is shown in Table 4 below.

Table 4: Organisational Level Model Feedback

	COMPANY	WEIGHTED / UN-WEIGHTED SCORE (PPP)	MATURITY INDICATOR	STRATEGIC ASSET ALIGNMENT	COMPETITIVE ADVANTAGE INDICATOR	COMPETITIVE ADVANTAGE SUSTAINABILITY INDICATOR
1.	November-Co	Un-weighted	29% - LEVEL 2	69%	81%	36%
		Weighted	33% - LEVEL 2	67%	77%	31%
2.	Whiskey-Co	Un-weighted	37% - LEVEL 2	63%	71%	70%
		Weighted	41% - LEVEL 2	65%	71%	73%
3.	Mike-Co	Un-weighted	31% - LEVEL 2	69%	76%	48%
		Weighted	36% - LEVEL 2	63%	71%	55%
4.	Betha-Co	Un-weighted	19% - LEVEL 1	52%	38%	26%
		Weighted	21% - LEVEL 2	52%	32%	21%

These results, as well as the detail from which it is made up from, were shared with the key contact persons in the various organisations who took part in the validation case studies. It was said earlier in this paper that the need for a model exists which can assist project-based organisations to assess the alignment of their Project Management Capability as a strategic asset, in order to ensure sustainable competitive advantage. From applying the proposed model to the cases mentioned above, and gaining the listed output, it can therefore be said that the new proposed model is validated, i.e. the model is functioning as it is intended to.

After the assessments were completed, the interviewees were asked to take part in a further verification questionnaire in which the design requirements were tested against the model outputs. Nine verification questions were developed and asked, to test the 27 design requirements. The detail of which is shown in Table 5 below.

Table 5: External Verification Questionnaires

DESIGN REQUIREMENT		VERIFICATION QUESTION	
Project Based Organisation			
1	OR1	VQ-A1	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant output as to what extent your organisation can be categorised as a PBO.
2	PR1		
3	IR1		
Project Management Maturity			
4	OR2	VQ-B1	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on the maturity of your organisation's PMC.
5	PR2		
6	IR2		
7	OR3	VQ-B2	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on how the organisation's PMM can be improved by referring to the improvement of specific Project Management Capabilities.
8	PR3		
9	IR3		
Strategic Assets			
10	OR4	VQ-C1	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on closely the organisation's PMC is aligned towards being a strategic asset.
11	PR4		
12	IR4		
13	OR5	VQ-C2	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output,
14	PR5		

15	IR5		or feedback, on how to improve the organisation's PMC towards being more aligned towards being a strategic asset for your organisation.
Sustainable Competitive Advantage			
16	OR6	VQ-D1	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on how competitive the organisation's PMC is within the industry you are operating in.
17	PR6		
18	IR6		
19	OR7	VQ-D2	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on how organisations can improve its PMC to be more competitive in the industry it operates in.
20	PR7		
21	IR7		
22	OR8	VQ-D3	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on how sustainable the PMC competitiveness is within the industry you are operating in.
23	PR8		
24	IR8		
25	OR9	VQ-D4	The model is able to guide you through the process of providing relevant data to do an assessment and give relevant quantitative output, or feedback, on how to improve the sustainability of the PMC competitiveness within the industry you are operating in.
26	PR9		
27	IR9		

The results from the external verification questionnaires (Table 6) show conclusively that the all the design requirements are met and that the model can be classified as externally verified.

Table 6: Verification Questionnaire Results

		Validators Responses					
	Validation Questions	Mr. CB (November)	Mr. WV (Whiskey)	Mr. JVW (Mike)	Mr. PB (Betha)	Average (%)	Result (Rank)
1.	VQ-A1	4	4	4	3	94%	1
2.	VQ-B1	3	3	3	4	81%	4
3.	VQ-B2	4	4	4	2	88%	3
4.	VQ-C1	3	4	4	3	88%	3
5.	VQ-C2	4	4	4	3	94%	2
6.	VQ-D1	4	3	4	3	88%	3
7.	VQ-D2	3	4	4	2	81%	4
8.	VQ-D3	4	4	3	3	88%	3
9.	VQ-D4	3	4	4	2	81%	4
	Average:	83%	80%	78%	90%	87%	

5.1 Further observations after testing the final model

The following observations were made after the case study tests were concluded:

1. It was observed that when it came to applying the Capability Competitive Ranking (model input), the more specialised industries found it hard to rank the capabilities against other industry role players. It is therefore proposed that if this data cannot be easily generated by the organisation under assessment, that the consultant do a further study into the industry competitors to supply information regarding the industry competitors.

2. In the process of scoring capabilities for the *Project Management Maturity* assessment, it was observed that the model takes in consideration capacities not yet in use (in other words to "aspire to") as well as what will the actions be to improve that capability until the next round of assessment. Currently, the model output only shows maturity level 1 – 4, which means that if a capability is score 0 on in the maturity continuum, that capability will show a maturity of Level 1 (and not 0), which can confuse the interpretation of that. The fact that the maturity level, or maturity score, is also given in terms of a percentage value, should solve a larger part of this problem.

3. In the process of gathering data for the input of the model, it was seen that because most of the organisations used in these case studies were relatively low developed in its *Project Management Capability*, there were no defined or structured *Program* or *Portfolio* functions in these organisations. This lead assessment challenges in terms of gaining data about the organisation's *programs* and *portfolios*. The solution to this challenge is directed at the quality of the facilitation process. It was seen that a good deal of time have to be spent on defining what "programs" and "portfolios" mean for these organisations. One way to get around this challenge is to score the weighting of these two functions lower (or in some cases even zero), therefore the "un-calculated" data given for these scores will not have such a large influence on the output data.

4. One of the high-level requirements of the model was that it had to be flexible and should be able to be adapted to the organisation under assessment. The organisations assessed for these case studies were very diverse, both in its industry, size and level of overall development. The model took that into consideration and supplied good output as to adding value specifically to the organisation under consideration.

6. CONCLUSION

This paper explored the research proposition that a framework can be developed and that a model can be designed according to this framework, to evaluate the contribution, strength or alignment of organisation's *Project Management Capability* as a *strategic asset*, and to use this information to assure *sustainable competitive advantage* for the organisation.

By applying a needs analysis to the research problems statement, a set of 27 design requirements were developed which were scrutinised by means of a rigorous literature study to form five *sub-models*, each addressing a set of design requirements. The five *sub-models* were then entered into a proposed framework for the model to form the first iteration of the proposed model.

As a means verification, five academic experts were asked to take part in the study by giving insights into the academic integrity of the study as well as to propose updates which will form part of the final iteration of the proposed model. With model being verified and updated according to the academic experts, the final iteration, Rev_01, were developed and validated externally by applying it to four organisations as real-life case studies. The four case study assessments also served as an external verification to assess if all design requirements were met.

From the feedback from the academic experts, the case study validation and the external verification, it can conclusively be said that the design requirements are met and that the model does indeed solve the research problem statement.

REFERENCES

[1] **Thiry, M. and Deguire, M.** 2007. Recent developments in project-based organisations. *International Journal of Project Management*, 25(7), pp 649-658.

[2] **Morris, P.W.** 2013. *Reconstructing project management*, John Wiley & Sons.

[3] **Bollinger, A.S. and Smith, R.D.** 2001. Managing organizational knowledge as a strategic asset. *Journal of knowledge management*, 5(1), pp 8-18.

[4] **Archibald, R.D. Peppard, J. Ward, J., 2003**. The Importance and Value of Project Management for Enterprises and Institutions. *Project Management Conference proceedings – Bogota, Colombia*, pp 1-17.

[5] **Kerzner, H.R.** 2001. *Planning for Project Management Maturity Model*, New York: John Wiley & Sons.

[6] **Wiewiora, A. Trigunarsyah, B. Murphy, G.D. Gable, G.G. and Liang, C.** 2009. *The impact of unique characteristics of projects and project-based organisations on knowledge transfer*.

[7] **Kerzner H.R. and Saladis FP.** *Value-driven project management*. John Wiley & Sons; 2011.

[8] **Thomas, J. and Mullaly, M.** 2008. *Researching the value of project management*. Project Management Institute.

[9] **Kwak, Y.H. and Anbari, F.T.** 2006. Benefits, obstacles, and future of six sigma approach. *Technovation*, 26(5), pp 708-715.

[10] **Charvat, J.** 2003. *Project management methodologies: selecting, implementing, and supporting methodologies and processes for projects*. John Wiley & Sons.
[11] **Willis, B.E.** 1995. APM project-management body of knowledge: the European view. *International Journal of Project Management*, 13(2), pp 95-98.
[12] **Snyder, C.S.** 2014. *A Guide to the Project Management Body of Knowledge: PMBOK* (®) Guide. Project Management Institute.
[13] **Kwak, Y.H. and Ibbs, C.W.** 2002. Project management process maturity (PM) 2 model. *Journal of management in engineering*, 18(3), pp 150-155.
[14] **Gareis, R. and Huemann, M.** 2000. *Project management competences in the project-oriented organisation*. The Gower handbook of project management. Gower: Aldershot, pp 709-721.
[15] **Atkinson, R.** 1999. Project management: cost, time and quality, two best guesses and a phenomenon, it's time to accept other success criteria. *International journal of project management*, 17(6), pp 337-342.
[16] **Cleland, D. and Ireland, L.** 2002. *Project management: Strategic design and integration*.
[17] **Kelchner, L.** 2017. *The Importance of Organizational Capability* | Chron.com. 2017. The Importance of Organizational Capability | Chron.com. [ONLINE] Available at: http://smallbusiness.chron.com/importance-organizational-capability-13295.html. [Accessed 20 September 2017].
[18] **Scott, J.** 2014. *Putting Business Capabilities to Work*. [ONLINE] Available at: https://www.omg.org/news/member-news/OMG-Putting-Cap-To-Work.pdf [Accessed 12 June 2018].
[19] **Peppard, J. and Ward, J.** 2004. Beyond strategic information systems: towards an IS capability. *The Journal of Strategic Information Systems*, 13(2), pp 167-194.
[20] **Langston, C.** 2013. Development of generic key performance indicators for PMBOK using a 3D project integration model, *Australasian Journal of Construction Economics and Building*, 13(4), pp 78-91.
[21] **Cooke-Davies, T.** (2002) The 'real' success factors on projects, International *Journal of Project Management*, 20(3), pp 185-190.
[22] **Backlund, F., Chronéer, D. and Sundqvist, E.** 2014. Project management maturity models - a critical review: a case study within Swedish engineering and construction organisations, *Procedia Social and Behavioural Sciences*, 119, pp 837-846.
[23] **Jugdev, K. and Thomas, J.** 2002. *Project management maturity models: the silver bullets of competitive advantage*, Project Management Institute.
[24] **Langston, C. and Ghanbaripour, A.N.** 2016. A Management Maturity Model (MMM) for project-based organisational performance assessment. *Construction Economics and Building*, 16(4), pp 68-85.
[25] **Artto, K., Martinsuo, M., Gemünden, H.G. and Murtoaro, J.** 2009. Foundations of program management: a bibliometric view, *International Journal of Project Management*, 27(1), pp 1-18.
[26] **Young, M. Young, R. and Romero Zapata, J.** 2014. Project, programme and portfolio maturity: a case study of Australian Federal Government, *International Journal of Managing Projects in Business*, 7(2), pp 215-230.
[27] **Hardgrave, B.C. and Armstrong, D.J.** 2005 Software process improvement: it's a journey, not a destination, *Communications of the ACM*, 48(11), 93-96.
[28] **Pane, E.S. and Sarno, R.** 2015 Capability Maturity Model Integration (CMMI) for optimising object-oriented analysis and design (OOAD), *Procedia Computer Science*, 72, 40-48.
[29] **Staples, M. Niazi, M. Jeffery, R. Abrahams, A. Byatt, P. and Murphy, R.** 2007. An exploratory study of why organizations do not adopt CMMI. *Journal of systems and software*, 80(6), pp 883-895.
[30] **De Souza, T.F. and Gomes, C.F.S.** 2015. Assessment of maturity in project management: a bibliometric study of main models. *Procedia Computer Science*, 55, pp 92-101.
[31] **mit, R. and Schoemaker, P.J.H.** 1993. Strategic assets and organizational rent. *Strategic Management Journal*, 14(1), pp 33-46.
[32] **Frame, J.D.** 2002. *The new project management: tools for an age of rapid change, complexity, and other business realities*. John Wiley & Sons.
[33] **Porter, M.E.** 2008. The five competitive forces that shape strategy. *Harvard business review*, 86(1), pp 25-40.
[34] **Ma, H.** 1999. *Anatomy of competitive advantage: a SELECT framework. Management decision*, 37(9), pp 709-718.
[35] **Porter, M.E.** 1980. *Competitive Strategy*, Free Press, New York, NY.
[36] **Barney, J.B.** 1991. *Firm* resources and sustained competitive advantage, *Journal of Management*, Vol. 17, pp 99-120.
[37] **Prahalad, C.K. and Hamel, G.** 1990. The core competence of corporations, *Harvard Business Review*, May-June, pp 79-91.
[38] **Goodstein, L.D., Nolan, T.M. and Pfeiffer, J.W.** 1993. *Applied strategic planning: How to develop a plan that really works*. McGraw Hill Professional.
[39] **Nonaka I.** 1991. The knowledge-creating company. *Harvard Business Review*, 6(8), pp 96-104.
[40] **Pemberton, J.D. and Stonehouse, G.H.** 2000. Organisational Learning and Knowledge Assets – an essential partnership. *The Learning Organization* 7(4), pp 184-193.
[41] **Stonehouse GH, Pemberton JD.** 1999. Learning and Knowledge Management in the Intelligent Organisation. *Participation and Empowerment: an International Journal* 7(5), pp 131-144.

[42] **Pemberton, J.D. Stonehouse, G.H. and Yarrow, D.J.** 2001. Benchmarking and the role of organizational learning in developing competitive advantage. *Knowledge and Process Management*, 8(2), pp 123-135.

[43] **Jugdev, K.** 2003. *Developing and Sustaining Project Management as a Strategic Asset: A Multiple Case*. Study Using the Resource-Based View. Unpublished doctorial dissertation. Calgary, Canada.

[44] **Porter, M.E.** 1996. What is strategy? *Harvard Business Review*, 74(6), pp 61-78.

[45] **Barney, J. B.** 2002. *Gaining and sustaining competitive advantage* (2nd ed.). Upper Saddle River, New Jersey: Prentice-Hall, Inc.

[46] **Thomas, G**. 2011. *A typology for the case study in social science following a review of definition, discourse, and structure. Qualitative inquiry* 17.6: 511-521.

[47] **George, A. L. and Bennett, A.** 2005. *Case studies and theory development in the social sciences.* Cambridge: MIT Press.

* References used in background research in the process of developing the various sub-models and design criteria.

INDUSTRY 4.0: A MYTH OR A REALITY IN SOUTH AFRICA?

M.T. Dewa[1,1], **D.Q. Adams**[2], T.G.**Tendayi**[3], **L. Nyanga**[4], **M. Gxamza**[5] **& L. Ganduri**[6]

[1-6] Department of Industrial and Systems Engineering
Cape Peninsula University of Technology, South Africa
[1]dewam@cput.ac.za, [2]adamsdq@cput.ac.za, [3]tendayit@cput.ac.za, [4]nyangal@cput.ac.za,
[5]gxamzam@cput.ac.za, [6]ganduril@cput.ac.za

ABSTRACT

There has been much discussion on the subject of the fourth industrial revolution, also known as Industry 4.0, over the past decade in academia, industry and media. It is generally agreed and documented that the shift towards Industry 4.0 started in Germany and rapidly spread to the rest of the developed economies such as the United States and Japan. What has received little attention or documentation, is the use of Industry 4.0 in developing economies such as South Africa. This paper attempts to shed more light on the state of preparedness of South African companies to embrace and implement Industry 4.0. A comprehensive literature study is conducted, from which the critical technological framework for Industry 4.0 is outlined. From this information, an empirical study is then performed through a survey of carefully selected companies in South Africa. From this survey, the preparedness of the investigated firms with regards to awareness and implementation of Industry 4.0 is identified. The paper then concludes with recommendations on the way forward if South African companies are to ensure that they don't get 'left behind the train' of Industry 4.0.

[1]The author is a Lecturer in the Industrial and Systems Engineering Department at the Cape Peninsula University of Technology, South Africa
[1]Corresponding Author

1. INTRODUCTION: BACKGROUND TO INDUSTRY 4.0

The pursuit of every professional industrial engineer is grounded in three main ideas, the idea that humankind throughout the ages man has searched to improve in all areas of their lives, personal, professional and spiritual. The second idea is that of doing things in the most optimal methods with what we have, and lastly the notion of searching for ways to do things better, faster and smarter. The world is changing and there is need to keep up to pace with the fast growing and rapidly changing environments.

This paper focuses on Industry 4.0 (I.4.0). The emergence of the I.4.0 concept began in Germany. As an introduction to the overarching topic, this section serves to shed light on the thinking of I.4.0. The German federal government first initiated “Industrie 4.0” as a strategic initiative that was adopted as part of the “High-Tech Strategy 2020 Action Plan” in November 2011 [1]. The concept started as a consolidated push to improve the countries manufacturing sector. Hermann et al. [2] examines this work and gives the topic a defined area with which to work within and aims to create design criteria for other researchers to follow on developing the concept further. According to the Hermann et al. [2], the increasing integration of the internet and connectivity into the industrial value chain has set the foundation for the next industrial revolution. Kiel [3] also highlighted that the fourth industrial revolution is a present hour reality built on the idea of connecting many things together and have them work in perfect balance. Fundamentally humankind is searching for the next improvement and Hermann et al. [2] believe it will come in the form of connection of humans, machines and processes. A revolution can be defined as the major change to the way thinking and doing something, or a substantial shift in the process of performing tasks. This can happen in small jumps or it can happen in the form of big jumps. It is often assumed that revolutions must come as large explosions. However, a revolution can be in the form of big ideas, and small tasks and improvements getting one closer and closer to that Big Idea. The idea and background of I.4.0 came from the grounding in previous industrial revolutions, as seen in Figure 1 below. Starting from water, steam and machines, to electricity and mass production, finally into with the dawn of Computers and the digital age of IT and machine automation.

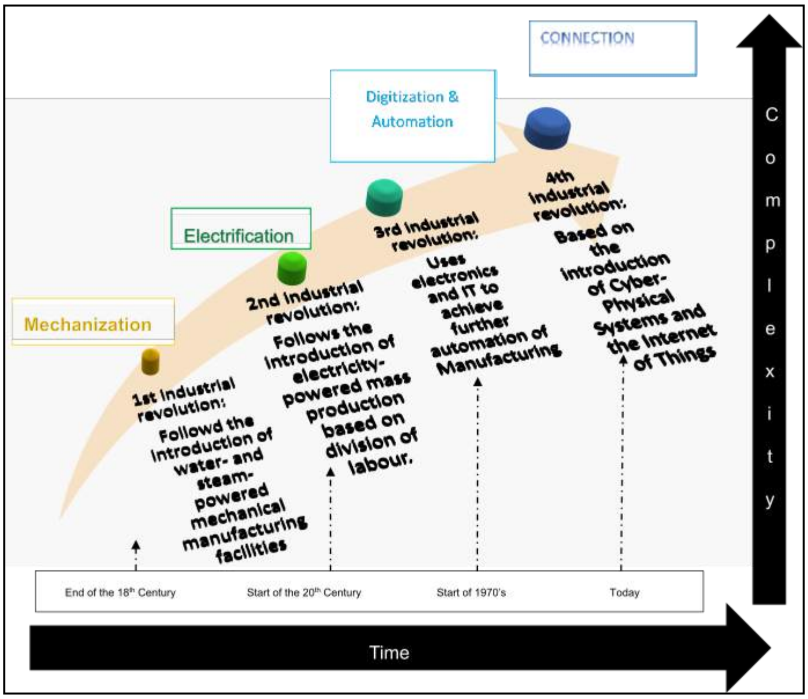


Figure 1: Overview of industrial revolutions: Source, Own Illustration based off [1]

1.1 Local relevant literature for South Africa

An investigation of available literature relevant to the South African context was conducted in order to identify what other authors are saying about the readiness of South Africa (SA) in terms of I.4.0. The literature covered focused on studies conducted in South Africa. Most work done to date refers to teaching, learning and education creation and development of frameworks within a learning factory environment [4]. Some reports on the subject were done for the Denel, Food and Beverage, Plastics, Construction, and Mining industries. These reports are published online in magazine articles. Calitz, Poiset and Cullen [5] do an analysis of the automotive sector of SA when surveying the field of 13 businesses and respondents provide positive signs that indeed the strategies and thinking of using collaborative robots is being investigated and indeed 2 working types can be seen. Other respondents provide confirmation of proof of concept machines being tested.

Dlodlo [6] does an analysis of rural needs in Zambia and SA,[7] and how the internet of things (IoT) can help alleviate some of the challenges by injecting productivity and growth in the agriculture sector. A special report published by Deloitte [8] on trying to ascertain if Africa ready for a "digital transformation", looks broadly at various companies in the automotive and manufacturing sector on the merging of real and virtual worlds. In the paper, we seek to provide a unique local perspective of the concept I.4.0. Empirical evidence for the South African manufacturing sector remains lacking in the literature, hence leaving a gap for potential researchers to explore. This also leaves businesses and reasearchers unable to gauge where SA as a nation stands with regards to the I.4.0 concept.

There seems to be a an ongoing research process uncovering the description and meaning of I.4.0 for SA as mentioned in the evidence above. However, researchers and practitioners still need to select and explore specific scenarios where I.4.0 can be implemented. In addition to this, a uniquely South African perspective on the readiness of SA for I.4.0 is yet to be determined. The paper will focus specifically on looking at the enabling technologies used in manufacturing firms, in order to establish an assessment tool in order to gain quantitaitve understanding of the current state of I.4.0 within the SA manufacturing industry.

2. RESEARCH METHODOLOGY

2.1 Aim, scope and research questions

The primary aim of the research was to develop a tool with which to assess whether the South African industry is ready for I.4.0 concept. The objectives were to identify Industry 4.0's different dimensions (critical success areas) that are being utilised in available work and the awareness of industry with state-of-the-art enabling technologies to facilitate this concept. Furthermore, another objective involved coming up with a method of rating and assessing companies in different sectors based on the parameters derived from identified dimensions earlier on. To fulfil the above aim, a three-point methodology was followed: (i) Literature search, to study Industry 4.0 and understand its practical aspects and requirements, (ii) Data collection from Industry focusing on Industry 4.0 infrastructure development and future plans, and (iii) Design conceptualisation of a tool (or assessment strategy). The outcome of the literature search was that there are certain crucial enabling I.4.0 technologies that need to be set in place before the concept can be developed further upon. Secondly, in order to precipitate this concept a tool for assessment of the readiness was needed. Warwick University [25] outlined a set of 6 Dimensions indicative pointers towards the areas upon which businesses should be rated.
The outcomes of these processes formed the basis for a survey of Industry partners to determine their orientation and progress toward Industry 4.0 development. The guiding research questions were:

1. What are the key Industry 4.0 technologies, concepts and features facing industry that need to be understood?
2. Are companies/industries aware of the Industry 4.0 concept?
3. How will organisations who are interested in the emerging I.4.0 know if they are ready or not?

The main focus was on the South African manufacturing sector as the assumption is that they would be most "ready" and 'aware' with regards to the I.4.0. concept, since it was initiated for the German manufacturing sector. However, a limitation encountered during the study was that of access to research participants. In most cases, it was a challenge for respondents to be open and share their companies' perspectives and action regarding the shift and strategic direction of thought. Another constraint encountered was that of the profile of the respondents participating in the study being limited to personnel in companies with alumni the department of Industrial and Systems engineering. The industry partner database of all companies where previous students have spent their 'in-service' year of work integrated learning module was utilised to identify these companies.

2.2 Method

A pilot study was conducted with 7 manufacturing companies, whereby senior level supervisors were consulted about the awareness of I.4.0 and use of digital technologies in their respective workplaces. The puropse of the pilot study was for refining the data collection instrument. The initial survey involved open-ended questions as described in Section 2.1. Using the open ended surveys and interviews, a readiness model was developed for the South African context. The aim f the model was measure and assess the companies' position with regards to implementing the I.4.0 technologies, methodologies, models and frameworks.

After the pilot study, a structured survey questionaire was sent out to 150 participants. The questionaire was designed in such a way that the participants would assess thier respective organisations for the purpose of determining readiness levels for the concept I.4.0. An examples of a question is illustrated in Figure 2 below.

Instructions:
For each of the 6 dimensions, use the readiness assessment criteria to identify your current level of readiness (0, 1, 2, 3 or 4) for each of the sub-dimensions.

An Outsider is a company that does not meet any of the requirements for Industry 4.0 of which Industry 4.0 is either unknown or irrelevant for them. A Beginner is a company which might not meet the requirements of Industry 4.0 but they are aware of it.

Figure 2: Survey Instructions

A five-point Likert scale was used in most cases for the responses to the questions for each dimension with 0 = outsider, 1 = beginner, 2 = intermediate, 3 = experienced and 4 = expert as illustrated in Figure 3.

Section B: Industry 4.0 Dimensions

Dimension 2 - Manufacturing Technologies

Technology	Level 0 Outsider	Level 1 Beginner	Level 2 Intermediate	Level 3 Experienced	Level 4 Expert	Company Level
Mobile devices (Cell phones or Tablets)	Never used	Rarely used	Often used	Very Often used	Always used	
Production or shop-floor Machines connected to the Internet	Never used	Rarely used	Often used	Very Often used	Always used	
Internet of Things platforms	Never used	Rarely used	Often used	Very Often used	Always used	
Radio Frequency Identification (RFID)	Never used	Rarely used	Often used	Very Often used	Always used	
Bar Code Technology	Never used	Rarely used	Often used	Very Often used	Always used	
QR Code Technology (add picture in appendix)	Never used	Rarely used	Often used	Very Often used	Always used	
Advanced Human-Machine Interfaces (HMI)	Never used	Rarely used	Often used	Very Often used	Always used	
3D Printers	Never used	Rarely used	Often used	Very Often used	Always used	
Smart/Intelligent Sensors	Never used	Rarely used	Often used	Very Often used	Always used	
Authentication and Fraud detection software	Never used	Rarely used	Often used	Very Often used	Always used	
Big Data Advanced Analytics	Never used	Rarely used	Often used	Very Often used	Always used	

Figure 3: Extract from survey questionnaire, Section B, Dimension 2

2.3 Scope of Study

The questionnaires were distributed to 150 participants across South Africa. 103 participants were from the Western Cape and 21 participants from Gauteng province, made up 83% of the participant pool. The companies were classified into ten main operating sectors, namely “Food & Agro-processing”, “Automotive”, “Chemicals”, “Information & Communication Technology”, “Electronics”, “Metals and Mining”, “Textile, Clothing & Footwear”, “Wholesale and Retail trade”, “Financial Services” and “Other” Which include “Consulting”, “Supply Chain Management”, “Aerospace” and “Research and Development”. Figure 4 below illustrates the percentage of firms who participated per province, while Figure 5 indicates the sub-sector percentage of firms which were involved in the study.

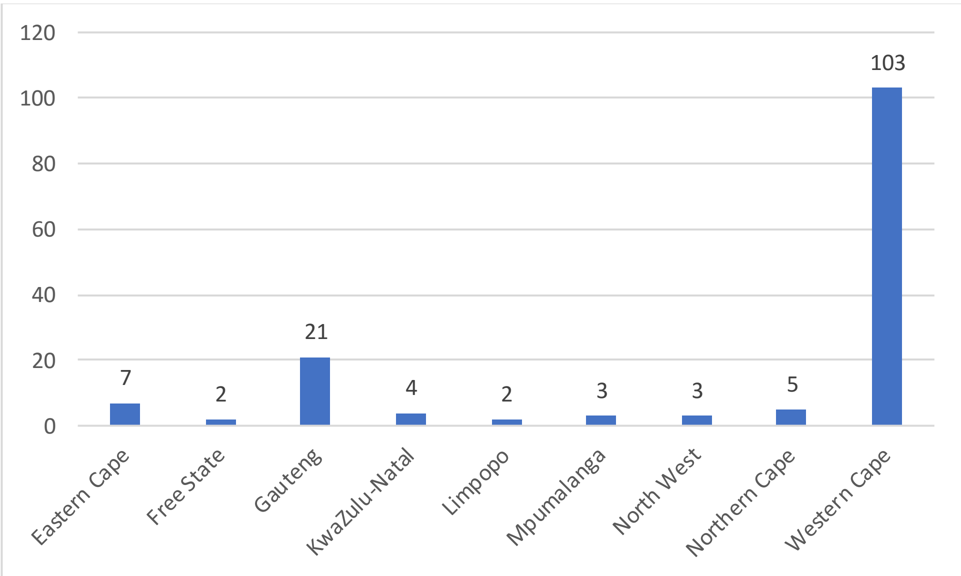


Figure 4: Company breakdown per province

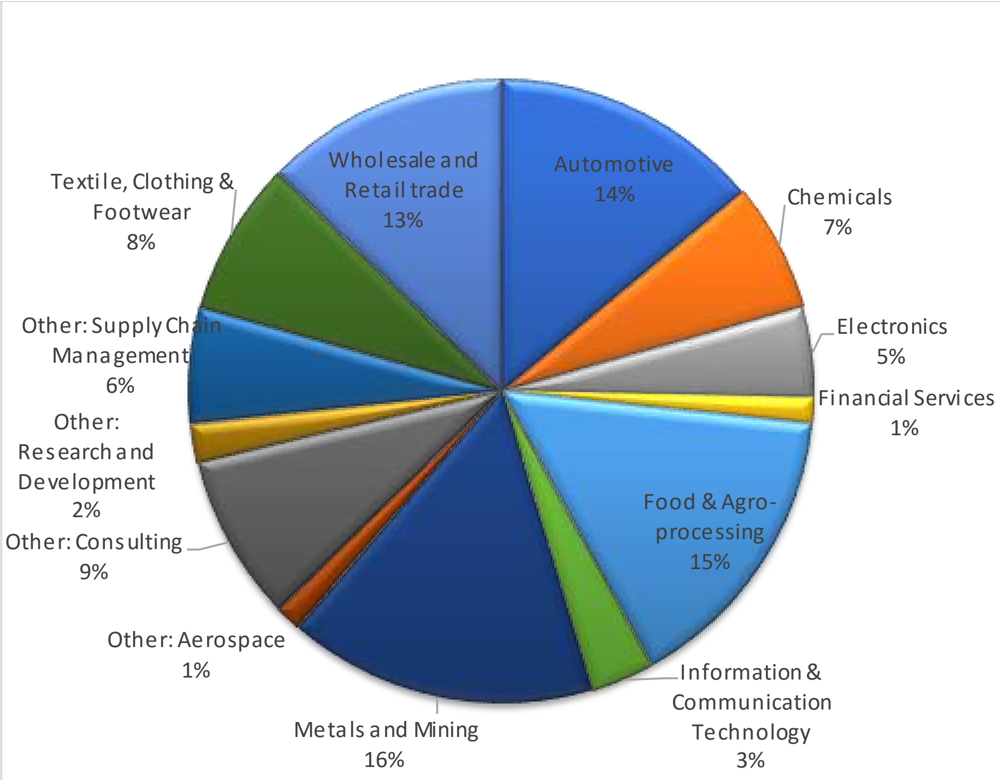


Figure 5: Sub sector percentage of firms

The companies visited were also grouped according to number of employees in the company with the following ranges; 0 - 50, 51 -100, 101 - 500, 501 - 1000 and more than 1000.

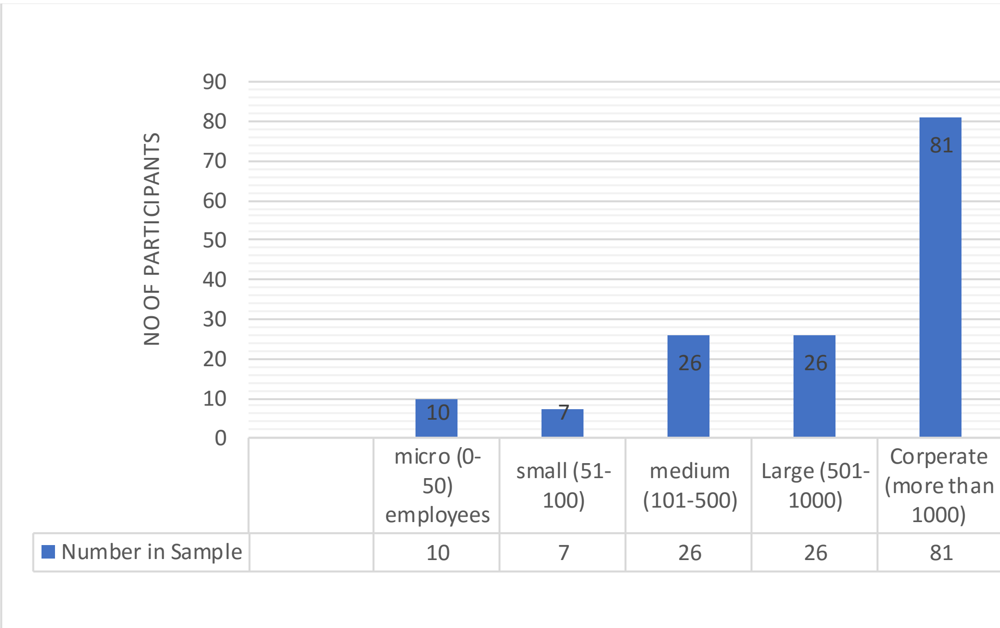


Figure 6: Company size data

The main group of participants were selected to be from a "Senior Management Level", 67%, with an excess of 10 years' experience in their given field. Details of the breakdown of the different levels of participants can be seen in Figure 7 below:

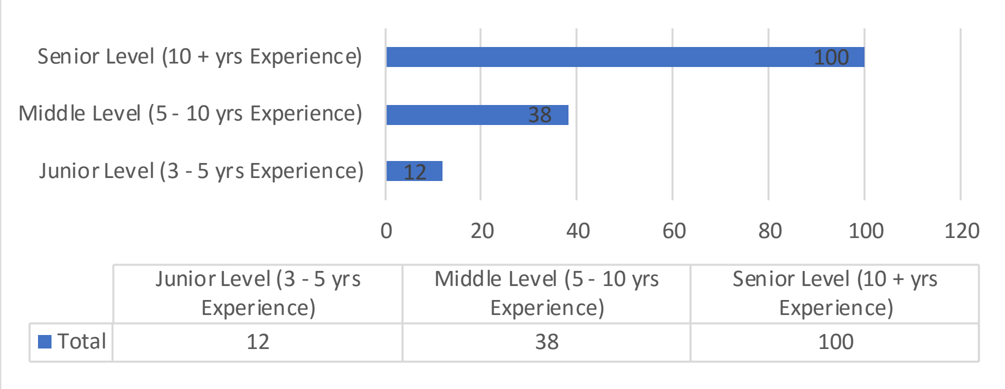


Figure **7: Management levels**

3. LITERATURE REVIEW

3.1 Search and data gathering strategies

The search strategy employed in the literature review involved first identifying the relevant data sources and key words. These sources included journals, conference proceedings, technical reports, articles from trade journals and company websites. The search field included a significant number of non-journal internet sources, and the search was conducted using a range of keywords and key phrases that could be relevant to the South African context and their potential impact on the manufacturing sector. Examples of these included, but were not limited to: "SA and Industry 4.0" and "Fourth industrial revolution impact on Africa".

3.2 Technology Readiness Levels

The National Aeronautical Space Administration introduced the technology readiness level (TRL) scale in the 1970s as a tool for assessing the maturity of technologies during complex system development. Since then, it has been used to make multi-million dollar technology management decisions [9]. The TRL scale has been embraced by the U.S. Congress' General Accountability Office (GAO), adopted by the U.S. Department of Defense (DOD) and other numerous organizations ([10],[11], [12] and [13]). The scale shown in Figure 8 the progression of a technology from a level where only basic principles have been observed and reported (TRL1) up to a level were the actual system has been proven through successful system and/or mission operations (TRL9). Most of the existing TRL models focus on specific criteria for manufacturing technologies and for concepts such as Industry 4.0 [14]. Examples of these models include Manufacturing readiness levels (MRL) created by the US Department of Defense [15], the Manufacturing capability readiness level (MCRL) created by Rolls-Royce [6] and the Manufacturing technology readiness level (MTRL) proposed by Peter [16].

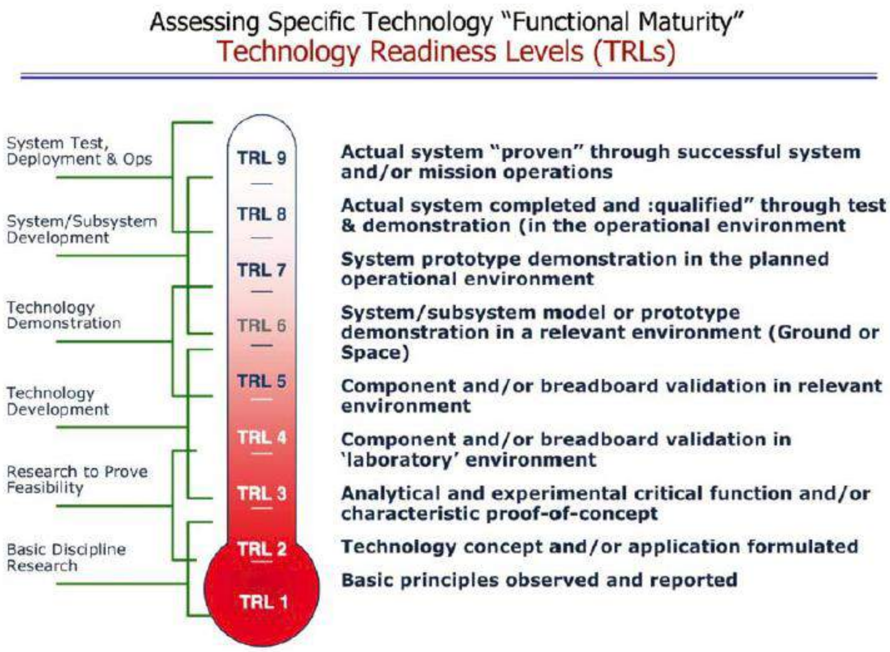


Figure 8 Overview of the technology readiness level scale [10][11]

A few models and tools for assessing readiness or maturity of I.4.0 have been proposed. Lichtblau et al [17] developed a self-check online tool which companies can use to measure their own Industry 4.0 readiness. The tool uses six dimensions of Industry 4.0 including 18 items to indicate readiness in 5 levels. The company's self-assessment profile is then benchmarked against the profile of leading Industry 4.0 companies and their target profiles. The barriers for progressing to the next stage are defined and advice is given on how to overcome them. Kagermann et al [18] have a model which focuses on the development of horizontal integration, in addition to vertical integration, through value networks and end-to-end digital integration across the entire value chain. They identified eight key priority areas and gave recommendations on how to implement them. Bitkom, VDMA and ZVEI [19] developed a reference architecture model for Industry 4.0 (RAMI 4.0) as a way of achieving a common understanding of what standards, use cases, etc. are necessary for Industry 4.0. The model contains key milestones focusing on enabling the companies to achieve horizontal integration via value creation networks over the entire life cycle. It also includes vertical integration, networked production systems, new social infrastructures for work and continual development of cross-sectional technologies. Pricewaterhouse Coopers [20] developed an online-self assessment with dimensions which focus on digital maturity in 4 levels. Unlike the model by Lichtblau et al [17], this assessment tool has 6 dimensions. Rockwell Automation [21] developed a five-stage approach to realise Industry 4.0 using a technology focused assessment in 4 dimensions. It gives no details about items and the development process offered.

Roland Berger Strategic Advisory Company developed a country ranking called the Roland Berger Industry 4.0 Readiness Index [22]. It comprises core categories such as Industrial Excellence and Value Network and sub categories such as production process sophistication, degree of automation and workforce readiness innovation intensity. The readiness index ranks European countries into four categories, namely forerunners, traditionalists, hesitators and potentialists. The Forerunners are characterised by a broad industrial base, modernized development-oriented business conditions and application of technologies. Traditionalists have a strong industrial base and still have a healthy structure. Hesitators on the other hand lack a reliable industrial base and are struggling, with government finances not being able to transform their economy. Finally, Potentialists have a former strong industrial base which has weakened during the recent years. Viharos et al [23] proposed a non- comparative Industry 4.0 readiness evaluation method that is fully personalised to the manufacturing company evaluated and independent of recommendations of international consulting firms. The Singapore Economic Development Board [24] developed a three building block, eight pillar and 16 dimension readiness index. Onur et al [25] developed a model which consists of 6 core dimensions and 37 sub-dimensions. Its purpose is to provide a simple and intuitive way for companies to start to assess their readiness and their future ambition to harness the potential of the cyber-physical age.

3.3 What other developing countries are doing

According to Gilchrist [26] some of the developing countries that are already preparing for and adopting strategies regarding Industry 4.0 are China and India. In addition to these countries, there are other developing countries such as Singapore, Thailand and Malaysia which are making inroads in this regard. A further distinction can be made between traditional developing countries such as SA, and more advanced developing countries like Singapore as explained by Wilson [27]. What follows is a brief discussion on some of the developments, together with challenges to implementation, that are taking place in these countries and the lessons that SA can draw from them.

Applications of Industry 4.0 go beyond the traditional manufacturing space such as in Singapore where the phenomenon was used to optimise resources within an Eco-industrial park as explained by Pan et al.[28]. In Malaysia, Barhin et al. [29] explore developments around robotic and automation product innovation using industry 4.0 principles. However, most of this innovation was found to be happening mostly amongst the big foreign companies based in Malaysia and not as much within the Small to Medium Enterprises who do not consider Industry 4.0 to be of great relevance to them. The Malaysian government is now actively engaged in addressing this through programmes such as the Eleventh Malaysia Plan.

A suggestion has been made by Zhou [30] for China to develop its own industrial development model which involves the parallel development of Industry 2.0, Industry 3.0 and Industry 4.0. The reason for this approach is based on the argument that it is unrealistic to expect that all industries can immediately achieve an Industry 4.0 upgrade. Thailand is another country that is going through a gradual transition to Industry 4.0 or what Jones and Pimdee [31] describe as Thailand 4.0 which they argue can be achieved by first investigating the legacy of Thailand 1.0, 2.0 and 3.0. Some significant obstacles to the implementation of Thailand 4.0/Industry 4.0 included what they called the middle income and inequality traps. It is the authors' premise that these challenges are indeed not unique to Thailand and can be observed in other developing countries as well.

It can be seen from the small sample of countries covered that Industry 4.0 has indeed taken off in the developing and so-called advanced developing countries. This has happened with varying degrees of success and with common obstacles to implementation being observed along the way. Valuable lessons can be learnt from the experiences of these countries as SA, which is also a developing country, embarks on the same journey.

3.4 Enabling Technologies

The Industry 4.0 framework, established by Reinhard et al. [32], is used in this paper to formulate the theoretical framework used. In this paper, different classes of currently available digital technologies, which can potentially improve manufacturing operations, are outlined. This model was specifically chosen due to its inclusion of a number of current technologies which fit into the I.4.0 framework.

3.4.1 Location detection technologies

With the Global Positioning System (GPS) technology freely and readily available nowadays, the tracking of object position is now possible. In the early 19th century, barcode technology was commonly employed to record and communicate locational information for objects. However, with the advancement of electronic gadgets, RFID technology is also growing in usage within the manufacturing sector [33]. These technologies allow Automatic Identification (Auto ID) and tracking of objects while collecting specific data of parts moving in an environment [33]. Location detection technologies like RFID devices are known to have many applications in supply-chain related problems, such as creating an Internet of Things framework [34], distributed manufacturing control [35] and mass customisation production [36]. The major drawbacks in the implementation of RFID technology have been as a result of proximity challenges, the effect of metal and water on RFID waves [35] and the high cost of implementation [34]. Due to the high implementation costs, SMMEs rarely use these technologies on the shop floor.

3.4.2 Internet of Things platforms

The IoT is the network of physical objects around us that contain electronic components, software, sensors and networking systems that allow these objects to exchange and acquire information. Hence, an IoT framework comprises of everyday physical objects with attached sensors which send big data streams to the internet for information analytics via a communications or networking channel [38]. To converge this world of devices together in a more efficient way, IoT platforms have emerged [39]. These are readily available online tools for linking sensors to the cloud with analytics capabilities as well. These include the Google cloud, Thingworx, Microsoft Azure, and Jasper (Cisco) platforms [40].

3.4.3 Mobile devices

More recently, mobile devices like smartphones and tablets have become ubiquitous in everyday life [41]. These gadgets have become readily available at low prices, causing a growth in usage in the 21st century. Modern smartphones have become programmable. They also come with a growing number of embedded sensors, such as a digital compass, an accelerometer, gyroscope, GPS, microphone and camera [42]. As a result, the range of data these devices can handle is enormous, including types like text messages, GPS, barcodes, QR-codes, pictures or images, audio and video [43]. Furthermore, nowadays, mobile devices can browse the internet, making them easily part of a global network of other objects. Mobile data can be accessed anywhere at any time, in near real-time [44]. Mobile devices are well-suited for applications in which the data collection is repeatedly done, conducted in a distributed way and a large percentage of the data types collected are quantitative in nature [45].

3.4.4 Augmented reality and wearables

Wearables and augmented reality are two technologies that look poised to empower the worker of tomorrow to be extraordinarily efficient and productive through contextual computing capabilities [46]. These applications are used for maximising the contextual awareness of an individual through providing real-time information, which is fundamental for decision-making. A majority of applications for this technology exist in the medical field where human monitoring systems for pulse rate, body temperature and exercise habits have been developed ([47] and [48]). However, the major drawbacks for the adoption of this technology include limited battery lifespan of devices, design and aesthetics of wearables, data privacy and management, and interoperability among solutions and vendors [49]. Furthermore, the cost of implementing such technologies is very high, while a majority of vendor-oriented integration issues remain unresolved [50].

3.4.5 Cloud computing platforms

According to Careterro and Blas [51], cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g. networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This model of technology is a broad field of study with numerous resources which allow the sharing of data over distributed systems. Google has the most commonly used cloud computing resources in the form of Google drive [52]. Microsoft Azure is another commonly used platform for Windows users [51]. Cloud computing platforms have a wide usage in many domains, including business [53], healthcare and manufacturing [54].

3.4.6 *Multilevel customer interaction and customer profiling technology*

A number of technologies, which enhance the interaction of customers with products and services that companies offer, are emerging on the market [55]. These solutions allow customers to define the product design during the conceptualisation stage and also order items online without having to visit a firm or store [56]. Application areas for this class of technology include the use of self-service technologies (SSTs), such as telephone banking, automated hotel checkout, and online investment trading, whereby customers produce services for themselves without assistance from firm employees.

3.4.7 *Big data analytics technology*

With the volume and speed of data generated by computers and the internet growing exponentially each year, the field of big data has grown and attracted much attention in academia [57]. The field involves the use of data streams for inferential decision-making. As a result, many domain applications for big data technology have emerged and these include; supply chain management [58], manufacturing [59] and healthcare [60]. A major drawback in the wide adoption of big data analytics technologies is the issue of data security [61].

3.4.8 *Smart sensors and 3D Printers*

With the ability to embed intelligence in systems made easier in the 21st century, smart sensors are also emerging on the market [62]. These devices can make decisions based on the data they obtain through instrumental recording. Smart sensors have been used widely in academia and research with applications in the transportation and logistics industry [63], manufacturing [64], security [65] and healthcare [66] amongst many other examples. Rapid prototyping has become a key stage in the product development cycle and 3D printers have emerged on the market to fill this gap. Their usage has been broad in fields of construction [67], manufacturing [68] and medicine ([69]; [70]; [71]; [72]and [73]).

3.4.9 *Advanced human-machine interfaces*

With the growth of embedded sensors in industrial equipment, human-machine interfaces for the real-time analytics of machinery performance have become a present hour reality [74]. Advanced human-machine interfaces are common on current industrial machines with numerous applications, which include motion study [75], and manufacturing [76].

3.4.10 *Authentication and fraud detection technologies*

The high rate of cybercriminal activities and need for security in specific fields make authentication and fraud detection technologies critical. The retail payments have the highest usage of these technologies [77]. Other applications are found in internet banking [78] and security systems [79].

4. READINESS LEVEL MODEL CONCEPT FOR SA CONTEXT I.4.0.

The models used by Warwick University [25] and Impuls [17] were used in developing the model used in this paper. An additional level 0 (outsider) was added to the Warwick model to cater for the respondents who did not know about Industry 4.0 or are not interested in implementing it. As for the Impuls model, Level 5 (top performer) was removed. This is because from our preliminary survey, we did not have any companies that had the technology to fully support I.4.0. Six dimensions of I.4.0 were selected namely, Products and Services, Enabling Technologies, Manufacturing and Operations, Strategy and Organisation, Supply Chain Integration and the Business Model. In each of these dimensions the respondents would score their company on a scale of 0 to 4 where 0 is Outsider, 1 is Beginner, 2 is Intermediate, 3 is Experienced, 4 is Expert. The average score for each company per dimension was then calculated then the overall Readiness score was then determined from those averages.

5. DISCUSSION OF RESULTS AND ANALYSIS

5.1 Results

The questionnaire survey was conducted with 150 participants from different provinces in South Africa (as illustrated in Figure 4). Of the 150 questionnaires sent out, only 30 were returned, making the response rate 20%. However, only 21 of the 30 returned questionnaire were in a usable format, with the other nine inadequately completed. Hence the nine incomplete questionnaires were disregarded and only 21 were used in the analysis.

5.1.1 *Enabling Technology*

This dimension assessed the frequency of use of the I.4.0 enabling technologies. According to the survey results shown in Figure 9, the most readily used are cloud computing platforms and mobile devices which are on level 3. Bar code technology leads the auto identification technologies in terms of readiness and usage. QR codes are next popular in our survey results coming in at level 2.42 readiness level. RFID is rarely used, possibly due to the high investment costs required to implement it. IoT and Big Data and 3D printers are rarely used according to the results. Internet of things could possibly be low because of lack of awareness, insufficient bandwidth infrastructure and insecurity with regards to cyber-attacks.

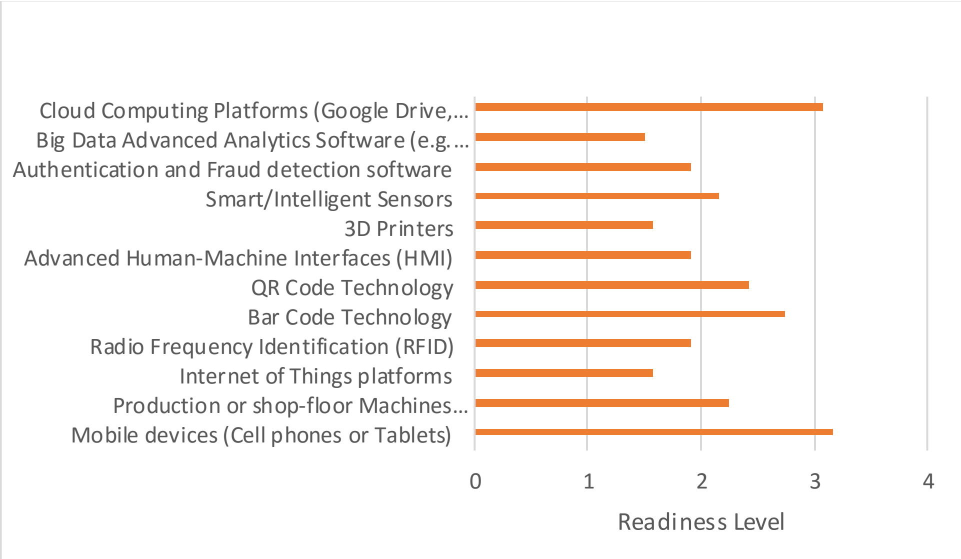


Figure 9: Enabling Technologies Results

5.1.2 Manufacturing and Operations

This dimension looks at the suitability of the manufacturing systems and operations for the implementation of I.4.0. Information Technology (IT) and data security is rated at level 2 as shown in Figure 10, although this is still fairly beginner level of I.4.0 readiness. Computer and IT solutions in the form of high specification computers, servers and firewalls have been invested in or partially implemented. These still cannot link with the basic machinery to invoke much change and integration to the manufacturing systems at large. The results also show M2M, autonomously guided work pieces, self-optimising process, and digital modelling are rated on the lower end of the assessment tool. This indicates that South African firms have basic machinery in their manufacturing systems, which are not yet capable of dealing digital modelling and other such advanced I.4.0 tasks.

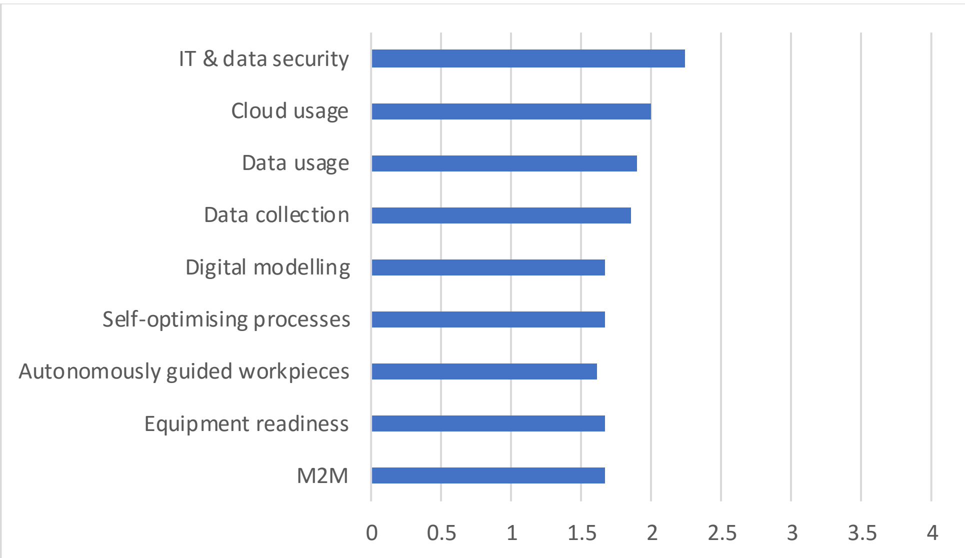


Figure 10: Manufacturing and Operations Results

5.1.3 Strategy and Organisation

This dimension analyses the suitability of the strategy of the organisation for the implementation of I.4.0. The collaboration sub-dimension assessed the level of the organisations with regards to their ability to interact inter-departmentally and also across different companies. The results in Figure 11 indicate is that there is interaction between departments but there is no cross-functional collaboration between them as indicated by the rating of level 2.4. The current rating can be justified by the fact that departments intrinsically interact on a day to day basis. The investment sub-dimension was rated at level imim1.3 and this can be attributed to the lack of value proposition perceived by businesses on implementing I.4.0.

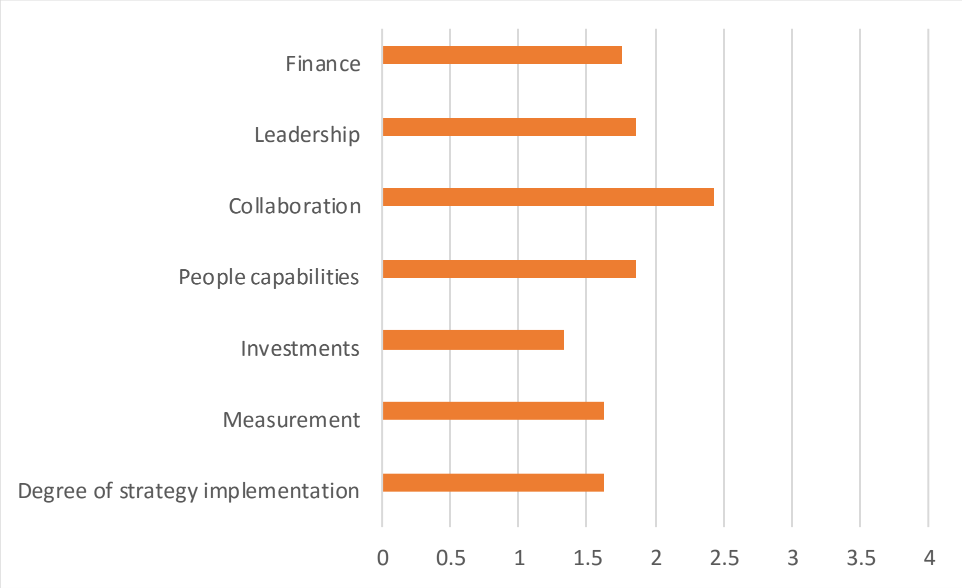


Figure 11: Strategy and Organisation Results

5.1.4 Overall Readiness Levels

The following results present the averages for all the 6 dimensions. From the results in Figure 12, it was observed that the aggregate readiness level starts high in the micro (0-50) range, then drops down to the lowest point in the medium (101-500) range. Based on these findings, our assumption is that the small micro company sizes have smart innovative and flexible processes, and when the company starts to have too many employees and staff the necessity for integration and I.4.0 enabling technologies the thematic starts to drop down. As the company size becomes 501 and more a company cannot function without, the internet, cloud-based solutions, mobile technology that is integrated into the entire system of business operations. A reason for this trend could be that small business owners are entrepreneurs who are trying to maximise on the opportunities that I.4.0 will bring to their businesses [80]. However, as they move on they start to experience challenges with resources to fully implement I.4.0. As with the big companies, these are mostly OEMS who have mother companies where I.4.0 is advanced and there is a strategic push for it. This is consistent with other developing countries such as Malaysia as mentioned in the literature above.

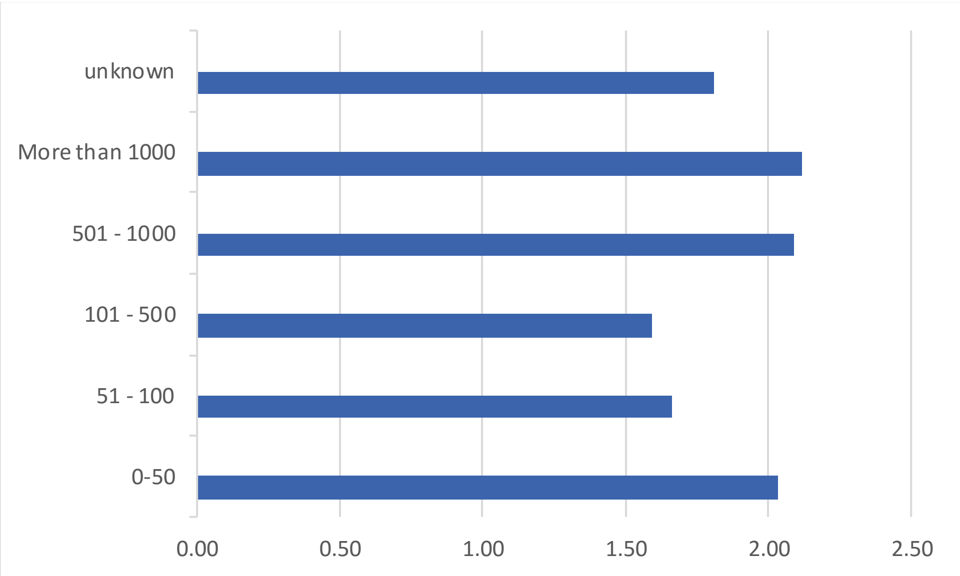


Figure 12: Average readiness level per company size

According to the findings, a readiness level score of 1.89 was obtained as shown by Figure 13. This score was calculated by using the average of the individual readiness levels from all respondents.

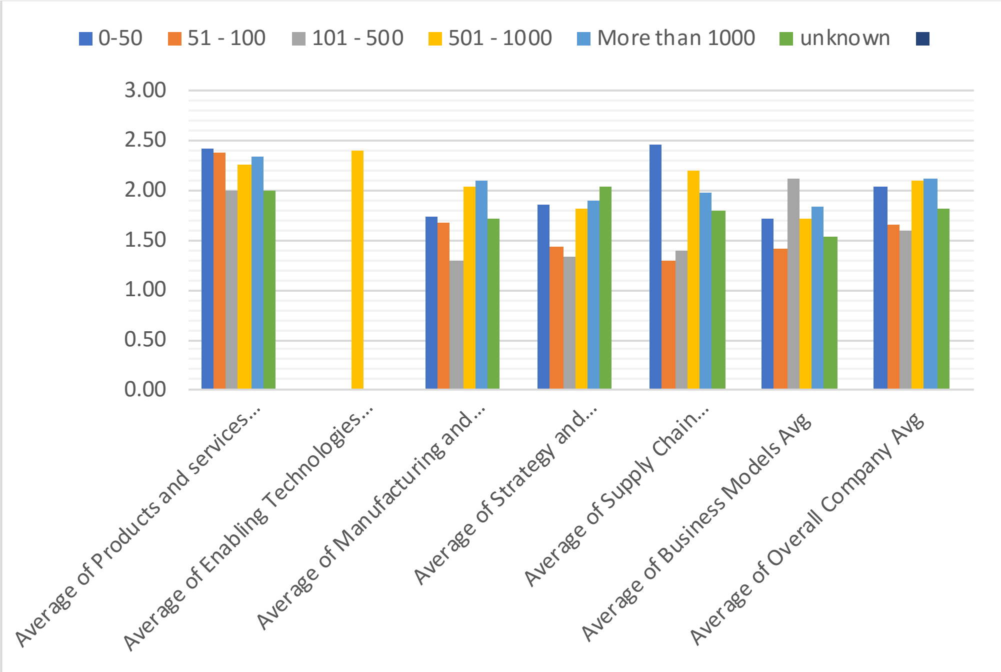


Figure 13: Overall averages per Dimension

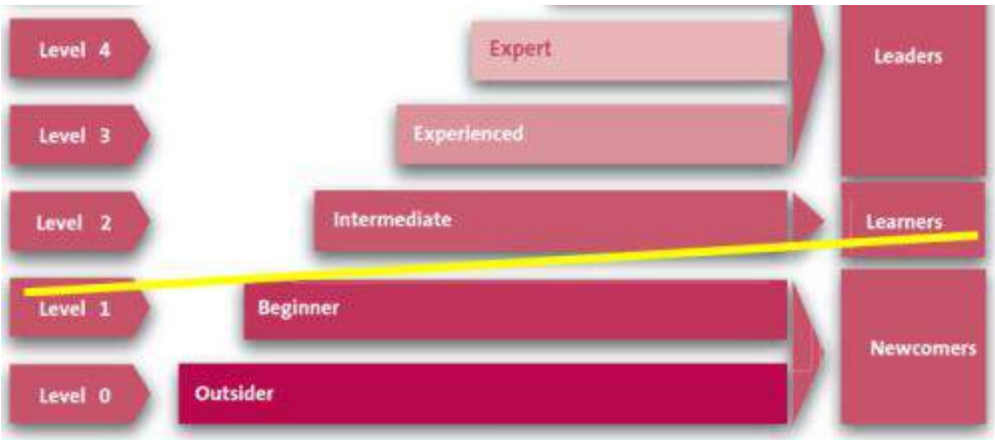


Figure 14: Overall Readiness of Respondents

The yellow line in Figure 14 indicates the overall readiness level according to the responses of the survey, and also in relation to all five levels. Level 0 - level 4. The companies are sitting on the boundary between Newcomers and Learners.

5.2 Analysis

As our Technology Readiness level results have shown South African firms are lagging behind in terms of implementation of I.4.0. One of the basis for the implementation of I.4.0. is the existence of enabling technologies. From the survey done on the frequency of use of enabling technologies, most of the respondents (12 out 21) did not answer the question. This could possibly be due to them not having the technologies, in the case of micro and small companies, or not understanding the use of the technologies, in the case of large companies. This could be the major reason why the companies scored so low in this survey.

Another possible reason could be lack of supporting government policies. Top Government officials have been publicly acknowledging and discussing the need for I.4.0. However, there is no evidence of tangible policies, programmes or engagement with industry. Most of the firms who are trying to implement I.4.0 are entrepreneurs who want to maximise on the opportunities provided by the concept.

6. CONCLUSION & RECOMMENDATIONS

The results of the study show that there is a lack of awareness concerning Industry 4.0 both in academia and in industry. There is therefore a great need to educate the community on the concept and formulate implementation strategies which are relevant to the SAn environment. Government at the top level of government has been heard to support the initiative.

Future work can be conducted on developing a decision-making framework for the implementation of Industry 4.0 in SA.

REFERENCES

[1] **Kagermann, H., Wahlster, W. and Helbig, J.** 2013. Recommendations for implementing the strategic initiative INDUSTRIE 4.0 - Final report of the Industrie 4.0 Working Group, Frankfurt/Main, 2013.

[2] **Hermann, M., Pentek, T. and Otto, B**. 2016, Design Principles for Industrie 4 . 0 Scenarios, *49th Hawaii International Conference on System Sciences*, pp. 3928–3937.

[3] **Kiel, D**. 2017, "What do we know about 'Industry 4.0' so far ?," in *Proceedings of the International Association for Management of Technology Conference*, 2017, no. Vol 26, pp. 0–22.

[4] **Sackey, S. M., Bester, A. and Adams, D.** 2017, Industry 4.0 learning factory didactic design parameters for industrial engineering education in South Africa, *South African J. Ind. Eng.*, vol. 28, no. 1, 2017.

[5] **Calitz, A.P., Poisat, P. and Cullen, M.** 2016, The future African workplace : The use of collaborative robots in manufacturing, *SA J. Hum. Resour. Manag.*, vol. 15, no. 1, pp. 1–11,

[6] **Dlodlo, N.** 2015, The Internet of Things in Agriculture for Sustainable Rural Development, no. April, 2015.

[7] **Mtsweni, J.** 2017, "Industry 4.0 meets Cyber 4.0," *Science Scope*, vol. 12, no. 3, pp. 56–59.

[8] **Deloitte**, "Industry 4.0 Is Africa ready for digital transformation?," 2016.

[9] **Olechowski, A., Eppinger, S.D. and Joglekar, N.** 2015 Technology Readiness Levels at 40: A Study of State-of-the-Art Use, Challenges, and Opportunities. *In: Proceedings of PICMET '15: Management of the Technology Age*. Portland, Oregon, USA, pp 0–28

[10] **Mankins, J.C.** 2009, Technology readiness assessments: A retrospective. *Acta Astronaut,* 65:1216–1223 . doi: 10.1016/j.actaastro.2009.03.058

[11] **Mankins, J.C.** 2009 Technology readiness and risk assessments: A new approach. *Acta Astronaut,* 65:1208–1215 . doi: 10.1016/j.actaastro.2009.03.059

[12] **Engel D, Dalton A, Anderson K, et al**, 2012, Development of Technology Readiness Level (TRL) Metrics and Risk Measures. In: Pacific Northwest Natl. Lab. https://www.pnnl.gov/main/publications/external/technical_reports/pnnl-21737.pdf. Accessed 3 May 2018

[13] **U.S. Department of Defense.** 2009, Technology Readiness Assessment (TRA) Deskbook. In: Technology. http://www.dod.mil/ddre/doc/DoD_TRA_July_2009_Read_Version.pdf

[14] **Ward, M.J., Halliday, S.T., Foden, J.** 2012. A readiness level approach to manufacturing technology development in the aerospace sector: An industrial approach. *Proc Inst Mech Eng Part B J Eng Manuf* 226:547–552 . doi: 10.1177/0954405411418753

[15] **US Department of Defense. 2012.** Manufacturing Readiness Level (MRL) Deskbook Version 2.4. http://www.dodmrl.com/MRL_Deskbook_V2.pdf. Accessed 5 May 2018

[16] **Peters, S. 2015,** A readiness level model for new manufacturing technologies. Prod Eng 9:647–654 . doi: 10.1007/s11740-015-0636-5

[17] **Lichtblau, K., Stich. V., Bertenrath, R, et al.** 2015, Impuls INDUSTRIE 4.0 READINESS. Aachen, Cologne

[18] **Kagermann, P.D.H., Wahlster, P.D.W. and Helbig, D.J.** 2013, Recommendations for implementing the strategic initiative INDUSTRIE 4.0. In: Final Rep. Ind. 4.0 WG. http://www.acatech.de/fileadmin/user_upload/Baumstruktur_nach_Website/Acatech/root/de/Material_fuer_Sonderseiten/Industrie_4.0/Final_report__Industrie_4.0_accessible.pdf

[19] **Bitkom, VDMA, ZVEI,** 2016, Implementation Strategy Industrie 4.0. 1–104

[20] **PricewaterhouseCoopers,** 2016, Industry 4 . 0 : Building the digital enterprise. https://www.pwc.com/gx/en/industries/industries-4.0/landing-page/industry-4.0-building-your-digital-enterprise-april-2016.pdf

[21] **Rockwell Automation,** 2014, The Connected Enterprise Maturity Model. http://www.rockwellautomation.com/rockwellautomation/innovation/connected-enterprise/maturity-model.page?

[22] **Roland Berger Strategy Consultants,** 2014, INDUSTRY 4.0 : The new industrial revolution. How Europe will succeed? https://www.rolandberger.com/publications/publication_pdf/roland_berger_tab_industry_4_0_20140403. pdf pp.16. Accessed 6 May 2017

[23] **Viharos. Z.J., Soós. S. and Nick. G., et al,** 2017 Non-comparative, industry 4.0 readiness evaluation for manufacturing enterprises. 1*5th IMEKO TC10 Work Tech Diagnostics 2017 - "Technical Diagnostics Cyber-Physical Era"* 181–187

[24] **Singapore Economic Development Board,** 2017. The Singapore Smart Industry Readiness Index. In: Singapore Gov. https://www.gov.sg/~/sgpcmedia/media_releases/edb/press_release/P-20171113-1/attachment/The Singapore Smart Industry Readiness Index - Whitepaper_final.pdf

[25] **Onur, A., Gibson, J., Godsell, P.J., et al** (2017) An Industry 4 readiness assessment tool. In: Univ. Warwick. https://warwick.ac.uk/fac/sci/wmg/research/scip/industry4report/final_version_of_i4_report_for_use_on_websites.pdf. Accessed 3 May 2018

[26] **Gilchrist, A.** 2017 Industry 4.0 Opportunities and Challenges of the New Industrial Revolution for Developing Countries and Economies in Transition. Available at: www.unido.org.

[27] **Wilson, P.** 2000 'The dilemma of a more advanced developing country: Conflicting views on the development strategy of Singapore', *Developing Economies*, 38(1), pp. 105–134. doi: 10.1111/j.1746-1049.2000.tb00873.x.

[28] **Pan, M. *et al*.** 2015 'Applying Industry 4.0 to the Jurong Island Eco-industrial Park', *Energy Procedia*, 75, pp. 1536–1541. doi: 10.1016/j.egypro.2015.07.313.

[29] **Bahrin, M. A. K. *et al*.** 2016 'Industry 4.0: A review on industrial automation and robotic', *Jurnal Teknologi*, 78(6-13), pp. 137–143. doi: 10.11113/jt.v78.9285.

[30] **Zhou, K.** 2015 'Industry 4 . 0: Towards Future Industrial Opportunities and Challenges', *12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pp. 2147–2152.

[31] **Jones, C. and Pimdee, P.** 2017 'Innovative ideas: Thailand 4.0 and the fourth industrial revolution', 17, pp. 4–35.

[32] **Reinhard, G., Jesper, V. & Stefan, S.** 2016. Industry 4.0: Building the Digital Enterprise. *2016 Global Industry 4.0 Survey*, pp 1-39.

[33] **Nambiar, A.N.** 2009. RFID Technology: A Review of its Applications. *Proceedings of the World Congress on Engineering and Computer Science*, 2, pp 20-22.

[34] **Ting, S.L., Tsang, A.H.C. & Tse, Y.K.** 2013. A Framework for the Implementation of RFID Systems. *International Journal of Engineering Business Management*, 5(9), pp 1–16.

[35] **Welbourne, E., Battle, L., Cole, G., Gould, K., Rector, K., Raymer, S. & Balazinska, M.** 2009. Building the Internet of Things Using RFID. *Internet Computing, IEEE*, 13(3), pp 48-55.

[36] **Barenji, A.V, Barenji, R.V & Hashemipour, M.** 2014. A Framework for Structural Modelling of an RFID-Enabled Intelligent Distributed Manufacturing Control System. *South African Journal of Industrial Engineering*, 25(2), pp 48-66.

[37] **Zhong, R.Y., Dai, Q.Y., Qu, T., Hu, G.J. & Huang, G.Q.** 2013. RFID-Enabled Real-Time Manufacturing Execution System for Mass-Customization Production. *Robotics and Computer-Integrated Manufacturing*, 29(2),pp 283–292.

[38] **Doukas, C.** 2012. *Building the Internet of Things with the Arduino*. USA: CreateSpace Independent Publishing Platform.

[39] **Ray, P.P.** 2016. A Survey of IoT Cloud Platforms. *Future Computing and Informatics Journal*, 1(1-2),pp 35–46.

[40] **Lamont, J.** 2017. IoT Platforms Create Value in Connectivity. *KM World*, 26(3), pp 3.

[41] **Trucano, M.** 2014. *Using mobile phones in data collection: Opportunities, issues and challenges*. [Online], Retrieved September 10 2015: http://blogs.worldbank.org/edutech/using-mobile-phones-data-collection-opportunities-issues-and-challenge.

[42] **Park, J.** 2015. Evaluating a Mobile Data-Collection System for Production Information in SMEs. *Computers in Industry*, 68, pp 53–64.

[43] **Vieira, D. & Vieira, D.** 2015. Empowering Mobile Users : Applications in Mobile Data Collection Empowering Mobile Users : Applications in Mobile Data Collection.

[44] **Jung, C.** 2011. Mobile Data Collection Systems - A Review of the Current State of the Field. Retrieved from http://humanitarian-nomad.org/mdc-research/

[45] **Satterlee, E., McCullough, L., Dawson, M. & Cheung, K.** 2015. *Paper-To-Mobile Data Collection: A Manual*. Retrieved August, 28, 2016, from https://www.fhi360.org/sites/default/files/media/documents/Paper_to_Mobile_Data_Collection_Manual_1.0.pdf.

[46] **Elvins, T.T. & Todd, T.** 1998. Augmented Reality. *ACM SIGGRAPH Computer Graphics*, 32(1), pp 11–13.

[47] **Chandra, P.** 2016. *Nanobiosensors for personalized and onsite biomedical diagnosis*. Stevenage, England: The Institution of Engineering and Technology.

[48] **Abbasi, Q.H., Ur-Rehman, M., Qaraqe, K. & Alomainy, A.** 2016. *Advances in Body-Centric Wireless Communication : Applications and state-of-the-art*. Stevenage, England: The Institute of Engineering and Technology.

[49] **PR Newswire.** 2016. *Office of the Future - Part 1: Wearables and Augmented Reality*. Retrieved August 13, 2015, from https://search-proquest-com.ez.sun.ac.za/docview/1803835661?rfr_id=info%3Axri%2Fsid%3Aprimo [2017, October 09].

[50] **Newman, N.J. & Clark, A.F.** 1999. Sulawesi: A Wearable Application Integration Framework. *In The Third International Symposium on Wearable Computers* (pp. 170–171). San Francisco, California: IEEE Computer Society.

[51] **Carretero, J. & Blas, J.G.** 2014. Introduction to Cloud Computing: Platforms and Solutions. *Cluster Computing*, 17(4),pp 1225–1229.

[52] **Jennings, R.** 2009. *Cloud Computing with the Windows Azure Platform*. Retrieved from http://lib.myilibrary.com.ez.sun.ac.za/Open.aspx?id=230684

[53] **Shroff, G.** 2010. *Enterprise Cloud Computing : Technology, Architecture, Applications*. Cambridge, UK:

Cambridge University Press.
[54] **Hung, M.H., Lin, Y.C., Quoc Huy, T., Yang, H.C. & Cheng, F.T.** 2012. Development of a Cloud-Computing-Based Equipment Monitoring System for Machine Tool Industry. *Proceedings of International Conference on Automation Science and Engineering* (pp. 962–967).
[55] **Meuter, M.L., Bitner, M.J., Ostrom, A.L. & Brown, S.W.** 2005. Choosing Among Alternative Service Delivery Modes: An Investigation of Customer Technologies. *Journal of Marketing*, 69(2), pp 61–83.
[56] **Meuter, M.L., Ostrom, A.L., Roundtree, R.I. & Bitner, M.J.** 2000. Self-Service Technologies: Understanding Customer Satisfaction with Technology. *Journal of Marketing,* 64(3), pp 50–64.
[57] **Rajaraman, V.** 2016. Big Data Analytics. *Resonance,* 21(8), pp 695–716.
[58] **Shahriar, A., Wamba, S.F., Gunasekaran, A., Dubey, R. & Childe, S.** 2016. How to Improve Firm Performance Using Big Data Analytics Capability and Business Strategy Alignment? *International Journal of Production Economics*, 182, pp 113–131.
[59] **Zhong, R.Y., Xu, C., Chen, C. & Huang, G.Q.** 2017. Big Data Analytics for Physical Internet-Based Intelligent Manufacturing Shop Floors. *International Journal of Production Research*, 55(9), pp 2610–2621.
[60] **Yichuan, W. & Nick, H.** 2017. Exploring the Path to Big Data Analytics Success in Healthcare. *Journal of Business Research,* 70, 287–299.
[61] **Zhenning, X., Frankwick, E. & Ramirez, G.** 2016. Effects of Big Data Analytics and Traditional Marketing Analytics on New Product Success: A Knowledge Fusion Perspective. *Journal of Business Research,* 69(5), pp 1562–1566.
[62] **Meijer, G.C.M. (Ed.).** 2008. *Smart Sensor Systems*. West Sussex, UK: J. Wiley & Sons, Ltd.
[63] **Moreu, F.** 2017. Railroad Bridge Monitoring using Wireless Smart Sensors. *Structural Control,* 24(2), pp 17.
[64] **Nihtianov, S.** 2017. New Trends in Smart Sensors for Industrial Applications - Part I. *IEEE Transactions on Industrial Electronics,* 64(9), pp 3.
[65] **Acho, L.** 2017. Event-Driven Observer-Based Smart-Sensors for Output Feedback Control of Linear Systems. *Sensors (14248220),* 17(9), pp 11.
[66] **Gaggioli, A., Pioggia, G., Tartarisco, G., Baldus, G., Corda, D., Cipresso, P. & Riva, G.** 2013. A mobile Data Collection Platform for Mental Health Research. *Personal and Ubiquitous Computing*, 17(2), pp 241–251.
[67] **Honiball, J.R.** 2010. The Application of 3D Printing in Reconstructive Surgery (Doctoral dissertation, University of Stellenbosch, South Africa). Retrieved from http://scholar.sun.ac.za/handle/10019.1/4207
[68] **Long, Y., Pan, J., Zhang, Q. & Hao, Y.** 2017. 3D Printing Technology and its Impact on Chinese Manufacturing. *International Journal of Production Research,* 55(5), pp 1488–1497.
[69] **Christensen, A. & Rybicki, F.J.** 2017. Maintaining Safety and Efficacy for 3D Printing in Medicine. *3D Printing in Medicine,* 3(1), pp 1.
[70] **Hirsch, J.D., Vincent, R.L. & Eisenman, D.J.** 2017. Surgical Reconstruction of the Ossicular Chain with Custom 3D Printed Ossicular Prosthesis. *3D Printing in Medicine*, 3(1), pp 7.
[71] **Leng, S., McGee, K., Morris, J., Alexander, A., Kuhlmann, J., Vrieze, T., McCollough, C.H. & Matsumoto, J.** 2017. Anatomic Modeling Using 3D Printing: Quality Assurance and Optimization. *3D Printing in Medicine,* 3(1), pp 6.
[72] **Costa, M.O., Nosach, R. & Harding, J.C.S.** 2017. Development of a 3D Printed Device to Support Long Term Intestinal Culture as an Alternative to Hyperoxic Chamber Methods. *3D Printing in Medicine,* 3(1), pp 9.
[73] **Radenkoic, D., Solouk, A. & Seifalian, A.** 2016. Personalized Development of Human Organs Using 3D Printing Technology. *Medical Hypotheses*, 87,pp 30–33.
[74] **Teramoto, K. & Onosato, M.** 2001. In-Process Visualization of Machining State with Sensor-Based Simulation to Support the Recognition Ability of Operators. *Human Friendly Mechatronics,* pp 389–394.
[75] **Inoue, K., Okuda, A., Tani, H. & Mae, Y.** 2001. Distributed Control of Hyper-Redundant Manipulator with Expansion and Contraction Motion for Obstacle Avoidance. *Human Friendly Mechatronics*, pp 315–320.
[76] **Jeong, J.W., Yeo, W.H., Akhtar, A., Norton, J.J.S., Kwack, Y.J., Li, S., Jung, S.Y., Su, Y., … Rogers, J.A.** 2013. Materials and Optimized Designs for Human-Machine Interfaces Via Epidermal Electronics. *Advanced Materials*, 25(47), pp 6839–6846.
[77] **Summers, B.J.** 2009. Fraud Containment. *Economic Perspectives,* pp 17–21.
[78] **Bignell, K.B.** 2006. Authentication in an Internet Banking Environment: Towards Developing a Strategy for Fraud Detection. *Proceedings of International Conference on Internet Surveillance and Protection,* pp. 23. Cap Esterel, France: The Institute of Electrical and Electronics Engineers.
[79] **News Bites.** 2014. CA Technologies Helps Card Issuers Combat Fraud and Increase Revenue with Next-Generation Authentication Solution.
[80] **Dewa M, Van der Merwe A F and Matope S. 2018.** Digitalisation of shop-floor operations in the South African Tool, Die and Mould-making industry. South African Journal of Industrial Engineering, Vol 29(2), August 2018, pp. 153 – 170.

A SIMULATION ANNEALING APPROACH TO DETERMINE THE EFFECTIVE TRAIN STATION LAYOUT WITH DIFFERENT PEAK PERIODS

T. Bele[1]*, T.B. Tengen[2] & B.B.B. Eboule[3]
[1,2,3]Department of Industrial Engineering and Operations Management
Vaal University of Technology, South Africa
[1]tshegofatsobele@yahoo.com, [2]thomas@vut.ac.za , [3]bengaebouele@yahoo.fr

ABSTRACT

On daily basis different types of passengers use the Johannesburg commuter rail system. This means that there should be a dynamic and heterogeneous element that should be controlled in the commuter rail system design. A standard approach to designing an effective flow layout is to use From-To-Chart technique from which one can improve on efficiency by minimizing backtracking and then use Block Diagram strategy to obtain the physical display. This has been employed and improvements were achieved in previous study. However, the data that considered was for an overall period. The issue that arises is when the commuter rail system experiences different peak periods and different volumes that sum up to the overall data. If different peak periods and their associated volumes are used to design the layout then the layout will have different efficiencies under the different conditions. This research paper uses the techniques borrowed from simulated annealing to compare different periods and different efficiencies to decide on the layout that is suitable for the commuter rail system under the different peak scenarios. Results shows how flexibility in the layout design can be managed.

*Corresponding author

1.1 INTRODUCTION

The United Nations Convention on the Rights of Persons with Disabilities [1], which was ratified by South Africa in November 2007, states that persons with disabilities include those who have long term physical, mental, intellectual or sensory impairments from which interaction with various barriers may hinder their full and effective participation in society on an equal basis with others.

The Passenger Rail Agency of South Africa launched a project [2] whose aims were to identify and address ways of integrating universally accessible designs into PRASA projects. This means that a commuter rail system should give equal opportunity and accessibility to services, products, systems and environments regardless of economic situations, social situation, religious or cultural background, gender and functional limitations. An approach is to employ modelling and simulation since they play a vital role in the modern life [3] as they mimic the behaviours of the real systems and they are essential in the effective design, evaluation and operation of new systems.

The first simulated annealing was performed by Kirkpatrick *et.al* [4]. Simulated Annealing is a stochastic optimization technique. Simulated Annealing is a method based on Monte Carlo Simulation to solve challenging combinatorial optimization problems [5]. Simulated Annealing is derived from analogy to determine the behaviour of a physical system by melting a substance and lowering its temperature slowly until it reaches a freezing point. The slow cooling implemented in the simulated annealing algorithm is interpreted as a slow decrease in the probability of accepting worse solutions as the solution space is explored. In metaheuristics, accepting worst solutions is a fundamental property as it allows more extensive search for the global optimum solution

Simulated annealing can be described as follows: for each optimization problem there's a set of feasible solutions with each solution S having a cost value obtained from the cost function *f(s)[6]*. The objective is to find a feasible solution with a minimum cost. Local optimum is found by finding the cost of an initial solution S generated from an initial combination of conditions or parameters and then repeatedly attempting to find a better solution by moving to a neighbouring set of conditions with a lower cost function.

it is possible to save significant time and cost in if facility layout challenges are solved and, furthermore, by also using simulated annealing to develop an efficient layout in a railways system [7]. In this paper the concept of simulated annealing was applied on the From-To-Chart technique, which is the optimization (or cost) function of determining effective flow pattern. The cost (i.e. efficiency in our particular case) of the current layout obtained from the overall data of the all peak periods was determined. The solution or cost of the next set of conditions was determined and a decision to move or stay with current solution was based on the efficiency obtained. Peak hours, passenger volumes, location of facilities and passengers' flow routes to these facilities were used as the control parameters to determine the cost of the flow pattern, with the intention of choosing the route or flow pattern whose total cost is minimized or highly effective layout or route.

At peak conditions, passenger volumes increase at Johannesburg train station. Therefore the rail station needs be designed to serve heavy time based, directional and spatial peaking periods. This paper proposes remedy that will improve passenger flow in a commuter rail system by using simulated annealing applied on From-To-Charts technique as the objective function to determine effective flow pattern that will cater for all the different peak hours that have different efficiencies so that an improved layout that will match the overall expectation is recommended

1.2 METHODOLOGY

In order to determine the needed facilities in a commuter rail system, Reeds Plant Layout Procedure[8] was followed. By following this procedure information on the type of passengers to be accommodated in a commuter rail system was established, who were children, the elderly, physically limited, auditory limited, inebriated limited, cognitive limited, linguistic limited, pregnant women, temporary limited, visually limited and paraplegics (people on wheelchairs[9]). As reported by Anastasia L. *et.al* [10], additional differentiating factors are passenger density, passenger attitude and route familiarity. When planning for passenger flow, it is important to take commuter conditions, evacuations, special events, off peak conditions and the types of passenger-flow into consideration as reported by the South Africa Rail Commuter Corporation [11]. That collected data was for the different peak periods per day whose sum gave the overall period data.

From-To-Chart technique [8] was then used to determine the different efficiencies and improve on them by minimizing backtracking, and hence, reducing high penalty points. Backtracking represents the movement of passengers forward and backward within the facility. Backtracking impacts flow by increasing distance moved and its associated cost, thus increasing the penalty point and therefore requires to be minimized. Back tracking values are found in the flow entries below diagonal line that divides the upper right part from the lower left part of the From-To-Chart. The penalty points are obtained by doubling the backtracking entries and then multiplying by the number of cells from diagonal line. Back tracking was minimised by proposing that those cells (i.e. flow between) with high penalty points should be placed closer to each other, and the From-To-Chart techniques applied again to obtain improved routes' efficiencies. These steps were repeated for all the different

scenarios (i.e. different peak periods) and the scenario with the higher efficiency was considered as the preferred layout.
The Block Diagram strategy [8] was used to propose the physical display of the improved train station layout. In order to applying the From-To-Chart technique, all processes or departments where processes take place are listed in the order that is identical to the required flow route across the column and down the row on the left-hand side of the chart. The listing should be in the direction of the overall flow's layout from beginning to end. Required next was a compilation of process sequence or route for a group of representative passengers. In general, one may consider flow volume, distances travelled, where to, costs etc. The flow of passengers through each process sequence or route was tracked and process chart was completed for each passenger type, followed by performing the From-To-Chart analysis for each data set.

The issue that arose was the fact that the commuter rail system experienced different peak periods and different volumes that summed up to the overall data. If different peak periods and their associated volumes were used to design the layout then there would be different efficiencies under the different conditions.

So, the skill of Simulated Annealing was used to compare the different efficiencies at different peak periods, with the objective of determining an effective facility layout that will accommodate for passengers with special needs. The validation process depended on parameters that need to be matched or compared [12]. The parameters set for this project are the peak periods, passenger volumes, location of facilities and passengers' flow routes to these facilities at the commuter rail system. As the fundamental concept of simulated annealing comes from statistical mechanics and combinatorial optimization, the layout was improved by occasionally accepting non-improving solutions to allow the algorithm to explore other regions of the solution space instead of just stopping at the first seemingly good solution that was encountered [8]. The data given in Table 1 below was collected for the different peak periods for different passengers. The data was collected over a period of two months from Passenger Rail Agency of South Africa managerial foot count reports and also by direct observations. The ABCDEFG flow route (i.e. a greater major of the customer were expected to enter the train station from the taxi rank at A, and then move to B which are ablutions that cater for normal passengers and the disabled passengers, then to C (Game shopping store), then to D (i.e. business lounges), then to E (Shosholoza premier class), then to F (medical centre) and then to G (which is the exit to train station platform from where passengers catch trains to various destinations) was the prescribed route that passengers have to follow in the train station.

A column-by-column combination was considered, and from-to-chart technique was applied to determine different efficiencies for different peak periods. The column-by-column combination was constrained by the following facts. Abled persons and children were bound by the requirements of their jobs e.g. children and abled persons had to go to school or work in the morning and come back in the afternoon. At mid-day passengers who required assistance felt that is was convenient for them to use the train station at that time for safety reasons and to be assisted in order for them to utilise train facilities and resources happily. Dormant period occurs at night where less passengers use the train station; however, this period layout will not affect the results obtained for the effective layout as the layout that effectively assists passengers during peak periods will still effectively assists passengers at this time.

Table 1: A number of passengers at train station

	06:00am-09:00am	**09:00am-12:00mid day**	**12:00midday -03:00pm**	**03:00pm-06:00pm**	**06:00pm-09:00pm**	**Total**
CHILDREN	90	68	34	98	10	300
ELDERLY	150	55	63	182	50	500
PHYSICAL LIMITED	143	75	53	121	88	480
COGNITVE LIMITED	43	91	130	60	26	350
LINGUISTIC LIMITED	70	40	60	20	10	200
AUDITORY LIMITED	20	12	32	24	12	100
VISUALLY LIMITED	20	64	40	15	11	150
TEMPORARY LIMITED	51	31	44	43	20	189
PREGNANT PERSONS	40	130	202	86	30	488
INEBRIATED PERSON	13	50	86	100	151	400
WHEELCHAIR (PARAPLEGIC)	43	62	82	52	26	260
ABLED PERSONS	600	200	300	650	200	1950

1.3 ANALYSIS OF RESULTS

Scenario 1

For the first peak period, which was from 06:00am to 09:00am, passenger volumes increased at train stations. In the morning peak, commuters moved with more urgency in order to arrive at work, school etc. in time. During this time what was noticeable with passenger flow was the fact that passengers used the same train station to reach the same destination, day after day and therefore were familiar with their habitual routes. This type of passengers did not pay attention to their surroundings all the times as they relied on their travelling experience. During this period, it was observed that inebriated, auditory and visually limited persons were less at the train station due to crowding since this time could cause discomfort and risks for these people.

The From-To-Chart technique whose principle has been explained above was used to measure the effectiveness of the prescribed sequence of flow, which gave 390.5/689.4 = 57% efficiency. The layout efficiency of the prescribed route ABCDEFG was calculated by dividing the sum of the total points by the sum of total of penalty points. The improved route was proposed by minimising back tracking i.e. by bringing departments with high penalty points closer to each other to get a new route.

Scenario 2

The second peak period occurred between 09:00am-12:00midday. During this time, it was when most of the people with special needs began to use the commuter rail facilities as there was less crowding. There was a decrease in volume of auditory, linguistic and temporary limited persons. These people in most cases relied on other passengers to commute effectively in train stations. It was also a time where there are less abled passengers who could possibly offer assistance to the abled passengers. However, there are a high number of physical, cognitive and pregnant people during this time at the train station. In this scenario the efficiency of the prescribed route was 66%

Scenario 3

The peak period which was between 12:00midday-03:00pm had an increased number of cognitive, pregnant and inebriated persons. However, because it was during lunch period for most people, the volume of people increased as most people were using food facilities, ATM machines etc. For children, whose majority was in schools, and the auditory people, whose hearing might be affected by the loud noise, were less in number at the train stations. Under this scenario the efficiency calculated using the From-To-Chart Technique was 71%.

Scenario 4

The fourth peak period, which was from 03:00pm - 6:00pm, passenger volumes increase at train stations. Commuters moved with more urgency in order to get home as most of them were from places such as work and school. During this time what was noticeable with passenger flow was similar to the first peak period where passengers used the same train station to reach the same destination, day after day and therefore are familiar with their habitual routes. This type of passengers might have not paid attention to their surroundings as they rely on their travelling experience.

It was observed that linguistic, auditory and visually limited persons were less at the train station at this time due to crowding. Commuting at this time may cause discomfort and risks for these people. The effectiveness of the prescribed flow was measured, and it gave 64% efficiency.

Scenario 5

The next peak period occurred between 06:00pm-09:00pm. There were less number of children, linguistic and visually limited people during this period. During this period, it was dark and it might have been a risk for this type of passengers to use the train facilities on their own. The From-To-Chart technique was used to measure the effectiveness of the prescribed sequence of flow, which gave 1673.1/2914.1=57% efficiency.

The overall system efficiency was measured using the From-To-Chart technique for the prescribed sequence of flow and the results obtained was 64% efficiency.

The recommended Efficiency

The efficiencies of the different peak information were compared through deviation analysis were the efficiencies obtained for each peak period, which are 57%, 66%, 71%, 64% and 57%, were compared with the overall efficiency of the train station. Results indicate that the efficiency obtained at the peak period 03-00pm-06-00pm should be used to design the layout of the commuter rail system as it matches the overall efficiency

1.4 CONCLUSION

The commuter rail system's layout needs to be effective in order to meet passenger's expectations. Managing passengers' expectations and improving their train station experience is crucial to the rail business where correlation between passengers and profitability has been widely accepted. Aligning to national imperatives by accommodating for passengers with special needs is important for train stations managers to further explore the train station service quality that will differentiate the success of the commuter rail system. The results that were obtained in this paper illustrate how flexibility in the layout design can be managed. This was achieved through the application of simulation annealing, where different solutions were explored and the From-To-Chart technique where an improved layout was achieved by minimising backtracking. Deviation analysis was employed to determine a cost effective to obtain different layout efficiencies and be able to compare them to a known standard. It is recommended that to redesign the layout of the train station the improved afternoon peak period layout should be used as it matches the overall layout efficiency.

1.5 REFERENCE

[1]. The United Nations Convention on the Rights of Persons with Disabilities Report (2006).
[2]. The Passenger Rail Agency Report (2014).
[3]. Pace, D. (2004). Modelling & Simulation Verification and Validation Challenges. 25:2.
[4]. Kirkparick, S, Gelatt, C and Vecchi, M. (1983). Optimization by Simulation Annealing. Journal of Science. 220 (4598)671-680, May.
[5]. Barretto, P, Chwif, L, Marcos, R and Moscato, L. (1998). A Solution to the Facility Layout Problem using Simulated Annealing. Journal of Engineering. 36:125-132.
[6]. Al-Khedhairi, A. (2008). Simulation Annealing Metaheuristics for Solving P-Median problem. Journal of Science.3. (28):1357-1365.
[7]. Momeni M, Sarmadi M, and Yaghini M. (2012). Solving Train formation problems using simulation annealing algorithm in a simplex framework. Journal of Engineering 48:402-416, March.
[8]. Bozer, Y., Tanchoo, J.M.A., Tompkins, J. and White, J. 2010. Facilities Planning. New York:Wiley.
[9]. Swart Pieter Project Manager-Park station 24 June.
[10]. Loukaitou-Sideris, A., Taylor, B.D., Turkey, C. 2015. Passenger Flows in Underground Railway Station and Platform.
[11]. South African Rail Commuter Corporation (2003) Special needs passenger: Policy for Commuter Rail South Africa. Final Report. SARCC South Africa.
[12]. Lautala P. and Pouryousef. 2015. Hybrid Simulation Approach for Improving Railway Capacity and Train Schedule. Journal of rail Transport Planning & Management 5:211-224.

USING AN ADAPTED SYSTEM DYNAMICS APPROACH TO DETERMINE THE LINKAGE BETWEEN ELECTRIC VEHICLE MARKET PENETRATION AND AFFORDABILITY

N.S. Pillay[1*], A.C. Brent[2] & J.K. Musango[3]

[1]Department of Industrial Engineering
Stellenbosch University, South Africa
Pillayna@eskom.co.za

[2]Department of Industrial Engineering &
The Centre for Renewable and Sustainable Energy Studies,
Stellenbosch University, South Africa
acb@sun.ac.za

[3] School of Public Leadership &
The Centre for Renewable and Sustainable Energy Studies
Stellenbosch University, South Africa
josephine.musango@spl.sun.ac.za

ABSTRACT

This paper focuses on an adapted process for system dynamics modelling based on industry experience and the successful implementation of system dynamics models within an electricity utility. The modelling process was demonstrated using a case study of battery electric vehicle (BEV) market penetration in South Africa and its substitution of internal combustion engine vehicle, as a function of affordability based on real disposable income. The results indicate that South Africans are living beyond their “income” constraints and purchasing far more vehicles than what their disposable income allows, with the situation worsening over time. The Gauteng province will have the largest potential to absorb BEVs (81,123) and the highest impact on residential electricity consumption (an additional 4,291 GWh) whilst the lowest is the Northern Cape province with 5,140 BEVs (an equivalent of 272 GWh). However, if disposable income is used as a parametric to determine the affordability of BEVs then there may be 80% less than the expected number of BEVs in terms of market penetration. To benefit from a reduction in carbon emissions in the transport sector, a renewables heavy supply mix would be required else there is not much benefit with South Africa’s current coal heavy supply mix.

[1] The author is enrolled for an PhD degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. Introduction

Eskom, a State Owned Corporation (SOC), is an electricity utility in South Africa, responsible for supplying approximately 96% of the electricity used in the country and more than 45% of Africa [1]. The nature of the utility's operating environment in the context of socio-economic, political and environmental changes requires an evolving business model with a dynamic planning approach [2] capable of understanding feedback behaviour across the electricity value chain [3]. To develop strategically sound business models for economic and environmentally conscious competitive advantage, advanced modelling tools and processes are required [4],[5].

Experience and extensive engagements between the organisation's stakeholders and the System Dynamics Centre at Eskom SOC resulted in an adapted system dynamic modelling process for the successful implementation of simulations to tackle complex system problems. This paper explains the conventional high-level modelling process and then the adapted system dynamics modelling process, using the results from the *E-StratBEV* system dynamics simulator. The *E-StratBEV* was developed to determine the linkage between battery electric vehicle (BEV) market penetration and affordability, based on disposable income in South Africa.

2. Energy Modelling Approaches

Mathematical energy modelling tools, focussed on the energy-economy-environment nexus, have been critical in support of strategic business model development [6]. There are numerous energy modelling methods, however, the main categories include econometrics [7], macro-economics [8], multi-criteria decision analysis [9], optimization [10] and simulation [11], [12]. The most applied simulation and modelling methods includes agent-based modelling, discrete event modelling and the system dynamics process [13]. System dynamics modelling was identified as being useful in strategy refinement and the transfer of insights as part of strategy implementation [14].

The system dynamics modelling process advocated by Sterman [15] was initially used as a standard framework for developing system dynamics models by the Eskom System Dynamics Centre, but this modelling process gradually evolved with experience gained through stakeholder engagements, which contributed to the successful implementation of the adapted system dynamics process within the power utility [16].

2.2 System Dynamics Modelling

System dynamics is characterised by a computerised approach, based on systems thinking principles and was founded by Forrester in the 1970s, who applied engineering concepts of feedback systems and digital simulation to understand "*the counterintuitive behaviour of social systems*" [17]. The type of problems, which require a system dynamics approach, would be those that have quantities that change over time and are dynamic in nature, and those which have feedback [18]. The system dynamics modelling process uses the premise that every real system, including business environments, could be explained in a mathematically-based approach using a series of equations, represented by interconnected flows or rates and storage levels (stocks). This is represented by Equation 1, where the state variable is the stock, which is based on the difference between the rate of change of the inflow and the outflow, where *dt* is the time interval for each computational step.

$$stock(t) = stock*(t - dt) + (Inflow - Outflow)*dt \tag{1}$$

The system dynamics modelling process requires several steps, as shown in Figure 1 [19].

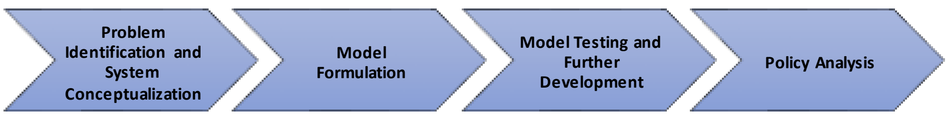


Figure 1: System Dynamics Modelling Process [19]

Following a detailed modelling process prevents the modeller from conceptual problems, and provides the necessary contextualisation of the system problem to be modelled.

2.3 Review of System Dynamics Modelling Processes

A summary of the conventional system dynamics modelling processes is provided in Table 1, with a list of sub-elements constituting each step.

Step 1: Problem Identification and System Conceptualization: This step includes defining the problem dynamically with or without the use of data, since the behaviour of a system variable over time follows a pattern that can be illustrated [19]. After identifying the problem, key variables and reference modes (historical data or patterns of behaviour of variables); feedback structures are identified. For this purpose, causal loop diagrams [20] illustrate the visual cause-effect and loops of role-playing variables related to the system problem [21]. Part of this step includes a clear model purpose. Sterman [15] refers to this step as *Problem Articulation*, which also includes defining time horizons.

Step 2: Model Formulation: This includes formulating rate equations and defining the variable parameters and initial values. Coyle [22] included influence diagrams in the second step, and Sterman [15] included various other tools, such as a model boundary chart (which indicates which variables are endogenous (outputs), exogenous (inputs) or excluded in the model structure), a sub-system diagram, a causal loop diagram, stock and flow maps and policy structure diagrams. The causal loop diagrams or system diagrams may be constructed by group model building, a facilitated participatory modelling method whereby stakeholders are guided through a brainstorming session to determine as many variables and relationships of these variables linked to the system problem [23].

Step 3: Model Testing and Further Development: Understanding model behaviour and sensitivity runs, refinement and reformulation, as well as validation, characterise this step. The iterative nature of the modelling process is emphasised where the iterations depend on the complexity of the system problem being modelled. Sensitivity analysis is based on 3 main categories: numerical (if parameters change with numerical values) [24], behavioural (model behaviour and pattern over time changes when parameters change) [25], and policy sensitivity (checking model runs against policy-based conclusions) [26].

Step 4: Policy Analysis: At this stage, sensitive policy parameters are identified, the ones which result in the most change in influencing the system and would tend to be the areas to be leveraged, besides the feedback loops which dominate system behaviour [27].

In Eskom SOC, the first system dynamics simulation was developed in 2010 and presented to the company's Executive Committee and Board of Directors, as part of a scenario-planning project. The simulation proved advantageous in understanding the underlying system structure through scenario analysis, to challenge and even change previous mental models, as identified by Schoemaker [28]. Over the last five years, it became apparent that the acceptance and successful implementation of system dynamics models relied on the modelling process followed with stakeholders, the effectiveness of engagements as well as the identification of non-intuitive leverage points. The adapted system dynamics modelling process resulted in several system dynamics models being implemented after development, and bridges the gap between model development and successful implementation and execution of the models. This paper explains each of the steps in the adapted system dynamics modelling process using a case study of electric vehicle market penetration in South Africa's provinces and its substitution of internal combustion engine vehicle (ICEV), as a function of affordability based on real disposable income.

Table 1: Summary of Steps in a System Dynamics Modelling Process

SOURCE	STEP 1	STEP 2	STEP 3	STEP 4	STEP 5
Richardson & Pugh III, 1989 [19]	**Problem identification and system conceptualization:** Define the problem, model purpose, list key variables and reference modes, and causal loop diagrams.	**Model formulation:** Rate equations and variable parameters.	**Model testing and further development:** Sensitivity runs, refinement, and validation.	**Policy analysis:** Dominant feedback loops, and the most sensitive parameters identified.	-
Sterman, 2000 [15]	**Problem articulation:** Define problem, key variables, time horizon, and historical reference modes.	**Dynamic hypothesis:** Identify theories around hypothesis, model boundary chart, sub-system diagram, causal loop diagram, stock and flow maps, and policy structure diagrams.	**Formulation of the model:** Specify structure & rules, estimate parameters, and test for consistency with model purpose.	**Testing:** Compare to reference modes, and robustness in extreme conditions.	**Policy Formulation and Evaluation:** Specify scenario, what if and sensitivity analysis, and determine synergies between policies.
Randers J. , 1976 [21]	**Conceptualization:** Definition of question asked, real-world behaviour, system boundary through feedback loops (CLD), and identifies system descriptors.	**Formulation:** Detailed model structure, and parameter specification.	**Testing:** Model assumptions, and model behaviour sensitivity to policies.	-	-
Coyle, 1996 [22]	**Problem recognition:** Who cares, and why.	**Problem understanding and system description:** Develop influence diagrams.	**Qualitative analysis:** Innovative ideas, and theories.	**Simulation modelling:** Use of software, and includes model testing.	**Policy testing & design:** Sensitivity testing resulting in insights and policy optimisation.

3. ADAPTED SYSTEM DYNAMICS MODELLING PROCESS

Eskom SOC developed an Integrated Strategic Electricity Plan (ISEP), which includes complex dynamics such as increasing generation availability alongside generation expansion, ensuring a financially and environmentally sustainable business, and a consolidation of socio-economic contributions to ensure economic development in the country and in Africa [29]. These documented complexities require extensive time to read through and filter out salient points so that mental models can be established by decision makers, based on perceived contextualisation. Mental models are unique to the individual and based on their theoretical knowledge, business experience and intuitive deductions [30].

Various modelling methods and processes are used to support the development of the organizational strategies [14]. System dynamics was introduced as an additional modelling process, incorporating group model building which could provide causal linkages and feedback loops for hundreds of variables and provide understanding of the system problem by allowing various sensitivity and scenario analysis [26].

The adapted eight step system dynamics modelling process includes some group model building elements [31] including system conceptualization, model formulation and decision making, with an emphasis on those practical aspects of project scoping, model communication and knowledge transfer necessary for the implementation of models and modelling solutions (Figure 2).

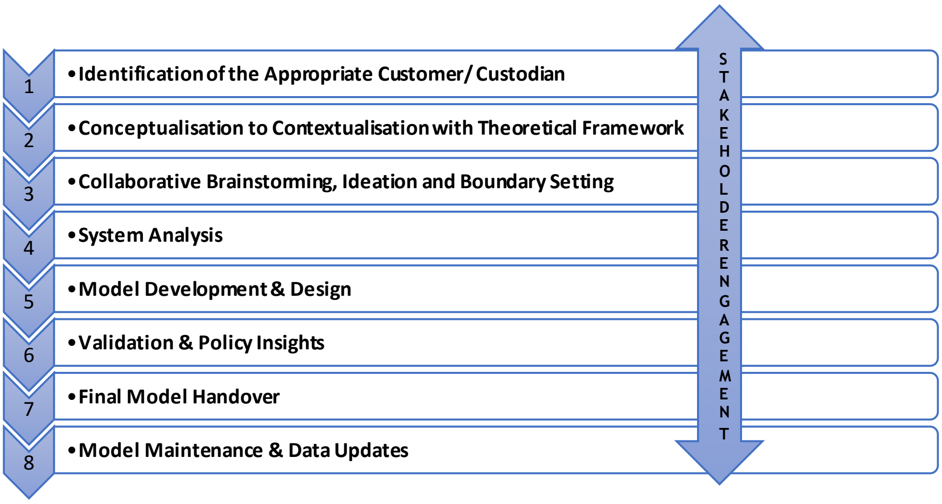


Figure 2: Adapted System Dynamics Modelling Process

3.1 Identification of the Appropriate Customer/ Custodian

A project may be initiated by a customer or be proposed by a system dynamicist (based on their experience, technical knowledge and understanding of the company's corporate objectives and business priorities). If the system dynamicist proposes a system dynamics model for development, they identify a potential custodian who is on a executive level with sufficient business influence. The custodian then nominates a technical owner who would be trained in using the completed tool, running the relevant scenarios and reporting the results to the senior decision makers. Generally, the initial challenges for a system dynamicist is being presented with broad scopes to deliver on specific results using a system dynamics model.

The system dynamicist has to engage with stakeholders (sometimes on multiple occasions) until a focussed question is defined to address a particular business system problem. Generally, the customer has conceptualised what they think should be developed through their mental models but experience difficulty in contextualising their ideas and formulating a focussing question. For customers/custodians with no prior system dynamics knowledge, successfully completed system dynamics tools are presented to the potential customer, relevant to their business interest with a discussion of how the results from the tools have been used for strategic cost benefit analysis or enhanced system understanding for improved decision making. Part of this step also involves identifying an interdisciplinary work group to engage with on a regular basis. The work group members could consist of the system dynamicist, the custodian and or customer, technical engineering members, environmentalists, financiers, subject matter experts etc.

3.1.1 Case Study

This study was initiated by the system dynamicist after discussions with external stakeholders (Postgraduate University supervisors) and proposed to the company. It was approved by the company, as sponsored further studies, since it aligned with the company's business strategy and corporate plan. This paper focusses on the approach which was taken to develop the simulator known as E-StratBEV, which was used to run scenarios and determine the linkage between BEV market penetration and affordability based on disposable income.

3.2 Conceptualisation to Contextualisation with the Theoretical Framework

In supporting the customer to contextualise the project requirements, they are advised to suggest typical graphical outputs or variables they would like to understand by running the model. This does not prescribe a preconceived result in terms of the emergent model solution since the trend or graphical result may be non-intuitive and unexpected but it allows reflection on the system variables which may have to be reported on, and establishing a focussing question. Part of contextualising includes establishing a suitable modelling timeframe. The timeframe provides insight in determining the resolution of data which would be required. Depending on the resolution of data, different data owners or workgroup members are identified for further liaising.

3.2.1 Case Study

The following focusing question and objectives were established for the BEV study:

Table 2: Focusing Question and Objectives of the Battery Electric Vehicle Study

Focusing question	Objectives
How does disposable income affect the affordability in the residential sector and market penetration of BEVs in South Africa?	• To develop electricity supply and demand side sub-modules, with a focus on the residential sector electricity consumption profile. • To factor in income distribution and consumer affordability causally with residential electricity consumption • To determine the charging requirements of BEVs after ICEV substitution and the related carbon emissions on a provincial level.

The timeframe for the simulator was 2015 [32] until 2050, to coincide with the timelines for the Green Transport Strategy [33], and the Integrated Energy Plan [34]. The South African provinces included the Eastern Cape province, the Free State province, Gauteng province, KwaZulu-Natal province, Limpopo province, Mpumalanga province, North West province, the Northern Cape province and the Western Cape province.

The system dynamicist then constructed a diagrammatic framework based on the operational and theoretical information linked to the system problem being modelled. This framework illustrated the high level system architecture map of the system problem and the related environment, in support of problem contextualisation (Figure 3). It did not display cause and effect relationships or directional quantities linked to the variables but included important upstream and downstream variables, driving forces and externalities specific to that environment. The value of such a map is that it helped with engaging with the stakeholders on a first pass and proved to be very effective. It assisted the system dynamicist in acquiring the necessary basic system problem understanding on a more technical level since it required extensive literature reviews before it could be constructed. It was, thereafter, through engagements, verified by the customer as being the correct final framework meeting the project specification. This step required high level assumptions to be agreed upon and possible proxies where no data was available for quantitative mathematical equations.

3.2.2 Case Study

In Figure 3, the electricity supply and demand modules were developed and then used to determine the reserve margin. The supply module comprised of various generation options (fossil fuel, nuclear, gas, hydro, renewables) and was used to calculate the carbon emissions within the coal heavy supply mix. The vehicle module was made up of BEVs (requiring electricity for charging) and ICEVs (which also contributed to carbon emissions). The disposable income in the residental sector was used to calculate vehicle affordability whilst various drivers were linked to the BEV module to assess influences on BEV market penetration.

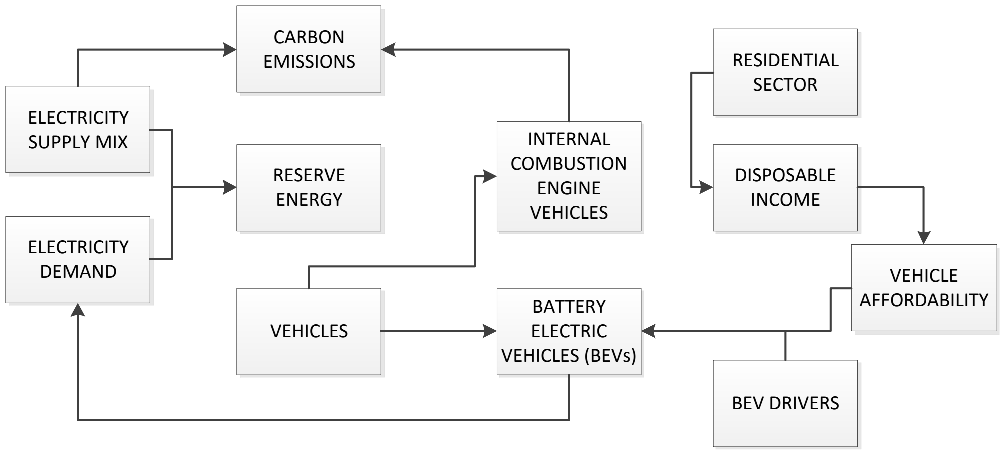


Figure 3: System Architecture Map

3.3 Collaborative Brainstorming, Ideation and Boundary Setting

This step is consultative and involves a group of role players, some of whom were identified to be part of the work group team in prior steps. The constitution of this collaboration and ideation team is critical and relates back to group model building exercises [23] since the output can be fairly subjective i.e. developing a causal loop diagram (CLD). The quality of subjective outputs such as the CLDs depend largely on the collective knowledge and experience of the participants hence the emphasis on carefully selected participants. The system dynamicist facilitates the interactive and carefully managed session and directs the discussion around those aspects related to the system problem, by steering the group clear of emotionally charged arguments. This process is usually started by using sticky notes, one per variable, and through group discussions the causal linkages are made between the variables on the sticky notes. The CLD is usually finalised in a follow up session with the group after the system dynamicist completes the first draft. CLDs may be revised and be part of an iterative dynamic process over the life of the project. At this stage, the system dynamicist is equipped with a very clear idea of the variables that would be included in the model boundary chart. The stakeholder discussions also help finalise the assumptions necessary for further work and those variables which may be excluded due to the required customer-defined project scope.

3.3.1 Case Study

Figure 4 illustrates the CLD for the BEV case study. For a big electricity supply-demand gap, the supply increases to close the gap, which then also results in less supply required (*Electricity Supply-Demand: Balancing Loop 1*). The Gross Domestic Product (GDP) served as a proxy for the economic health of the country. When the economy experiences growth (in the services, transport, agriculture, forestry and fishing, industrial sectors), the gap between supply and demand of energy becomes smaller. When the gap gets smaller, the GDP may be influenced by balancing factors which slow down economic growth (*Economic Growth: Balancing Loop B2*). In South Africa this supply is met by the predominantly coal-based base load. An increase in GDP usually results in employment which supports the drive to a more equal income distribution. In the residential sector, this means more average real disposable income per household. An increase in real disposable income means consumers can afford purchasing motor vehicles, which supports transport sector growth and the overall economy (*Income Distribution: Reinforcing Loop R1*). An increase in ICEV sales results in more carbon dioxide emissions. In the national drive to decarbonize our environment, an increase in carbon emissions in the transport sector incentivizes consumers to purchase more electric vehicles. The introduction of BEVs increases the demand for electricity required for charging and increases the supply-demand gap (*Transport Sector: Reinforcing Loop R2*).

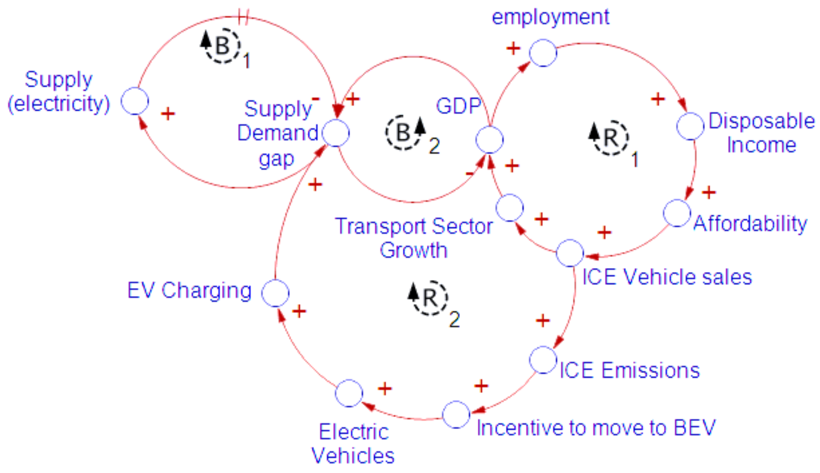


Figure 4: Causal Loop Diagram

The Model Boundary Chart (MBC) (see Table 3) lists the exogenous, endogenous and excluded variables for this study.

Table 3: Model Boundary Chart

EXOGENOUS VARIABLES	*ENDOGENOUS VARIABLES*	*EXCLUDED VARIABLES*
• Electricity supply Options (MW) • Residential electricity demand (GW) • Disposable income (R) • BEV import taxes (R) • Registered ICEVs (Number)	• Carbon Emissions (Mtons) • BEV charging requirements (kWh) • Affordable BEVs (Number)	• Infrastructure • Weather • Politics

3.4 System Analysis

Experience has shown that there are many organizational misconceptions that the system dynamics modelling software constitutes the system dynamics modelling process. The reality is that significant time should be spent on problem identification and contextualisation, as well as data and system analysis, before commencing with the structural design of the system dynamics model. The analysis could include statistical methods or programmable codes to determine patterns or relationships or simply involve a process of ordering and simplifying both qualitative and quantitative data into a time resolution suitable for importing into the system dynamics model. The empirical data may have gaps which may require classical regression, time series decomposition, least squares approximation, numerical interpolation or exponential smoothing or a combination of data analysis techniques. Perpetual linear growth trends are dismissed on the premise that real system elements have biophysical constraints and if the appropriate time period has been selected, these trends tend to plateau and reflect the carrying capacity of the system. This step is also critical since it can help the system dynamicist establish if any integration errors or incorrect structural linkages have been made which may result in large variances in the results once the modelling software is used. Preliminary calculations have also assisted in the initiation and development of smaller system dynamics models, which have been used as stand-alone tools for some customers.

3.4.1 Case Study

In this study, due to the huge volume of hourly BEV data, preliminary data mining was necessary to filter the 2015 BEV data obtained from the Eskom-Nissan BEV pilot study [35] so sub-routines (MS Excel Macros) were written to allow filtering and ordering of the data before further mathematical computations. The historic real disposable income on a national level was used to determine the future trend nationally using Equation 2, where x is time.

$$Future\ Real\ Disposable\ Income = -27.165 \times x^3 + 1{,}732 \times x^2 - 4{,}824 \times x + 888{,}155 \qquad (2)$$

3.5 Model Development & Design

The model structure may be designed to represent component configurations on the power plant, causally and mathematically linking the variables identified in the CLD, the MBC and the system architecture map. Once the model has been developed, the Beta version is usually handed over to the model owner and customer. They engage with the interface and run various scenarios, a process which does not require in-depth knowledge of the model structure. It has been found that only when the model is run by the model owner, do they fundamentally tune into the model and pick up elements which they would like to have changed, despite having various demo's by the system dynamicist during previous meetings. The additional "tweaking" and refinement of the simulation is then concluded.

3.5.1 Case Study

The empirical number of registered ICEVs per year in the South African provinces was obtained [36] from year 1999 until 2017 and an S-curve (Equation 3) was built into the model structure to determine the future trend of actual ICEVs until year 2050 shown in Figure 5.

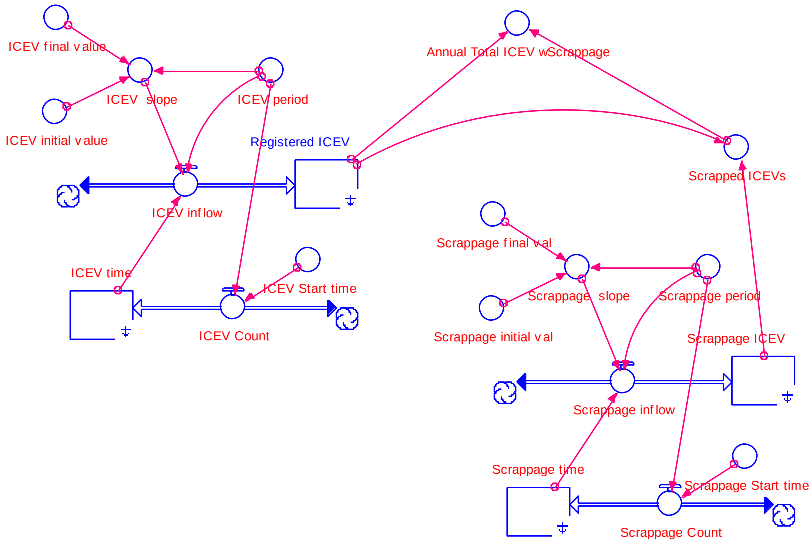


Figure 5: Model Structure of Actual Motor Cars

Equation 3 allows for asymptotic conversion to lower values, by specifying a negative value for U_1, or a positive stabilizing non-zero value by retaining a positive value for U_1.

$$P(t)=U_0+\frac{U_1}{1+exp[-c(t-t_0)]} \qquad (3)$$

where P is the dependent variable and $P(t)$ is a function of time t; U_0 is the zero offset; U_1 is the ultimate increase (or decrease) above U_0, modelled using a S-curve; c is a growth rate exponent that determines the maximum slope of the S-curve; and t_0 is the time at which the maximum slope is reached (inflection point).

The scrappage rates built into the structure obtained from derivations by Bento et al. [37] using data from 1987 to 2014 and the average vehicle age. For this study an average vehicle age of 11 years was used [38] and a corresponding scrappage rate of 13.84%.

Once the actual ICEVs per province was calculated, the average monthly income per proportion of households with a motor vehicle was used to calculate the affordable number of ICEVs using the percentages in Table 4, sourced from the South African Institute of Race Relations [39]. The percentages were applied for all the provinces however, the disposable income and the number of households were specific to each province.

Table 4: Monthly Income and Proportion of Households with ICEVs

Monthly Income Category	*Proportion of Households with a Motor Vehicle*
Up to R799	2.8%
R800-R1 399	4.5%
R1 400-R2 499	2.9%

R2 500-R4 999	9.3%
R5 000-R7 999	20.7%
R8 000-R10 999	44.7%
R11 000-R19 999	75.2%
R20 000	145.3%

Source: Van Heerden [40]; IRR, Eighty20, XtracT based on AMPS [39]

The gap between the actual and affordable ICEVs was linked to an ICEV Correction Factor which accounted for the umbrella of influences e.g. vehicle financing schemes, which result in consumers purchasing more ICEVs than they should be able to afford.

The target number of BEVs was 233,700 BEVs from 2019 until 2040 for South Africa using the GDP parametric (South Africa's GDP is on average 0.0057 of the world GDP), and a global BEV target of 41 million by 2040 [41]. GDP was used to make the provincial distributions due to BEVs still being classified as luxury goods, if drivers and incentives were already in place to make BEVs more affordable then disposable income could have been used as a measure. Thereafter BEV substitutions with the actual ICEVs were made on a provincial level and the resulting impact on residential electricity consumption and carbon emissions calculated. The ICEV Correction Factor was used to adjust the BEV market penetration to what the affordable number per province was expected to be.

3.6 Validation & Policy Insights

In this step, the final validation of the model is carried out. Besides the work group members, any internal and external parties with an interest in the model are contacted to run through the model scenarios and calibrate according to experience and new information which may emerge, a process which allows for theoretical and empirical consistency checks [42][43]. Empirical consistency includes comparing the simulation results to historical data and ensuring closeness of fit. Validation is also carried out by evaluating other models, results and assumptions on related work and comparing the project results.

3.6.1 Case Study

Figure 6 shows the difference between the actual and affordable number of ICEVs in 2015 and expected in 2030 and 2050. Gauteng province has the highest average difference between Actual and Affordable ICEVs by 2050 (4.252 mill), followed by the Western Cape province (1.545 mill) and KwaZulu Natal province (1.174 mill). All the other provinces have smaller differences between the Actual and Affordable ICEVs (between 0.108 mill and 0.519 mill). The reasons for this difference could include access to many readily available vehicle finance schemes such as balloon payments (which allows for reduced monthly instalments for the period of the credit agreement with an inflated final instalment due to the capital amount not being settled), fixed interest rates (credit agreement linked to the dynamic change in the prime interest rate) and instalment sale agreements (credit agreement that allows consumers to spread the capital amount plus interest over a set period).

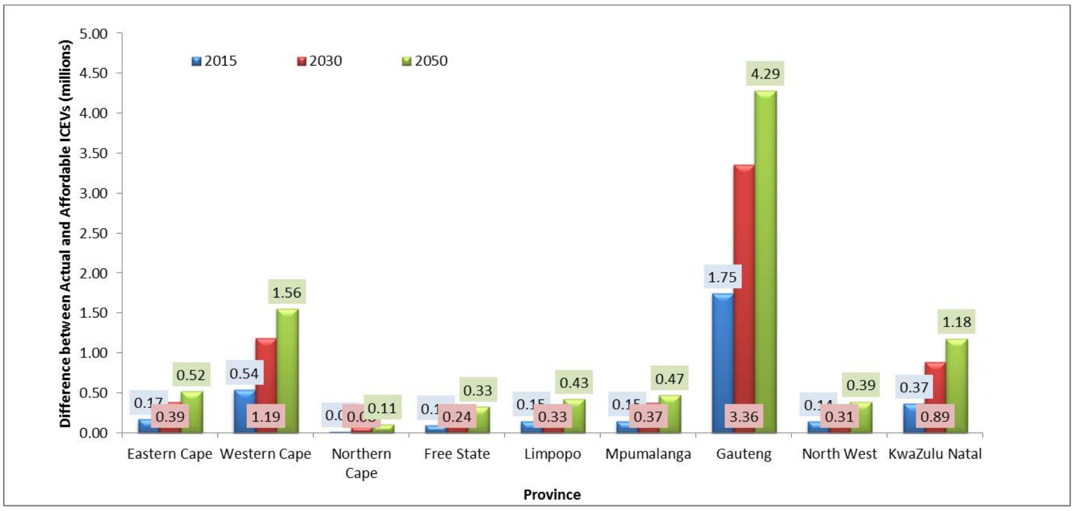


Figure 6: Provincial Distribution of ICEVs (Actual and Affordable) per Province

Figure 7 shows the BEV distribution per province after substitution with ICEVs and then after adjusting with the ICEV Correction Factor to obtain an expected number of BEVs based on affordability. Gauteng province would be expected to have the highest number of BEVs (81,123) followed by KwaZulu Natal (37,220) and the Western Cape (32,220).

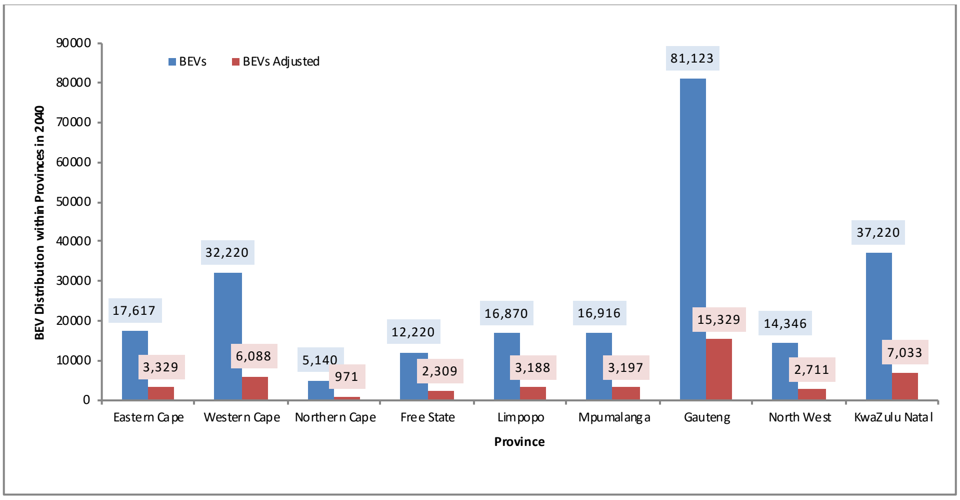


Figure 7: Provincial Distribution of BEVs with Adjustments Based on Affordability

The adjusted BEV distributions were about 80% less than the direct substitution with ICEVs based on original target.

Figure 8 shows the impact of the BEVs on the residential electricity consumption. Gauteng province has the highest BEV impact on residential consumption (adding an additional 4,291 GWh), followed by KwaZulu Natal province (an additional 1,969 GWh), then the Western Cape (an additional 1,704 GWh).

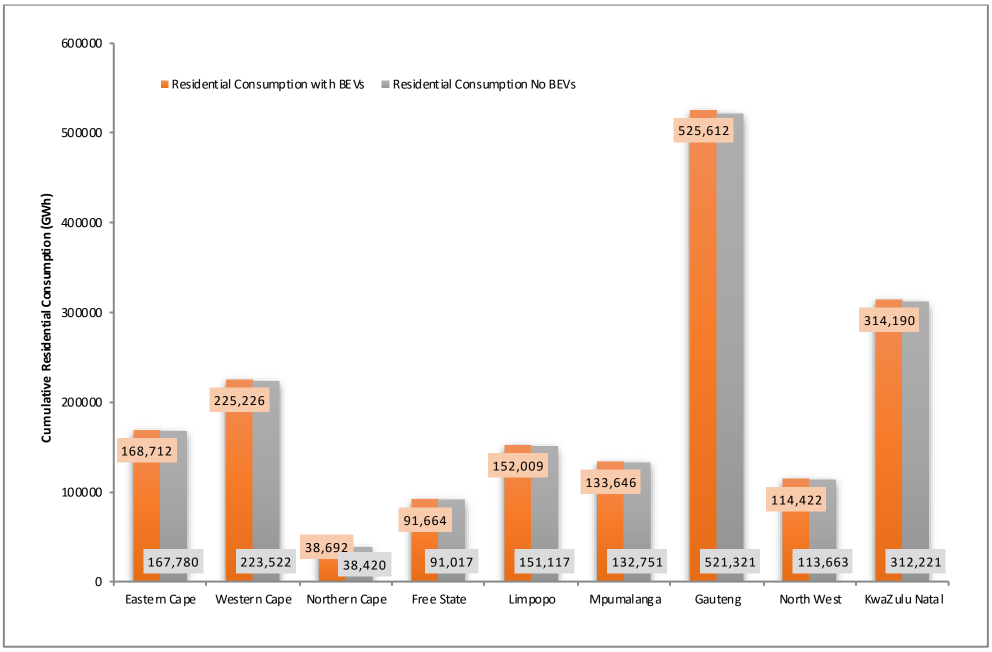


Figure 8: Cumulative Residential Electricity Consumption (2019-2040)

The cumulative change in carbon emissions from 2019 until 2040 was calculated on a provincial level, Table 5.

Table 5: Cumulative Change in Carbon Emissions (Mton) from 2019 - 2040

Province	Energy Sector	Transport Sector
Eastern Cape	0.39	-0.34
Free State	0.28	-0.19
Northern Cape	0.13	-0.18
Western Cape	0.70	-0.80

Gauteng	1.80	-2.00
KwaZulu Natal	0.78	-0.65
Limpopo	0.41	-0.20
Mpumalanga	0.39	-0.30
North West	0.35	-0.21

The net effect in terms of carbon emissions increasing in the energy sector and decreasing in the transport sector is negligible, most likely due to the carbon heavy supply mix and the fact that emissions calculations in the transport sector were from tank to wheel and not well to wheel.

3.7 Final Model Handover

The model handover stage is a formal step to ensure that the simulation results are checked against the original scope of the project and have been completed. Various training sessions are arranged with the model owner to ensure independent running of the model. The handover stage is officially minuted and signed off by the system dynamicist and the customer. Post the handover stage, the customer generally identifies additional minor model changes or additions which could enhance the optimal running of the tool.

3.7.1 Case Study

For this study, the *e-StratBEV* was handed over to the Eskom eMobility team to use in further scenario analysis.

3.8 Model Maintenance & Data Updates

If the amendments identified by the customer, post the final handover stage, require significant structural model changes, then a new project is again started, however, if the changes are minor such as data updates or quick model changes, then these are covered under Model Maintenance. A record is kept of the date of model completion and the interval dates/frequency when the maintenance should take place. This step minimizes potential frustration the customer's side and ensures long term use of the tool.

4 CONCLUSIONS

The adapted system dynamics modelling processes, which included elements of group model building necessary for strategy implementation, proved to be an effective and rigorous modelling process for the practical development and implementation of system dynamics tools for use in an electricity utility. Certainly, although the modelling method provided some quantitative insights into regional variations of the impact and expectations of BEV market penetration in South Africa, it still provides more value if used for exploratory modelling and descriptive comparative scenario analysis as opposed to predictive system behaviour down to the last decimal place.

Looking at the gap between actual and affordable vehicles, nationally, it would appear that South Africans are living beyond their "income" constraints and purchasing far more vehicles than what their disposable income allows, with the situation worsening over time, due to mechanisms encouraging consumers to live on credit, such as vehicle finance schemes. The results indicate that the Gauteng province and the Western Cape province have the largest potential to absorb BEVs (and the highest impact on residential electricity consumption) whilst the lowest is Limpopo province, based on GDP distributions. Future projects to build BEV charging infrastructure should consider these dynamics. To benefit from a reduction in carbon emissions in the transport sector, a renewables heavy supply mix would be required.

5. ACKNOWLEDGEMENTS

The author would like to thank Eskom SOC, Research Testing & Development for permission to make use of data and to provide technical reviews on the project progress; as well as the Eskom System Dynamics Centre of Expertise for support on model validation.

6. REFERENCES

[1] Eskom. (2017, May). *About Electricity*. Onttrek April 12, 2018 uit Eskom: http://www.eskom.co.za/AboutElectricity/FactsFigures/Documents/GX0001GenPlantMixRev19.pdf

[2] Ervural, B. C., Evren, R., & Delen, D. (2018, March). A multi-objective decision-making approach for sustainable energy investment planning. *Renewable Energy*, 387-402.

[3] Krishnan, V., Ho, J., Hobbs, B. F., Liu, A. L., McCalley, J. D., Shahidehpour, M., et al. (2016, May). Co-optimization of electricity transmission and generation resources for planning and policy analysis: review of concepts and modeling approaches. *Energy Systems, 7*(2), 297-332.

[4] Dyner, I., & Larsen, E. R. (2001, November). From planning to strategy in the electricity industry. *Energy Policy, 29*(13), 1145-1154.

[5] Foley, A. M., O Gallachoir, B. P., Hur, J., Baldick, R., & McKeogh, E. J. (2010, December). A strategic review of electricity systems models. *Energy, 35*(12), 4522-4530.

[6] Yan, Q., Wang, X., Balezentis, T., & Streimikiene, D. (2018, March 21). Energy-economy-environment (3E) performance of Chinese regions based on the data envelopment analysis model with mixed assumptions on disposability. *Energy & Environment*.

[7] Rafique, M. M., Ali Shakir, M., & Chohan, G. Y. (2018, January). Statistical and Econometric Estimation of Current and Projected Energy Utilization in Pakistan. *NFC IEFR Journal of Engineering and Scientific Research, 5*(2).

[8] Perez-Quiros, G. (2018, Januray-March). Discussion of the paper “Deciding between alternative approaches in macroeconomics”. *International Journal of Forecasting, 34*(1), 139-141.

[9] Baudry, G., Macharis, C., & Vallee, T. (2018, January 1). Range-based Multi-Actor Multi-Criteria Analysis: A combined method of Multi-Actor Multi-Criteria Analysis and Monte Carlo simulation to support participatory decision making under uncertainty. *European Journal of Operational Research, 264*(1), 257-269.

[10] Bairamzadeh, S., Saidi-Mehrabad, M., & Pishvaee, M. S. (2018, February). Modelling different types of uncertainty in biofuel supply network design and planning: A robust optimization approach. *Renewable Energy, 116*(Part A), 500-517.

[11] Garwood, T. L., Hughes, B. R., Oates, M. R., O'Connor, D., & Hughes, R. (2018, January). A review of energy simulation tools for the manufacturing sector. *Renewable and Sustainable Energy Reviews, 81*(Part 1), 895-911.

[12] Axelrod, R. (1997). Advancing the Art of Simulation in the Social Sciences. In R. Conte, R. Hegselmann, & P. Terna (Red.), *Simulating Social Phenomena. Lecture Notes in Economics and Mathematical Systems. 456*, ble. 21-40. Springer, Berlin, Heidelberg.

[13] Moon, Y. B. (2016, August 22). Simulation modelling for sustainability: a review of the literature. *International Journal of Sustainable Engineering, 10*(1), 2-19.

[14] Snabe, B., & Grobler, A. (2006, September). System dynamics modelling for strategy implementation—case study and issues. *Systems Research & Behavioural Science, 23*(4), 467-481.

[15] Sterman, J. D. (2000). *Business Dynamics: Systems Thinking and Modelling for a Complex World.* McGraw-Hill.

[16] Pillay, N. S. (2014, April). Using System Dynamics to Explore Gini Coefficient Parametrics. Cape Town, South Africa: University of Cape Town.

[17] System Dynamics Society. (2016, July-December). In Memorium Jay W. Forrester 1918-2016. *System Dynamics Review, 32*(3-4), 185-186.

[18] Forrester, J. W. (2009, July). Learning through System Dynamics as Preparation for the 21st Century. MIT.

[19] Richardson, G. P., & Pugh III, A. L. (1989). *Introduction to System Dynamics Modeling*. United States of America: Pegasus.

[20] Systems Thinker. (2011, February). Building Blocks:Guidelines fir Causal Loop Diagrams. *22(1)*. Pegasus.

[21] Randers, J. (1976). A Framework for Discussion of Model Conceptualisation. *The 1976 International Conference on System Dynamics* , 414-452. Norway: System Dynamics Society.

[22] Coyle, R. G. (1996). *System Dynamics Modelling: A practical approach* (Vol. 1). Florida, USA: Chapman & Hall/CRC.

[23] Vennix, J. A. (2000, February). Group Model Building: Tackling Messy Problems. *System Dynamics Review, 15*(4), 379-401.

[24] Maly, T., & Petzold, L. R. (1996, February). Numerical methods and software for sensitivity analysis of differential-algebraic systems. *Applied Numerical Mathematics, 20*(1-2), 57-79.

[25] Hekimoğlu, M., & Barlas, Y. (2010). Sensitivity analysis of system dynamics models by behavior pattern measures. *The 28th International Conference of the System Dynamics Society*. Seoul: System Dynamics Society.

[26] Forrester, J. W., & Senge, P. M. (1979, June 8). *Tests for building confidence in system dynamics models*. Onttrek April 17, 2018 uit Clexchange.org: http://static.clexchange.org/ftp/documents/roadmaps/RM10/D-2926-7.pdf

[27] Richardson, G. (1991). System Dynamics: Simulation for Policy Analysis from a Feedback Perspective. In P. A. Fishwick, & P. A. Luker (Reds.), *Advances in Simulation* (ble. 144-169). New York: Springer.

[28] Schoemaker, P. H. (1995). Scenario Planning: a Tool for Strategic Thinking. *Sloan Management Review, 36*, 25-40.
[29] Eskom SOC. (2017, March 31). *Eskom's Integrated Report*. Onttrek April 17, 2018 uit Eskom: http://www.eskom.co.za/IR2017/Documents/Eskom_integrated_report_2017.pdf.
[30] Jones, Natalie A; Ross, Helen; Lynam, Timothy; Perez, Pascal; Leitch, Anne. (2011). *Mental Models: An Interdisciplinary Synthesis of Theory and Methods*, Ecology and Society,16(1), Resilience Alliance Inc.
[31] Andersen, D. F., & Richardson, G. R. (1997, Summer). Scripts for Group Model Building. *System Dynamics Review, 13*(2), 107-129.
[32] Parker, C. (2017, August). Assessing the Viability of e-Mobility: Assessing the Viability of e-Mobility. Johannesburg: Frost & Sullivan.
[33] DOT. (2016, August). *Draft Green Transport Strategy: (2017-2050).* Retrieved January 17, 2018, from Chemical & Allied Industries Association: https://www.caia.co.za/wp-content/uploads/2016/08/Draft%20GTS%20Revised%20v5.pdf
[34] DOE. (2016, November 22). *Integrated Energy Plan (IEP).* Onttrek January 19, 2018 uit Department of Energy: http://www.energy.gov.za/files/IEP/presentations/Integrated-Energy-Plan-22-Nov-2016.pdf
[35] Langley, P. (2016). Electric Vehicle Results. *Report 1789591*. Johannesburg, South Africa: Eskom Research Testing and Development.
[36] ArriveAlive. (2017). *Registered Vehicles*. Onttrek October 30, 2017 uit Arrive Alive: https://www.arrivealive.co.za/document/Nov_PDFS/Nov2006_RegVehs_LicPrDP_Fuel_Dist1.pdf
[37] Bento, A., Roth, K., & Zuo, Y. (2016, January 18). *Vehicle Lifetime Trends and Scrappage Behaviour in the U.S. Used Car Market*. University of California [Online] http://faculty.sites.uci.edu/kevinroth/files/2011/03/Scrappage_18Jan2016.pdf. [2018, June 6].
[38] DOT. (2004, September). *Age of Vehicle*. Arrive Alive [Online] https://www.arrivealive.co.za/document/04AgeofVehicle.pdf. [2017, November 11].
[39] IRR. (2014). Eighty20, XtracT based on AMPS 2014. *Individual January-December 2014 data*. Johnson, R. M. (2014, March 24-25). *Integrated Resource Planning with PLEXOS*. Onttrek April 17, 2018 uit Energy Exemplar: https://energyexemplar.com/wp-content/uploads/publications/2014_EnergyExemplar/Energy%20Exemplar%20Morocco%20Integrated%20Planning%20V2.pdf
[40] Van Heerden, G. (2016). Assets & Inome 2016. IRR.
[41] Meyer, P. (1994). Bi-Logistic Growth. *Technological Forecasting and Social Change, 47*, 89-102.
[42] Randall, T. (2016, February 25). *Here's How Electric Cars Will Cause the Next Oil Crisis*. Retrieved June 12, 2018, from Bloomberg New Energy Finance: https://www.bloomberg.com/features/2016-ev-oil-crisis/
[43] Sterman, J. D. (1984, Winter). Appropriate summary statistics for evaluating the historical fit of system dynamics models. *Dynamics, 10*(2), 51-66.

DEMAND FORECASTING FOR NETWORK CAPACITY PLANNING IN ELECTRICAL UTILITIES – A REVIEW OF EXISTING METHODS CONSIDERING THE EVOLVING TECHNOLOGIES OF THE ENERGY ARENA

J. Breedt[1*], L. Louw[2], I.H. de Kock[3]

[123]Department of Industrial Engineering
Stellenbosch University, South Africa
[1]Jana.Breedt@eskom.co.za, [2]louisl@sun.ac.za, [3]imkedk@sun.ac.za

ABSTRACT

Planning for sufficient energy resources in a country is of paramount importance to ensure sustainable development of the economy and prosperity of its citizens. In South Africa the national utility, Eskom, is tasked to create a balance between the electricity demand and the supply thereof. Forecasting the electricity load on the networks to supply the country demand becomes an important task to ensure that capacity planning does not constrain potential growth, and neither does it construct overinvestment to compromise feasibility of implementation. The landscape of energy utilization is currently experiencing rapid evolution in technology and poses significant challenges to the way the electricity demand forecast needs to be done. Technology is evolving to provide more efficient, cost effective and reliable alternative energy sources than the conventional methods used in the past. Improved electricity efficiency and user behavior plays a significant role in future electricity demand requirements. This paper provides a comparative literature review on current forecasting methodologies to provide insight to which of these methods can be utilized in the future. A set of requirements is concluded on to identify the most relevant and effective forecasting methodologies to improve accuracy on forecasting electricity demand into the technology advanced future.

[1] The author is enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa
*Corresponding author

1. INTRODUCTION: AN EVOLVING POWER SECTOR

The turn of the millennium indicated the start of a new era in which the fundamentals of the power sector began to change. For most of the past century the electric utilities of the world was mostly stable, fairly predictable and commonly operated in a vertically integrated supply chain. Supply of electricity requirements was mostly operated by conventional power stations (mostly coal fired, gas or hydro generation plants) which was then transferred through a high voltage transmission grid to a distribution substations where it is transformed to lower voltage distribution networks for residential and industrial uses. The utilities had mostly predictable load growth and therefor predictable sales which was governed and periodically adjusted to keep utilities reasonably profitable. A significant change in technology within the power sector changed the stable state of the power sector. New technologies for generation of electricity, especially on the customer side of the value chain started to emerge. The term 'prosumer' was derived from the original word customer, which indicates a customer that can use energy and simultaneously play an active participant in the market by supplying energy back into the networks [1]. This changes the configuration of the network operations significantly.

Throughout the ages, industrialisation brought technological leaps and this led to paradigm shifts which today is referred to as "industrial revolutions". This took us from mechanisation (the 1st industrial revolution), to the extensive use of electricity (the 2nd industrial revolution), then the widespread digitalization (the 3rd industrial revolution), and then the latest revolutionary changes in technology advancement brought us to a new fundamental paradigm. With terms such as the "Internet of things", revolutionary energy generation and advanced digitalisation of industries and "smart" technologies, it brings us to the 4th industrial revolution that is now looming [2].

The world economic forum in collaboration with Bain and Company did an excessive study on the future of technologies influencing the grid. They argue that the fourth industrial revolution builds on the legacy of the digitalisation of the third industrialisation where multiple technologies are combined to takes us to a paradigm shift in economy, business, society and the individual. This paradigm shift involves transformation of entire systems and in particular also within the electricity landscape where the scope of the power sector is becoming more complex than ever before, with rapidly changing technologies in power generation and storage thereof, emerging innovative business models and shifting regulatory landscapes. It is imperative to support a reliable, economically competitive and environmentally sustainable electricity system as a corner stone for a modern society into the future [3].

In the book, Future of Utilities and Utilities of the future [1], it was highlighted that there is a growing concern as to which end of the scale the electric power sector would tilt. Some experts is of the opinion that the utilities as they are known currently is at a death spiral form where there is no escape which can bring the end of utilities as they are currently known. Specialists on the other side of the spectrum are of the opinion that the world is at the start of a new evolution and that the many challenges the sector face should be embraced. A revolutionary future is envisaged even though significant changes is required within the power sectors' network operations, business models, how it is regulated and its culture. A recent study by Accenture indicated that by 2025 the annual revenues of US utilities may potentially be $48 billion lower than they would have been otherwise due to the rapid growth of distributed energy generation sources and gains in energy efficiency. The same study indicates demand of the grid to lower by 15% relative to current status quo. European utilities are likely to experience the same, coupled with a current decline in conventional demand growth. Other studies concluded that the rapidly falling prices of Photo Voltage (PV) technologies will ensure an increase in uptake of this technology in both commercial, industrial, agriculture and domestic use. It is further projected that globally fossil fuels might be dominated greatly by renewable energy generation by 2050 [1].

The U.S Energy Information Administration (EIA) publishes a yearly International Energy Outlook Overview with global energy predictions up to 2050. The EIA classifies the world countries into Organization for Economic Cooperation and Development (OECD) members and non-members (non-OECD). According to EIA the non-OEC countries indicate a stronger demand growth in energy needs as these countries show strong economic growth, increased access to marketed energy and quickly growing populations (these include African countries, China, India, Brazil to name a few). These countries face challenges with both forecasting the growing demand and engineering systems around evolutionary new technologies. On the other side of the equation is the OECD countries (OECD Europe, Canada, United States, Japan, Australia and New Zealand amongst the top players) where energy demand is on a decline or a slower growth due to increase in energy efficiencies. The going concern on these networks is changing of structures and regulation to include new technologies and systems, rather than plan for demand growth [4].

These facts are all an indication that renewable and alternative energy sources is showing great possibility in changing the conventional power delivery methods, and possibly the demand on the utility electricity grid. Taking all these fundamental paradigm shifts and increasing technological evolution in consideration, the question emanates - how does this influence the design of the current utilities?

Increasingly, utilities will be facing challenges with network designs evolving into decentralised from centralised, on how customers will be connected to the grid, whether they will be customers or prosumers, and how these structures would be governed and regulated to still provide a resilient and profitable organisation. This is all aspects that form the future scenarios of how we can design electricity grids going forward. However for the foreseeable future a prominent, and very important function is still needed as the starting point of all these aforementioned planning activities: forecasting the network demand. Such a demand forecast is complex and increasingly difficult to formulate taking the amount of uncertainties and external factors such as technological advancement, political stability, population growth of the country and governing structures of the power sector into account.

Form a local perspective the focus then moves to the South African utility, Eskom Holdings SOC Ltd. (Eskom). The forecasting of the electricity demand as a fundamental input to the network capacity planning processes should be defined and refined to obtain the most suitable method(s) for forecasting the future. Such methods is needed to forecast electricity demand to create equilibrium between generation sources and the national demand within the electricity supply chain. A supply chain once vertically integrated and now finding a notion towards horizontal integration. These forecast needs to serve as input to network development plans that will enable necessary infrastructure development for economic growth empowerment.

This article provides a brief literature review on forecasting techniques available for long term demand forecast to assist in determining the most suitable approach in forecasting future demand, taking in consideration the range of external factors created by advanced technology development and changing of energy sources from conventional to alternative sources. The study aims to enable Transmission network capacity forecasters in Eskom to provide more accurate and appropriate forecasting for adequate network development plans.

2. BACKGROUND ON FORECASTING DEMAND FOR THE SOUTH AFRICAN ELECTRICITY GRID

2.1 Current Transmission Demand Forecasting in Eskom

Eskom is the current custodian and main contributor to South Africa's base electricity load and plays a fundamental role in meeting South Africa's energy requirements. Eskom currently supplies 97% of the country's demand and therefor as the principal transmission and distribution licensee in South Africa, Eskom is responsible for developing and maintaining the country's transmission and distribution infrastructure as well as South Africa's interconnections with the Southern African Power Pool. The transmission grid planning forecast is produced as input to the Transmission Development Plan (TDP), which essentially contributes to the capacity planning of the transmission networks in order to fulfil the vision set by the National Development Plan (NDP) for the energy sector of South Africa. The NDP is a governmental document that sets growth targets for all South African sectors. The NDP's main aim is to eliminate poverty, reduce inequality and grow the South African economy. The current edition is aimed at growth targets to be achieved by year 2030 [5].

Transmission network studies require long term forecasts with at least a ten year horizon, however a twenty-year plus horizon is preferred. A Transmission demand forecast has the purpose to identify future load growths and highlight possible network constraints in need of expansion [6]. Taking in consideration that the energy sector is undergoing a paradigm shift, it is imperative that the forecasting needs to adapt to provide accurate results.

The system in which Eskom operates its production can currently be classified as a balanced vertically integrated supply chain as most of its production comes from owned subsidiaries. This gives Eskom the competitive advantage of currently being the monopoly energy provider to Southern Africa. However, capacity constraints from 2007 coupled with economic downturn created cause for panic. Alternative energy sources and advanced technologies in renewable generation become a prominent trigger to the formation of a horizontally integrated supply chain going into the future, in order to sustain market share and competitive advantage. Eskom facilitates a renewable integration programme where independent power producers (IPP's) can receive licences and generate power to use and sell back on to the national grid [6].

It is imperative that any forecast is done within the right context given the purpose it should serve. Within Eskom, a number of forecasts exist, each driven to satisfy a specific business need. However, it is important that although different forecasts serve different purposes and business needs, such forecasts should be aligned and managed from a holistic perspective to not only achieve the overall objectives of the utility, but to also adhere to the electricity demand of a country, and in specific, align to the goals set, in the case of South African utility Eskom, these goals were set by government in the NDP [5].

The Eskom transmission demand forecast is currently done by using a top down approach where South Africa's total electricity demand is taken into account within a mathematical model and balancing algorithm. Methodologies are continuously updated to forecast the demand. The methodology currently consists of both quantitative and qualitative forecasting techniques with the S-curve methodology applied predominantly. The methodology used in Eskom was developed by Dr DF Payne in 2004 and is continuously researched and improved to suit the growing complexity of the energy utility operations [7].

Figure 2-1 gives a brief overview on the current forecast process followed in the Eskom transmission demand forecast process. A top down approach is followed where national load is disaggregated to the transmission (Tx) station points within the networks. Spatial demand areas are translated to load capacity needs per Main Transmission Substation (MTS) for capacity design engineering purposes. A supply and demand model is used to better understand future expected demand growths in the complex and highly uncertain environment. A mathematical model is then used to solve a non-linear programming problem when allocating loads to substation nodes. The demand per forecast area is validated by evaluating relevant market intelligence and external information such as economic analysis and other growth factors [7]. Bottom up forecast validation is then done with alignment that is done on distribution forecasts and interaction with the distribution operating areas. This interaction is facilitated by use of the Geo Spatial Forecasting (GLF) forecasting tool used by the distribution forecasters for the bottom up forecast

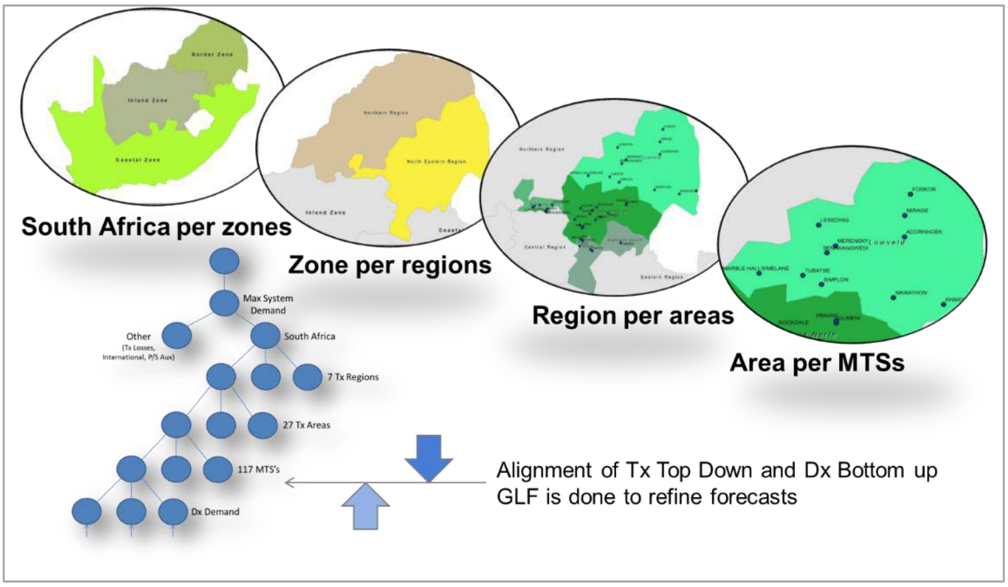


Figure 2-1: Current top down methodology for Eskom Transmission Demand Forecast

A schematic design was derived to show the context of the current Eskom supply chain, and can be seen in figure 2-2. Prior to 2008 the Eskom system was a closed supply chain vertically integrated with only the conventional coal, nuclear, single hydro plant and selected co-generation from key customers. Additional to this the Direct Current (DC) line built from Caborra-bassa was also accounted for within the supply system.

The national grid is split into distribution networks (Dx) ranging from 22kV to 132kV networks, which is then connected to the larger transmission networks (Tx), designed for 257kV to 765kV lines to transport power over larger distances.

Figure 2-2 is a schematical representation of the network conectivity landscape until recent changes in generation resources and adding of the independent power producers programme (IPP) to the national grid. The diagram shows the simplistic verticaly integrated supply chain. Generation supply (mostly conventional) is connected to the grids, which is in turn connected to the national demand. This is a fairly uncomplicated and linear landscape for which forecasting was fairly predictable by using trends of the past and qualitative and quantitative approaches to determine the forecast.

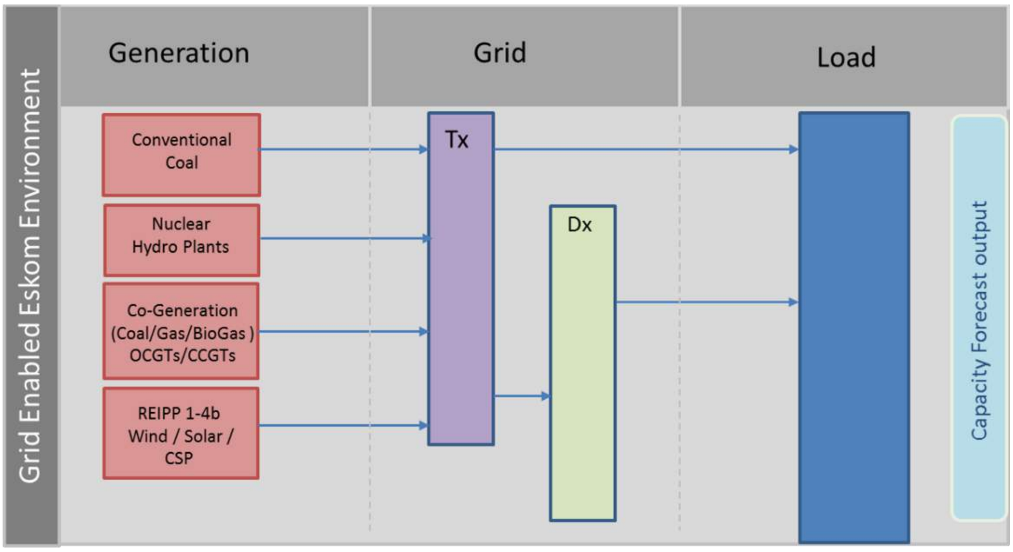


Figure 2-2 : Network connectivity diagram indicating the vertically integrated supply chain without unconventional a generation source

2.2 Future considerations within the Transmission Demand Forecasting in Eskom

In his article, 'Gridlocked', Hedden [8] investigates the future scenarios of the South African Grid. He discussed the loadshedding South Africans experienced between 2008 and 2014 , and argues that this was seen as an electricity crises at the time. A significant amount of pressure was put on the South African Department of Energy (DoE) to search for ways to increase generation capacity and implement demand side management programmes to reduce annual load usage. New coal-fired station power plants are in construction, and oil and gas are explored as a possible energy source. New Independent power suppliers (IPP's) are contributing towards energy supplies. This is currently increasing the blurring of the line between consumers to prosumers. Hedden therfor states that the electricity sector in South Africa is changing exponentially with the influence of renewable energy, IPP's and small-scale embedded generation (SSEG).

When considering a smarter grid scenario into the future, Hedden [8] forecasts that there will be a need for the grid to enable renewable and embedded generation sources in such a manner that the benefits is distributed by all users of the grid. A higher penetration of renewable energy benefits will increase economic growth and sustainability.
Such a scenario implicates that the future of the Eskom grid will increasingly include current initiatives such as the renewable programs (REIPP), gas (for both combustion power stations and residential and industrial use), micro and privately owned grids, self-generation, electric vehicles and more alternative energy sources.

It can be argued that the future forecasting scope of grid capacity demand forecasting might change, considering the impact of self-generation or off grid capabilities provided by alternative energy sources (such as gas) and alternative generation of energy (such as renewables). The network connection landscape is then expanded with a section that can be either connected to the grid on temporary basis or be deflected and functiononing totally off-grid. The alternating users will typically make use of PV technology or wind turbines for energy generation, however if the source (such as wind or sun) is unavailable beyond the current storage capability of the generator. These customers would return to the utility for energy needs and should therefor still be accounted for in the base load predictions. Currently most of the official generation sources feeds back into the national grid and can be seen as an import or supply onto the system.

As technology evolves and electricity tarriffs increases, customers are moving increasingly towards grid deflection where they utilise generation technologies in order to function independently from the utility grid. It becomes evident then that a certain portion of the load of the country is not supplied by the local grid, and is seen as self-sufficient and a form of grid deflection can be seen. Currently this is a very small portion of the national load, however this is fast changing when taking in consideration that the 4th industrial revolution is at our front door and the technologies that it brings changes. The advanced technological improvements and the sustainability thereoff will have an impact on the proportion of grid serviced load and the customer differentiation that will cause future load forecasting to look substantially different. Figure 2-3 gives a representation of the network conectivity landscape taking self generation, grid deflection and unconventional generation sources in consideration. The figure shows where there is posibility for excess generation to be evacuated through the utility grid and therefor changes the capacity forecasting scope, as indicated by the arrows following back into the grids. When this is the case, the forecaster should now take in consideration not only the spatial load to be supplied, but also the excess generation to be evacuated through the grid. The excess generation that needs to be evacuated will require capacity on the networks, and therefor needs to be included intor the capacity forecast per transformer.

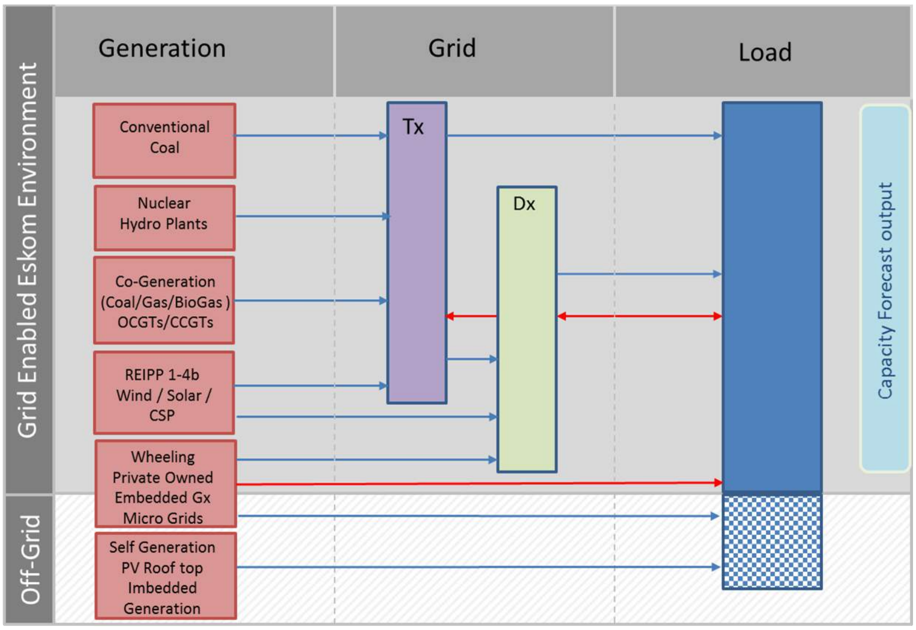


Figure 2-3: Network connectivity diagram indicating a horizontally integrated supply chain with unconventional generation sources included.

Figure 2-4 gives a schematic prediction on the author's sentiment on how the forecasting scope for national grid capacity development might differ from the current moving into the future. The perception is that currently there is a portion of customers fully relient on the national grid, a portion that is seen as variable and can be selfsuffiecient at certain times of the day, however they return as soon as ther generation reliability decreases. Then lastly there is a small portion currently off grid. It is foreseen that in future these segments will change in proportion to one another, however it is still unpredictable to what extend they will change. This creates further uncertainty with the forecast of capacity for utility planning purposed.

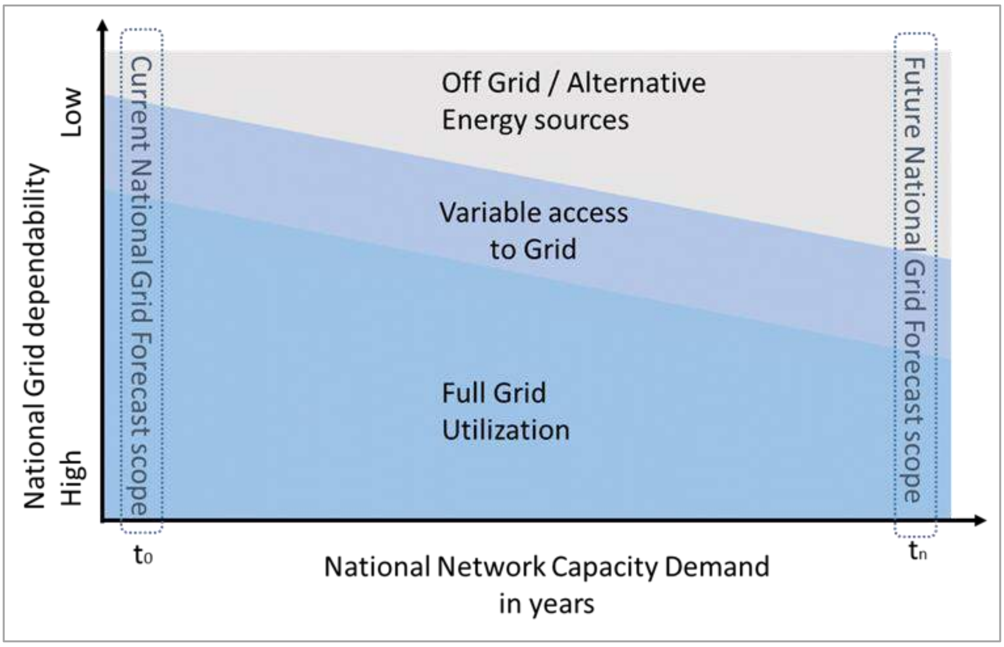


Figure 2-4: Schematic prediction on the forecasting scope into the future

Kessides, Bogetic and Maurer [9] identifies a number of current and forthcoming issues in the South African electricity sector where they highlight the difference in location of load centres and future generation sources. It can be argued that once the alternative energy sources reaches mature levels of production, it might have the need to feed back into the national grid in order to "sell" energy supply to the national utility. If we consider the South African generation landscape, where conventional power supply is currently located in the north of the country (close to the current load centres), and most renewable energy natural resources such as wind and sun is most prevalent to be set up in the south of the country (far from the major load c entres), a need for transportation, import and evacuation of generation becomes a great grid planning c onsideration [9].

Figure 2-5 shows that both demand drivers and generation sources can have an impact on the grid demand forecast going into the future. First of all it is the spatial area load that needs to be forecasted, then secondary effects on the required grid capacity caused by excess generation should be considered.

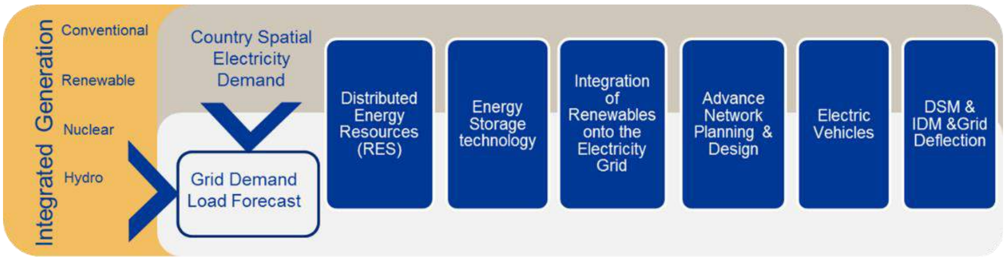


Figure 2-5: Identified possible influencing factors to the Grid Demand Forecast

There is a great need to evaluate the current available forecasting methodologies and accompanying forecasting techniques to ensure optimal forecasting for the growing complexities of the power sector. The following section will give a brief overview on available technologies.

3. A LITTERATURE REVIEW ON FORECASTING METHODOLOGIES

3.1 Demand Forecasting Defined

Dr Hans Levenbach [10] has devoted his career into finding appropriate demand forecasting practices to facilitate demand planning within any supply and demand domain. In his book Change & Chance embraced, he defines demand forecasting as a process with the main objective to predict future events or conditions. He further explains that forecasting attempts to predict change in the presence of uncertainty. He argues that there is much more surrounding forecasting than using past trends to predict the future. This could be agreed to if the changes within the power sector is evaluated. The future will certainly change from the past conventional power sector. The challenge is to distinguish what combination of techniques is available in literature to combine into a framework for the technological advanced future. Care should be taken to distinguish which areas of the past can still be utilized qualitatively to apply to similar growth areas into the future [10].

3.2 Forecasting electricity demand

Similar to most manufacturing industries where supply and demand needs to be balanced, the power sector can be related to a supply and demand model. In a load forecasting case study done by Hong, and Shahidehpour [11] they express an opinion that electric utilities run the power grid, to deliver energy. It is in simple terms the business of turning electrons into electricity for consumption by the user, a simple demand and supply problem. In the current electrical industry, energy cannot be massively stored with the technologies at hand. Therefor electricity cannot be considered as a product that can be stored on a shelf and sold for future use, it is rather a product that is sold at real time and the supply and demand should be balanced every operating moment. This increases the complexity of forecasting the product and ensuring adequate demand can be met at all times.

With the future in mind, it can then be argued that with technological advancement electricity storage might influence how the demand and supply will be balanced. It can be agreed that the power system planning of electricity demand and supply is seen as a non-linear system with a number of influencing factors and role players rather than a simple linear system with inputs and outputs. Hong further reasons that load forecasting is used through all segments of the electric power industry, including generation, transmission and distribution network planning, and in the sales and revenue planning arena. He stresses the importance of load forecasting, as inaccurate load forecasts might result in financial burden [11].

3.3 Evaluation of forecasting methods and techniques available

When deciding on which forecasting methods and techniques to use, the forecaster should take care to analyze what is the purpose and goal of the forecast and what data is available. Chambers, Mullick and Smith [13] published a paper on how to select the right forecasting technique. They substantiated that the forecaster's selection of a method depends on many factors such as the context of the forecast, the relevance and availability of data, the accuracy of the forecast, the time period for which the forecast should be applied and timelines for analysis. Chambers *et al*. [13].

From a selection of articles on forecasting techniques, it can be reasoned that forecasting techniques can be grouped into eras of development. The first group of techniques can be grouped into "Pre-Personal Computer" area which pertains strongly to techniques used mainly before the 1980's. As the computer industry ramped up and the personal use of computers became more accessible, a new set of forecasting techniques was launched, these where technological advanced and computer driven to allow increased processing capability. Spatial forecasting techniques was also introduced in this era of intelligent methods. Thirdly the latest group of methods can be classified as strategic or smart grid era techniques leading us to intelligent and scenario based forecasting technique.

Hong *et al.* [11] explains the word "technique" as a group of models that fall in the same family, such as Multiple Linear Regression (MLR) models and Artificial Neural Networks (ANN). A forecasting technique is explained by Levenbach as a simplified representation of a real world situation [10]. On the contrary, "methodology / methods" represents the general solution framework that can be implemented with multiple techniques [11]. Figure 3-1 shows a summary of the technique categories identified.

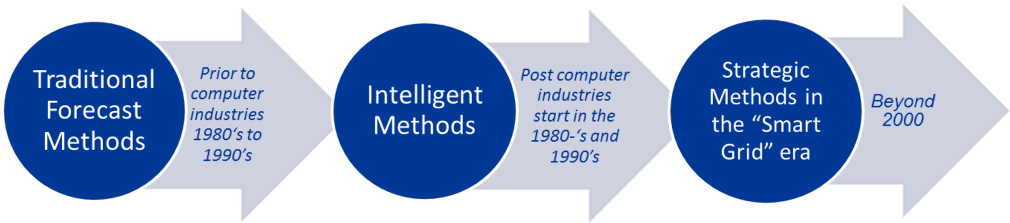


Figure 3-1: Forecast Techniques grouping according to era

The forecasting techniques found in literature will now be discussed and grouped accordingly. Various literature sources were sourced. Sioshansi [1], Levenbach [10], Hong *et al.* [11], Chambers *et al.* [13], Lempert [14].

3.4 Traditional Forecast Methods (Prior to computer industries 1980's to 1990's)

The very first energy forecast was done when Thomas Edison provided streetlamps to Pearl Street Station in New York in 1882. Hong highlighted that equation was simple, and the amount of lights was counted to calculate the demand to meet. Following the development of the energy implementation in more than streetlights, Hong describes that engineers had to make use of tables and figures to calculate forecasts as no statistical packages where invented yet [11].

The traditional forecast methods are grouped by Chambers et al. [13]as three basic types – Quantitative and qualitative techniques, time series analysis and projection, and causal models. Each of these has a number of techniques relating to the set of models. Some of these models used no computer processing, whilst others were at the very start of computer involvement.

3.4.1 Quantitative and Qualitative Methods

Quantitative techniques is mostly suited for short term forecasts where rigorous data analysis is done mostly by statistical and deterministic approaches. Regression methods and statistical analysis is often used where good current data is available. Chambers *et al.* [13].

Qualitative methods are more often used where there is not much data available and the expert opinion needs to be applied. This method is sub divided by Chambers *et al.* [13] into the following forecasting techniques:

3.4.1.1 Delphi Method

With this method a panel of experts is used to analyze and interrogate information. This is often done by questionnaires, interviews or by panel discussion. This method has a fair to good reputation on forecasting accuracy for both short and long term forecasting. It is typically applied to forecasts of long-range and new-product sales, as well as forecasts of margins. Chambers *et al.* [13]

3.4.1.2 Market Research

This technique includes the systematic and formal process to test hypothesis about real markets. This technique is excellent for use in short term and fair to good for use in long term forecasting. In this technique Chambers *et al.* describes that a considerable amount of market data is needed. Chambers *et al.* [13]

3.4.1.3 Panel Consensus

This is one of the most widely practiced qualitative method. This technique assumes that a panel of experts can derive at a better forecast than one person. Panel members is free to contribute their opinion on the forecast. A disadvantage of this technique is that the forecast is often influenced by social factors and my not reflect a true consensus. This is not a good method for long term forecasting according to Chambers *et al.* [13]

3.4.1.4 Visionary Forecast

This technique relates to the Delphi ad Panel Consensus techniques, however is rather based on personal visions and judgment. It is therefore not a very reliable forecasting technique and is not widely used for either short or long term forecasting. Chambers *et al.* [13]

3.4.1.5 Historical Analogy

This is a comparative analysis that uses past trends to introduce similar growth patterns into the future. This is not a strong technique to use for short term forecasting, however it is fair for use in long term forecasting. It is often used for new product forecasts. Chambers *et al.* [13]

3.4.2 Time Series

Levenbach [10] defines a time series as a set of historic data which is chronologically ordered such as energy consumed per hour for an extended period. The following techniques can be summarized under time series models.

3.4.2.1 Moving Average

Chambers *et al.* [13] describes moving average as a time series where each point of a moving average is the arrhythmic or weighted average of a number of consecutive points in the series. A number of data points is selected in order to eliminate effects of seasonal or irregularity. It is a very poor technique to use for long term forecasting. Chambers *et al.* [13]

3.4.2.2 Exponential smoothing

This technique is proven to be similar to the moving average by Chambers *et al.* except that more recent data points are allocated more weight. Descriptively the forecast is the same as the past, with some proportions of the past error. There is many variations of exponential smoothing according to Chambers *et al.* [13]

3.4.2.3 Box-Jenkins

Box-Jenkins technique is explained as a mathematical model that is fitted to a time series and is optimal in the sense that it assigns smaller errors to history than any other model. This is one of the most accurate traditional statistical routines and is excellent for short term forecasting, however it is very poor for the use in long term forecasting. Chambers *et al.* [13]

3.4.2.4 Trend Projections

This technique is one of the very first used and fits a trend line to a mathematical equation and then projects it into the future by means of the used equation. There are a few different types of these trends such as polynomial, logarithmic, and so on. This is a good technique to use in long term forecasting. Chambers *et al.* [13]

3.4.3 Causal Methods

Chambers *et al.* argues that causal models takes into account everything known about the forecasting system and utilizes information such as sudden events that could have influenced the data. They further argue that causal models are the best for predicting long term forecasts. Chambers *et al.* [13]

3.4.3.1 Regression Models

It is stated that these techniques relate strongly to sales and other economic variables. This technique estimates an equation using the least squares techniques. Relationships are primarily analyzed statistically and quantitative conclusions can be made. This is very good for short term forecasting however it is unsuitable for the use in long term forecasting. Chambers *et al.* [13]

3.4.3.2 Econometric Models

Ghods and Kalantar [15] describes the econometric approach as one that combines economic theory and statistical techniques for forecasting electricity demand. The approach approximates the relationship between energy consumption and factors influencing consumption. Ghods *et al.* [15]. It allows for multiple interdependent regression equations that describes the economic and profit activity. If excellent data is available this can be a good technique to use in long term forecasting.

3.4.3.3 Input-Output Models

Chambers *et al*. [13] describes these models as techniques used to describe the flow of a process where input parameters are transformed to certain outputs. These models can be enhanced to create economic input-output models where this technique is combined with the econometric models. This is argued as an excellent technique for long term forecasting.

3.4.3.4 Life-Cycle Analysis

This is explained by Chambers *et al*. as an analysis and forecasting done for a new product development growth rates based on S-curve techniques. This is a good technology for both long and short term forecasting [13].

3.5 Intelligent Methods (Post computer industries start in the 1980-'s and 1990's)

Hong states that computer applications started to ramp up after the 1980's and a significant amount of research was placed into computer based techniques as well as spatial load forecasting. This introduced the concepts of load forecasting regarding where the load was coming, when it will be anticipated and how much load growth is anticipated.

3.5.1 Trending, Simulation and Hybrid methods

Trending methods use mathematical functions to fit the past load growth to the future load growth. Hong *et al*. [11] explains that this technique is mostly applied by using polynomial regression where the load is transferred to a function of trend and cycle. He further explains that simulation methods can also be excellent for long term use and is a technique that attempts to model the load growth patterns at play. Furthermore Hong *et al*. [11] explains how Hybrid models can be created to combine both trending and simulation to further accuracy on the forecast.

3.5.2 Artificial Intelligence

Hong *et al*.[11] explains that artificial intelligence models are based on artificial intelligent techniques such as Artificial Neural Networks (ANN), fuzzy logic, and black-box models to name a few.
Hong *et al*.[11] elucidates that ANNs have been extensively used for load forecasting since the 1990's. ANN is a soft computing technique that does not require the forecaster to explicitly model an underlying physical system. The system is enabled to learn from historical data. Mapping can be established between the input variables and the electricity demand, which can then be adopted for prediction.

Fuzzy logic is used for short term forecasting and can be used for network forecasting, and not ideal for long term forecasting as summarized by Suganthi and Samuel [16]. They further explain that neural networks can be used successfully to model electrical networks and the associated load points [16].

3.6 Future Strategic Methods in the "Smart Grid" era

Hong explains how the power sector is going through a grid-modernization process where technologies are changing and network planning needs new consideration. New forecast areas such as Demand-response and Renewable-Generation forecasting will need to be investigated going into the future. Hong [12].

The great amount of uncertainty in the future power sector leads us to the following forecasting techniques.

3.6.1 Systems Thinking and Complex Theory

Stoker and Fick [17] delivered a paper which analyses the current situation in respect of South Africa's electricity situation and asks the question if Systems Thinking, more specifically as it pertains to the theory of Complex Adaptive Systems, would make a contribution to improved Electrical Energy decision making. The authors concludes that the mental framework which Systems Thinking provides, and specifically the theory which explains the behavior of Complex Adaptive Systems, could make a positive contribution towards achieving quicker coherence in respect of South Africa's future Electricity System. Stoker *et al*. [17]

3.6.2 Scenario Planning

Amer, Dai and Jetter [18] produced a review on scenario planning. They highlight that the use of scenario planning techniques is very useful in an era where uncertainty, innovation and change are prominent. They reason that scenario planning stimulates strategic thinking and helps to overcome thinking limitations by allowing multiple outcomes. They further argue that scenario planning is a valuable tool that can assist organizations to prepare for possible eventualities and make them more flexible and innovative. Creating different scenarios delivers an overall view of the environment and highlights the interactions among several trends and events in the future.

They reason that scenarios can generally be used for any time frame, however it is most useful in long term forecasts. It has been proven that at corporate level scenario planning approaches was more popular and agile in larger companies and generally used for forecasting horizons of 10 years or more. Amer *et al.*[18].

3.6.3 Practical applications of strategic methods in literature

A few examples is available where scenario planning techniques was used in the power sector and Eskom specific. Hedden [8] produced three scenarios in his article 'Gridlocked' where he took South Africa from now to 2050 with his scenarios. He differentiated the three scenarios as the 'Current path scenario' where Eskom stays at the monopoly of the energy market, then he introduces the 'Efficient Grid scenario' in this scenario optimization of the grid is considered with investment in electricity generating capacity and ensuring efficiency of network operations. He then concludes with a 'Smart Grid scenario' where network efficiency and generation capacity is accompanied by integrated energy and grid planning to include renewables and decentralized electricity sources.

Another strategist in scenario planning is Cilliers [19] who investigated scenarios into the year 2035 with his article on South African futures. Cilliers [19] takes political circumstances in consideration and then proposes three scenarios. The first is a very successful Scenario called 'Mandela Magic' that alludes to a GDP growth rate of 5.1% pa. His second scenario is more to the conservative side, however still produces a 3.8% pa growth. And then he conclude with a lower scenario where he argues that the political instability South Africa is constantly experiencing will lead to a scenario called 'Nation Divided', with a nominal growth rate of 2.6% pa. Cilliers [19].

4. FACTORS INFLUENCING A DEMAND FORECAST

Other than the methods for creating forecasts, there is a number of factors external to the forecasting parameters that has an influence on the forecast. Some of these factors can be listed:

4.1 Technology and innovation

Musango and Brent [20] produced a study on technology in society. They argue that technology assessment is not new, however it is more relevant than ever in todays' energy market. They acknowledge that technology innovation is seen as a competitive advantage. They conclude that energy analysis tools is needed and should be researched further to enable the governing structures to identify and incorporate technological change into their plans. Musango [20].
In McKinsey's Quarterly report, Nyquist [21] reason that technology outliers should be monitored in order for the business to expect change and effectively deal with the change it might bring [21].

4.2 Energy efficiency

Energy efficiency is described by Hedden [8] as a parameter that is not necessarily linear to capacity requirement. He states that an increase in efficiency can lead to a decrease in cost which then in turn lowers the energy intensity. However, due to lower costs, demand usage can in fact go up. Energy efficiency and the change it can bring should be monitored closely.

4.3 Battery Storage

Mills, Barbose and Seel [22] produced a study on how to plan for distributed disruption, with the focus on incorporating solar into utility planning. They argue that the rapid growth of photovoltaic will continue to grow and that this will have critical implications for utility planning processes. They further explain how the storage component on this technology can affect the size and type of future infrastructure needed. This can have an immense effect on how we need to forecast and plan for future networks. As affordability of storage increases and technology advances, the effect will become apparent. Mills *et al.* [22].

4.4 Regulatory changes and economic stance

The world economic forum did a comprehensive study on finding operating models for the future. [4] They specify that governance is an important role player in how we drive change and agility. They state that Governance refers to the organizational accountabilities and the critical decisions that needs to take place by putting economic favorable models in place.
Cilliers [19] argues in his article that the heart of the economic failures lay at the inability of government, labor and business to chore around common growth visions. It is therefore inherently important that regulatory institutions and government work together to create a coherent vision for development of the power sector to enhance economic growth. He highlights how historical political climates in the different years where causes for failures and successes alike. Cilliers [19].

5. THE FUTURE OF FORECASTING IN A GROWING UTILITY

This article highlights the amount of uncertainty currently experienced in the power sector. There is a definite need to create a process for choosing, evaluating and executing the most accurate combination of techniques and tools available, whilst taking the ever changing environment into account.

A few forecasting models is available in literature to apply. These needs to be researched further and the applicability should be linked to the above mentioned techniques.

Suganthi *et al.* [16] elaborates on some energy models for demand forecasting, and one prominent model used is the LEAP model. The long-range Energy Alternatives Planning system was developed by the Stockholm environmental Institute and is a bottom up forecasting model. This is not ideal for the National Demand forecast that takes a top down approach from national perspective to transmission network points [16].

The initial review of the current electricity supply chain and the future challenges with advanced technology led to a set of initial research questions:

1. How will the demand forecast for capacity planning be influenced going into the new Power System Paradigm with the effects of Distributed Energy Resources and advanced technologies?
2. Which forecast techniques of the past is still applicable and agile enough to apply?
3. How can new complex thinking and current forecasting techniques be combined to guide forecasters to improve accuracy in the uncertain future?
4. What will the impact of unconventional generation be on the spatial forecast models, when considering evacuation of power from one area to another?

A set of requirements was identified and it is suggested that a model be designed to create a unique modelling approach for forecasting into the future. These aspects should be researched further and combined into a final forecasting framework for implementing within the current demand forecasting process:

1. Determine all possible influencing factors which can have an impact on the demand forecast, such as demand response and sustainable renewable generation forecasts.
2. Group impacting factors to be tracked and continuously revised for inclusion into forecast.
3. Evaluate advanced implementation models on long term electricity forecasting available from literature, aid in selection of the most suitable needs.
4. Combine available techniques and models to create custom framework to aid in building forecast scenarios. Framework suggesting optimal forecasting technique for identified forecast should be:
    a. Structured and include method selection criteria to facilitate forecasting.
    b. Should include methodology on incorporating renewable and decentralized generation sources into grid forecast.
5. Investigate tools to use for user interface to facilitate in combining different results to provide the forecaster with a holistic view of the influencing factors for the forecast.
6. Combine selected techniques with future scenarios to enable agile and robust forecast fitted to the scenario framework Add next – structured literature review to be done on application of techniques and models in long term energy forecasting

6. CONCLUSION

As the preceding discussions advocate, it is imminently clear that the electrical power sector is experiencing a paradigm shift and undergoing rapid change. It is important that the forecasting processes, models and techniques available is used optimally to ensure accuracy of the forecast as input to the planning of the electrical grid. A brief comparative evaluation was done on available and applicable techniques to gain better understanding in the type of models available from literature. Literature showed a strong notion towards strategic thinking and implementation of scenarios based forecasting. There is a need to further research and create a cohesive methodology that will be translated into a framework that optimally combines the most appropriate techniques for forecasting the demand of a determined scenario. The scenarios will be mostly dependent on the technology deployment, energy efficiencies and data available for use of the appropriate technique. It is suggested that an in-depth study is done to create a long term forecasting framework for demand forecasting of the electrical networks in South Africa.

7. REFERENCES

[1] Edited by Sioshansi, F. P. (2016) Future of Utilities Utilities of the Future. 1st edn. Elsevier B.V.

[2] Lasi H, Fet- tke P, Kemper H-G, Feld T, Hoff- mann M (2014) Industrie 4.0. WIRT- SCHAFTSINFORMATIK. doi: 10.1007/ s11576-014-0424-4.

[3] Astarloa, B, Kaakeh, A., Lombardi, M. (2017) 'The Future of Electricity: New Technologies Transforming the Grid Edge', World Economic forum, (March), p. 32. Available at: http://www3.weforum.org/docs/WEF_Future_of_Electricity_2017.pdf%0Ahttps://www.weforum.org/reports/the-future-of-electricity-new-technologies-transforming-the-grid-edge

[4] EIA (2017) 'International Energy Outlook 2017 Overview', U.S. Energy Information Administration, IEO2017(2017), p. 143. doi: www.eia.gov/forecasts/ieo/pdf/0484(2016).pdf

[5] National Development Plan 2030 | South African Government (2010). Available at: http://www.gov.za/issues/national-development-plan-2030 (Accessed: 16 June 2017).

[6] Eskom Holdings SOC Ltd official Website, accessed 11/05/2018 http://www.eskom.co.za/OurCompany/CompanyInformation/Pages/Company_Information.aspx

[7] Payne, D. F. (2004) 'Modeling of different long term electrical forecasts'.

[8] Hedden, S. (2015) 'Gridlocked: a long-term look at South Africa's electricity sector', African Futures, 15(September), p. 23. Available at: https://issafrica.s3.amazonaws.com/site/uploads/AF15_2.pdf

[9] 'Current and Forthcoming Issues in the South African Electricity Sector Ioannis N . Kessides (DECRG) Zeljko Bogetic (AFTP1) Luiz Maurer (AFTEG)' (2006), (March).

[10]Levenbach, H. (2017) Change & Chance Embraced : Achieving Agility with smarter forecasting in the Supply Chain, Copyright. Edited by F. copyeditor David Coen. Agile Forecasting. doi: 10.1007/978-1-4899-7578-2_5.

[11]Hong, T. and Shahidehpour, M. (2015) 'Load Forecasting Case Study', p. 171.

[12]Amer, M., Daim, T. U. and Jetter, A. (2013) 'A review of scenario planning', Futures. Elsevier Ltd, 46, pp. 23–40. doi: 10.1016/j.futures.2012.10.003.

[13]Chambers, J. C., Mullick, S. K. and Smith, D. D. (1971) 'How to choose the right forecasting technique', Harvard Business Review, 49(4), pp. 45–70. doi: 10.1225/71403.

[14]Lempert, R. J. (2003) Shaping the Next One Hundred Years: New Methods for Quantitative, Long-Term Policy Analysis. doi: 10.1016/j.techfore.2003.09.006.

[15]Ghods, L. and Kalantar, M. (2011) 'Different methods of long-term electric load demand forecasting; a comprehensive review', Iranian Journal of Electrical and Electronic Engineering, 7(4), pp. 249–259. doi: 10.1109/ICIT.2008.4608469.

[16]Suganthi, L. and Samuel, A. A. (2012) 'Energy models for demand forecasting — A review', Renewable and Sustainable Energy Reviews. Elsevier Ltd, 16(2), pp. 1223–1240. doi: 10.1016/j.rser.2011.08.014.

[17]Stoker, P. and Fick, J. (2014) 'Applying systems Thinking in shaping South Africa's Electricity System',Incose International Symposium, s4,pp. 437-452. doi: 10.1002/j.2334-5837.2014.00032.

[18]Amer, M., Daim, T. U. and Jetter, A. (2013) 'A review of scenario planning', *Futures*. Elsevier Ltd, 46, pp. 23–40. doi: 10.1016/j.futures.2012.10.003.

[19]Cilliers, J. (2015) 'South African Futures 2035 - Can Bafana Bafana still score?', Institute for Security Studies Papers, (282), p. 32. Available at: https://www.issafrica.org/uploads/Paper282.pdf.

[20]Musango, J. K. and Brent, A. C. (2011) 'Assessing the sustainability of energy technological systems in Southern Africa: A review and way forward', Technology in Society, 33(1–2), pp. 145–155. doi: 10.1016/j.techsoc.2011.03.011.

[21]Nyquist, S. (2015) 'Peering into energy ' s crystal ball', McKingsey Quarterly (July 2015) pp. 1–7.

[22]Mills, A., Barbose, G., Seel, J., Dong, C., Mai, T., Sigrin, B., Lawrence, E., Orlando

[23]Z, *et al*. (2016) 'Planning for a Distributed Disruption: Innovative Practices for Incorporating Distributed Solar into Utility Planning', (August), pp. 1–33. Available at: https://emp.lbl.gov/sites/all/files/lbnl-1006047-ppt_0.pdf%5Cnhttps://emp.lbl.gov/publications/planning-distributed-disruption.

IMPLEMENTING A REMOTE CONDITION MONITORING SYSTEM FOR SOUTH AFRICAN GOLD MINES

J.A. Stols[1], C. Cilliers[2] & J.F. van Rensburg[3]

[1]North-West University, CRCED-Pretoria, Pretoria, South Africa
astols@researchtoolbox.com

[2]North-West University, CRCED-Pretoria, Pretoria, South Africa
ccilliers@researchtoolbox.com

[3]North-West University, CRCED-Pretoria, Pretoria, South Africa
jvanrensburg@researchtoolbox.com

ABSTRACT

The South African gold mining industry is currently under a lot of financial pressure due to electricity tariff increases, strikes, increased minimum wages and volatile commodity prices. Besides decreasing electricity consumption, mining companies have started looking towards improved maintenance strategies to reduce the operational costs of mines.

There are a great variety of maintenance strategies being used in different industries worldwide. In the South African mining sector, maintenance strategies are however not being implemented according to any specific process. This also applies to the condition monitoring of machines, which serves as the foundation of a properly structured and efficient maintenance strategy. There are different possible causes for this, which are different for each mine. Some mines lack proper infrastructure for monitoring machine conditions, while others simply do not have the required technical resources to implement and maintain a proper condition monitoring system.

Specialised condition monitoring systems are also expensive to implement and maintain, while requiring continuous manual user input to analyse and identify machine deterioration. This paper will therefore focus on developing an implementation process for a basic remote condition monitoring system on South African gold mines. This process will include the implementation of an alarm notification process that can be used to structure the maintenance strategy.

The developed process was used to deploy the same basic condition monitoring system on different South African gold mines within the same mining group. The resulting system was further used to centralise condition monitoring data from these mines for easier access and simplified reporting.

This work was sponsored by ETA Operations

[1] Mr J.A. Stols holds a M.Eng degree in Mechanical Engineering from North-West University. He is currently a PhD student at the North-West University's Centre for Continued Research and Engineering Development (CRCED) in Pretoria.
[2] Dr. C. Cilliers holds a PhD in Mechanical Engineering from the North-West University. He is currently a post-doctoral student at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.
[3] Dr. J.F. van Rensburg is a registered professional engineer and holds a PhD in electrical engineering from the North-West University. He is currently a lecturer at the North-West University's Centre for Research and Continued Engineering Development (CRCED) in Pretoria.

1. INTRODUCTION

The world is witnessing one of the most significant technological evolutions in the history of mankind. The fourth industrial revolution is the result of the digitalisation brought on by the third industrial revolution. It is the fusion of technologies that connects the physical and digital worlds to optimise operational efficiency and reliability. This is hinged on the analysis of high volumes of data, which is the result of low-cost sensors and the decreasing price of computer processing power. This revolution is also evolving at such a rapid pace that it is disrupting all industries worldwide, including the mining industry. One of the most effected aspects of the mining industry, is the operational reliability of critical equipment through effective maintenance.

1.1 Background

Maintenance costs in the mining industry can account for 20 – 50% of the total operational costs of a mine [1]. The primary cause of lost production time and increased maintenance costs can be attributed to early component failures [2]. This means that optimising maintenance strategies through an efficient condition monitoring system can lead to significant savings in terms of both maintenance costs and production losses.

South African gold mines rely primarily on five types of mine support systems (MSS's) to keep underground working environments safe and enable sustainable mining operations. These MSS's are mine dewatering-, mine ventilation, compressed air-, refrigeration- and winder systems [3]. Due to a lack of SCADA integration of winder systems on most South African gold mines, they will not form part of the condition monitoring implementation process.

Since these systems are of such importance and mainly powered by machines, the failures of these machines must either be avoided, or attended to immediately when they occur. The most expensive maintenance strategy is a run-to-failure or corrective maintenance strategy and it should therefore be avoided as far as possible. Preventative maintenance on the other hand can either be time based or condition based and is done at specific intervals or when certain condition-based criteria are met. Maintenance resources on South African gold mines are also limited, which would make condition based maintenance more desirable since it is less resource intensive [4].

Condition based maintenance ideally requires a condition monitoring system to support it and improve its efficiency. However, condition monitoring equipment are very expensive to maintain and requires specialist knowledge to operate. The MSS machines are exposed to very harsh operating environments, which makes the use of specialised condition monitoring equipment to continuously monitor machine conditions in gold mines impractical.

There is thus a need for a more structured approach to condition monitoring in the South African mining industry. The best way of accomplishing a standardised condition monitoring process is by following a generic implementation process for a condition monitoring system. A process will thus be developed to implement a basic remote condition monitoring system on South African gold mines. Such a process needs to have minimal implementation costs that can put gold mines under more financial pressure. The system will therefore only use existing monitoring infrastructure and a minimal data logging resolution. The proposed data communication for the basic condition monitoring system that will be implemented is illustrated in Figure 1 below.

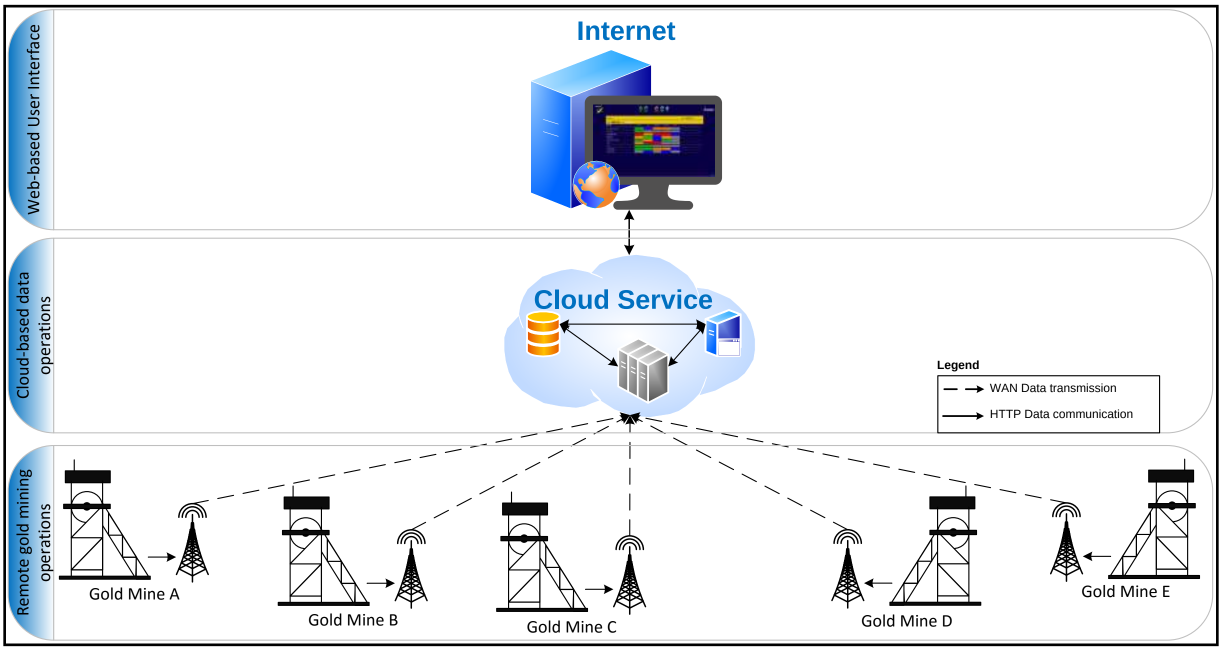


Figure 1: Condition monitoring data communication

The layout in Figure 1 shows that the system can be divided into three sections, which are the gold mines, the data processing and user interface. In the first section condition monitoring data is collected on individual gold mines by data loggers. The second section represents a centralised data processing server that receives, processes and archives collected condition monitoring data from various gold mines over a secured network. Lastly, section three illustrates a web-based user interface that uses the processed condition monitoring data to present machine and system conditions visually.

The condition monitoring data from the different gold mines will be sent over a secure GSM network to a central server for storage and processing as shown in Figure 1. It is however important to know how and where condition monitoring fits into the different types of maintenance strategies before developing a suitable implementation process.

1.2 Maintenance strategies

Maintenance can be categorised into two main categories, namely corrective maintenance and preventative maintenance. Corrective maintenance is done when a machine fails and is done to return it to its correct state. This type of maintenance cannot be planned or anticipated [5]. Corrective maintenance can further be sub-divided into two categories, namely deferred and immediate maintenance. Preventative maintenance can be defined as the care and servicing of equipment or facilities by personnel to keep them in a satisfactory operational state. Preventative maintenance is done either to prevent imminent failures or to keep them from developing into major failures. This type of maintenance can be planned according to its occurrence, either based on the condition of machines or predetermined intervals. Condition monitoring can therefore be considered as a supportive component of preventative maintenance [4].

Condition monitoring systems on South African MSS's would enable maintenance personnel to actively monitor machine conditions and enable them to do preventative maintenance. It would also notify maintenance personnel directly of machine parameters that are exceeding their predetermined limits. This would help mine personnel evaluate the system's health by monitoring the individual machines that make up the MSS's and prevent production losses due to machine failures. These MSS's are primarily driven by specific machines and configurations for each system.

Compressed air is generated on the surface by compressors ranging in size from 1 MW to 15 MW of installed capacity [6], [7]. Deep level gold mines have at least two main ventilation fans on the surface, with one constantly running. These ventilation fans can range in size from 100 kW to 3 MW each [8]. Refrigeration systems on deep-level mines use refrigeration plants, or fridge plants, to chill water to between 3 °C and 6 °C [9]. Mine dewatering systems use multistage centrifugal pumps to extract excessive water from mining areas and prevent flooding [10]. Winders are used in the South African mining industry to transport mine workers or reef vertically between mining levels and to the mine's surface [11]. The machines involved in these MSS's need to be properly maintained to ensure uninterrupted and sustainable mining operations.

The machines are monitored by sensors that transmit parameter values to the centralised Supervisory Control and Data Acquisition (SCADA) computer of each mine via field IO modules and a PLC [12]. Most of these sensors are installed upon the commissioning of such a machine, since it legally required [13]. A control system, such as a SCADA system, can be defined as one or more devices that are used to manage and control the function of other devices. SCADA systems are responsible for the collection and transfer of information to a central location. They are also used to do analyses and for the control of some processes while displaying the relevant information to an operator [12].

A typical control system consists of three basic components, which are a master SCADA station, remote assets like PLC's and a communication medium [12]. The SCADA computer can share process data with other computers over an OPC (Open Platform Communications) connection. OPC is a set of standards that allows for the reliable and secure data communication of devices from different vendors [14].

1.3 Generic CM system components

SCADA systems are not typically being used in the mining industry specifically for condition monitoring purposes, but for the remote control of processes and monitoring of process data. These SCADA systems along with other infrastructure can be incorporated into a basic condition monitoring system with minimal additional capital expenditure. A generic implementation process is however required to do this properly, which will also help identify shortcomings in existing infrastructure. A layout of the basic components required to implement a condition monitoring system is shown in Figure 2 below.

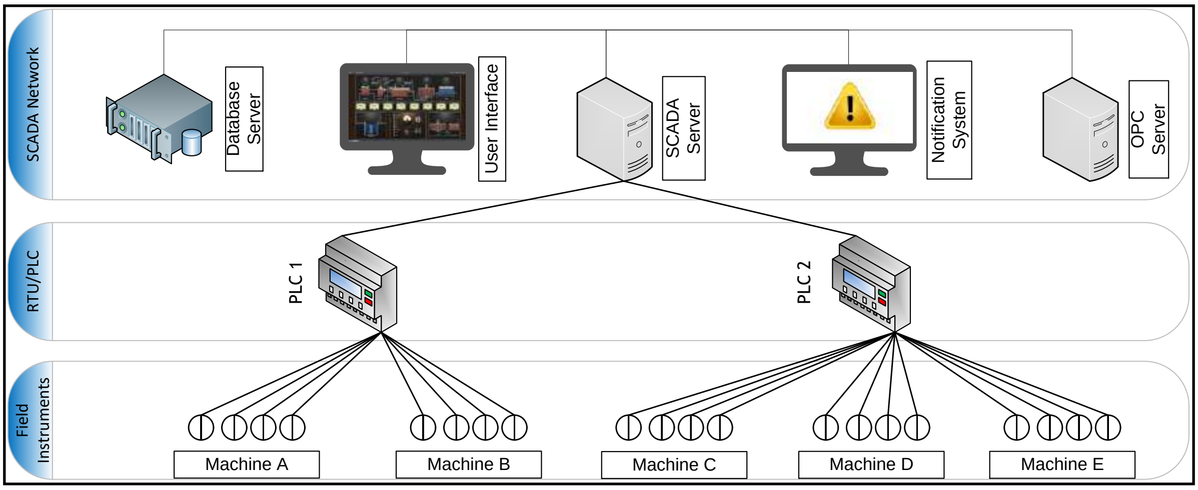


Figure 2: Generic components needed for condition monitoring

The configuration of the basic components shown in Figure 2 is not specific, and similar configurations can also be used. A user interface is required to view current or historic data, which can be done either locally or remotely. Additionally, a notification system is required to notify personnel of undesired circumstances and a database is needed to store historical data for future reference and reporting. This layout makes a condition monitoring system look rather simple, but the compatibility of software packages poses several challenges poses several challenges.

The condition monitoring systems from various mines can be linked by centralising the condition monitoring data. Instead of linking the condition monitoring systems of such mines directly to each other and solving compatibility problems from different SCADA systems, it is simpler to utilise a centralised system that can receive data from these different systems as different formats and process it to be archived as one database.

Data transfer between applications from different developers, can already make this troublesome for computers on the same network. This communication problem only becomes more complicated when the data needs to be communicated between remote computers on different networks. Even though SCADA systems can do condition monitoring, due to inter-application compatibility problems, it would be simpler to implement a system that can obtain data from various SCADA systems on different gold mines and archive it centrally. This can be attributed to the fact that the active SCADA systems on the different gold mines are not necessarily from the same software company. These companies don't provide easy integration of SCADA systems from their competitors as part of a marketing strategy. This system however needs to be implemented and configured in the same way for all condition monitoring sites to ensure consistent data quality. This generic implementation process will be discussed in the following section.

2. METHODOLOGY

2.1 Overview

This process will be developed and tested for the implementation of a condition monitoring system on the MSS's of South African gold mines. An overview of this process is given in Figure 3 below.

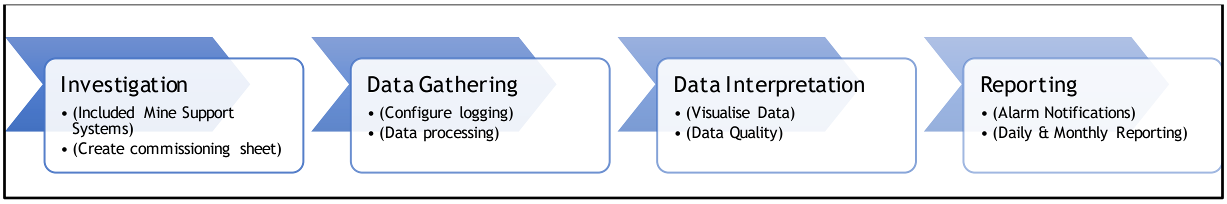


Figure 3: Implementation process overview

This implementation process consists of four parts as shown in Figure 3. The investigation step focuses on determining the scope of the machines to be monitored and documenting the available machine parameters that need to be monitored. The second step consists of configuring the logging of the monitored parameters and sending the recorded data, at regular intervals, to a centralised server to be processed and archived. The following step in the process is to configure a visual representation of the monitored parameters and to verify the data quality. The fourth and final step in the implementation process is to configure the condition monitoring system feedback or outputs in terms of alarm notifications and automated reports. These steps will be discussed in more detail in the following sections.

2.2 Investigation

Before starting the implementation of a condition monitoring system, it is important to determine the scope of the machines that will be monitored. It is recommended to start by implementing the condition monitoring system on the most important systems of the mine. In the case of South African gold mines, this would be the five mine support systems, since they are essential for the effective and uninterrupted operation of gold mines. The primary deciding factors in which MSS's are to be monitored, are the availability of instrumentation and the overall system availability. It is thus important to determine which machines in each system will be monitored and what parameters will be used.

For the purpose of this study, only machines with existing monitoring infrastructure will be included in an effort to minimise the implementation of a remote condition monitoring system. The SCADA computer can thus be used to determine the scope of the condition monitoring system by checking which machines are currently monitored and what parameters are available for each. All this information needs to be documented for future reference. The document created to contain this information is referred to as a commissioning sheet.

The commissioning sheet should also contain the parameters' measurement locations and types in addition to their corresponding machines and MSS's. The tag names used on the SCADA computer to monitor these parameters must be documented as well as the parameter logging interval. It is important to note that obtaining data directly from PLC's can add unnecessary traffic on the control network and might and influence its reliability negatively. Since tag naming conventions vary significantly from mine-to-mine and even for different dates of creation, it is recommended that a unique naming convention is used to record parameter values. This will make it easier to identify parameters from their names irrespective of their original tag names or origin. An example of the commissioning sheet for two mine ventilation fans is used to illustrate the resulting layout of the commissioning sheet. This is shown in Figure 4 below. (DE -Drive End, NDE – Non-Drive End)

Machine	Measurement Type	Measurement Location	SCADA Tag Name	Condition Monitoring Tag Name
Ventilation Fan 1	Vibration (mm/s)	Motor DE	vfan1_mot_de_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Motor DE Bearing Vib_V
		Motor NDE	vfan1_mot_nde_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Motor NDE Bearing Vib_V
		Fan DE	vfan1_fan_de_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Fan DE Bearing Vib_V
		Fan NDE	vfan1_fan_nde_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Fan NDE Bearing Vib_V
	Temperature (°C)	Fan DE	vfan1_fan_de_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Fan DE Bearing Temp_T
		Fan NDE	vfan1_fan_nde_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Fan NDE Bearing Temp_T
		Motor DE	vfan1_mot_de_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Motor DE Temp_T
		Motor NDE	vfan1_mot_nde_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Motor NDE Temp_T
	Status (-)	Running Status	vfan1.sfp_mp_diginput_1.dig11.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Running_Y
	Current (A)	Motor	vfan1_mot_current.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Current_I
	Energy (W)	Motor	vfan1_total_active_power.value	Gauteng Gold_Mine B_Ventilation_MainFan1__Total Active Power_J
Ventilation Fan 2	Vibration (mm/s)	Motor DE	vfan2_mot_de_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Motor DE Bearing Vib_V
		Motor NDE	vfan2_mot_nde_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Motor NDE Bearing Vib_V
		Fan DE	vfan2_fan_de_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Fan DE Bearing Vib_V
		Fan NDE	vfan2_fan_nde_vibration.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Fan NDE Bearing Vib_V
	Temperature (°C)	Fan DE	vfan2_fan_de_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Fan DE Bearing Temp_T
		Fan NDE	vfan2_fan_nde_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Fan NDE Bearing Temp_T
		Motor DE	vfan2_mot_de_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Motor DE Temp_T
		Motor NDE	vfan2_mot_nde_temp.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Motor NDE Temp_T
	Status (-)	Running Status	vfan2.sfp_mp_diginput_1.dig13.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Running_Y
	Current (A)	Motor	vfan2_mot_current.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Current_I
	Energy (W)	Motor	vfan2_total_active_power.value	Gauteng Gold_Mine B_Ventilation_MainFan2__Total Active Power_J

Figure 4: Commissioning sheet layout example

Literature has shown that nearly half of all machine failures can be accredited to bearing failures [15]. One of the popular methods of condition monitoring in bearings is vibration monitoring. There are also well-accepted standards like ISO 10816 available for vibration monitoring. Bearing temperature monitoring can be seen as the traditional method of condition monitoring and can provide vital information about machine and bearing conditions [15]. Therefore, the initial measurement types for condition monitoring will be vibration and temperature, but this can be expanded once the condition monitoring system has been implemented. However, for the purposes of this article, only these two measurement types will be used. The measurement locations must also be system specific to ensure that similar machines are monitored similarly. Once the parameters have been documented the parameters must be recorded and processed, which will be discussed in the data gathering section.

2.3 Data Gathering

In the development of this process, software data loggers have been used to record parameter values. Individual data loggers should be used for the different MSS's. This will ease the maintainability of the data loggers in the future. A logging interval of thirty minutes has been chosen for data logging, which should still enable the identification of deteriorating machine conditions. The average parameter value for every thirty-minute interval is calculated and logged to provide a more accurate indication of parameter values.

Even though vibration monitoring is being done at much higher frequencies, an increasing vibration will still be detectable at these logging intervals. This will minimise the amount of data that needs to be processed and stored without significantly compromising the data quality. The logged parameter values are sent as e-mail attachments to a remote server for processing and to be archived every thirty minutes. This means that the processing server will receive forty-eight e-mails daily from each condition monitoring site. It is important to note that machine conditions are monitored in real-time for alarm notification purposes, but data is only being logged at thirty-minute intervals for reporting. This means that when a parameter exceeds its limit, an alarm notification will be triggered on the condition monitoring server without the need for processing by the centralised server. This is illustrated in Figure 5 for one parameter being logged by the software data logger.

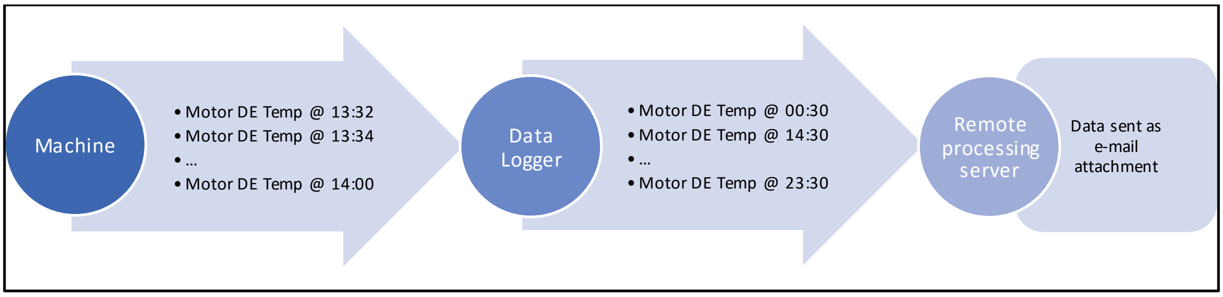


Figure 5: Example of data aggregation by software data loggers

This processing server extracts the parameter values and sorts them accordingly before they are archived in a database for reporting purposes. It is also important to note that data accuracy is not as important for condition monitoring as for performance monitoring and data reliability should be prioritised above accuracy. This means that it is more important to be able to replicate measurements that it is to get accurate measurements [16]Centralising the condition monitoring data from different remote sites, makes it easier to access and to use for reporting on the condition monitoring of these remote sites. One of the reporting methods are to present the parameter values visually on a web-based user interface. This will be discussed as part of the data interpretation process in the next section.

2.4 Data Interpretation

Presenting parameter values visually, will help simplify the interpretation of logged parameter values. Parameter values can be presented as line graphs, which provide a simple indication of its recent behaviour. The limit for each parameter is also indicated on these line graphs so that the user can evaluate its condition. These parameter limits may also be used to quantify the machine and MSS conditions in terms of the distance between the parameters and their limits.

It is important to verify that the values being logged for each parameter are correct. There are a couple of easy ways to identify suspicious parameter values. The first is static parameter values, which means that the value of the parameter does not change significantly irrespective of the state of its machine. This is also valid for a parameter value that is constantly "0" even when the machine is running. Another example is parameter values that are unreasonably low or high, like a bearing temperature that is negative or an excessive vibration. In cases of incorrect data being monitored, the configuration of the parameter logging and processing is verified and followed by an inspection of the relevant instrumentation if necessary. Once the parameter data are correctly being recorded and visualised, it should be used to provide feedback on the condition of the monitored machines and MSS's. This will be discussed in the next section.

2.5 Reporting

One of the simplest ways of providing feedback is by sending notification alarms to relevant maintenance personnel when a parameter exceeds its limit. This can be done either via SMS or e-mail and ensures that personnel can be instantly notified of parameter limit violations and take corrective action. The notification system should be configured so that personnel are notified timeously but not be annoyed by these notifications. The notifications should also be able to accommodate parameter fluctuations, so it is important to define a set of conditions that needs to be satisfied before an alarm notification is triggered.

An example of such conditions is that the machine must be running, the parameter must exceed its respective limit for five consecutive minutes and a notification for this violation must not have been sent during the past twenty-four hours. It has been found that the preferred notification channel by mine personnel is by SMS, which makes it easier for them to get notifications no matter where they are or what time it is. Automated condition monitoring reports should also be configured for the different condition monitoring sites.

Reports must be configured to be generated automatically and sent to mine personnel daily. This can be seen as official notifications to mine personnel about faulty machines. It can also be used to plan scheduled maintenance and provide feedback on problems that were corrected. These daily reports include a list of machine parameters that exceeded their limits during the previous day as well as a list of machine parameters that got close to their parameter limits. Line graphs are also included in the report for the machine parameters that exceeded their limits to give personnel a quick overview of the parameter's recent behaviour. In an effort to optimise the implementation process, it was implemented on the four MSS's of a South African gold mine as a pilot study.

3. CASE STUDY

3.1 Pilot Study

The monitored systems were mine dewatering-, mine ventilation, refrigeration and compressed air systems.

3.1.1 CM System Layout

The condition monitoring system that was implemented, had a very unique layout. This layout is shown in Figure 6 below.

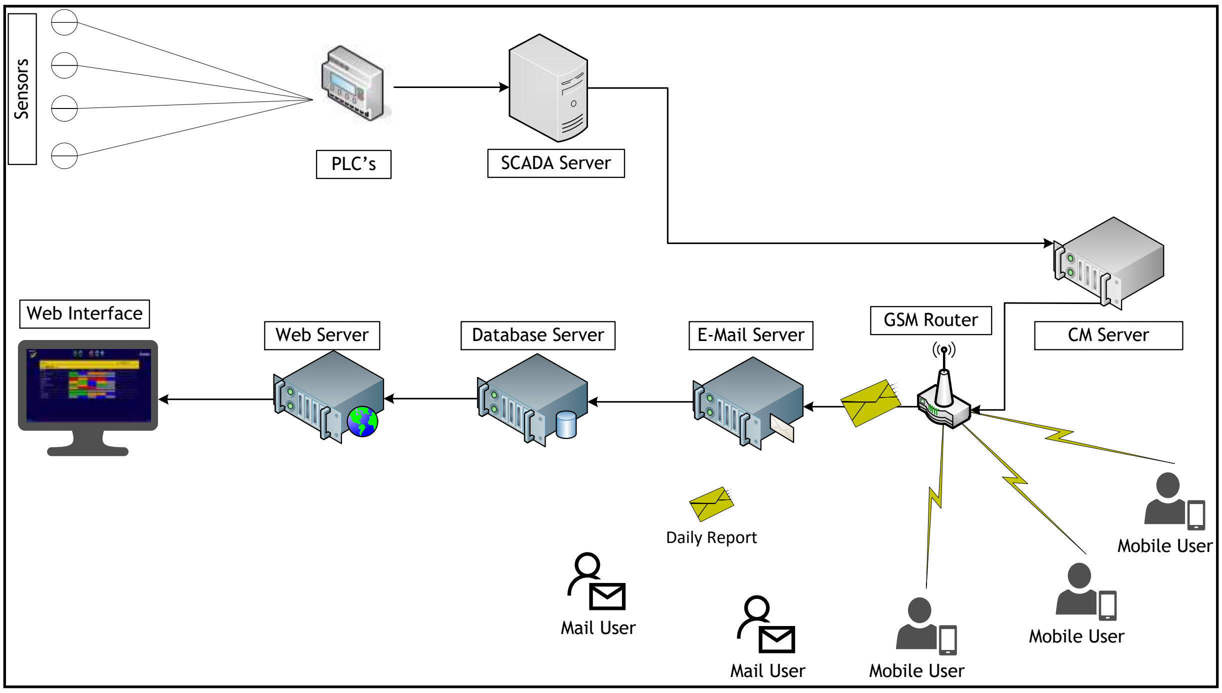


Figure 6: CM System Layout

As shown in Figure 6, machine parameter values are measured by sensors and communicated to Programmable Logic Controllers (PLC's). These PLC's are all connected to the SCADA computer of the mine, which is used to monitor and control mine operations from a central location (Control Room). The CM server uses a secure mobile network to send alarm notifications via e-mail or SMS to mine personnel when machine parameters exceed their specified limits. This GSM router is also used to send the recorded parameter values obtained from the SCADA computer as e-mail attachments to a central server to be processed and archived.

The database server extracts the logged parameter values, sorts and archives them in a database for further use. The database server also transmits the sorted parameter values to a web server, which is used to present the parameter values visually on a central web interface.

3.2 Group implementation

The condition monitoring system implementation process was optimised from experience gained on the pilot study gold mine and deployed on the MSS machines of additional gold mines within the same South African mining group. The following sections will discuss the implementation process on all the condition monitoring sites by using Gold Mine B as an example.

3.2.1 Investigation

Investigations indicated that Gold Mine B had four mine support systems that needed to be included in the condition monitoring system. The ventilation system consisted of four ventilation fans, the dewatering system consisted of nineteen water pumps, the compressed air system consisted of five compressors and the refrigeration system consisted of thirteen fridge plants. The number of parameters monitored under each MSS is shown in Figure 7 below.

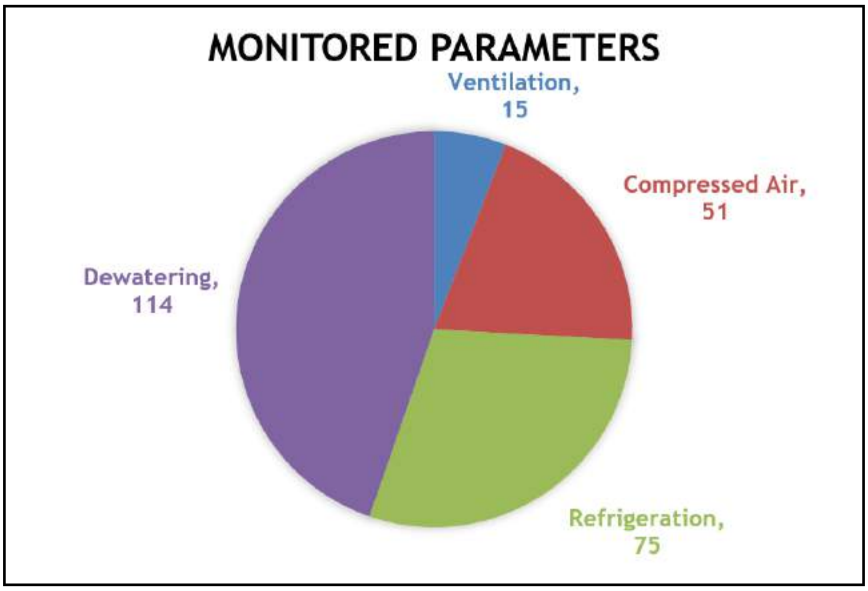


Figure 7: Parameters included in condition monitoring system

These parameters were all included in the commissioning sheet and assigned new parameter names by using a naming convention. If the monitored parameter's tag name is changed on the mine's control system for some reason, the data loggers needs to be updated to take this into account. This commissioning sheet was then used to configure a data logging platform. The investigation process was the same for all the mines in the mining group.

3.2.2 Data Gathering

Parameters were grouped and logged separately according to their MSS. The half-hourly average value of all the monitored parameters were recorded for six different gold mines and sent to a central server for processing. This resulted in a total of 1.2 million parameter values being processed monthly, while around 800 parameters where compared to their individual parameter limits every half hour.

3.2.3 Data Interpretation

The data quality of the logged parameters was evaluated, and the parameter values were configured to be displayed on a web-based user interface. An overview of the visual presentation of the MSS conditions is shown in Figure 8 below.

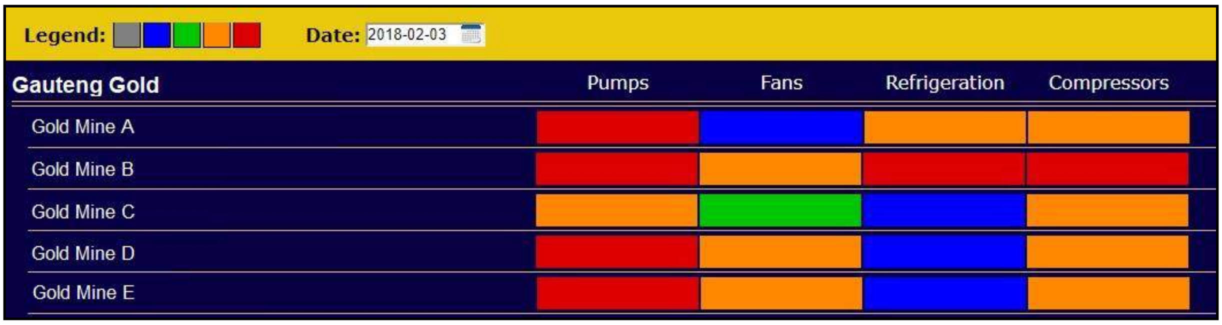


Figure 8: Condition monitoring overview of mines

The colour for each MSS is determined by the status of the machine parameter within the MSS that has the highest value. The different colours indicate different parameter states. A blue colour as shown in Figure 8, indicates that no parameter limit has been defined for the machines within the MSS, while the green colour indicates that all the machine parameters within the MSS are well below their respective limits. If the system is represented by a red colour, it means that one or more of the parameters has exceeded its defined limit, while an orange colour means that one or more of the parameters are getting close to their defined limits.

These recorded parameters are visually presented on a web-based interface, which gives mine personnel quick and easy access to historical parameter data.

3.2.4 Reporting

The condition monitoring system's function can be observed in an incident where a compressor's bearing started failing. This triggered an alarm notification, which resulted in maintenance being done on the compressor. The problem was solved by replacing the bearing, which resulted in a decreased vibration reading. This is illustrated by the parameter trend in Figure 9 below.

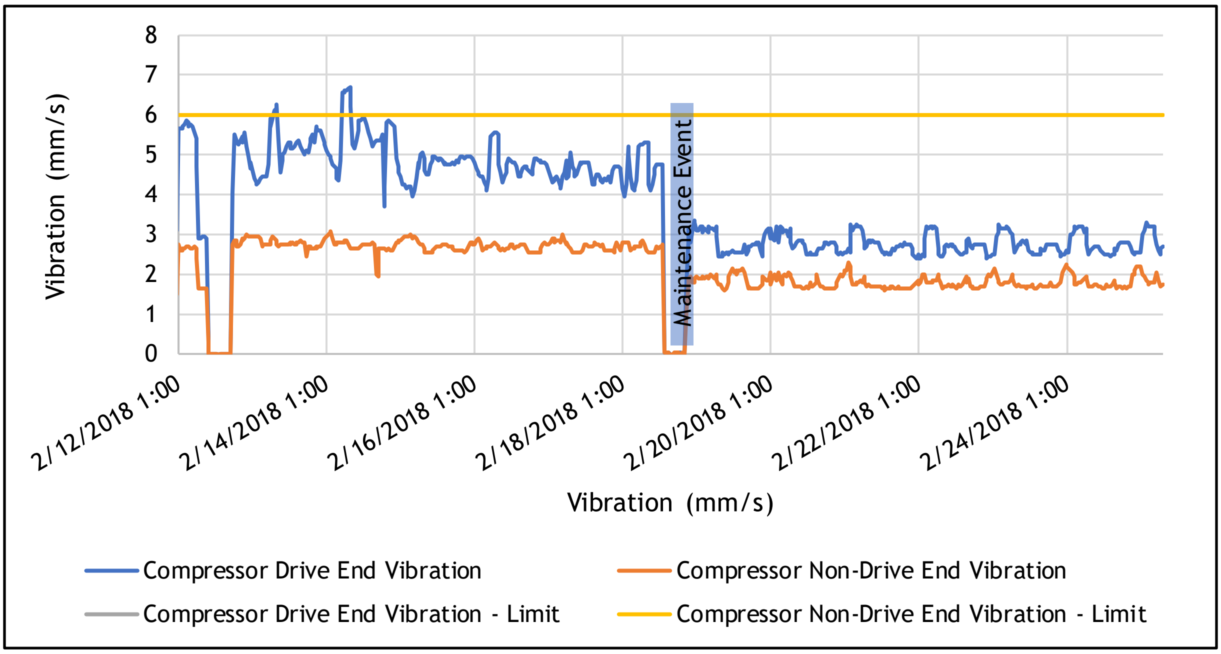


Figure 9: Example of condition monitoring system function

The data in Figure 9 represents two vibration limits, although only one is visible. The is because they are identical and the line representing one limit is covering the line representing the other. Each time a recorded parameter exceeds the specified limit, it is referred to as a parameter violation. The purpose of this is to quantify a machine condition, and there is no need to classify a machine as faulty more than once a day. These violations are limited to one occurrence per day for each parameter. The total monthly parameter violations for the MSS's of each mine was recorded and are shown in Figure 10 below.

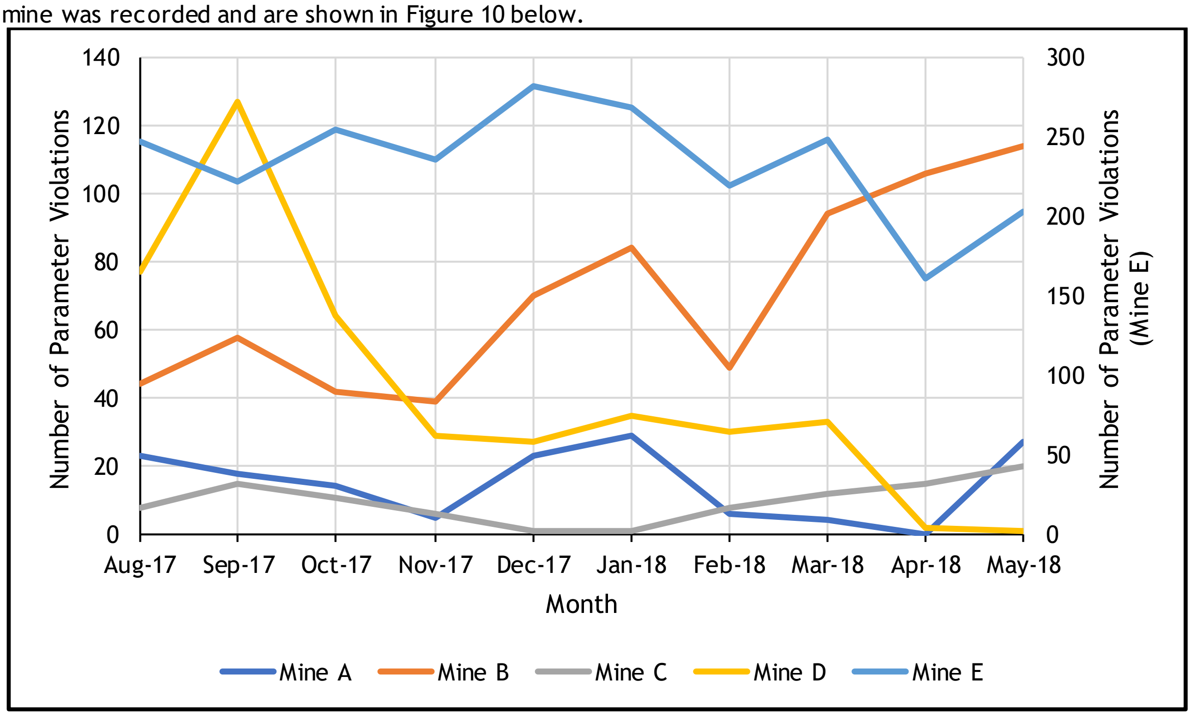


Figure 10: Monthly parameter violations per mine

The changes shown in Figure 10 can be quantified in terms of a monthly parameter violation count and a total reduction in the number of monthly parameter violations during the illustrated period

Although there was an initial increase in the total number of violations, this might be due to the condition monitoring systems becoming fully operational. Some mines even had an overall increase in monthly critical exceptions. This might be attributed to the fact that the condition monitoring system can only notify mine personnel about faulty machines but cannot force maintenance to be done more timeously. These mines might belong to the same mining group, but it is important to note that each mine has its own maintenance team, which is responsible for maintenance of machines and systems. The maintenance teams on these mines would also differ in terms of resources, since the maintenance requirements of each mine is unique.

The data represented in Figure 10 can be broken into more detail to show the condition of individual systems in terms of monthly parameter violations per system. The condition monitoring system can be used to track system specific improvements in a mine's maintenance strategy. An example of such an improvement is shown in Figure 11 below.

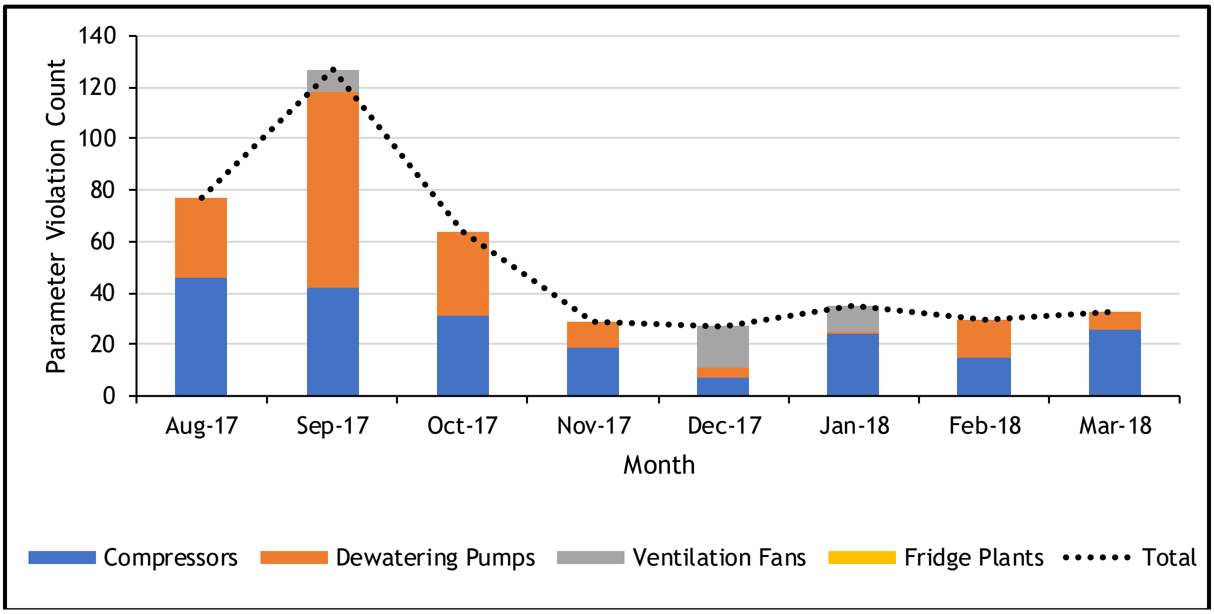


Figure 11: Example of an improving maintenance strategy

From the data in Figure 11, even though there was an initial increase in parameter violations after the implementation of the condition monitoring system, the monthly parameter violations decreased to almost none. The dotted line represents the total monthly parameter violations for all the systems during each month. This can be attributed to persistent and sufficient maintenance actions. Another reason for this is that due to an infrastructure change on the mine, its support systems became more efficient and had a reduced workload.

Unfortunately, the number of parameter violations can also increase on a month-to-month basis for individual systems due to a failing maintenance strategy. This is indicated in Figure 12 below.

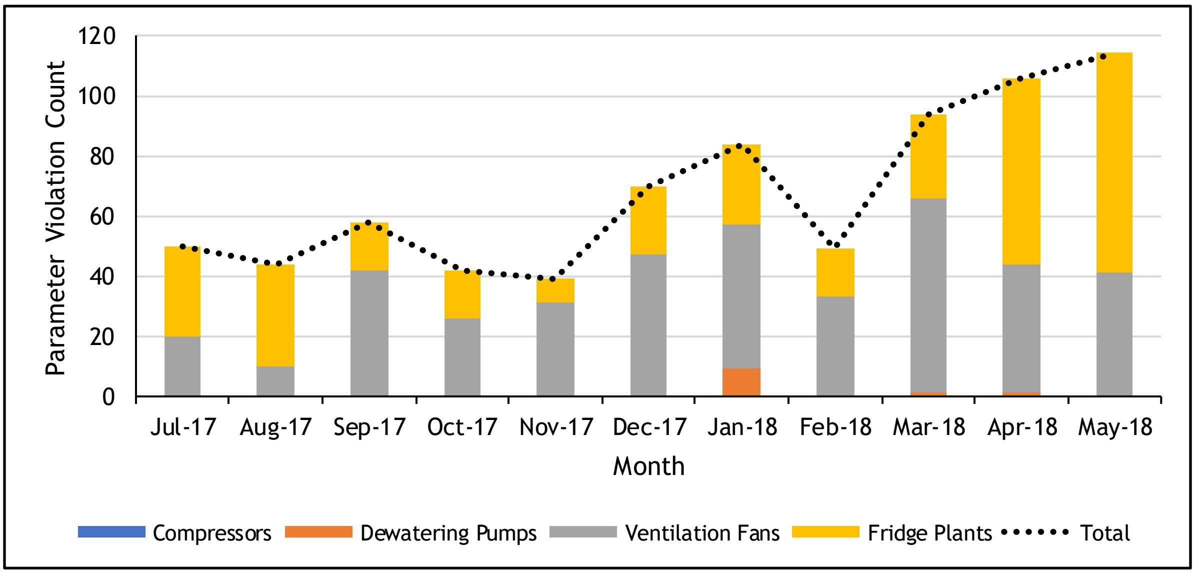


Figure 12: Example of a poor maintenance strategy

The data in Figure 12 shows that the number of parameter violations by the ventilation system stayed relatively constant, while the parameter violations caused by the fridge plants increased significantly after an initial decrease. This has caused the overall MSS machines' operating conditions to deteriorate and might indicate a deteriorating interest in proper maintenance by mining personnel. Since each mine has its own maintenance responsibilities, this can be seen as a sign of inadequate management, or the need for a more structured approach to maintenance strategies to ensure consistency on different mines.

A general decline in parameter violations, as shown in Figure 11, indicates that maintenance personnel on mines started doing more preventative maintenance and thus avoiding machine parameters exceeding their predetermined limits. This indicates that there has been a definite impact on the overall condition of MSS machines and their maintenance strategies by the condition monitoring system. There were however also some obstacles that were encountered during the implementation of this condition monitoring system. The most significant obstacles will be briefly discussed in the next section.

4. PROBLEMS FACED DURING IMPLEMENTATION

4.1 Investigation

A small group of people were initially tasked with implementing the condition monitoring system on these gold mines to speed up the process and distribute the workload. Since there was no defined naming convention for parameters being logged for condition monitoring, some parameters were poorly named, or the names were not descriptive enough. This kind of problem can be addressed and prevented by following a generic naming convention for all the parameters monitored by the data loggers.

The documentation process also led to new problems. Some personnel did not document additional parameters being logged or the defined parameter limits used for some notification alarms. Others did not update the required documentation after changes were made to the data logging platform. This is another indication that a structured implementation process can assist the management and optimisation of existing maintenance strategies. There was also no complete list of available machine parameters, which meant that the data logging platform had to be updated regularly to correct faulty configurations or add new machine parameters to be monitored.

One of the most important parts of the implementation process is obtaining a tag list or list of monitored parameters before starting any data gathering. This will be used to evaluate and identify parameters that are important for condition monitoring. It was thus found to be essential for everyone working on the data gathering to update the relevant documentation frequently and to distribute these documents to the relevant people.

4.2 Data Gathering

A lack of experience in some areas also led to data loggers being configured incorrectly or alarm notifications not functioning correctly. There were also problems on the condition monitoring server itself. Some unused parameters were logged to build a more complete database for future reference. This caused the data logging platform to stop responding (frozen computer application) and caused either static parameter values to be logged or no parameters being logged altogether. On some occasions, the connection to the SCADA computer was lost and because it was unmonitored, it led to further data losses. These issues can be minimised by having qualified people maintain both the software and hardware components of the condition monitoring system.

Furthermore, the condition monitoring server failed to start up correctly after a power failure, which meant that data was not being logged. Due to the server not staring up correctly, remote connectivity was also not available to identify the cause of data loss. This meant that the server had to be restarted locally before data logging could resume. There was also an incident where a hardware component on the condition monitoring server failed and prevented the server from starting up without manual intervention. This caused data loss on several occasions before the problem was identified. This can be prevented by using an uninterruptible power supply (UPS), which will prevent data loss during power failures and protect equipment from electrical power surges.

4.3 Data Interpretation

There were some instruments that became faulty and caused unrealistic values to be logged without being noticed immediately. This was easily identified on the visual data graphs and could be corrected. The user interface made it easy to track and identify data loss or poor data quality incidents. Problems like this emphasises the importance of checking data quality to identify faulty instrumentation or configuration errors.

4.4 Reporting

A GSM router with a prepaid sim card was used to send data via e-mail, alarm notifications via SMS and provide remote access to the condition monitoring server for troubleshooting and maintenance purposes. This led to some problems like the data bundle or SMS bundle getting depleted unexpectedly. The depletion of the SMS bundle prevented essential alarm notifications from being sent, while the depletion of the data bundle had more significant consequences. A depleted data bundle meant that half-hourly data could not be sent to the central server for processing and remote connectivity for troubleshooting was also not possible. Another common problem was that due to poor mobile signal, remote connections were very slow and data log file sizes were very limited. It is thus recommended that a reliable system is used to transmit data and alarm notifications.

The alarm notifications were also initially configured to trigger too frequently, which ended up annoying mine personnel more than being informative. There were other occasions where alarms were triggered for machine parameter while the machine was switched off. This was remedied by decreasing the minimum alarm frequency and developing a list of conditions that had to be satisfied before alarm notifications could be triggered, as discussed in the previous section.

5. CONCLUSION

In this paper we have seen that there are five essential MSS's for deep level gold mining in South Africa and that each of them is driven by large motors. It was found in literature that unexpected maintenance can lead to significant operational costs and production losses. There is thus a need for active condition monitoring on the machines of MSS's and real-time notifications of parameters with excessively high values.

A condition monitoring system implementation process was therefore developed and optimised on the MSS machines of various South African gold mines in order to monitor the condition of MSS machines and encourage preventative maintenance. Although the total number of parameter exceptions increased initially, there was a general decline since the implementation. This means that the condition monitoring on MSS machines resulted in more efficient maintenance strategies on South African gold mines.

6. REFERENCES

[1] R. L. M. Kotze and J. K. Visser, "An analysis of maintenance performance systems in the South African mining industry," *South African Journal of Industrial Engineering*, vol. 23, no. November, pp. 13–29, 2012.

[2] K. Farthing, "Mining Weekly - Asset management software reduces expenditure," *Mining Weekly*, 2009. [Online]. Available: http://www.miningweekly.com/print-version/reducing-expenditure-through-asset-management-software-2009-06-19. [Accessed: 11-Jan-2018].

[3] A. J. Schutte, "An integrated energy efficiency strategy for deep mine ventilation and refrigeration," Ph.D. Thesis, North-West University, 2013.

[4] G. Niu, B. S. Yang, and M. Pecht, "Development of an optimized condition-based maintenance system by data fusion and reliability-centered maintenance," *Reliability Engineering and System Safety*, vol. 95, no. 7, pp. 786–796, 2010.

[5] B. S. Dhillon, *Engineering Maintenance*. New York, Washinton D.C.: CRC Press, 2002.

[6] J. I. G. Bredenkamp, "Reconfiguring mining compressed air networks for cost savings," M.Eng. Dissertation, North-West University, Potchefstroom, 2014.

[7] A. J. M. Van Tonder, "Automation of compressor networks through a dynamic control system," Ph.D. Thesis, North-West University, 2014.

[8] A. Chatterjee, L. Zhang, and X. Xia, "Optimization of mine ventilation fan speeds according to ventilation on demand and time of use tariff," *Applied Energy*, vol. 146, pp. 65–73, 2015.

[9] G. E. Du Plessis, L. Liebenberg, E. H. Mathews, and J. N. Du Plessis, "A versatile energy management system for large integrated cooling systems," *Energy Conversion and Management*, vol. 66, pp. 312–325, 2013.

[10] H. Zhang, W. Zhang, L. Lv, and Y. Feng, "Effect of fissure water on mechanical characteristics of rock mass," *Mining Science and Technology*, vol. 20, no. 6, pp. 846–849, 2010.

[11] J. C. Vosloo, "Control of an underground rock winder system to reduce electricity costs on RSA gold mines," M.Eng. Dissertation, North-West University, 2006.

[12] R. J. Robles and M.-K. Choi, "Assessment of the Vulnerabilities of SCADA, Control Systems and Critical Infrastructure Systems," *International Journal of Grid and Distributed Computing*, vol. 2, no. 2, pp. 27–34, 2009.

[13] Department of Labour, "Occupational Health and Safety Act, No. 85 of 1993," *Government Gazette*, no. 181, p. 27, 1993.

[14] "What is OPC? - OPC Foundation." [Online]. Available: https://opcfoundation.org/about/what-is-opc/. [Accessed: 22-May-2018].

[15] W. Zhou, T. G. Habetler, and R. G. Harley, "Bearing Condition Monitoring Methods for Electric Machines: A General Review," *2007 IEEE International Symposium on Diagnostics for Electric Machines, Power Electronics and Drives*, pp. 3–6, 2007.

[16] ISO 17359:2002, "Condition monitoring and diagnostics of machines - General guidelines," *International Organization for Standardization*, vol. 2002, no. 50, 2002.

DEVELOPING A MODEL TO OVERCOME THE ORGANISATIONAL COMMUNICATION DEFICIENCIES BETWEEN STAKEHOLDERS IN THE ASSET MANAGEMENT INDUSTRY

N.L. van Rooyen[1*] & J.L. Jooste[2]

[12]Department of Industrial Engineering
Stellenbosch University, South Africa
[1]ninalou93@gmail.com, [2]wyhan@sun.ac.za

ABSTRACT

In providing Asset Management (AM) services, strategic partnerships are established for reaching organisational objectives and to obtain a competitive advantage. For this relationship to be optimal, effective communication is essential. Stakeholders collaborating in this relationship however, have diverse backgrounds, knowledge and experience and are not always effective communicators. The problem is the likelihood that service delivery objectives are not being met due to communication deficiencies. Although research about effective communication is available in other service providing industries, such as aviation and medicine, it remains a challenge in the AM services industry. This paper identifies the root causes of communication deficiencies in the AM services industry and presents a model, the COMMPAS model, that could support AM stakeholders in overcoming these deficiencies.

[1] The author was enrolled for an MEng (Industrial) degree in the Department of Industrial Engineering, Stellenbosch University, South Africa

*Corresponding author

1. INTRODUCTION

Effective communication is an essential element to obtain full cooperation in a partnership to reach the partnership objectives and achieve competitive advantage [1]. Feliu-Mojer (2015) [2] argues that effective communication skills are no longer perceived as soft skills and are becoming one of the core professional skills that are expected to be present in a professional.

In a study of the top communicators of organisations carrying job titles such as chief executive officer, executive director, president and so forth, none of the organisations studied was efficient in communication [3]. This proves that, even though organisations have found ways to transmit orders, share information and any form of communication, it remains to be something that can be improved upon and add value. Even though the importance of communication skills in professionals is realised, training opportunities for communication skills for professionals are still limited [2].

Similarly, communication is identified as a challenge in the asset management (AM) industry [4]. Numerous solutions have been developed to support organisations to improve their own communication efficiency. Such solutions address communication issues in the line of strategic planning, organisational effectiveness and customer- and employee relations [3]. The nine Knowledge Areas of Project Management Body of Knowledge (PMBOK) can be used as an example, where one of the knowledge areas is Project Communication Management, the SBAR situational briefing model and using vision to improve organisational communication. There are, however, limited academic resources that address communication challenges in the PAM industry, specifically communication challenges between asset management service providers or consultants, asset owners and contractors. The purpose of this study is to explore the opportunity of addressing these challenges and to develop a model to support PAM organisations with improved communication, enabling them to reach their service delivery objectives.

2. THEORETICAL FRAMEWORK

The qualitative Hutter-Hennink [5] research design is followed for this research. The process of the qualitative research cycle consists of three interlinked cycles, namely the design cycle, the ethnographic cycle and the analytic cycle. The specific research methodology adapted from the Hutter-Hennink design is illustrated in Figure 1Figure . The research design is based on the first, design cycle of the qualitative research cycle. The cycle commences with formulating the research questions, to establish a focus in the research and give direction to the search for literature and theory. The second data collection cycle is based on the ethnographic cycle and consists of selecting a fieldwork approach. The fieldwork approach covers *how* data will be collected as well as *which* research methods will be used. For this study, the data collection cycle consisting of implicit activities, such as in-depth interviews and focus group discussions. In the third and final analytic cycle the model is constructed. This cycle comprises of developing codes to describe and compare, to categorise and conceptualise and to develop a theory. Activities such as code development and data preparation are performed by using the Atlas.ti qualitative data analysis and research software. Each of the three cycles is presented in the rest of the paper.

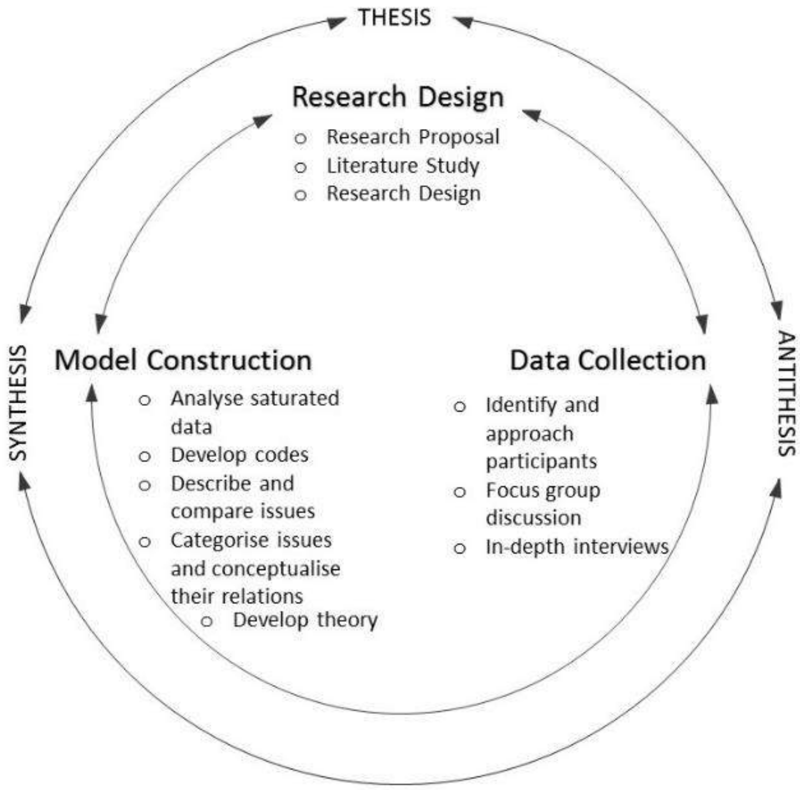


Figure 1: Adapted Hutter-Hennink Qualitative Research Cycle

3. LITERATURE REVIEW

The scholarship of organisational communication is vast and includes models such as: Laswell Formula [6]; Shannon-Weaver Model of Communication [7]; Osgood-Schramm Circular Model [8]; Dance's Helical Model [9]; Berlo model [10]; Barnlund model [11]; the McGreavy, et al. [12] approach to Organisational Communication as a Complex System; using vision to improve organisational communication [13] ; and the Situation, Background, Assessment, Recommendation (SBAR) situational briefing model [14].

The majority of these organisational communication models have linear representations of communication which skews the emphasis towards the sender [15]. More recent models, such as using vision to improve organisational communication, however states that effective communication is better achieved by creating a shared context and using receiver-oriented approaches to communication. The problem however with most of the existing models are the predominant focus on communication basics, while there are little indications about how to support organisations in overcoming communication deficiencies.

Two models which do provide support for organisational communication beyond the basics, is the SBAR situational briefing model and the use of vision to improve organisational communication. A summary about each of these models is provided:

According to Leonard [14] the SBAR model produces the best results for managing communication in a complex environment. The SBAR model is a situational briefing model that originates in the medical industry. The model is effective since it is easy to remember, it enables structured conversations which are necessary in critical situations where attention and action is urgent, and it sets expectations for what will be communicated ([14]; [16]). The SBAR model consists of four elements, namely; situation, background, assessment and recommendation, which are explained at the hand of medical examples as it is originally illustrated by Leonard [14]:

- Situation: "Dr Preston, I am calling about Mr. Lakewood, who is having trouble breathing."
- Background: "He is a 54-year-old man with chronic lung disease who has been sliding downhill, and now he is acutely worse."
- Assessment: "I do not hear any breathing sounds in his right chest. I think he has a pneumothorax."
- Recommendation: "I need you to see him right now. I think he needs a chest tube."

The advantages of the SBAR model are threefold; critical information is transmitted in a predictable structure; the standard communication method creates familiarity in how people communicate and critical thinking skills are developed as the person initiating the communication needs to provide assessment of the situation and what they think the solution is prior to starting the conversation.

A different approach to organisational communication is to use vision to create a shared context to improve communication [13]. Kelly [13] questions the classic communication models such as the Shannon-Weaver and Berlo Model based on the challenges Quirke [15] offers against these models. Quirke argues that a primary reason for communication deficiencies is that managers and supervisors underestimate the central role of the receiver in the communication process and argues that more context should be shared for a receiver to decode a message. Quirke lists many reasons why context is traditionally not shared, one for example is that the emphasis is on task focused communication to get immediate results which results in context being left out.

It is evident from literature that before the emergence of vision-directed communication, barriers to organisational communication were categorised into interpersonal barriers and organisational barriers [15]. Interpersonal barriers include: the perception and perceptual selection process; semantics; channel selection; and inconsistent verbal and non-verbal communication. Organisational barriers include: physical distractions; information overload; time pressure; technical and in-group language; status differences; task and organisation structure requirements; and absence of formal communication channels. The vision-directed approach delivers a solution to these barriers [15].

The vision-directed communication model addresses both the interpersonal and organisational barriers. Kelly [13] supports this by stating that this approach: builts a shared context through by reinforcing the organisation's values and beliefs; minimise communication barriers and reconceptualised problems as opportunities; and finally, the approach uses artefacts, symbols and ceremonies, which in itself are messages that are present even when the sender is absent.

4. DATA COLLECTION AND ANALYSIS

The data collection for the study is conducted in the form of focus group discussions and in-depth interviews. The analysis of the data is performed by developing inferences and codes, describing and comparing codes, and finally categorising and conceptualising the codes.

Strong relations were established with a PAM organisation, through which eligible research participants were identified. Participants have been invited to participate in the study by means of a face-to-face introduction or a phone call. Participants who agreed to contribute to the study were collaborated with further by means of emails until a time had been confirmed to meet face-to-face (if possible) or via Skype, to conduct the in-depth interview or focus group discussion. The recruited participants covered profiles such as, Business Area Manager (BAM), Field Engineer, Project Coordinator, Call Centre operator, Asset Care Engineer and Contractors. There are language and cultural differences within these subgroups as well.

Focus group discussions and in-depth interviews with the selected research participants produced a dataset from which communication deficiencies are identified. Inferences are developed to represent issues raised by research participants. These inferences are analysed and refined by using the Atlas.ti software to develop a list of codes that serve as markers of evident issues which arose from the participant data gathered. A summarised list of the issues based on the coding process is provided with an explanation to clarify each issue:

- *Cloud volume of information* - Communication deficiencies do not exist as a result of too little information, but rather due to too much information that is stagnating at the resources, such as the software that is used to capture data of the physical assets, and not shared with stakeholders.
- *Engagement* - Employees are not making an effort to communicate effectively due to a lack of engagement.
- *Feedback* - No feedback is given when work is done.
- *Incomplete Information* - Miscommunication and confusion is a result of an order, an assignment or feedback that is given but the message is incomplete.
- *Interdependent* - The inter-dependency of stakeholders leads to a delayed decision-making process.
- *Lack of Experience* - New employees do not have experience about what essential information to share with whom. The difference in work experience amongst the diverse workforce also leads to messages that are interpreted differently by each of the role players in the communication system.
- *Lack of Knowledge* - Employees lack knowledge regarding communication skills and also knowledge of each other's fields of expertise, and therefore they misunderstand each other.
- *Language and Culture* - Misunderstandings occur due to differences in vocabulary and manner of communicating, which stem from different languages and cultures.

- *Role definition* - Unclear role definition becomes a communication deficiency when employees are not sure to whom to send information, or do not know that an order is relevant to them when they receive it.
- *Update* - Employees are not always sharing new information as it becomes available, which leads to communication deficiencies due to stakeholders not always having and receiving the most up-to-date information.
- *Urgency* - Urgent information is not conveyed timeously, which means the urgency is either not comprehended or not appreciated.
- *Volume of Priorities* - Too many priorities lead to role players being too busy to communicate with each other properly or at all, or they do not react to a request from a colleague as it is not of high priority to them personally.
- *Volume of Role Players* - Too many role players become a deficiency when the number of people that must constantly be kept informed is too great and too many stakeholders are dependent on the authority or contribution of another, whether it is a contribution of time, money or intellect.

The listed issues are described and compared, categorised and conceptualised to expand the understanding of the issues. By grouping the issues, similar root causes are investigated to develop insights of how the issues are integrated and how they influence each other, or which issues stem from the same root cause. The conceptual model illustrated in Figure 2.

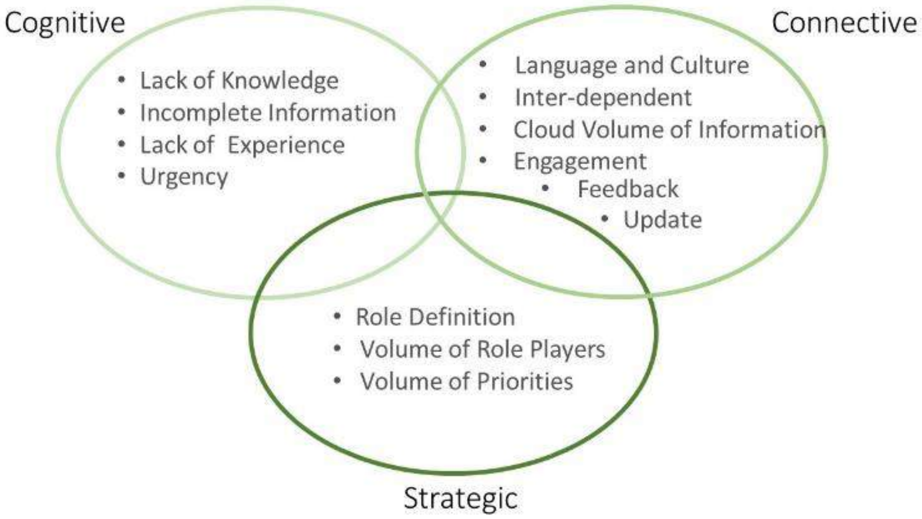


Figure 2: Conceptual Model as preparation for final Communication Model.

Conceptualisation is done by investigating the categories developed from analysing the issues, to devise higher-order categories from the data. The categories are further developed into three domains, namely the cognitive, connective and strategic domain. This is the foundation on which the COMMPAS model is built.
The concepts that are developed are rooted in the data, thus each of the three domains represents a group of communication problems. Each domain, together with the group of issues that are associated with each domain, are explain:

The *cognitive* domain represents barriers of effective communication that relate to the act or process of knowing. Some employees simply lack the knowledge of either how, what, when or with whom to communicate. This domain covers either the fact that knowledge is lacking in some way, or how the acquisition of knowledge is integrated into the system.

The issues associated with the cognitive domain are:

- Lack of Knowledge: From the data, it is evident that employees do not always have knowledge about the various areas within the organisation. It is imperative that the knowledge level in an organisation is managed.
- Incomplete Information: Information shared is often incomplete, which leads to stakeholders misinterpreting a message. This result in work instructions not being completed or being completed incorrectly due to the stakeholder being uncertain of what is required of them.
- Lack of Experience: The lack of experience identifies areas for learning opportunity to continuously improve communication skills, as well as areas where knowledge should be expanded.
- Urgency: Immediate information is often needed to make a decision, which leaves employees frustrated when they are unable to proceed due to the required information not being immediately available.

The *connective* domain addresses barriers of effective communication that are rooted in the relationship between role players and the responsibilities that they have towards each other. What connects actors in the communication system is activities such as feedback and teamwork. If these activities are not performed properly, the result is a breakdown in communication.

The issues in the connective domain are:

- Language and Culture: The different ways of communicating and different terminologies that are being used results in miscommunication.
- Interdependent: Stakeholders are interdependent on each other for obtaining information. The interdependence for information becomes an issue when urgent decisions need to be made or action needs to be taken, thus hindering progress.
- Cloud Volume of Information: The information needed to make decisions and to execute work is available but gets lost along the connections between stakeholders. It is either not shared by the stakeholder that has the information, or it is not attended to by the stakeholder that receives the information.
- Engagement: Engaged employees and actors in the communication system are essential. Disengaged employees, tend to not contribute to the link between role players by not giving feedback or not taking part in the team context.
- Feedback: The action of sending new information when a task is completed is not being followed thoroughly and results in complications. The act of feedback strengthens the link between actors and ensures that the optimal amount of information is shared with relevant actors.
- Update: Role players are not mindful of the importance of keeping all relevant stakeholders up to date.

The *strategic* domain represents deficiencies that are rooted in the strategies of an organisation and how they align their resources with their organisational goals.

The issues in the prospective domain are:

- Role Definition: Stakeholders are uncertain of their roles when it comes to more detailed responsibilities, which results in uncertain and less motivated employees.
- Volume of Priorities: The number of priorities that stakeholders must consider impede them from integrating continuous communication into their activities.
- Volume of Role Players: The number of role players in asset management service makes communication complex. A system needs to be in place relating to who communicates with whom and who should be kept up to date by whom.

5. THE COMMPAS ORANISATIONAL COMMUNICATION MODEL

The construction of the *Communication in Managing Physical Assets* (COMMPAS) model is based on the domains described in the conceptual model. The domains are integrated and interdependent to increase the probability of effective communication between AM role players. The COMMPAS model is illustrated in Figure 3. The model consists of the following components: the outer component consists of the culture and vision of communication that encompasses and informs the entire model and which is implemented and conserved by means of training, which is the middle component, and the SBAR communication technique as core component.

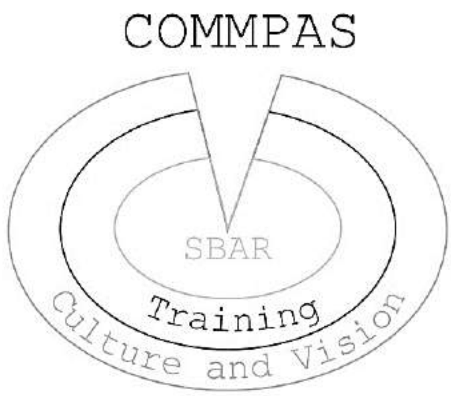


Figure 3: COMPASS Organisational Communication Model

The elliptic form of the model illustrates the three organisational elements (culture, vision and training) that must be encouraged and developed continuously. The model is implemented inward, starting with the outer circle. Once implemented, the three elements of the model support and further establish each other.

5.1 Outer component: Culture and Vision.

The first and most important step is to develop a communication vision that establishes and governs a culture of communication in the organisation. This is the foundation of effective communication in a PAM organisation that employs a diverse work-force with different skills, languages, cultures, knowledge capacities and experience. The content of the vision would be specific to the organisation. It must essentially entail a plan and standard for communication in the organisation, which states the importance of *effective* communication to the organisation, for example that feedback about a job is a priority and should be approached with the same importance as the job itself.

5.2 Middle component: Training.

Training is the tool by which the organisation's communication vision and standards, as well as the practical conduct regarding communication such as the SBAR model, is instilled. Developing the training programme follows a three step needs assessment:

- Conduct an organisational analysis to determine where the training is needed in the organisation and which organisational goals they aim to address through the training;
- Conduct a task analysis to determine what the training will cover, i.e. what must the trainee learn to be an effective communicator according to the standards and practice (conduct/vision) of the organisation; and
- Conduct a person analysis, to determine who needs training, what their current knowledge capacity is to address the gap between what they currently know and what they are required to know to identify what they need training for i.e. what they are supposed to be able to do after having completed the training.

This framework is used to guide organisations to develop a personalised and effective training programme. The content and participants of the training programme should be specific to the organisation.

5.3 Core component: SBAR model.

The SBAR technique for communication is the core element of the model. It is a practical tool that standardises the communication process within an organisation, to support role players to communicate effectively. Confusion and uncertainty regarding what information to include when sending a message is eliminated. The lack of shared experience that leads to miscommunications amongst role players, is eliminated by the SBAR technique that includes comprehensive information in a concise message. The receiver of the message can interpret it more easily, due to their familiarity with the method of communicating, which increases the probability of the receiver decoding the message as it was encoded by the sender.

The foundation for effective communication is set by means of a culture and vision of communication, it is carried over and implemented by means of training and it is practised at the hand of the SBAR technique.

6. CONCLUSIONS AND RECOMMENDATTIONS

The partnership between a PAM service provider and their clients and contractors respectively, is a strategic relationship that is established to benefit all stakeholders. The relationship represents a set of organisational objectives that can only be met as a result of an effective synergy. A critical success factor for an effective synergy, is open and effective communication between the partners or role players in the relationship [17].

Communication deficiencies are the result of root causes that could be summarised into three categories: cognitive deficiencies, connective deficiencies and strategic deficiencies. To address these deficiencies, the solution must include: a platform for effective knowledge transfer, a shared vision by all role players in the communication system to understand their role in the system and the importance thereof, and a technique or practical strategy by which each employee can effectively communicate important information. Extensive research and data analysis serves as evidence of these statements.

The COMMPAS model is an integration of the three solutions that address the three categories of root causes. The three elements in the model are: a communication vision and culture of communication, training, and the SBAR communication technique. The model is a continuous process of these three elements, which could be seen as concentric circles that work inwards and outwards to complement each other.

First, a vision of culture and of communication should be established in an organisation. It is important for role players in the communication system to understand the importance of communicating information and of keeping each other up to date. In a complex environment with a diverse work-force, it is imperative to have a shared vision for communication. Secondly, how the culture and vision of communication is carried through to the organisation is through training, which has proved to be the most effective technique to transfer knowledge. Thirdly, at the core of the model, is the SBAR communication technique, which serves as the simple, practical framework to support all role players in the communication system to communicate effectively. This situational briefing technique eliminates the issue of different role players not having the same knowledge and experience, as it provides a concise and comprehensive framework to transfer critical information.

This study was successful in identifying the root causes of communication deficiencies and delivering a model that supports PAM service providing organisations in overcoming these deficiencies. The model has the potential to be further developed into a universal model for all service providing organisations or to be developed into an organisation-specific solution for communication deficiencies.

Although the contributions made by this research accomplishes the research objectives, however, recommendations to further develop and improve this research is necessary. The dynamics of the relationship between a PAM service provider and their contractors and clients respectively, is a complex relationship. The service provider would be dependent on the buy-in from both the client and the contractor for the COMMPAS model to work optimally. This dynamic could be further investigated to develop a strategy that enables the service provider to extend the COMMPAS model to their partnerships.

Implementing a new communication model results in changes in an organisation. Change management, specific to implementing a communication model, could be investigated.

An opportunity for further research and development of the COMMPAS model is to develop an implementation guide or strategy to support organisations in successfully implementing the COMMPAS model, with minimum additional time and costs. Further a more detailed framework for training material that is employee-specific is required. In other words, training should address the specific knowledge that an employee who transfers the knowledge at the hand of a training method requires and which suits the employee's learning abilities.
It is evident that training is a successful technique to transfer knowledge. However, in a complex environment with different stakeholders and different knowledge capacities, it becomes difficult to teach the same material to all stakeholders using the same method. Training methods could be developed that suit the knowledge capacity and learning ability of each stakeholder or sub-groups of stakeholders and enables all stakeholders to obtain the required knowledge.

7. REFERENCES

[1] **Barnard, C. 1968.** *The functions of the executive.* Harvard University Press

[2] **Feliu-Mojer, M. 2015, August.** *Effective Communication, better science.* Retrieved from Scientific American: https://blogs.scientificamerican.com/guest-blog/effective-communication-better-science/

[3] **Dozier, D., Grunig, L., & Grunig, J. 2013.** *Manager's guide to excellence in public relations and communication management.* Routledge.

[4] **Jackson, E. 2016, January.** *Communication as challenge in the PAM industry.* Cape Town, Western Cape.

[5] **Hennink, M. & Hutter, I. & Bailey, A. (2011).** *Qualitative Research Methods.* Sage.

[6] **Lasswell, H. 1948.** The structure and function of communication in society. *The communication of ideas*, 215--228.

[7] **Weaver, C. S.1949**. *A Mathematical Model of Communication Urbana.* IL: University of Illinois Press, 1949.

[8] **McQuail, D. a. 2015.** *Communication Models for the study of Mass Communication.* Routledge.

[9] **Dance, F. 1967.** A helical model of communication. *Human Communication Theory, New York, NY: Holt, Rinehart and Winston*.

[10] **Berlo, D. K. 1960.** The Process of Communication. *New York: Holt, Rinehart and Winston*.

Berlo, D. K. (1960). The Process of Communication. *New York: Holt, Rinehart and Winston*.

[11] **Velentzas, J. a. 2014.** Communication cycle: Definition, process, models and examples. *Proceeding of the 5th International Conference on Finance, Accounting and Law (ICFA" 14)*, 117--131.

[12] **McGreavy, B., Lindenfeld, L., Bieluch, K., Silka, L., Leahy, J., & Zoellick, B. 2015.** Communication and sustainability science teams as complex systems. *Ecology and Society*, 2.

[13] **Kelly, D.** 2000. Using vision to improve organisational communication. *Leadership and Organisation Development Journal*, 92--101.

[14] **Leonard, M. a. 2004.** The human factor: the critical importance of effective teamwork and cmmunication in providing safe care. *Quality and Safety in Health Care*, i85--i90.

[15] **Quirke, B. 1995.** *Communicating corporate change: A practical guide to communication and corporate strategy.* McGraw-Hill.

[16] **Powell, S. 2017.** SBAR: it's not just another communication tool. *LWW*.

[17] **Jooste, J. 2014.** *A Critical Success Factor Model for Asset Management Services.* Stellenbosch: Stellenbosch University.

OPTIMAL SCHEDULING OF PROJECT ALLOCATION BY MEANS OF A WEIGHT-PARAMETER CALCULATION ALGORITHM

P.v.Z. Venter[1*]

[1]School of Industrial Engineering
North-West University, South Africa
12330825@nwu.ac.za

ABSTRACT

A common scheduling task is the allocation of projects amongst a number of recipients. Various techniques can be employed to assist such a decision-making process, which could even include simplistic 'first-come-first serve' or random allocation methods. However, when recipient satisfactory-levels are deemed relevant, allocation methods should typically incorporate personal preferences.

This paper presents two models to address project allocations. Firstly, a mixed integer linear programming model is formulated, which optimises client satisfactory levels relative towards the entire population. A more preferred project is indicated by a higher numerical value and the model optimised accordingly. Potential problems that typically arise are a 'cluster' of personal preferences. When a significant number of recipients prefer relative corresponding allocations, a simplistic numerical optimisation model may prove insufficient to yield suitable solutions.

The second proposed model is a novel pseudo-code algorithm that calculates weight parameters to be used in the optimisation matrix. When individual projects can be grouped in a way that relevancy is established between various choices, non-zero positive and negative values can be determined for choices outside the selected preferences.

The applicability of the optimisation model in conjunction with the algorithm is demonstrated through a real-life project scheduling case study.

*Corresponding author

1. BACKGROUND

Project or resource scheduling is an everyday occurrence. In the engineering industry resources are scheduled over long planning periods, where large machinery and man power need to be allocated for specific tasks at certain destinations over time. Excessive research is performed in order to determine how production costs and operational times can be minimised through optimal mathematical solutions.

A common engineering plant problem is the scheduling of personnel and tasks during pre-planned shut-down periods. For preventative maintenance purposes large sections or production trains are taken off-line and serviced. It is important to realise is that this typically entails a vast number of equipment, ranging from small valves to entire reactors. A limited number of resources are available to complete all of these tasks. Schedules need to be drawn up that indicate what equipment need to be serviced by whom.

Some considerations are that not every personnel member can work on all equipment types. Furthermore, according to law there is a maximum time limit that any person is allowed to work (16 hours), before he or she must withdraw from that day's work. Decision planning typically includes which personnel will work day or night shifts and for what duration. These are just some considerations that need to be taken into account when shut-down scheduling planning is performed. The aim should always be to limit the shut-down duration, since some production processes are off-line.

A typical scheduling problem that occurs not only in the engineering sector, but everyday life, is project or resource allocations. If a number of tasks or projects need to be assigned amongst more than one recipient, some form of allocation rationale is required. When resource allocation is a teller at a supermarket, the recipient needs to stand in a queue and wait his or her turn, i.e. first come first serve. For work project allocation the same principle may be applied, however, some recipients may feel deprived of their preferred choices, since it is not always possible to enquire before other personnel. It might be more sensible to randomly allocate projects to the recipients, however, this may result in instances where there exist a mutual feeling amongst the recipients that 'project-swops' are preferred. When personal preference plays a role it might, therefore, be more sensible to mathematically determine the best possible way to allocate projects amongst different recipients.

This paper proposes two models that can be incorporated to determine optimal project scheduling, subject to allocated weight. A mixed integer linear programming (MILP) is formulated that optimises project allocations amongst any number of recipients. The model uses a weight matrix in the optimisation process. The second, novel, model formulation in is a unique pseudo-code algorithm that can be used to determine the weight matrix for the MILP. This algorithm uses the recipients' personal preferences and dislikes in combination with knowledge of how certain project fields are related.

The next section discusses some relevant literature. Section 3 provides the proposed MILP and novel pseudo-code formulations. A real life case study is presented in Section 4, demonstrating the applicability of the two models for an everyday scheduling problem. The paper is concluded in Section 5.

2. A LITERATURE SURVEY

Numerous scheduling algorithms and philosophies are found in literature. A variety of complexities and solving techniques are addressed and a few are reported.

The authors of [1] investigated resource scheduling, based on a autonomous decision structure where the bidders demand a combination of dependent objects. It was reported that complex resource scheduling problems could be handled with this method. An optimised scheduling algorithm was proposed by [3]. The possibility to allow strong computer-linked systems to utilise resources to a maximum extend was attempted. An improvement with comparison to other methods was achieved. The authors of [4] incorporated 'In Grid' and peer-to-peer (P2P) computing environments for resource allocations. A Java simulation toolkit was developed to assist the difficulties experienced for a Grid environment.

A robust genetic algorithm for single-mode resource constrained project scheduling was proposed by [5] The working ability of this algorithm was displayed through a number of computational experiments. The absense of complete information and uncertainty for scheduling tasks were addressed by [6]. The authors performed a survey on fundamental scheduling approaches under uncertainty.

A multi-agent system for energy resource scheduling is proposed by [2] for a power system with distributed resources. They applied a distributed intelligent multi-agent technology in an attempt to make the power system more reliable through energy resource scheduling. This allowed for a reduction in operational costs. A joint problem of project selection and task scheduling were formulated by [7]. They addressed scenarios where numerous alternative projects can be followed. Their objective was to maximize the net present value (NPV) of profit.

This paper proposes optimal scheduling of resources, not through extensive computing power, but through calculating a weighted-matrix, taking possible relevancy between various parameters into consideration.

3. OPTIMISATION AND ALGORITHMIC MODELS

In this section the formulation of two models are proposed. The first is a MILP optimisation model. This model maximises the weight summation between allocated resources, i.e. optimising the distribution thereof. For every recipient, a unique weight is assigned towards each possible project or resource to be allocated. A positive weight indicates a preference towards a specific project, whereas a negative value relates to a dislike. A zero value indicates neutrality; therefore, no indication is available towards any preference with regards to that project or resource allocation.

The second proposed model is a novel pseudo-code algorithm. This code should be utilised to determine the weight-matrix for the MILP. Underlying relevancies or dependencies are taken into account in order to calculate the maximum possible weight entries to be used for the optimisation matrix. These weight calculations allow negative numbers for the instances where there is a dislike to receive a specific project or resource allocation.

3.1 Optimisation model formulation

The formulation of the MILP for optimal resource scheduling follows:

The index set of all recipients is given by $I^R = \{1, 2, \ldots, |I^R|\}$, whereas the index set of the number of resource subsets is denoted by $I^S = \{1, 2, \ldots, |I^S|\}$. The $1 \times |I^R|$ vector $\boldsymbol{F}$ indicates the number of resources to be allocated for each recipient $r \in I^R$. The quantity of resources within each subset is given by the $I^S \times 1$ vector $\boldsymbol{Q}$. The weight-parameter matrix is therefore defined by $\overline{\boldsymbol{W}}$=$|I^S| \times |I^R|$.

Binary decision variables are defined, where $y_{sr} = 1$ if a resource from the index set $s \in I^S$ is awarded to recipient $r \in I^R$ and zero otherwise.

The objective of the MILP is to optimise the combined allocated resource weights amongst the recipients, so that:

$$maximise \sum_{s \in I^S} \sum_{r \in I^R} y_{sr} w_{sr}, \tag{1}$$

subject to

$$\sum_{s \in I^S} y_{sr} = f_r, \quad \forall r \in I^R, f_r \in F \tag{2}$$

$$\sum_{r \in I^R} y_{sr} \leq q_s, \quad \forall s \in I^S, q_s \in Q \tag{3}$$

The objective function given by (1) is to optimise the total summation for the weight factors of the allocated resources or projects. Constraint set (2) ensures that each recipient is allocated the predetermined number. Note that this is not an inequality. As a result resources or projects will be allocated even if it is associated with a negative weight. The final constraint set, (3), ensures that the number of resources or projects that are allocated do not exceed the availability thereof.

Solving the MILP from the previous section will yield the optimal resource allocation between the recipients. However, some of the recipients may have 'clustered' preferences. If a significant number of recipients prefer the same type of resource or project, the weight-matrix may comprises non-zero (and zero) entries in corresponding columns between recipients. This will result in a number of zero weighted allocations amongst recipients. The MILP will result in a number of feasible solutions without any indication which recipient should receive a non-zero allocation. Furthermore, there is no indication as to how the zero entries should be distributed amongst those recipients. To address this problem a novel pseudo-code algorithm to determine weight allocations is formulated in the following section.

3.2 Pseudo-code algorithm for weight allocations

The formulation of the novel pseudo-code algorithm for the weight allocations follows in this section. It is important to note that the aim is to determine a more diverse weight distribution, in an attempt to comprise unique $w_{sr} \in \overline{\boldsymbol{W}}$ coefficients. If a different weight allocation vector can be determined for each recipient, an increased probability exists for finding an optimised solution.

The same parameter definitions from the MILP are used and not re-defined. Additional definitions requirements for the pseudo-code follow:
The index set of the different number of weight allocation arguments is given by $I^A = \{1, 2, \ldots, |I^A|\}$. The number of options (decisions) that a recipient is allowed for each questionnaire (argument) is given by D_a where $a \in |I^A|$ for the $D_a \times 1$ vector P^a. The load factor, for each argument from the decision $d \in D_a$, to all recipients are given by the $1 \times D_a$ vector $\boldsymbol{L^a}$. The recipient's decision matrices are defined by $\overline{\boldsymbol{R}}^{\boldsymbol{a}}_{\boldsymbol{rd}}$ where $a \in I^A, r \in I^R$ and $d \in D_a$. The matrices for all sub-

sets under each weight allocation for every factor is defined by $\overline{\boldsymbol{S}}_{sl}^{a}$ where $a \in I^A$, $s \in I^S$ and $l \in D_a$. An external predetermined weight allocation matrix for each subset $a \in I^A$ is defined by $\overline{\boldsymbol{W}}_{0}^{a}$.
The pseudo-code now follows:

$$for\ a = 1 \dots |I^A| \tag{4}$$
$$for\ r = 1 \dots |I^R| \tag{5}$$
$$for\ s = 1 \dots |I^S| \tag{6}$$
$$for\ d = 1 \dots D_a \tag{7}$$
$$for\ l = 1 \dots D_a \tag{8}$$
$$if\ r_{rd}^{a} = s_{sl}^{a} \tag{9}$$
$$w_{rs}^{ad} = p_{d}^{a} l_{d}^{a} \tag{10}$$
$$else\ w_{rs}^{ad} = 0 \tag{11}$$
$$\overline{\boldsymbol{W}} = \sum_{a \in I^A} \sum_{d=1}^{D_a} \overline{\boldsymbol{W}}^{ad} \cdot \overline{\boldsymbol{W}}_{0}{}^{a} \tag{12}$$

Equations (4) to (12) allow for any number, positive or negative, decision weights to be added. Each weight source allows for multiple weights within a source (Eq. 7), whereas Eq. 8 further allows additional weights with respect to the initial source. Note that the $\boldsymbol{L}^{a}$ contains default values of 1 if none are provided. It is important to note that Eq. 10 verifies if any relevancy exists between the coordinate with the matrix and the respected preference weight allocations.

4. CASE STUDY

This section demonstrates the working abilities of the two combined proposed models, from the previous section. The models are applied to a real life project allocation problem.

A group of Mechanical Engineering final year students need to be allocated a fourth year design project. There are 73 students and each must be assigned a single project out of 73 possibilities. In previous years projects were randomly allocated, with a typical end result of a significant number of students displeased with their proposed projects. It was not uncommon for any students to try and swop projects in an attempt to find a task more fitting for their interests and abilities. It must be noted that the Mechanical Engineering field comprise a significant spectrum of sub-fields.

In later years students were awarded the opportunity to communicate their top three choices and these were moved around in a spread sheet, by visual inspection, in an attempt to allocate projects 'as good as possible'. Since some projects were more popular, either due to the challenge or the study leader in charge of it, these projects typically received significant more interest.

The 2018 project allocations were performed by implementing the two proposed models of this paper. The MILP formulation was used to optimise (maximise) the total combined weights related to the 73 allocated projects. Each student indicated the top three preferred choices, together with the single most unwanted project to be assigned. Irrespective of the magnitude of weights chosen for these four project choices, the MILP results yielded multiple outcomes, where a great number of students were awarded a project allocated to a zero mark.

In order for the MILP to yield a maximum summation of weights to allocated factors, it is imperative for the $\overline{\boldsymbol{W}}$ matrix to consist of as few as possible neutral numbers, i.e. coordinates carrying zero weights. The proposed pseudo-code from the previous section was incorporated. In order to be able to determine weight allocations for various projects per student, more information was required on personal preferences. The following information was provided was available. Note that no other information was available for any of the students.

- A description of all 73 projects.
- The study leader for each project.
- The different research groups within the Mechanical Engineering department.
  - Information was given with regards to typical research projects attempted by each group.

With regard to the project leaders, each person was allowed to assign a minimum of one and a maximum of three research groups relevant to every single project. The three groups were chosen in order from the most to least relevant towards each project. Take note, this information was not available to the students.

Taking the above mentioned into consideration, the following information was acquired by each student:

- Indicate the preference from first to third project to be assigned.
- The single most disliked project, i.e. to be avoided.
- The preference of three research groups in decreasing order.
- The three research groups to be avoided; in decreasing order.
- The top three preferred study leaders.
- The bottom three study leaders to work under.

The following information is provided with regards to weights chosen to represent individual choices. Further note that these data were also not made available for the students.

In Table I parameters are supplied for weight calculations by the algorithm defined in the previous section. Note that these main-weights are given from most to least influential, either in a positive or a negative sense. The three project preferences are 25, 15 and 10, whereas a single project the student wishes not to receive is penalised by 25 (-25). It is important to note that weight allocations are typically always subject to the modeller's discretion. If the positively assigned values between choices one and three are basically the same, a third best choice is not 'heavily' penalised compared to the most preferred choice. For such an instance the MILP may solve for fewer instances where the top choice is awarded and more instances where the second or third option is allocated.

The preferred research groups follow. Note that the positive weight allocations are significantly higher than the preferred projects. It is important to note that a student will typically not display interest in a specific project, without preferring conceptually relevant work. Further, note the fraction awarded with respect to the group choices. A research group interest, as indicated by a student, differs from a higher to a lower interest, whereas research groups allocated to a project by a study leader also varies in relevancy. Therefore, there exists nine possible (3x3) outcomes, i.e. from a first group choice with that group being the most relevant to a project, down to the third group choice with a third relevance rating. If a student's third research group preference is the first group allocated to a specific project, it will not make sense to add a weight of 50 towards the project, but rather a weight of 50x0.8 (first research group x third preferred research group).

The preferred project leaders follow. Note that the positive weight allocations are slightly higher than the preferred projects. Even though the outcome is to determine a best-suited project allocation, it is important to note that a student will typically not display interest in a specific project, without preferring the study leader. The negative study leader weights, however, are more significant. The rationale is that a student ending up with a non-preferred study leader will, typically, be counterproductive and should be avoided.

As mentioned, a weight of 25 is assigned for a student's first project choice. If the student, furthermore, choose the relevant research group as first choice, another 50 will be added to total 75. The study leader not preferred is penalised heavily, i.e. with 78 points. If a student chooses a project and group as first choice, but indicating that the allocated study leader should not be assigned (a highly improbable situation), the combination of weighs is just below zero (-3). This weight ensures that every student has a negative value assigned towards receiving any project from the undesired study leader.

Table I. Weight parameters used by the algorithm for calculations.

	Weight allocations		
Preferences	**1st**	**2nd**	**3rd**
Project	25	15	10
No project	-25	-	-
Project leader	30	25	20
No project leader	-78	-65	-52
Research groups	50	40	30
No research groups	-50	-40	-30
1st research group fraction	1.0	0.9	0.8
2nd research group fraction	1.0	0.9	0.8
3rd research group fraction	1.0	0.9	0.8

It was furthermore mentioned earlier, weight allocations are subjectively assigned, however, typically with some rationale behind it. Furthermore, it is important to stress that the choice of weight allocations are equally applicable to all students and therefore unbiased to any individual. It should be noted that the aim is to optimise the process of dividing projects amongst a number of students. The final result can therefore be interpreted as the 'on-average most satisfactory feeling' within the group and not for a single individual.

The parameters from Table I are now assigned to the pseudo-code algorithm, in order to demonstrate how these dependent weight allocations can be utilised within the generic five-times for-loop. It should be noted from the information provided that 73 out of 74 projects should be allocated amongst $|I^R| = 73$ students. Note, however, that $|I^S| = 48$, since some selected projects allow more than one student. All relevant data are provided in the appendix, in order to reproduce the results that follow.

4.1 Project weight allocation

For the algorithm; when $a = 1$ all calculations are with regards to the preferred project allocations. From Table I it follows that $D_1 = 3$ and that $l_1^1 = 1$. The matrix $\overline{\boldsymbol{O}}^1$ is a 73×3 ($|I^R| \times D_1$) parameter input source for the MILP, set-up from the students' choices, where r_{rd}^1 indicates the d^{th} project choice for student $r \in I^R$. A project allocation that a student wishes to avoid is under $a = 2$. Positive project leader allocations are determined when $a = 3$, whereas $a = 4$ is for undesired supervisors.

For the preferred research group allocations $a = 5$. Note therefore that $\boldsymbol{L}^5$ contains non-unity values. Negative research group weight allocations are calculated for matrices $\overline{\boldsymbol{W}}^{(6\,1)}$ to $\overline{\boldsymbol{W}}^{(6\,3)}$ for $a = 6$. The matrix $\overline{\boldsymbol{W}}_0^a$ is determined by calculating each individual student's combined average over all the graduate modules and constant for all $a \in I^A$. Note that $\overline{\boldsymbol{W}}_0^a$ is a 73×48 ($|I^R| \times |I^S|$) matrix where each row's entries equal that of the associated student's averaged value.

4.2 Decision data

The following decision data are available from the students. Some graphs are given, depicting histograms of chosen data.

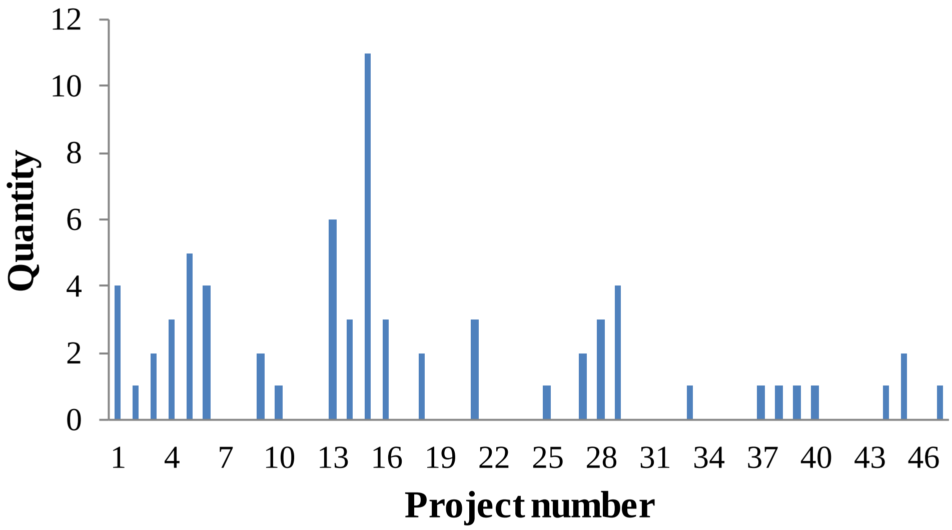


Figure 1. Distribution of first project choices.

Figures 1 to 3 show the positive preferences for the students first to third project choices. Note that the 'Quantity' on the y-axis refers to the number of times a student indicated the preference of the corresponding 'Project number' from the x-axis. From the graphs it follow that 26 indications were provided that a student wishes to be allocated to *Project 15*. Note, furthermore the similarities between the three graphs, indicating that a typical scheduling solution might prove difficult to solve, if only these data are taken into account.

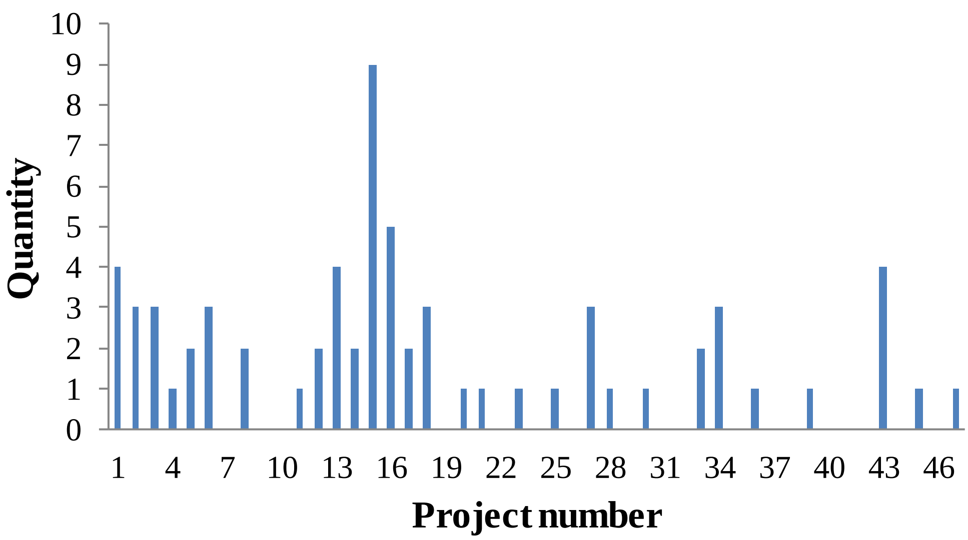


Figure 2. Distribution of second project choices.

To demonstrate this, the data from Figures 1 to 3 were firstly used for project allocations. Utilising these profiles, the MILP solves to allocate only 46 non-zero project weights. As a result, 27 students or 37% is awarded a project that where no interest was indicated. This will typically result in a number of displeased students, since no indication is available on how the 27 projects should be allocated.

Taking the data selection into account, as prescribed above, another eleven graphs can be provided. However, only two more histograms are plotted, which are the preferences for first study leader and research group.

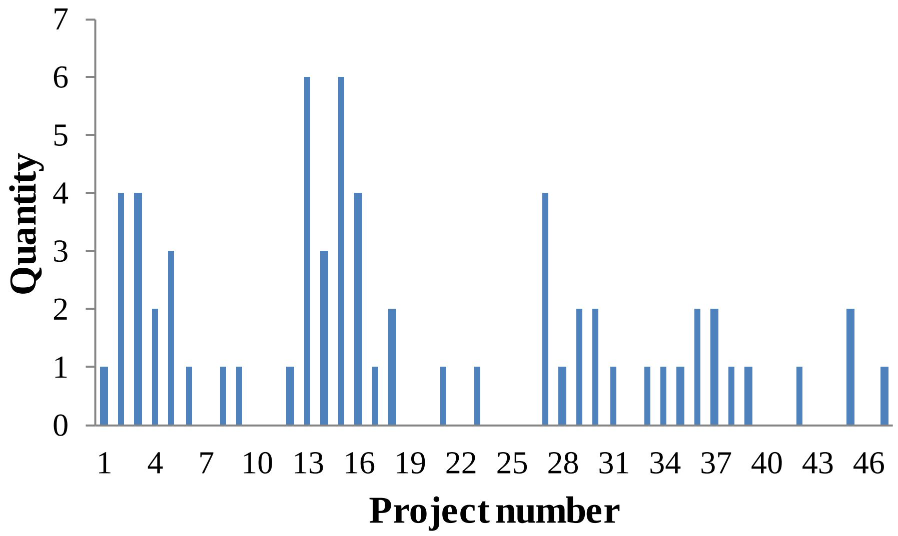


Figure 3. Distribution of third project choices.

The most preferred project leader and research group choices are given in Figures 4 and 5. From Figure 4 is clear that one project leader is the top choice for 22 of the 73 students (30%). This, furthermore, emphasises the difficulty of a typical scheduling problem, i.e. numerous decision making entities comprising likewise preferences. It is important to note that there are 15 study leaders and that *Project leader 6* can only accommodate a single student, *Project 15*, which is the most popular first, second and third choice. If no relevancy can be found with *Project 15*, 25/72 (35%) of project preference indications are lost for the decision making process. This emphasises the importance of finding conceptual similarities amongst the projects (resources).

It is important to note that the single project from *Project leader 6* is part of *Research groups 1* and *12*. Therefore, even though 22 students desire to work under *Project leader 6* and 26 indicate a preference for *Project 15*; only six students chose the research groups relevant to this project as a first choice. Also note that *Research group 1* is the most relevant for *Project 15*.

By taking all the information into account, each student's weight factors are calculated for every project. Note that the default weight factor is set at zero. These weights are now utilised by the MILP and solved for. Optimisation results are obtained and indicate that 71 projects are allocated toward a positive weight. The two remaining projects are allocated to zero weights. It is important to note that not a single project is allocated amongst a negative weight.

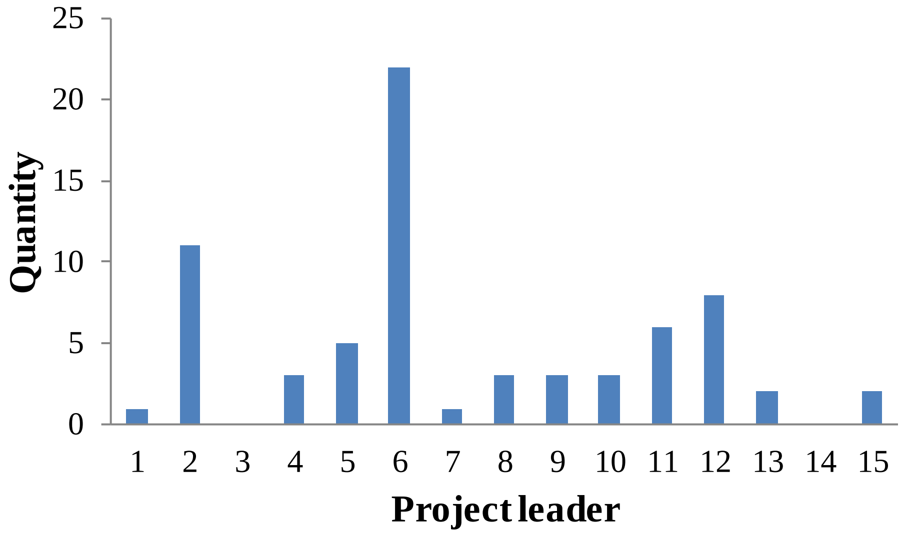


Figure 4. Distribution of most preferred project leader choice.

The additional information acquired from the students allowed for a substantial number of non-zero weight calculations by the algorithm. On average a student was assigned a positive weight for approximately 26 (36%), a negative for 28 (38%) and zero for 20 (26%) of the projects. Further note, the two students who were not awarded a project with a positive weight, received a neutral (zero) allocation.

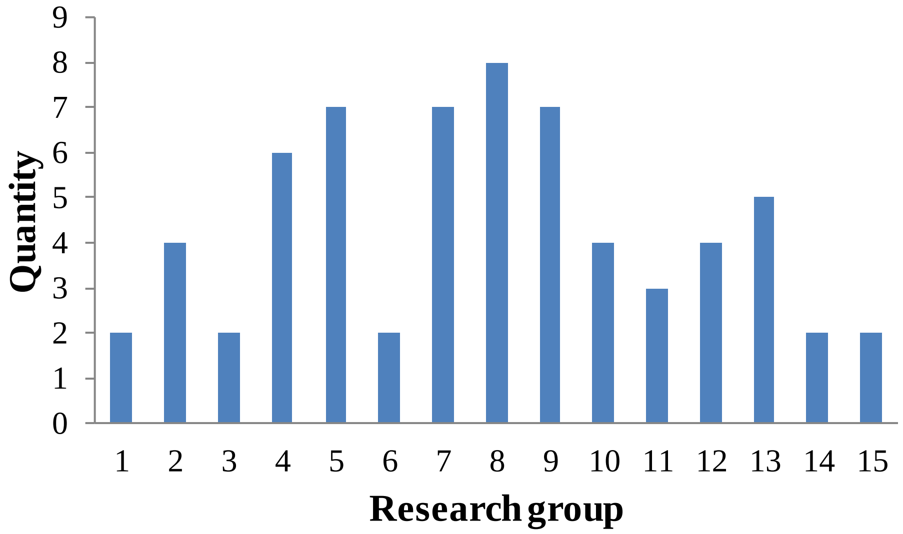


Figure 5. Distribution of first research group choices.

This case study shows not only the relevance, but also the necessity of the proposed MILP and novel pseudo-code algorithm. This algorithm, however, does imply that additional input parameters are required, and furthermore, an understanding of the data. Without the knowledge that each project comprises a study leader and the relevance of the research groups, the various weight allocations are not possible. This case study, therefore, emphasises the need to not only understand how to solve a problem mathematically, but be able to interpret all the relevant data in order to set up a model with a higher probability of yielding sensible outcomes.

5. SUMMARY AND CONCLUSION

This paper investigated the scheduling of resources or projects amongst recipients. Scheduling is an everyday occurrence with a variety of methodologies that can be applied to yield outcomes for this process. Improved or optimal scheduling may be required from a monetary, time or even recipient satisfactory level's point of view.

Two models were formulated in this paper. The first is a mixed integer linear programming model that incorporates binary decision making variables to solve for optimal resource or project scheduling. The MILP requires input parameters that contain a measure of preference. These preferences, however, might at instances not be sufficient to solve for satisfactory results. A novel pseudo-code algorithm was therefore proposed. This algorithm can be used to calculate preferences or weights for resources not initially indicated by the recipients. Dependencies between different resources or project are explored during the calculation process.

A real life case study was presented where 73 from 73 projects needed to be allocated amongst 73 students. When only, as accustomed, an indication of the top three preferred project allocations were available, the MILP could only solve for 63% of the students to receive some level of preference.

By incorporating additional information, utilised by the weight calculation algorithm, the MILP was re-solved. Optimal results indicated that 71 out of 73 (97%) of the students were allocated a project with a positive weight, where 3% received a neutral (zero) weight.

The results demonstrated the relevancy and importance of the proposed models. It is, however, required to gain additional data from the recipients and, furthermore, fully understand the nature of the data.

REFERENCES

[1] **Kutanoglu, E. & Wu, S.D.** 1999 On combinatorial auction and Lagrangean relaxation for distributed resource scheduling, IIE Transactions 31, pp. 813-826.

[2] **Zhong, H., Tao,. K. & Zhang, X.** 2010. An Approach to Optimized Resource Scheduling Algorithm for Open-sourceCloud Systems, The Fifth Annual ChinaGrid Conference, IEEE, pp. 124 -129.

[3] **Buyya, R. & and Murshed, M.** 2002. GridSim: a toolkit for the modeling and simulation of distributed resource management and scheduling for Grid computing Concurrency, Computat. Pract. Exper 14, pp. 1175–1220.

[4] **Alcatraz, J. & Maroto, C.** 2001. A Robust Genetic Algorithm for Resource Allocation in Project Scheduling, Annals of Operations Research, 102, pp. 83–109.

[5] **Herroelen, W. & Leus, R.** 2005. Project scheduling under uncertainty: Survey and research potentials, European Journal of Operational Research, 165, pp. 289–306.

[6] **T. Logenthiran, T., Srinivasan, D. & Khambadkone, A.M.** 2011. Multi-agent system for energy resource scheduling of integrated microgrids in a distributed system, Electric Power Systems Research 81, pp. 138–148.

[7] **Chen, J. & Askin, R.G.** 2009. Project selection, scheduling and resource allocation with time dependent returns, European Journal of Operational Research, 193, pp. 23–34.

THE EFFECT OF AVERAGING; WHY STOCHASTIC OPTIMISATION APPROACHES ARE IMPERATIVE

P.v.Z. Venter[1*], S.E. Terblanche[2], M. van Eldik[3]

[123]School of Industrial Engineering
North-West University, South africa
[1]philipzventer@gmail.com, [2]fanie.terblanche@nwu.ac.za, [3]martin.vaneldik@nwu.ac.za

ABSTRACT

A common engineering practise is to implement averaging of production-related data during decision-making processes. Average values of a data set are typically applied to determine an operating set-point or the sizing of equipment for investment purposes. In a production environment, where production-related outputs are within a sufficient small standard deviation, making use of average parameter levels may prove to be satisfactory. This may, however, not be suitable towards engineering production plants where more erratic fluctuations are observed.

This paper demonstrates how the averaging of production data in a fluctuating process environment may result in sub-optimal design and investment decisions. A case study is presented where results from an investment optimisation model, which incorporates historic profiles, are compared to possible investment outcomes based on average parameter levels. Results indicate that typical averaged production profiles may yield sub-optimal investment decisions with over-optimistic expected incomes.

*Corresponding author

1 BACKGROUND

Energy efficiency and better resource utilisation are frequently addressed all over the globe, especially in the engineering sector. With the exhaustion of energy resources, it is imperative for engineering plants to operate at efficient levels, both from an ethical and monetary point of view. Energy efficiency through lower resource utilisation will not only benefit the environment, but it will promote the profitability of an engineering company. For every engineering plant, operational and strategic goals should typically also include cost reduction initiatives. These costs may, for instance, include investments in the most profitable and therefore suitable equipment [1].

Investment decision making is typically an important managerial task. Informed decisions need to be made from suitable information. Any investment choice should rely on in-depth research and, where applicable, a sufficient modelling process. The aim of such a model should typically be to determine what the expected effect of the proposed change will be on the production outcomes, either with respect to quality or net income. If a simulation model is formulated, results can be generated and compared in order to determine the best-suited decisions.

The accuracy of any simulation model depends, however, on its ability to capture reality, which may include the assumptions used in the formulation process. Therefore, the quality of outcomes generated from a simulation model is dependent on the accuracy and relevancy of these assumptions, in combination with the objective function.

An important investment decision guideline, which might be overlooked within the assumptions, is the typical process flow conditions that the equipment experience on a day-to-day basis. In a steady-flow production environment, process conditions typically do not deviate from the average or expected values under standard operations. This is applicable for engineering plants where the flow rates and chemical compositions to the production processes are continuously under steady-state conditions. However, even for investment decisions under steady-state conditions, the best outcomes are not necessarily trivial to compute.

This paper investigates how investment decisions are influenced when different assumptions for an optimisation model are incorporated. The optimisation of an engineering plant that operates in a fluctuating resource environment has been presented [1]. This stochastic mixed integer linear programming (MILP) model optimises the net present value (NPV) of power (electricity) generation turbine investments in a fluctuating resource availability environment. Results show that in a fluctuating production environment, the investment decisions are biased with regards to modelling assumptions. The results emphasise the shortcomings of utilising averaged values within the simulation process that can lead to sub-optimal investment decisions.

In the following section a literature review on the modelling of energy recovery plants is provided. Section 3 provides information regarding an investment problem under fluctuating resource conditions, together with various investment decision making techniques. A real-life case study is presented in Section 4, followed by a summary and conclusion.

2 MODELLING OF ENERGY RECOVERY PLANTS

Engineering process plants typically produce numerous by-products throughout its production chains. These by-products are either dispersed off or utilised. For this paper, a by-product that is used as an energy source in other plant processes will be referred to as an energy resource.

A common practise for engineering process plants is to utilise self-generated resources for steam production in boiler houses. For process plants, these resources are in most cases in the form of usable off-gases. The production of off-gases typically forms part of continuous production processes, where some varying quantities are utilised as energy sources for the company. Due to the continuous nature of these process plants, residual off-gases must be flared into the environment, wasting its energy recovery potential. This results in a negative environmental impact, since off-gases typically comprise carbon-based molecules that results in carbon dioxide emissions. It should be noted, however, that due to pipeline pressure control for these open systems, off-gas flaring is always required. Flaring avoids oxygen enriched air to enter a pipeline network filled with combustibles, which poses serious operational risks.

Steam, in varying flow quantities, is continuously used for numerous processes throughout some engineering companies' production chains. It is, however, critical to ensure that steam demands are always met. Excess steam and steam generated from burnable off-gases may potentially be utilised for power generation. If such favourable conditions exist for an engineering plant it may be sensible to invest in power generation turbines, also known as energy recovery [1].

Figure 1 shows a simplistic layout of a typical engineering plant. Several raw materials are fed to the process plants where end-products are eventually produced. For some of these process plants off-gases are generated that are consumed in other processes. The off-gases not utilised in any plant processes, i.e. residual off-gases, are either used to generate steam in boiler houses or flared to the environment [2-7]. Excess steam is directed to turbines, which are coupled to generators, to generate electricity.

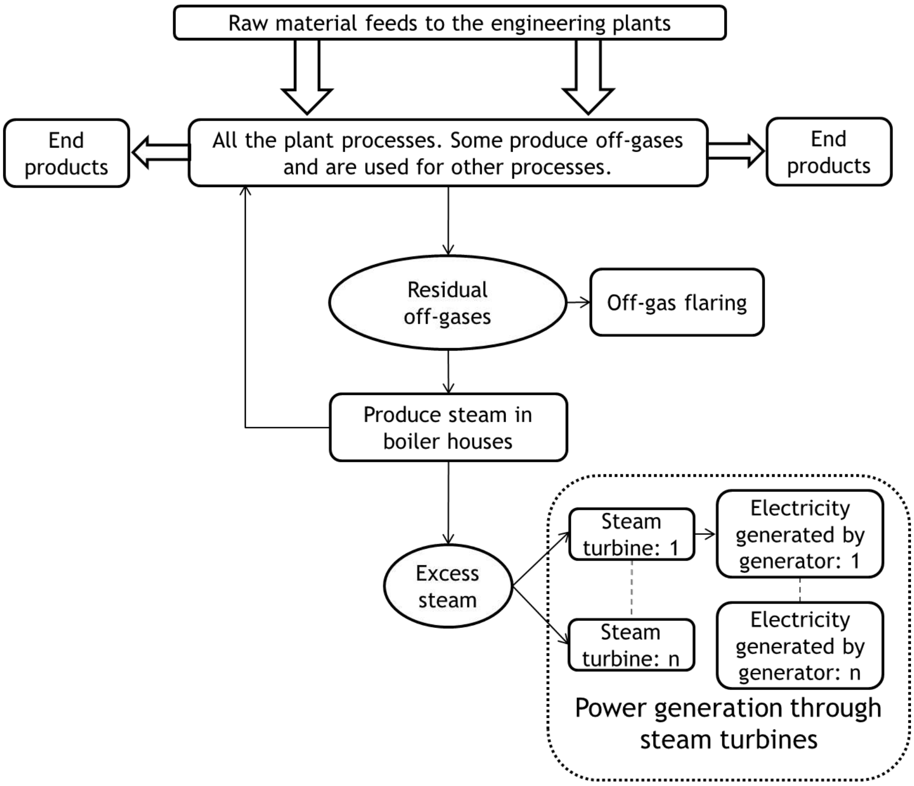


Figure 2. A simplistic layout of a typical engineering plant.

Variation in the quantities of raw material that feed to a process plant, or deviation of the chemical qualities thereof, may result in fluctuating off-gas production over time. In principle, fluctuating energy recovery is not necessarily problematic, unless power generation potential goes to waste, due to inefficient resource utilisation. It should be noted that due to the continuous operating nature of an engineering plant, unutilised steam cannot be stored for later use and its power generation potential will consequently go to waste [2-5].

If energy recovery takes place under fluctuating resource (steam) conditions, the turbine investment decision-making process might prove challenging. Note that the additional electricity generated has typically no influence on the engineering plant's production outputs and is small in comparison to plant electricity usages. As a result, supply and demand side need to be aligned for the energy recovery plant under consideration.

Various optimisation investment models exist for co-generation or energy recovery plants. These models are typically concerned with multiple power generation mediums and the minimisation of fixed and operational costs in order to determine optimal investment choices [8-15] More specifically, the optimisation models of [8-12] focus on linking energy demand with supply and make use of average-based estimates over time for optimal investment purposes.

Decision framework optimisation with integrated costing and performance evaluation is addressed by both [16] and [17], where an average weighted method is incorporated in the modelling process. For optimisation of civil engineering systems, the author of [18] used average values for the availability of structural monitoring data.

Several energy efficiency models are available in the literature. The authors of [19] proposed a mathematical optimisation formulation for resources operating under a fluctuating environment with a degree of uncertainty. This model may be used to determine optimal resource distribution for an energy recovery plant, by considering any number of operating turbines. The authors of [20] demonstrated the practical relevance of the formulation.

The model suggested in [1] optimises the NPV of turbine investment options, with the possibility of including investment and procurement of natural gas as an additional energy resource. Results indicated that such an expensive commodity can, at times, increase power generation profits in a fluctuating resource environment.

3 AVERAGING FOR INVESTMENT PURPOSES

As mentioned in Section 2, it is not uncommon to use averages for planning purposes in the engineering sector. This section provides some detail regarding averaging assumptions for the decision-making process within an energy recovery plant operating under fluctuating steam availability.

Expert opinion may provide an intuitive value for excess steam that is available for power generation. Although experience shouldn't be underestimated, it should still be used in combination with statistical and mathematical modelling techniques in order to support the investment decision-making process [21]. A numerically determined average of excess steam may provide an indication of expected steam flows in future, which in turn may be used to determine the turbines to be procured. When plant operations are of such a nature that only slight flow variations are observed over time, such an approach could be justified. However, when energy resources are present under fluctuating conditions, this methodology may prove to be insufficient [21].

To demonstrate the effect of averaging within the context of energy investment decision-making, consider a fictitious company operating under fluctuating conditions. Assume that two historic excess steam profiles exist and that these two steam profiles, i.e. Profile A and Profile B, are representative of what could be expected by the engineering plant. These profiles are plotted in Figure 2, together with the average over all the available data points (i.e. the dashed line at a level of 47 ton/h).

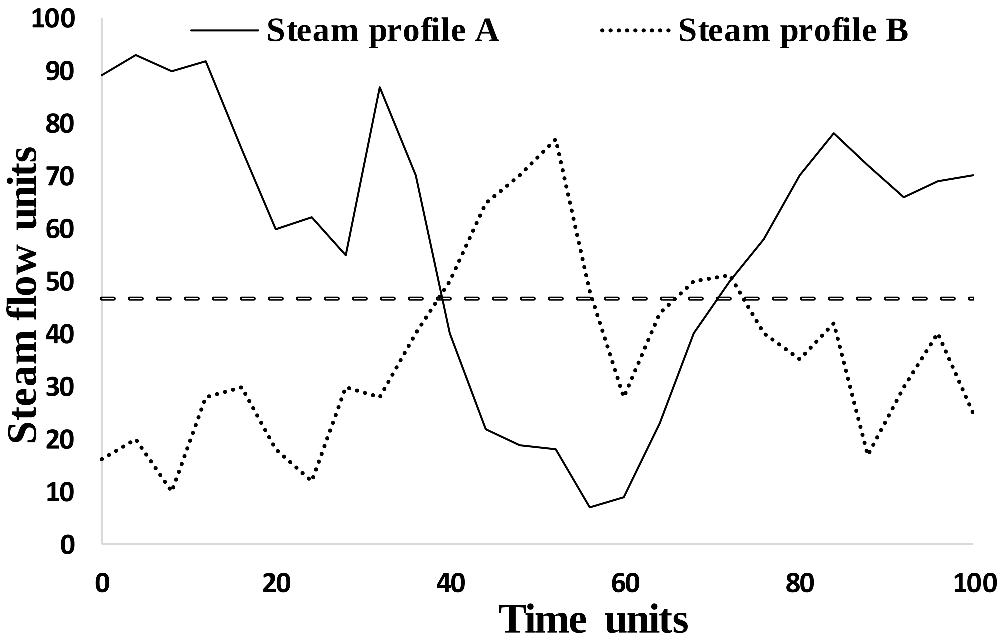


Figure 2. Two hypothetical steam profiles, i.e. A and B, with an average steam value.

Both Profiles A and B reveal the fluctuating nature of the company's excess steam availability over time, as can be observed from the graph in Figure 2. Assume the company explores the possibility to invest in energy recovery and that one of three turbines can be procured. Operating parameters, in terms of maximum and minimum allowable steam limits in ton per hour are provided in Table 1 for a small, medium and large capacity turbine.

Table 1. Three turbines' operating parameters in terms of maximum and minimum allowable steam limits.

Turbine	Maximum limit [ton/h]	Minimum limit [ton/h]
Small capacity	50	16
Medium capacity	60	18
Large capacity	75	22

All three turbines have equivalent efficiencies. It is assumed, only for explanatory purposes, that if steam is available above a turbine's minimum limit, it will be operational. Therefore, during each time period that the steam availability drops below the minimum allowable limit, the turbine experiences a trip. A trip is defined as when a turbine shuts down involuntarily. During a trip phase no steam can be utilised for power generation by that turbine.

The average value between the two profiles, also plotted in Figure 2, is just below 47 ton per hour. Using this value as investment benchmark, the small capacity turbine should be procured, since it can utilise steam up to 50 ton per hour. If the future steam flows do not fluctuate then all potential power generation will be realised by the small capacity turbine. Therefore, an assumption of average expected steam flow, indicates the best investment choice will be the small capacity turbine, with the expectation of operating at 47 ton per hour at any given time.

However, if in future either Profile A, or B realises, then from Figure 1 it is evident that steam above 50 ton per hour will go to waste, due to the turbine's maximum limit. Furthermore, both profiles consist of time periods where steam availability is lower than 16 ton per hour, indicating turbine trips. From this it can be concluded that for energy recovery under fluctuating resource availability, turbine investment decision-making should not necessarily be based on assumptions of average flow conditions.

An alternative data analysis approach is to consider an average profile over time. The average steam profile can be determined between profiles A and B and used for investment decision-making. The average profile between that of A and B is plotted in Figure 3. Note that the average profile captures some fluctuations, when compared to the averaged horizontal line in Figure 2. The operating limits of the best suited turbine, which is the turbine of medium size capacity, are also plotted on Figure 3. Note how the fluctuating average profile stays within the operating limits of the turbine. Thus, fluctuating power generation is anticipated, however, complete steam utilisation is expected with no unplanned trips.

If the average steam profile is used by this fictitious company for turbine investment planning, the medium capacity turbine should be procured. The procured turbine, however, will operate under steam profiles of similar nature to Profiles A and B. To graphically demonstrate a true representation of what could be expected for the medium size turbine, Figure 4 is provided.

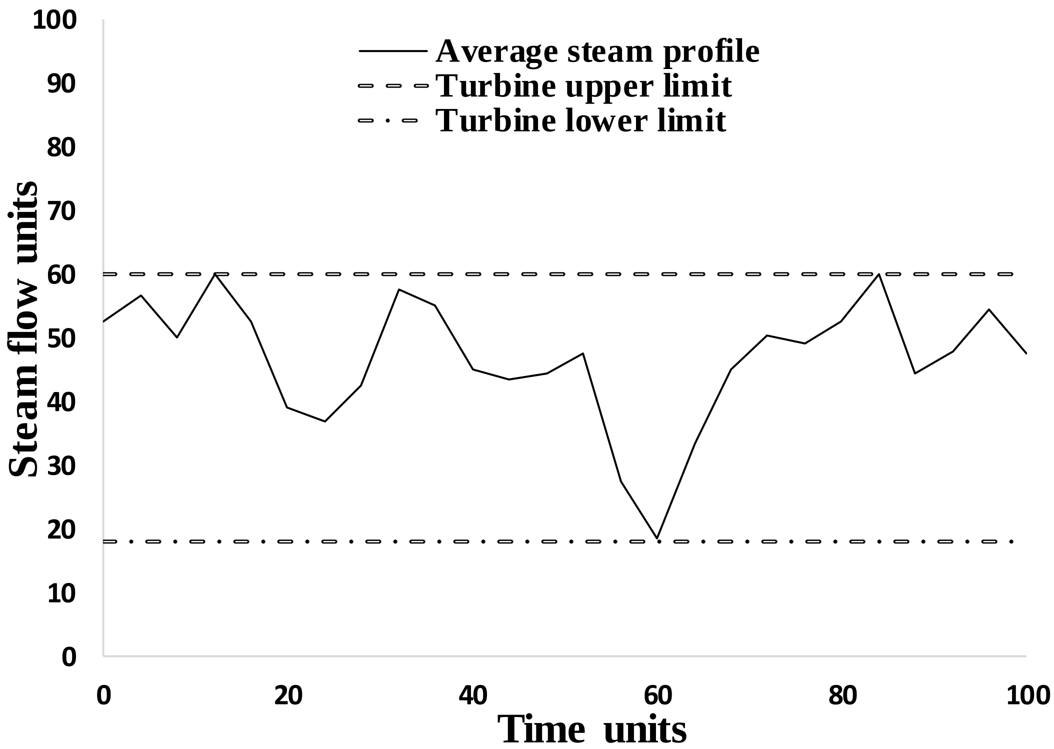


Figure 3. The average steam profile between profiles A and B over time, together with the operating limits of the medium capacity turbine.

The operating limits of the medium capacity turbine are plotted in Figure 4 together with Profiles A and B. Take note that for both profiles there are time periods in which the medium capacity turbine cannot utilise all or any of the steam. As a result, some power generation above its maximum limit and all below its minimum limit will be lost. This is due to averaging that can, to some extent, smooth out periods of high and low steam availability. Even though an average profile may be a good approximation of a fluctuating steam flow environment, it may be a misrepresentation of reality.

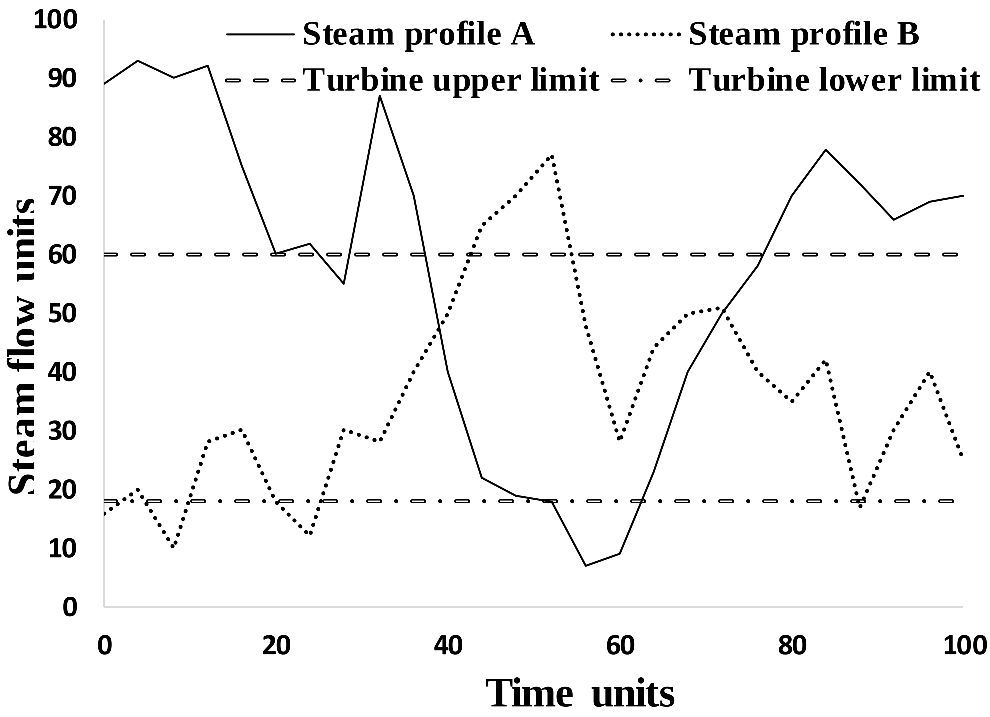


Figure 4. The operating limits of the medium capacity turbine under the average steam profile are plotted over time, together with steam profiles A and B.

Figure 5 shows steam profiles A and B together with the operating limits of the large capacity turbine. Note that even for this turbine, time periods exist during which steam availability is above its maximum limit. Comparing Figures 4 and 5 it can be observed that the medium capacity turbine allows for one less trip and additional power generation between the minimum flow limits of 16 ton/h and 18 ton/h. More power generation is achieved for the larger turbine between the maximum flow limits of 60 ton/h and 75 ton/h. Therefore, the large capacity turbine will incur more power generation.

From the above example it is evident that data analysis assumptions during the decision-making process, may have a significant influence on investment decisions. It is observed that averaging under uncertainty might yield sub-optimal investment choices and that the nature of process flows need to be accounted for. From this fictitious example, it follows that the large capacity turbine should be procured; however, what about other turbine options that were not included as part of the investigation? Another turbine may be more suitable than the large capacity one, or perhaps a combination of smaller capacity machines. Turbines are expensive machinery and long term investments; therefore, a decision-maker should strive to determine the "optimal" procurement option.

Investment decisions within an energy recovery context are primarily based on the most cost effective combination of turbines, which will maximise NPV. The decision trade-off is that an investment can be made in either a less expensive single large generating capacity turbine, or in a number of smaller capacity turbines which may be costlier. For the latter case, however, only partial generation capacity may be lost, compared to a significant loss in power generation capacity when a single large turbine trips.

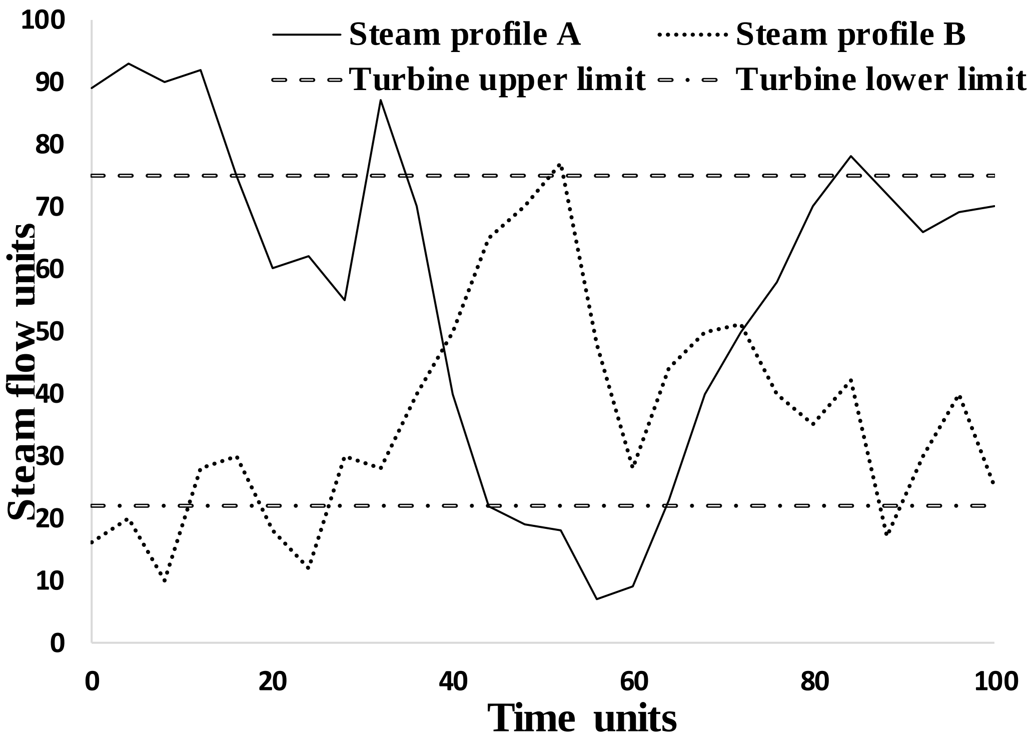


Figure 5. The operating limits of the larger capacity turbine with steam profiles A and B plotted over time.

To demonstrate these trade-off decisions, a case study follows which demonstrates what the long term monetary outcome will be under different assumptions for investment decision - making. For these comparisons to be made an optimisation model is required. To formulate a realistic optimisation model, a turbine restart that follows a trip occurrence must be understood.

Turbines are high speed rotating machines designed to operate continuously and only to be tripped for maintenance intervals, typically once every two to three years under perfect working conditions.

Protection measures are in place to prevent a trip occurrence near a restart. One such method is a restart waiting time. Before a tripped turbine can restart, sufficient steam must be available for some time. This sufficient available steam is defined as an interval leading to the start-up, where the tripped turbine can be operational without affecting the operational or start-up status of the remaining machines.

To demonstrate this concept for a single turbine, consider Figure 6. Let the turbine in Figure 6 be operational at 0 hours. At 20 hours steam availability drops below the machine's minimum limits so that it trips. The turbine, however, is only restarted at 80 hours, for none of the preceding time periods comprise an interval of sufficient length where the turbine could have continuously been operational. A typical waiting period found in industry is 15 hours.

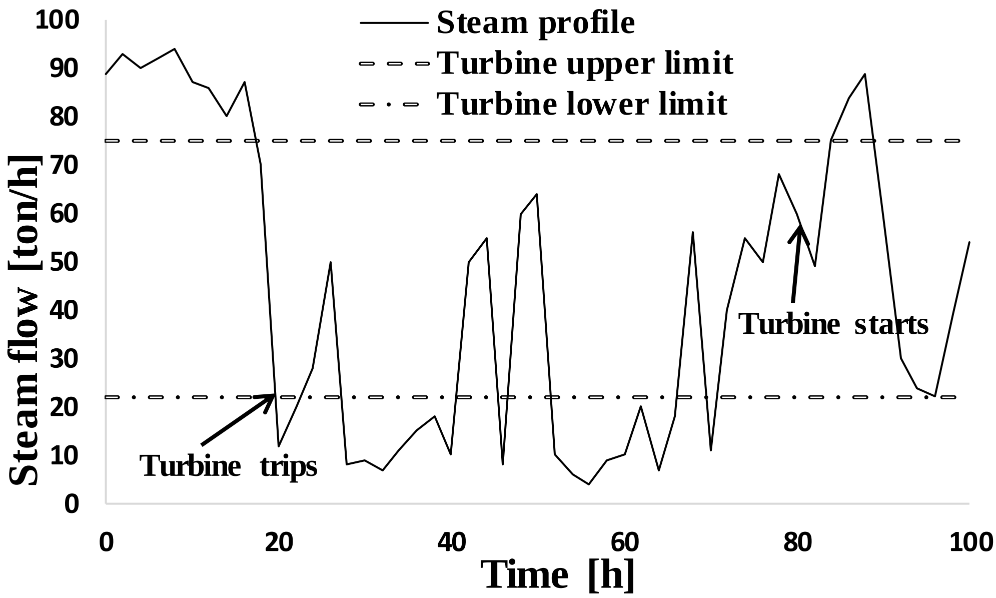


Figure 6. A hypothetical steam profile with a single operational turbine. The turbine trips at 20 hours and is restarted at 80 hours.

4 CASE STUDY RESULTS

The results presented in this section were generated by applying the Mixed Integer Linear Programming (MILP) model proposed in [1]. Steam flow data were obtained from an engineering company where steam is generated from off-gases. The data include steam production and utilisation measurements. Figure 7 shows production and usage data for a typical month as experienced by this company. The fluctuating steam production and usage patterns can clearly be seen on the graph. Assume that this company plans to invest in power generation turbines for energy recovery.

The fluctuating nature of excess steam is evident from Figure 7. A significant number of time periods exit when sufficient excess steam is available. There are, however, time periods during which insufficient excess steam is available that will typically result in turbine trips. These time periods can be observed between 200 hours to 350 hours, around 550 hours and 800 hours.

For all results to follow it is assumed that this company currently pays and will, therefore, safe R 500.00 for every MW hour generated. This value will increase by an annual inflation of 5.00%. The nominal annual income interest rate is also assumed to be 5.00 %. Take note that changes in inflation and income interest rates will influence the NPV calculations and is, therefore, only chosen for demonstration purposes. For purposes of demonstration, the waiting time period after a turbine trip is accepted as 15 hours, which is a typical industry standard.

The investment decision-making process is now addressed, incorporating assumptions and methods as discussed in this paper. Table 2 provides information on 10 turbines that are available for the decision-making process. Each turbine's capacity, upper-, lower limits and procurement costs are given. The steam to power conversion rate is chosen at a constant 5.5 ton per hour per MW for every turbine. It should be noted that there exist various turbine manufacturers around the globe. As for any machinery, the design specifications and costs and quality differ between manufacturers. These 10 turbines are, therefore, only used for demonstration purposes.

It is customary in most investment practises to determine payback periods and use that as decision benchmarks. The objective of this paper, however, is to compare results between different assumptions for the modelling process. Results are, therefore, compared with regards to an NPV over a fixed time period. This time period is chosen to be five years.

Table 2. Operating parameters and procurement costs of the 10 turbines used for simulation scenarios.

Turbine	Capacity [MW]	Upper limit [ton/h]	Lower limit [ton/h]	Cost (R 10^6)
1	3	16.5	8.5	36
2	5	27.5	12.5	50
3	10	55.0	20.5	95
4	15	82.5	28.5	140
5	20	110.0	35.5	180
6	25	137.5	42.5	220

7	30	165.0	49.5	260
8	35	192.5	56.5	280
9	40	220.0	63.5	320
10	45	275.0	77.5	400

The first method discussed in Section 3 is based on intuition. This process cannot be modelled and will therefore not be discussed any further. The second approach is to incorporate the numerical average value of the available steam. For the data set at hand this value is 164.0 – 46.3 =117.7 ton/h, which is the difference in averages for produced and utilised steam. Note that for the data set given in Figure 7 these average values are 159.3 – 60.6 = 98.8 ton/h, indicating that during that month less than average steam was available for potential energy recovery.

If only the data from Figure 7 are used during the decision-making process, steam availability of 98.8 ton per hour is expected. From this and the information from Table 2 the logical choice will be to invest in Turbine 5, the 20 MW machine. However, by taking the entire data set into account the expected steam availability is on average 117.7 ton per hour. This will result into procuring Turbine 6, the 25 MW machine. This stresses an important aspect; a sufficiently large data set is required that represents all or most scenarios that may be encountered in future. This paper, however, does not address sample sizes and it is therefore assumed that sufficient data are available for this exercise.

The MILP is solved over a five-year period for a steam profile that comprises a constant flow value of 117.7 ton per hour. As expected, Turbine 6 the 25 MW machine is determined to be the optimal investment choice. Results indicate that on an annual basis, power will be generated at 21.4 MW without any unplanned turbine trips. The expected NPV for a five-year term is determined to be R 236.3 million.

The third decision-making approach discussed in Section 3 is to utilise an average steam profile. The available steam data are combined to compute an average monthly profile. This average monthly profile is plotted in Figure 8. Note from this graph how a fluctuating nature of the steam profiles is captured. However, when Figure 8 is compared to Figure 7 extreme fluctuations are not captured within the average profile.

The MILP is now solved for the average profile. Since this single profile from Figure 8 is assumed to represent all excess steam profiles, a repeat thereof can be used as input parameters towards the model. The rationale is that a single month's steam profile is not sufficient for long term planning when expensive equipment is to be procured. However, when this profile is repeated, considering that the rate of power generation from a turbine is only dependant on its fixed design parameters, the timeline can be extended.

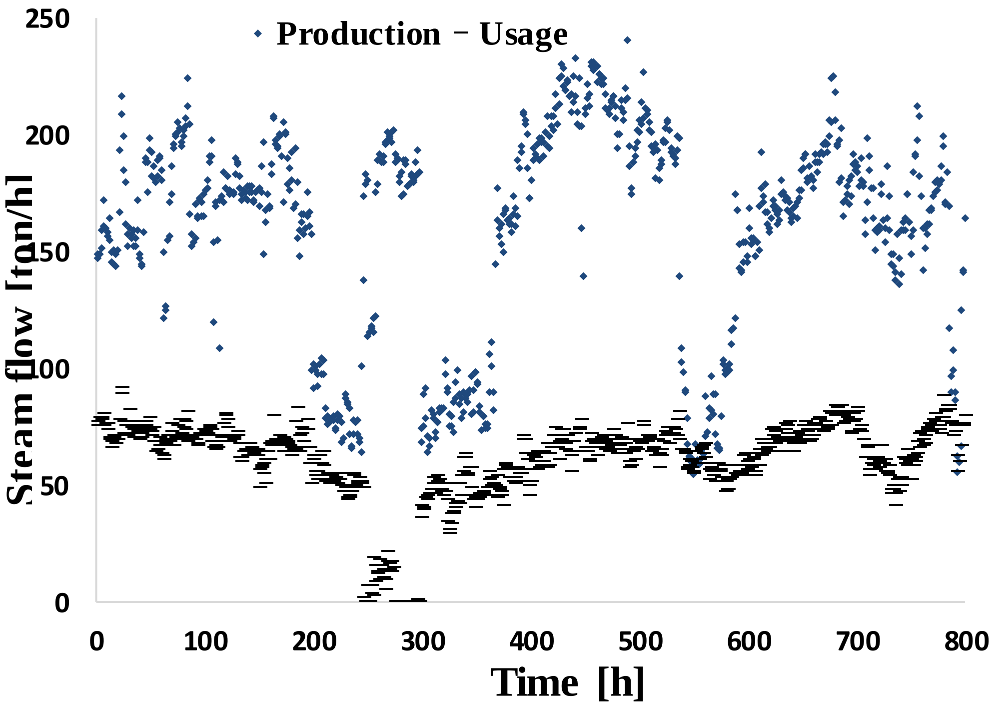


Figure 7. Measured steam flow data over 800 hours for steam productions and utilisations for the company.

Solving the MILP for the 10 turbines from Table 2, optimal results indicate that a single Turbine 6 should be procured. Therefore, even though an average steam profile is used rather than the average value, optimal results indicate an equivalent investment decision. These MILP results also indicate that no unplanned turbine trips are expected. Power generation, however, is predicted at 21.0 MW, which is 1.9% lower than predictions under a constant flow. At an NPV prediction of R 228.5 million this relates to a 3.3% lower net income expectation over a five-year period.

It is important to note that even though an average steam profile results in a lower expected power generation and NPV, optimal investment results are equivalent. Furthermore, if only a single average value is used, an optimisation model might not be required for investment decision making, since expected results can be determined from standard time-series calculations. It is therefore required to model detail flow profiles and compare the results, in order to make an informed decision.

The MILP is now solved for all the data without taking averages between different steam profiles. With the objective to optimise NPV over a five-year period, results show that two turbine investments are required. For an optimal NPV the company should invest in Turbines 2 and 5, the 5 MW and 20 MW machines. Note that the combined capacity is 25 MW, which equals the capacity of both previous investment decisions. Firstly, an important observation is that average approaches do not yield outcomes that significantly differ from optimal results; however, it does not yield equivalent investment decisions.

Optimal results indicate that the 5 MW turbine will, on average, produce 3.3 MW annually and experience 33 trip occurrences. The 25 MW will operate at an average of 15.9 MW with 34 trips per annum. Therefore, combined power generation of 19.2 MW with 67 trips is expected over a 12-month period. The five year NPV is R 179.0 million.

The true optimal NPV is 24.4 % below the expected NPV under an average flow condition and 21.7 % lower than that of an average profile. Expected optimal power generation is 10.3 % and 8.6 % lower than the two average based modelling approaches.

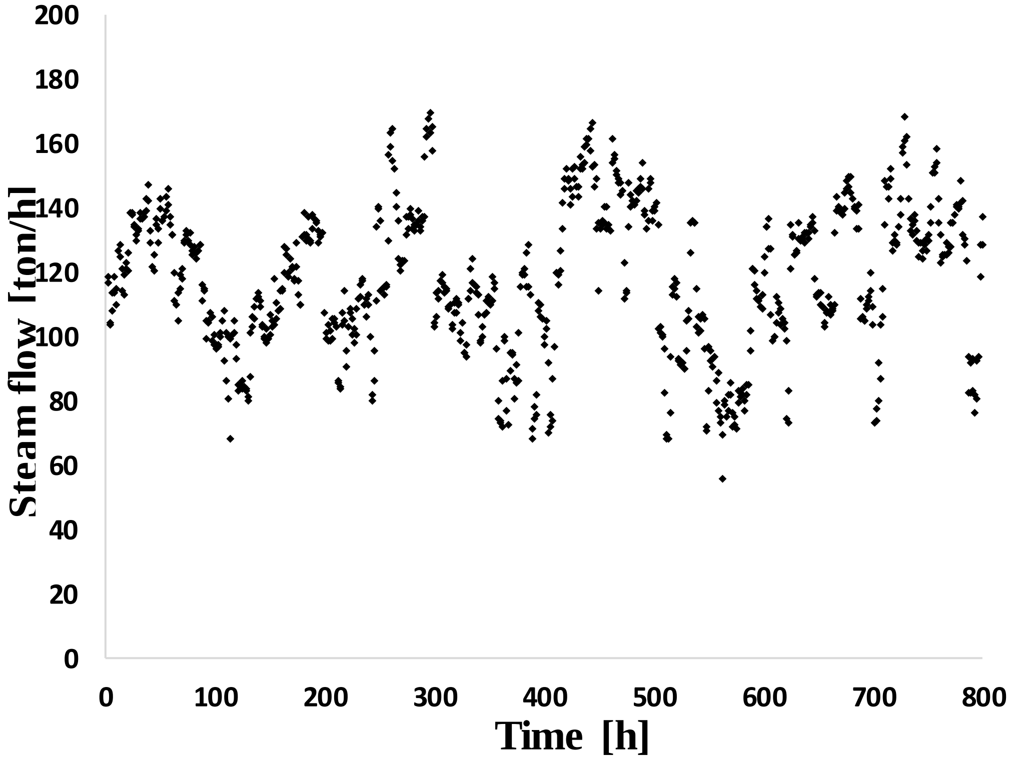


Figure 8. The average monthly available steam profile for the company.

From the results it is evident that under average assumptions, unrealistic higher power generation and NPV's are expected, and furthermore, sub-optimal decision-making results. Therefore, a stochastic optimisation approach is required that takes into account multiple steam profiles that are representative of stochastic future realisations.

The MILP is now solved for multiple steam profiles, where only Turbine 6 may be procured. The rationale is to simulate the expected investment outcomes of this sub-optimal turbine investment decision. If Turbine 6 is procured it is important to know what the true expectations should be and how those compare with the proposed, as well as the optimal outcomes.

Simulation results for Turbine 6 indicate that, on average, 52 trips are expected per annum with an average power generation rate of 18.2 MW. Note that 15 less trips are expected when compared to optimal results; however, with 1.0 MW less power generation than optimality. Further note, these results indicate 52 more trips per annum with between 2.8 MW and 3.4 MW less power generation when compared to results under average-assumptions. The NPV is predicted at R 166.6 million, which R 12.4 million or 6.9% below optimality.

Comparing the multiple profile simulated results for Turbine 6 with those under average-assumptions, the following observations are made:

- If Turbine 6 is procured a prediction of 52 unplanned trips will realise compared to zero expected.
- Power generation will be 13.3% to 15.0% less than expected.
- The true NPV of the investment over five years will be 27.1% to 29.5% below the expected value.

The most significant result is that the NPV is almost 30% below what is expected. It is important to note that accurate NPV values are not only relevant, but critical in the decision-making process. Therefore, inaccurate assumptions may result in other project investment not undertaken, due to a 'false-sense' of expected incomes or savings.

To conclude, this case study demonstrates that it is imperative to apply stochastic-based optimisation models in order to assist with the investment decision-making process.

5 SUMMARY

This paper investigated investment decision making under fluctuating resource availability. For engineering process plants where sufficient excess steam exists, power generation turbines can potentially be invested in. These turbines allow for energy to be recovered in the form of electricity that would otherwise have gone to waste.

The results in this paper were produced by applying a stochastic mixed integer linear programming model from literature. Assumptions with regard to input steam profiles were investigated through a real-life case study. Three steam profile instances were investigated of which the first two instances were generated under an average-assumption. The third data instances comprised multiple steam profiles which represented stochastic scenarios of potential future realisations.

This paper demonstrated that under fluctuating resource availability, the investment decision-making process should be based on stochastic optimisation approaches. By relying on average-based input parameters, unrealistic expectations are created which may be costly to a company in the long run.

REFERENCES

[1] **Venter, P.v.Z., Terblanch, S.E. & van Eldik, M.** 2018. Turbine investment optimisation for energy recovery plants by utilising historic steam flow profiles, Energy 155(2018), pp. 668-677.

[2] **Stoecker, W.F.** 1989. Design of thermal systems 3rd edition, McGraw-Hill.

[3] **Ganapathy, V.** 1991. Waste heat boiler deskbook. Lilburn GA(US); Fairmont Press, Inc.

[4] **Hewitt, G.F., Shires, G.L. & Bott, TR.** 1994. Process Heat Transfer. Begell House.

[5] **Turns, S.R.** 2006. Thermalfluid Sciences An Integrated Approach. Cambridge University Press.

[6] **Sonntag, R.E., Borgnakke, C. & van Wylen G.J.** 2003. Fundamentals of Thermodynamics, SixthEdition. Wiley & Sons, Inc.

[7] **Kotas, T.J.** 1995. The exergy method of thermal plant analysis. Krieger publishing company.

[8] **Henning, D.** 1997. Modest - an energy-system optimisation model applicable to local utilities and countries. Energy, 22(12), pp. 1135-1150.

[9] **Beihong, Z. & L. Weiding.** 2006. Multi-objective, multi-period optimization of biomass conversion technologies using evolutionary algorithms and mixed integer linear programming (MILP). Energy 38, pp. 189-195.

[10] **Fazlollahi, S., Mandel,P., Becker, G.F. & Marechal, F.** 2012. Methods for multi-obejtive investment and operating of complex energy systems. Energy 45, pp. 12-22.

[11] **Fazlollahi, S., & Marechal, F.** 2013. An optimal sizing method for cogeneration plants. Energy andBuildings 50, pp. 1504-1513.

[12] **Fumero, Y., Corsano, G. & Montagna, J.M.** 2013. A mixed integer linear programming model for simultaneous design and scheduling of owshop plants. Applied Mathematical Modelling, 37, pp. 1652-1664.

[13] Gamou, S., Yokoyama, R. & Ito, K. 2002. Optimal unit sizing of cogeneration systems in consideration of uncertain energy demands as continious random variables. Energy Conversion and Management 43, pp. 1349-1361.

[14] Hui, C.W. & Natori, Y. 1996. An industrial application using mixed-integer programming technique: a multi-period utility system model. Energy Conversion and Management, pp. 1577-1582.

[15] Lozano, M.A., Ramos, J.C. & Serra, L.M. 2010. Cost optimization of the design of chcp (combined heat, cooling and power) systems under legal constraints. Energy, 35, pp. 794-805.

[16] Fletcher, H.D. & Smith, D.B. 2004. Managing for value: developing a performance measurement system integrating economic value added and the balanced scorecard in strategic planning. Journal of Business Strategies 21, pp. 1–17.

[17] Tsaia. W.H. & Hung, S.J. 2009. A fuzzy goal programming approach for green supply chain optimisation under activity-based costing and performance evaluation with a value-chain structure, International Journal of Production Research 47(18), pp. 4991–5017.

[18] Frangopol, D.M. 2011. Life-cycle performance, management, and optimisation of structural systems under uncertainty: accomplishments and challenges. Structure and Infrastructure Engineering 7(6), pp. 389–413.

[19] Venter, P.v.Z. 2018. Investment and operational optimisation of an energy recovery engineering plant, PhD Thesis, North-West University.

[20] Venter, P.v.Z., Terblanch, S.E. & van Eldik, M. 2015. Off-gas power generation optimization using a mixed integer linear programming model. ORSSA 2015, pp. 81-90.

[21] Venter, P.v.Z., Terblanch, S.E. & van Eldik, M. 2015. Improving Power Generation from Fluctuating Off-Gas Productions, *ICUE 2015, pp. 225-230.*

BALANCING OF PRODUCTION LINE FOR CAR ENGINE COOLING SYSTEM COMPONENTS FOR INVENTORY REDUCTION

K.P.Lusiba[1*] & M. Dewa[2]

[1]Department of Industrial Engineering
Durban University of Technology, South Africa
kydiva.lusiba37@gmail.com

[2]Department of Industrial Engineering
Durban University of Technology, South Africa
mendond@dut.ac.za

ABSTRACT

Line imbalances result in decreased workstation and worker efficiency. The case study company is an international automotive component supplier striving to reduce waste in the form of waiting time and inventory. This paper presents an analysis of the balance of the plant and provide solutions for reduction for the inventory and waiting time at the furnace. The methods used include Pareto chart for evaluating and prioritising the products and the largest part of the total volume, analysing product mix and allocation to work centers, and generation of balancing chart fractals. Recommendations were then proposed to balance the lines so as to reduce the waiting time of the brazing furnace as well as inventory between the furnace and the assembly lines.

*Corresponding author

1. INTRODUCTION

Assembly lines are flow oriented production systems for industrial production of high quantity standardized commodities and line-balancing strategy can be used to make production lines flexible enough to absorb external and internal irregularities [1]. The research was conducted for a manufacturer of car heat exchangers, specialising in vehicle air conditioning and engine cooling systems. The factory consists of three different stages of manufacturing and preassembly of components. The case study company based in South Africa is an international automotive supplier and has 13 lines feeding the brazing furnace with products. These products can be categorized into three brazing families, which need different heating profiles at brazing temperatures when passing through the brazing furnace. The furnace then feeds subsequent nine assembly lines with these brazed products. The brazing furnace has, in general, per product, the lowest cycle time. Upstream, the furnace often waited for products to braze, whereas downstream, there is a lot of work in progress between the furnace and the downstream assembly lines. This implied that there could be a line imbalance induced by the three brazing product families. Waiting time and inventory obstruct a smooth material flow during production and are thus waste. The aim of the paper is to analyse the balance in the plant and to find the reasons for the inventory behind the furnace and the waiting time of the furnace, and thus propose recommendations to abate waste through line balancing.

2. LITERATURE REVIEW

2.1 Line Balancing

Line balancing is a technique utilized to minimize imbalance between workloads and workers in order to achieve the required run rate [2]. Consequently, the analysis of the line must be in terms of assembly process, layout of workstations, and the cycle time of the workstations. A multiple activity chart can be utilized to measure and assess interrelationships between operators and machines, and could be described as an operator and machine chart. The task time variation is primarily due to human's instability with respect to work rate, skilfulness, motivation as well as the failure sensitivity of complicated processes [3]. These sources of inconsistency are minimized by controlling the moving cost of machine and men. The worker working time and variation of operator and machine cycle time results in line imbalance. Furthermore, line imbalance is created by the change-over time for mixed model line, which is essential to apply lean techniques. Based on customer orders and call offs, the number of operators and machine at a workstation are increased or reduced to reduce line imbalance. Man and machine flexibility is achievable through open information and material flow in the production process. Balancing chart is a tool to analyse a process and is often used in combination with a value stream map (VSM) [4]. Every workstation's cycle time is represented by a bar on a bar chart, and in this study, cycle time is referred to as the time between two units coming out of one line as finished goods [5].

The process of assembly line balancing involves three steps that include initially deriving cycle time by taking the units required (demand or production rate) per day and divide it into the productive time available per day. The second step is to calculate the theoretical minimum number of workstations. This is the total task duration time (the time it takes to make the product) divided by the cycle time. Thirdly, the line is balanced by assigning specific assembly tasks to each workstation. An efficient balance is one that will complete the required assembly, follow the specified sequence, and keep the idle time at each work stations to a minimum [6].

2.2 Product Mix and Allocation to Work Centres

Product mix is the total number of products that an organisation can manufacture or offers to its customer, and mixed-model assembly line can produce various products from the same assembly line, thereby providing flexible production that is aligned to variable demand [7]. Heuristics can be used to balance production lines that manufacture different product models in order to mitigate capacity constraints at workstations and escalate balancing efficiency [8]. The assembly line needs to balance so that there is minimum waiting of the line due to different operation time at each workstation. The sequencing is therefore, not only the allocation of men and machines to operating activities, but also the optimal utilization of facilities by the proper balancing of the assembly line. The allocation of work elements to work centres is crucial for responding to changes to product mix so as to introduce agility to assembly lines [7].

2.3 Sources of Waste

Eliminating waste requires constant effort at cost reduction to maintain continuous profits in manufacturing. The major technique to reduce costs is to produce products in a waste free environment or system. There are different ways to analyse and implement cost reduction, from the start of designing all the way through to manufacturing and sales. One of the goals of the Toyota Production System, however, is to locate waste and eliminate it. It is possible to uncover a very large amount of waste by observing team members, equipment, materials and organization in the actual production line. In every case, waste never improves value; it only increases cost [9].

There are eight sources of waste that include inventory, over-production, defects, waiting, transportation, motion, incorrect use of staff and their skills, and over processing. Inventory is described as any excess products held in stock that is not directly needed by the customer. Therefore, the stock on hold, if not in use in production, takes up a lot of space, resulting in increased inventory cost and high possibilities of having obsolete stock while stored [10]. Most companies that have implemented lean ensures that they install IT systems to control their inventory to ensure that money is not wasted on unnecessary resources, component or products [11]. The smooth, continuous flow of work through each process ensures that excess amounts of inventory are minimized. If work-in-process develops because of unequal capabilities within the process, efforts need to be made to balance the flow of work through the system. Inventory ties up assets such as cash and real estate and often requires additional handling which requires additional labour and equipment [12].

Over-production – Over-production take place when an organisation produces more than what the customer needs. This means producing parts that are not ordered or producing more than what is needed at that time. Over-production is the worst of all the types of waste since it has a knock-on consequence of increasing all the other types of waste. Over production amplifies costs of inventory, defective parts, unnecessary transportation, waiting time and avoidable motion [13]. With waste of waiting, most activities in manufacturing are dependent on the processes that are downstream and upstream, when resources, equipment, information or labour hold-up the production line for some reason, production cost will increase and the time is wasted and that adversely impacts on profitability [14]. Defects are described as parts that do not meet the customer specifications. Defective parts result in reworks and rejects that leads to costly production processes. Defects are triggered by poor production processes as an outcome of machine breakdown or human error. Reworking or reprocessing takes extra time and hence increases the finished product cost. Rejecting or discarding products incurs extra costs and excessive resources usage that affect the production line [15].

Transportation waste is described as any material, objects, parts and finished goods movements that are unnecessary leading to waste of time, resources and money [14]. Unnecessary transport is generally compared to unnecessary movement, which falls under motion waste and that results to parts damage and failure of product. Motion relates to the additional steps taken by operators and equipment thereby wasting operator's effort and time effort. Unnecessary movement results from poor basic procedures and training, and poor work process design [16]. Incorrect use of staff and their skills, in some cases, can also yield waste. If the skills and abilities of staff are not utilized properly, the result is loss of business time, failure to use operator's skills and suggestions, leads to less or discarded opportunities of improvement and education. Staff needs to be integral to the entire manufacturing process [17]. The operators on the shop floor can come up with ideas in which the ideas can be used to eliminate types of waste. Engagement can improve the processes and develop the staff's skills continuously to avoid over-processing. Over processing is the unnecessary steps taken in production process. This also refers to extra processes such as re-working and re-processing [13]. The other meaning of over-processing is producing products or an output extremely exceeding the customer's specifications that are required. In most cases this takes place due to malfunctioning machinery, faults during re-processing, poor information transferring, ineffective methods and not using the customer requirements as a point of reference (that includes in-house customers downstream in the process).

3. CASE STUDY BACKGROUND

The factory consists of three different stages of manufacturing of components, and these include core building, brazing, and assembly, and generally, it is a norm that operators are responsible for the quality of individual work.

3.1 Plant Overview

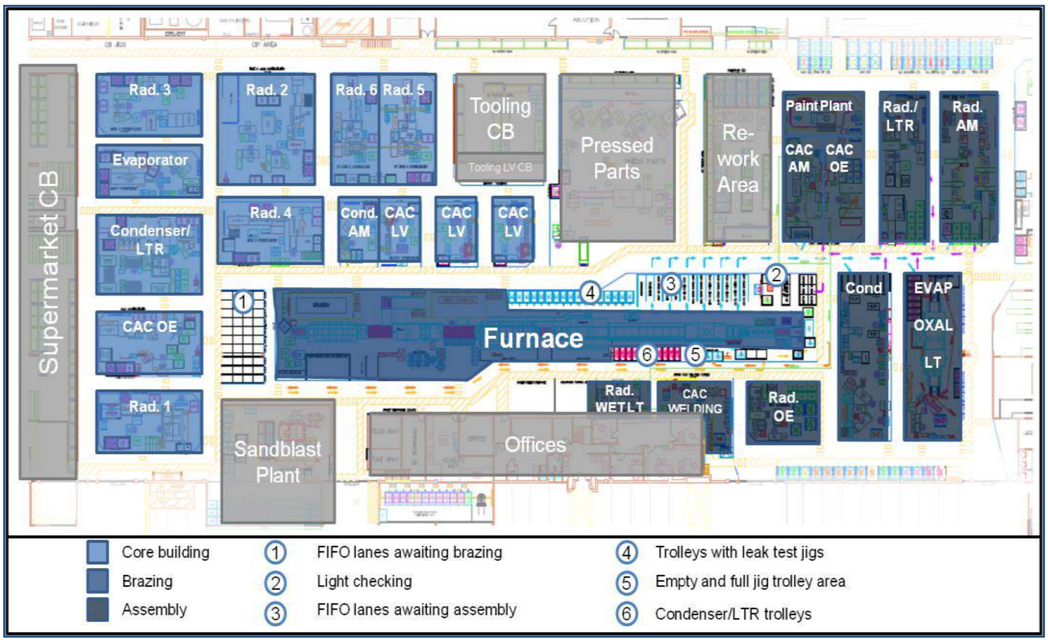


Figure 1: Plant Layout

The layout in Figure 1 shows an overview of the case study plant layout. Highlighted in blue are the core building lines, with a total of thirteen stations, from which a matrix is built, that is a combination of different small components assembled together. All the small components are provided from the core building supermarket, which is next to the core building (CB) lines on the plant layout. Lines one to six produce similar radiators while the remaining seven lines produce different components – condensers, Low temperature radiators (LTRs), Evaporators (Evaps) and Charge Air Coolers (CACs).

3.2 Production Process Flows

Figure 2 shows the full production process flow at the case study company. It commences with in bound logistics by which raw materials are brought in to core building and assembly lines. It also shows the linkages between core building, brazing and assembly, as well as reworks before the end product is sent to the finished goods warehouse.

3.2.1 Sub-Production Process Flow: Core Building

Figure 3 shows the sub-production line of core building, which is the first step of production on the engine cooling plant. This is the first stage of assembling single components into one subassembly, building of the matrix which is stacking fins and tubes together. Core building is where the headers or manifolds are hard-pressed into the matrix. Evaporators require a third step as the bended tubes have to be fitted on the manifolds. End of core building parts are packed on the jig for brazing and placed on a trolley, where the parts will be moved to the next station which is brazing.

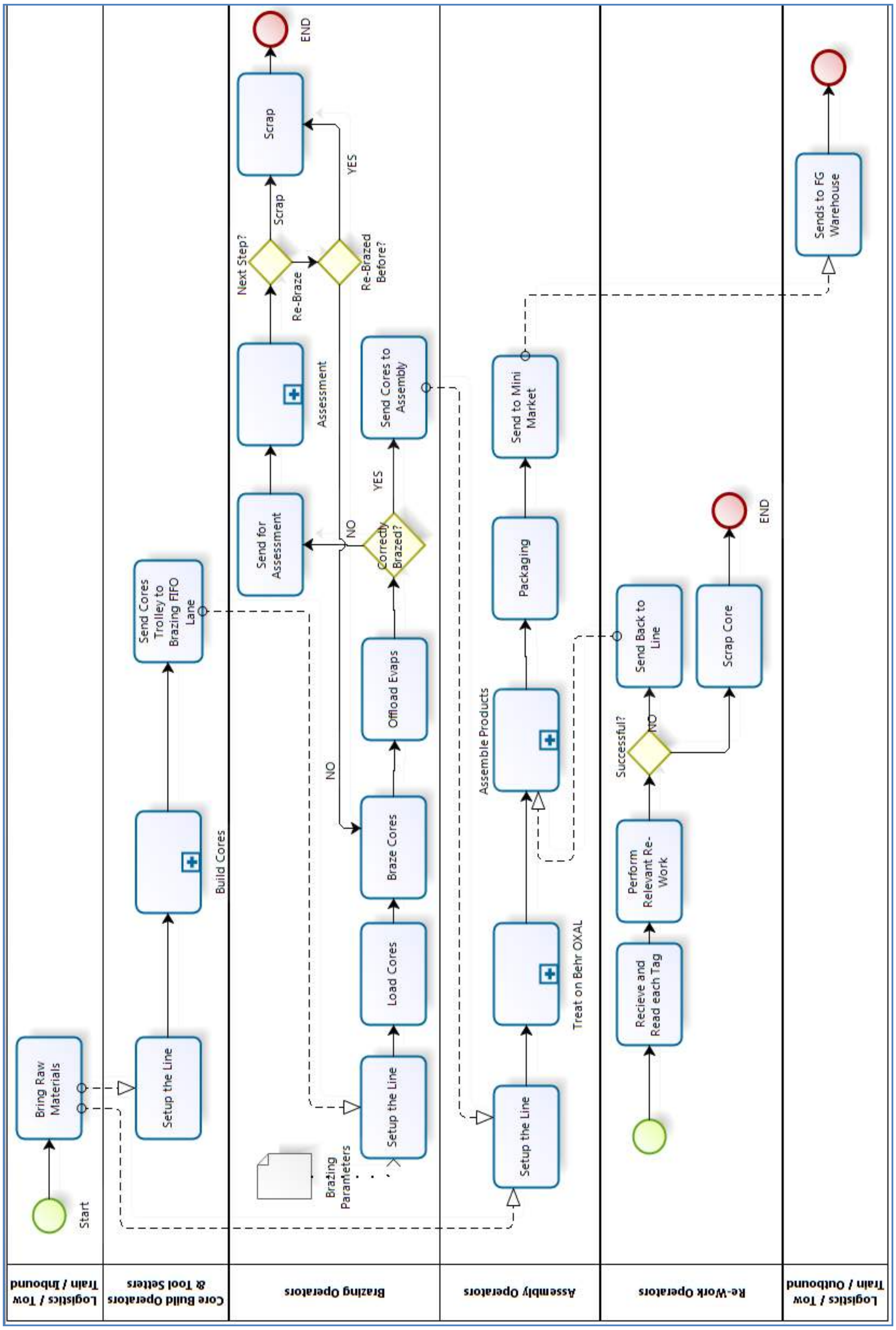


Figure 2: Full Production Process Flow

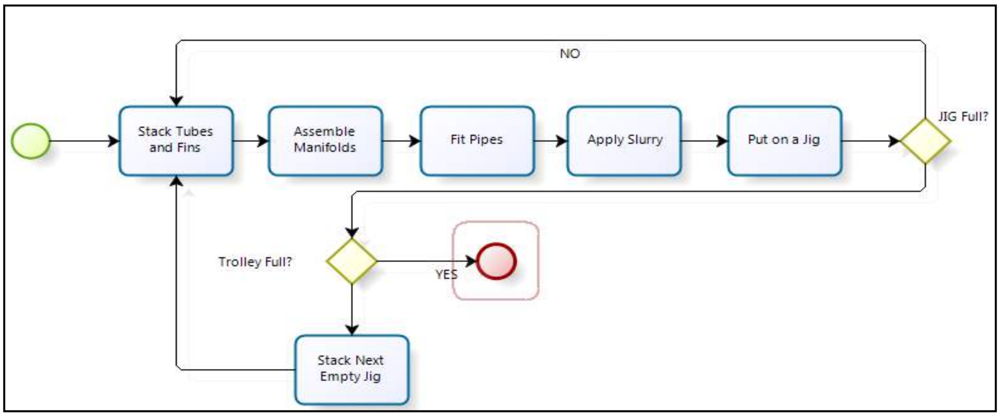


Figure 3: Sub-Production Process Flow: Core Building

Figure 4 also shows the assembly of components and building a matrix on the core building line.

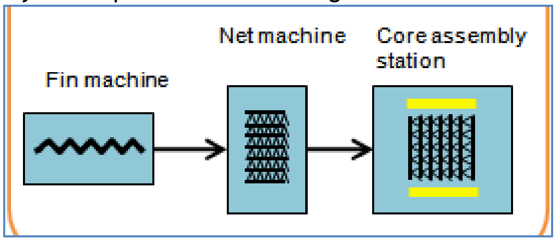


Figure 4: Core Building line - Assembly of components, building a matrix

3.2.2 *Sub- Production Process: Brazing Process*

Figure 5 shows the sub-production line of the brazing line which is the second step of production.

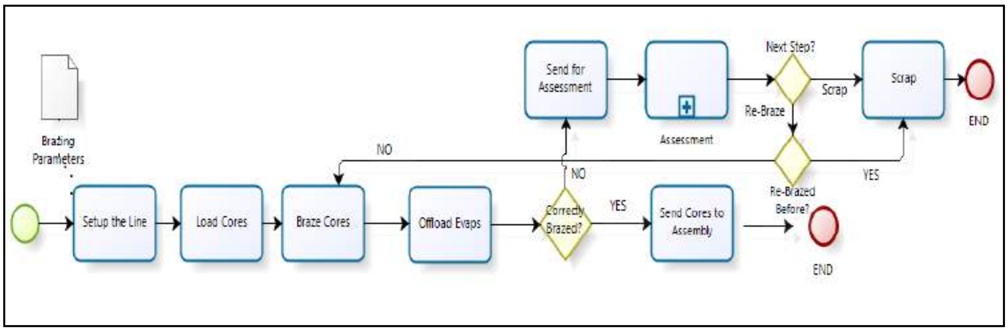


Figure 5: Sub- Production Process: Brazing Process

All parts go through this this step of production, where the brazing operation is carried out in a continuous flat belt furnace under controlled nitrogen atmosphere. The parts are subsequently kept on the First-in-first-out (FIFO) lane that is aside the assembly line. Components that have an aluminium silicon cladding are joined permanently at a lower melting temperature. The brazing furnace is loaded continuously and gaps of more than 1 m on the conveyor belt must be filled with dummy heat exchangers in the manual fluxing zone in order to avoid temperature fluctuations inside the brazing furnace. Stabilization of the brazing furnace temperature at the beginning of production is also accomplished by loading a minimum of 15 dummies. Also, after finishing with production, 10 dummies are loaded into the furnace. There are products that do not need the spray flux treatment as they use Silflux, which is flux that is already applied as a coating on the tubes of those products. These products include LTRs and some types of evaporators. For that reason, they are put on the conveyor just after the fluxing station.

There has to be a waiting time of 15 minutes after brazing Silflux parts, before the next spray flux part can be brazed and a 20 minutes waiting time for the other way round. This is because Silflux and spray flux parts cause different levels of humidity in the furnace. In addition to that, LTRs are exceptional in terms of the necessary belt speed. In order to prevent corrosion of the product later in its lifecycle the brazing process requires a very low belt speed. This leads to the following three brazing profiles, the different brazing profiles and Silflux and non-Silflux products are not the only constraints regarding the furnace. Most of the products need a product-specific brazing jig, which is only available at a limited number of units. As a consequence, core building cannot supply the furnace or the FIFO lanes in front of the furnace with unlimited numbers of products as they will run out of jigs. That being the case, those products have to be brazed in order to release jigs.

3.2.3 Sub-Production Process Flow: Assembly

The last work station for engine cooling is the valve assembly station where a valve is assembled to the core. The next step is the leak testing followed by final inspection and the product is packed into the box.

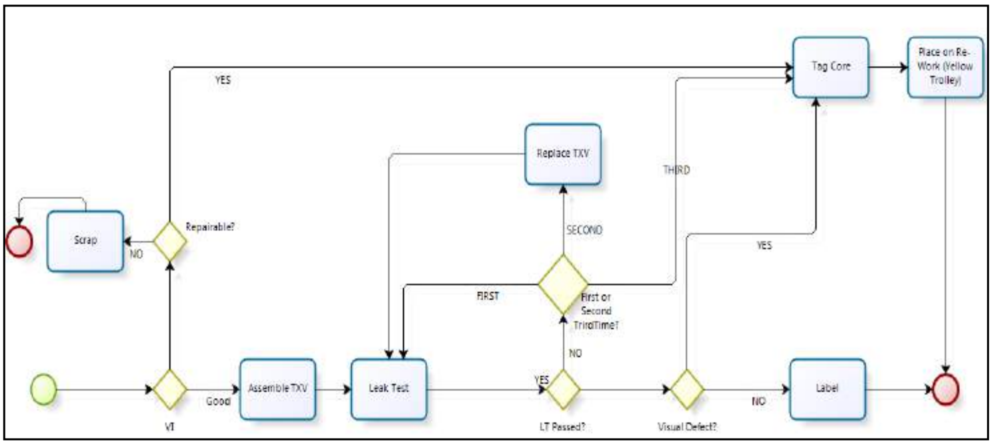


Figure 6:Sub-Production Process Flow: Assembly

This process happens just after brazing, the products are packed on post braze FIFO lanes and then collected there for the next step which is assembly line. Products without manifolds pass light check before that. The assembly process consists of crimping machines and air leak testers as shown in Figure 7. In addition, the CAC OE line has a paint station, as some CACs require coating and LTRs which do not require crimping just go through leak testing. Evaporators need more steps; they pass the “oxal station” where they get chemical treatment. The evaporators subsequently go to the valve assembly station where a valve is assembled to the core. The next step is the leak testing which is the last workstation, then final inspection and the product is packed into a box.

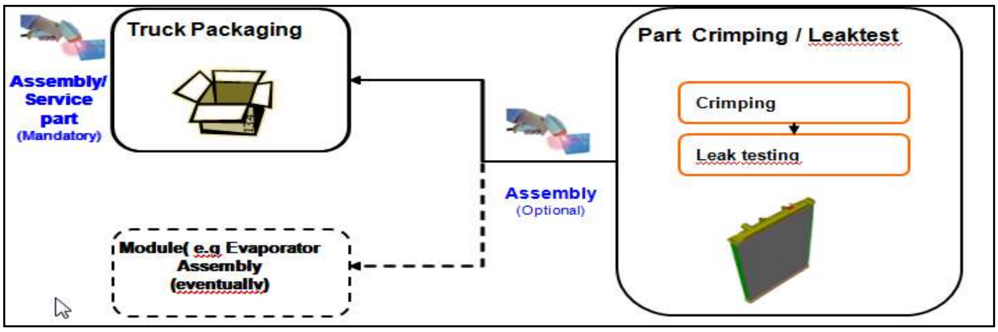


Figure 7: Assembly process -Crimping machines and air leak testers

3.2.4 Sub-Production Process Flow: OXAL

This process flow shows the sub-production line of the Oxal station which is the first step of the assembly line, only evaporators will pass the “Oxal station” where the product is chemically treated. The product is then moved to the valve assembly station where a valve is assembled to the core.

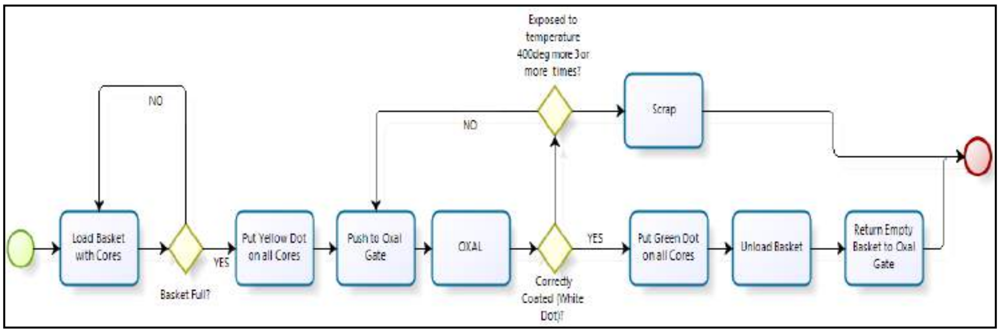


Figure 8: Sub-Production Process Flow- Oxal

Most of the previously described lines were designed to produce high volume parts. Very low volume aftermarket parts run on the three CAC LV lines. One operator per line builds the cores manually at the three CAC LV lines. These lines are very flexible since they consist of interchangeable units. One core builder table and one fin machine and both units are on wheels. Consequently, changeover times are very short since there are few assembly steps for fins, meshes and cores. There is no change of tools at the CB and no change of form rolls at the airway machine as the entire unit is moved in or out of the line. However, the cycle times are longer compared to the HV lines due to the manual processing. The CAC LV lines, the Condenser LV line is also entirely manual, but there is a slight difference in the sense that the core builder is not on wheels so there is a little less flexibility.

4. METHODOLOGY

The method used was evaluating the products and the largest part of the total volume, analyzing the product mix, movement and balancing analysis in the production line in order to be able to see the imbalance in the system. Balancing charts were generated helps to analyse the balance in between the three fractals in the case study plant. Firstly, the generation of the chart is depicted, then, the chart is used to analyse the current situation considering from 6 months before.

5. RESULTS AND DISCUSSION

This section addresses the line balancing analysis for the production fractals. In order to conduct a line balancing analysis, it is crucial to appreciate the product mix and allocation to work centers as well as the variety of parts produced by the case study company.

5.1 Product Mix and Allocation to Work Centers

There are several products that run on different lines and there are aftermarket products with very low volume and high volume parts that go to Original Equipment (OE) customers. A Pareto chart was prepared so as to get an overview of the different parts and volumes and then the allocation of the products to the lines was visualized. The Pareto chart aids to evaluate which products make up the largest part of the total volume, as the EC plant produces lots of low volume aftermarket parts and high volume OE parts.

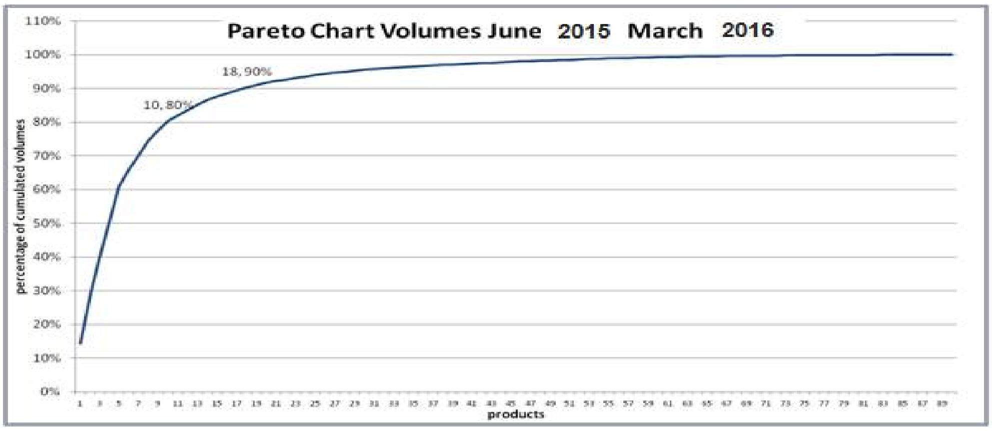


Figure 9: Pareto Chart for Engine Cooling Volumes

The volumes for the Pareto chart are taken from the heating fractal part numbers as the same product might have different assembly part numbers. The chart shows that there are 90 different brazed products and ten of those make up 80% of the total volume. Those ten are mainly CACs, Evaps, LTRs, one OE condenser model and the OE radiators. The allocation of products to work centres is crucial for ensuring high utilization of machines and assuring greater throughput. As already indicated, high volume products are processed at the OE radiator assembly line. In order to get an overview of the products and on which line they run, a matrix with a possibility to filter was created. It shows which products run on which core building lines and then are brazed on which profile.

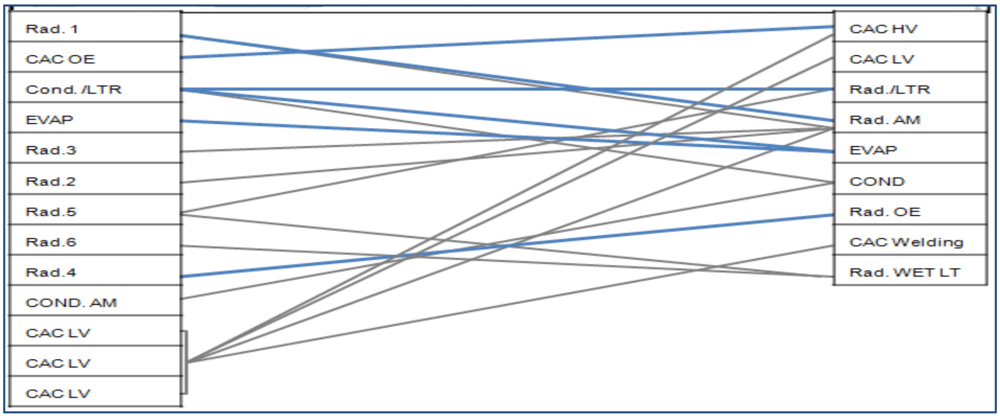


Figure 10: From Core Building to Assembly Lines without Considering Heating Families

Figure 10 summarises the relations for product families, without taking different brazing profiles into account. The routes for the high volume OE parts that could also be identified through the Pareto chart are displayed in blue, whereas the routes of low volume parts, produced once a month or less are shown in grey. As displayed, the HV lines for the CAC and the Rad. OE resp. Rad. 4 are dedicated to one product category and the assembly lines are fed by only one-part building line if the very low volume that goes from the LV Hand core builders to the CAC HV assembly line is disregarded. More complex is the situation for condensers, LTRs and evaporators since the Cond/LTR line produces LTRs are high volume and go to the Rad./LTR assembly line and low volume condensers go to the Cond. assembly line and one high volume condenser model that is leak tested at the evaporator line. The evaporator pre-assembly line only feeds the evaporator Assembly line. Figure 10 displays more grey links between core building and assembly but those routes are not described in detail as those parts are only produced once a month or less. A balancing chart was compiled in order to ascertain if there was an imbalance in the plant and establish that causes furnace idleness and high inventory level.

5.2 Generation of balancing chart for product families

The balancing chart was used to indicate the relations between the three product families during core building, brazing and assembly; and depicted as bars that consist of the net cycle time as well as additional times for defects, scrap, downtime and changeovers so that their sum results in gross cycle time. In addition, the customer takt time is displayed in the chart to see the gross cycle time in relation to customer demand.

There are multiple lines with multiple products that go to core building and assembly, which results in varying cycle times. All those parts go through the furnace but due to varying sizes of the products, the parts are brazed while stacked upon each other and the different brazing profiles cycle times vary too. As a result, the net cycle time of the balancing chart is calculated using the average cycle times of the products weighted by their volumes. The cycle times per part number and work centre were recorded on the SAP system. These cycle times are a result of time studies using videos of the process and an MTM study based on the filmed material since operators had a tendency to slow down if observed by human beings.

In addition to taking the weighted average of the cycle time for CB and Assembly, the result was divided by the number of lines that are active at the same time as there is not only one CB or Assembly line running at a time. Then the net cycle time bars for CB and Assembly indicate how many minutes one core would require. The number of lines running simultaneously is determined by the number of teams working in Assembly or CB and additional time for defects, scrap, changeovers and other downtime.

Defects occur at core building and the affected part is reworked immediately at the line by removing the damaged part e.g. a defective airway or a tube is replaced with a new part and parts that are not brazed properly are re-brazed. If parts at assembly do not pass leak testing, they are sent to a separate rework area; the defective parts get reworked and then sent back to the assembly line where they have to be leak tested again. Therefore, defective parts at core building, brazing and assembly reduce the output of the fractal, which means they increase the cycle time. Parts that go for scrap have to be rebuilt, which means a scrapped part from assembly has to be replaced which means the parts has to go through the same process again, a new core from core building, get brazed and assembled. Therefore, scrap influences the previous fractals as well because they are only detected in the assembly fractal.

Downtime records are taken per line at assembly and core building and the percentage for the balancing chart is also calculated using the weighted average of the downtimes per line and the corresponding volumes. This is not necessary for the furnace as this is only one machine with one downtime percentage. For CB and Assembly, the additional time for downtime is calculated as the quotient of the setup time and the lot size of the product. The weighted average of those times is then added onto the net cycle time, for the furnace, the percentage of downtime due to changeovers is used. Customer takt time for brazing is different from other fractals CB and

Assembly customer takt time (CTT) because of the different shift patterns and the difference in number of working hours, takt time to the frequency of a part or component must be produced to meet customers' demand.

The furnace runs two shifts without breaks, but, however, it is not fed with cores in the last hour before the furnace goes to standby-mode since it takes an hour until work-in-progress reach the other end of the furnace. Therefore, 7.5 hours per shift multiplied by two shifts equals the available time per day for the furnace. Brazing run two shifts having half an hour of break per shift and ten minutes for a continuous improvement process meeting. Therefore, for CB and Assembly there is 7.33 hours multiplied by three shifts of available time per day. The available time per day multiplied by the working days of the considered period divided by the volume of the period.

5.3 Plant balancing charts analysis

The creation of a balancing chart, as depicted in Figure 10 by using volumes, available labour and working days resulted in the chart shown in Figure 11 and the time units for net cycle time, additional time for changeover, downtime and scrap is minutes.

BALANCE 2016 JUL		CB	BRAZING	ASSY
NET CYCLE TIME (=VA / TEAMS)		0.442	0.299	0.330
ADDITIONAL TIME SCRAP AND DEFECTS		0.011	0.014	0.020
ADDITIONAL TIME DOWNTIME		0.087	0.024	0.052
ADDITIONAL TIME CHANGEOVER		0.069	0.016	0.036
GROSS CYCLE TIME (= NET CT + ADDITIONAL TIME)		0.608	0.353	0.439
PROCESS SPECIFIC CUSTOMER TAKT TIME		0.553	0.377	0.553
total demand for period	54 919			
working days in period	23			
operating hours per shift		7.33	7.5	7.33
shifts		3	2	3
teams (How many lines are active at the same time?)		4	1	3.66

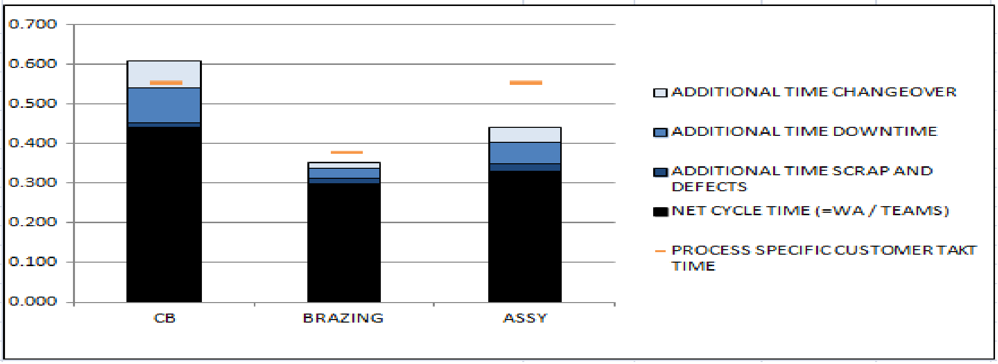


Figure 11: Balancing Chart July 2016

The three bars represent the cycle times of core building, brazing and assembly, net cycle time and the additional times on top of it. The orange horizontal lines represent customer takt time for the fractal. At first sight, one would perceive that core building was not able to meet customer demand, as the gross cycle time is higher than CTT. However, this is not the case because the additional times for downtimes and changeovers do not carry great weight. This is because core building consists of 13 lines but currently only four of them are manned at a time. Therefore, the operators work on other lines during a changeover and during longer breakdowns. Consequently, the 0.069 minutes of additional time for changeovers does not have to be taken into consideration. The cycle time of core building depends on how many lines are running at the same time, which is dependent on the labour. Therefore, the quotient of the cycle time and the CTT does not result in the loading of the 13 core building lines. Core building is able to meet customer demand and it is capable of producing at a rate of one core every 0,539 minutes not taking into account the additional times for changeovers. Then the furnace is consuming parts at a rate of one core every 0,377 minutes but only running two shifts. In order to compare the two cycle times, the cycle time of core building is multiplied with two thirds:

- Cycle times for CB: 0,539min *(2/3) = 0,359 min
- Cycle times for furnace: 0,377 min

The results demonstrate that core building has the capacity to feed the furnace with enough cores. The balancing chart in Figure 12 shows that the brazing furnace is not fully loaded as there is a gap in-between CTT and the gross cycle time of brazing. After brazing, the products go to assembly. In contrast to core building, the downtime for changeovers is relevant. Indeed, changeover times are generally shorter as in most cases just the leak testing jigs and crimp tools have to be exchanged but the operators help doing the changeover so that they cannot work on another line in the meantime. Even so the gross cycle time of assembly is lower than the time of core build. As a consequence, assembly must be able to process all the parts that are produced at core building and heated afterwards.

The analysis of the balancing chart revealed that there should not be a lot of inventory awaiting brazing. As the balancing chart only displays cycle time for core building and assembly as a whole it could be that there are parts that have a very short cycle time for core building but need more efforts to be assembled so that they queue in front of assembly. Or else there might be multiple core building lines feeding one assembly line which is then not able to cope. But only the evaporator assembly line can be fed by two HV core building lines when the evaporator line produces one special condenser and evaporators yet these products may not account for the bulk of the cores awaiting assembly. In addition, the comparison of the cycle times of assembly and core building yields that the CT of CB is always higher or equal to the CT for assembly of the specific product.

Therefore, it was noted that the high inventory level could be possibly as a result of the assembly lines that do not produce at the rate of net cycle time even if there is no downtime or defect. This possibility was investigated by performing a value stream analysis to unveil the reasons for the large core queues behind the furnace. The VSM revealed that the radiator assembly line had a shortage of the manning only two operators working instead of three operators, which would be the ideal situation. The CAC and Rad/LTR lines had neither breakdown and nor manning shortage, but the leak testers were not fully utilised to their full capacity. The leak testing process was the bottleneck of the lines, instead of immediately unloading and loading, the leak testers, when they finished one product they performing other operations instead of loading the CAC and Rad/LTR lines. For instance, it was noted that at the CAC OE line the operator was packing the products in the cardboard box for shipping and at the same time the operator of the crimping station stopped working as the FIFO lane in -between crimping and leak testing was full. After more than five minutes the crimping operator would decide to unload the leak tester that was waiting to be unloaded for that duration, thereby making it impossible to reach the SAP cycle time.

Since it was noticed that the operators were not full-time utilising the capacity of the bottleneck workstations, a workshop for the operators was proposed and conducted. A revised future state balancing chart was then generated and Figure 12 shows the revised balancing chart for the period of time from August to December 2016. Compared to the chart from July the gross cycle time of core building has slightly decreased whereas the CT of the brazing furnace has increased due to increase in the volumes of a parts that has a relatively short CT at assembly but needs a slow brazing programme e.g. LTRs. The CT for assembly also increased. It was apparent that core building must be able to provide cores for brazing in time as the additional time for changeovers actually is not as high as indicated in the chart. Moreover, the assembly process should be able to process what comes out of the furnace as core building still has a higher cycle time than assembly.

BALANCING CHART 2016 AUG - DEC		CB	BRAZING	ASSY
NET CYCLE TIME (=WA / TEAMS)		0.391	0.310	0.335
ADDITIONAL TIME : SCRAP AND DEFECTS		0.011	0.015	0.021
ADDITIONAL TIME : DOWNTIME		0.104	0.024	0.058
ADDITIONAL TIME CHANGEOVER		0.096	0.016	0.042
GROSS CYCLE TIME (= NET CT + ADDITIONAL TIME)		0.602	0.365	0.456
PROCESS SPECIFIC CUSTOMER TAKT TIME		0.549	0.375	0.549
Total Demand For Period	235 330			
Working Days In Period	98			
Operating Hours per Shift		7.33	7.5	7.33
Shifts		3	2	3
Teams (How many lines are active at the same time?)		4	1	3.66

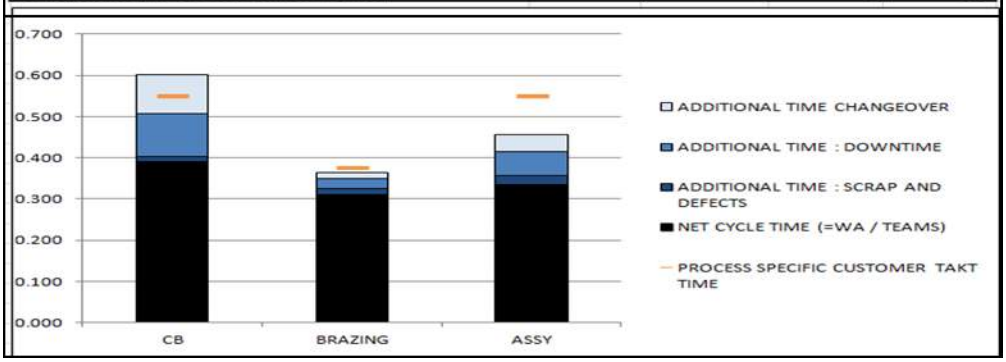


Figure 12: Balancing Chart August - December 2016 with time values in hours

It was noted that capacity constraints can result in line imbalance and hence it is crucial to ensure that all workstations have enough capacity so as to abate excessive work-in-progress inventory. There are changes in terms of cycle times and customer takt times compared to the balancing chart from July and the balancing chart may need to change in the future, as a function of the product demand.

6. CONCLUSION

The allocation of work elements to work centres is crucial for responding to changes to product mix so as to introduce agility to assembly lines. Balancing charts can be used to improve a process through inventory reduction and other forms of waste such as idle time. Recommendations were proposed to ensure that workstations have enough capacity so as to abate excessive work-in-progress inventory and assure balanced lines so as to reduce the waiting time of the brazing furnace as well as inventory between the furnace and the assembly lines. In addition, in case of an imbalance, corrective measures should be implemented as soon as possible to avoid production losses.

REFERENCES

[1] **Becker, C. and Scholl, A.** 2006. A survey on problems and methods in generalized assembly line balancing. European journal of operational research, 168(3), pp.694-715.

[2] **Lam, N. T., Toi, L. M., Tuyen, V. T. T. and Hien, D. N.** 2016. Lean Line Balancing for an Electronics Assembly Line. Procedia CIRP, 40 (Supplement C): 437-442.

[3] **Nash, A. and Poling, S. R.** 2008. Mapping the total Value Stream: A Comprehensive Guide for Production and Transactional Processes.

[4] **Sundar, R., Balaji, A. N. and Kumar, R. M. S. 2014**. A Review on Lean Manufacturing Implementation Techniques. Procedia Engineering, 97: 1875-1885.

[5] **Thomopoulos, N. T.** 2014. Assembly Line Planning and Control, Switzerland

[6] **Sharma, N., Matharou, G. S., Kaur, S. and Gupta, P**. 2013. Lean Manufacturing Tool and Techniques in Process Industry. International Journal of Scientific Research and Reviews 2(1): 54-63 Available: www.ijsrr.org (Accessed 17/08/2016).

[7] **Groover, M.P.** 2016. Automation, production systems, and computer-integrated manufacturing. Pearson Education India.

[8] **Reginato, G., Anzanello, M.J., Kahmann, A. and Schmidt, L**. 2016. Mixed assembly line balancing method in scenarios with different mix of products. Gestão & Produção, 23(2), pp.294-307.

[9] **Monden, Y.** 1994. Toyota production system: an integrated approach to just-in- time. London: Chapman & Hall.

[10] **Chen, L. and Xu, J.** 2015. Optimization of soft beverage inventory managementin practice for SMEs: A case study of JN Ltd. In China

[11] **Bhasin, S. 2015**. Lean Process Requirements. In: Lean Management Beyond Manufacturing. Springer, 51-102

[12] **Thangarajoo, Y. and Smith, A.** 2015. Lean thinking: An overview. Industrial Engineering & Management, 4 (2)

[13] **Santos, J., Wysk, R. A. and Torres, J. M.** 2014. Improving production with lean thinking. John Wiley & Sons.

[14] **Tesfaye, A.** 2014. MAnufacturing Wastes Measurement For Productivity Improvement In Ethiopia Plastic Industry.

[15] **Hirano, H.** 2016. JIT Implementation Manual--The Complete Guide to Just-In-Time Manufacturing: Volume 2--Waste and the 5S's. CRC Press.

[16] **Karodia, A. M., Cowden, R. and Kum, F. D.** 2014. The impact of training and development on employee performance: A case study of ESCON Consulting. Singaporean Journal of Business, Economics and Management Studies, 3 (3): 72-105.

[17] **Morgan, J. and Brenig-Jones, M.** 2015. Lean Six Sigma for Dummies. John Wiley & Sons.

CONTINUOUS PROCESS IMPROVEMENT ON HEAT EXCHANGER PLANT THROUGH VALUE STREAM MAPPING

K.P. Lusiba[1*] & M. Dewa[2]

[1]Department of Industrial Engineering
Durban University of Technology, South Africa
kydiva.lusiba37@gmail.com

[2]Department of Industrial Engineering
Durban University of Technology, South Africa
mendond@dut.ac.za

ABSTRACT

Lean manufacturing is an optimum approach for the reduction and elimination of waste within an organisation. The case study company is based in South Africa and produces heat exchangers through main processes which include pre-assembly, core building, brazing and final assembly. A plant walkthrough revealed an inefficient portrayed by inventory stacked awaiting final assembly and frequent delays at the brazing furnace as it waited for material from core building. This paper proposes a framework for continuous process improvement through the deployment of value stream mapping. Process analysis was conducted using value stream analysis and it was revealed that operators were not fully utilizing the capacity of the bottleneck workstations. Two instead of one planning points and inefficiency at assembly were identified as root causes of the high work in process level. Furthermore, work in process build-up was caused by the furnace that was run two shifts while the preceding assembly and core building were running three shifts. Recommendations were made for continuous process improvement and a roadmap for reduction of waste was proposed.

*Corresponding author

1. INTRODUCTION

Globalization has brought many advantages to globally operating companies such as a greater choice between different suppliers. The buying company can obtain the optimum product for an optimum price to exploit this opportunity, while the supplier faces intensified competition from other suppliers. As a result, it is crucial for suppliers to be competitive and this can be accomplished through the reduction and elimination of waste within an organisation. A value stream map is one of the best tools used to map the process and reveal the hidden waste in the system. The problem at the case-in-point plant was portrayed by inventory stacked awaiting final assembly and frequent delays at the brazing furnace as it waited for material from core building. This paper focuses on the implementation of value stream mapping for assessing the value and non-value adding activities in the manufacture of car heat exchangers components and proposes continuous process improvement opportunities for the automobile component manufacturer.

2. LITERATURE REVIEW

The seven wastes defined by Japanese Toyota manager Taichi Ohno include waste of inventory (stock on hand); waste of processing; waste of movement; waste of overproduction; waste of time on hand (waiting); waste in transportation; and waste in making parts that are defective [1]. Continuous improvement practice is one of the Lean manufacturing basics that focus on elimination of waste or elimination of non-value adding steps in an organisation. Value-stream mapping is a process of mapping entire materials and information needed for the manufacture of a specific product and how the information flows throughout the production system [2]. Value stream mapping is basically representing the value stream data into a map or diagram, in which the current or future state of the production system is represented. As the term indicates, a current state value stream map (VSM) represents the flow of information and materials in the current process while future state VSM shows the future state, where there is a reduction or elimination of waste from the system.

Value stream mapping is one of the visual tools that is utilized in lean manufacturing for identifying and analyzing all the production activities, from planning to shipment of the products. A VSM map makes it easy to obtain improvement opportunities with high impact on the entire production process [3]. The VSM tool examines the current and future process map layout of the value chain, and this makes it feasible to keep records of the current state and the actual process that will be enhanced. The VSM is a chart that consists of symbols and icons that explains the two different types of flows. The first flow is the information flow from logistics planning, that contains the customer call offs or orders that precede the production process. The second flow is the material flow in production, which is taking into consideration all the production steps that are required to produce good parts, until the part is sent to the customer. Performance measurement is allocated to all manufacturing processes that reveal and display the status of the existing processes and generally including, setup time, cycle time and baseline shift, number of operators in a shift, scrap rate, machine availability, efficiency, and machine downtime. As soon as the performance measurement methodologies are in place and drawn according to the value stream map, it becomes easier to pinpoint opportunities for improvement and prioritising according to the impact they have on cost reduction, increasing flexibility and enhancing quality and productivity. Lastly, a future state is drawn in order to aid with visualising the future process when the opportunities of improvement are implemented [4].

A value stream is a set of defined actions needed to get a specific product through the three critical tasks of management in any organization, and these include management of information, physical transformation and problem solving [5]. VSM is a visual technique used to map the flow of information and material required to manage the activities carried out by manufacturing organizations, suppliers and distributors to send finished goods to customers. VSM aids the identification of value adding and non-value adding activities in a value stream [4]. A succeeding step in VSM is developing a future state map supported by the improvements generated from the current state. The information availability in the VSM assist and confirms the decision of implementing lean techniques and can in addition keep the organizations motivated during the actual implementation phase with the aim of getting the desired outcome.

VSM is referred to as one of the important tools of lean approach and is utilized to identify value-adding and non-value adding (wasteful) activities in the system, but when the VSM tool is not applied correctly, it can complicate the process of waste identification, leading to misunderstandings and assessment errors, and weakening the future continuous improvements initiatives [4]. Process mapping highlights the different steps in the processes and underline where the value is added. Waste can be classified in line with the generally accepted types of waste. Value is the step where the expected outcomes for the customer are defined [6]. Questions that can be asked to define value include; Is this step changing the form or product character? Does this step meet an exact customer specification? If the answer is "yes" then more economic ways to perform this step are explored. Methods time measurement (MTM) is one of the methods that can be used to identify value in the process [7].

3. METHODOLOGY

The case-in-point manufacturer uses lean tools but, however, there are loop holes in the system that leads to poor product quality, high regulatory non-compliance risk, poor process flow and high lead times. A value stream mapping was therefore required to display the material and information flows and assessing the value and non-value adding activities in the manufacture of automotive heat exchanger components. Material supply and the finished goods warehouse were also taken into consideration in order to close the cycle. As representative products, a product of each of the high volume core building lines was chosen and the key figures for the map at core building and assembly include number of operators; part number of representative product; SAP cycle time; observed cycle time during line walk; throughput time; setup time; and number of shifts. Value analysis was then conducted for the core building process, after which an assessment of several continuous improvement initiatives was done to identify which initiatives would reduce inventory costs and waste from the plant. Three proposals were developed as roadmap for reduction of waste to improve the situation in the plant.

4. RESULTS

4.1 Current State Value Stream Map

A waste walk was conducted at the case study company in order to accurately map the current state. Figure 1 shows a value stream map that was developed for the Engine Cooling (EC) plant. Manufactured products are delivered to the customers once a week, excluding the radiators and the low temperature radiators (LTR), because these products are shipped on daily basis. Customer orders are sent to planning department, from which a plan for core building is issued. This is indicated in the VSM with an arrow from planning to core building and second plan is issued for assembly line in which the sequence can vary from the core building sequence. A second arrow on the VSM goes from planning to assembly line. Cores kept behind the furnace/ brazing line are counted daily and this information is used to issue a third plan for brazing. These three production sections that do not work hand in hand when it comes to planning, but all parts start at core building (initial assembly), brazing (pre-heating process to combine the parts together) and last step assembly (final assembly includes crimping & leak testing).

The observation at the assembly line revealed that the operators could not achieve the SAP cycle time at three of the four regarded lines, though there was no occurrence of breakdowns or defective/scrapped parts. Only the operators for the evaporator line were able to achieve the cycle time as shown in Figure 2.

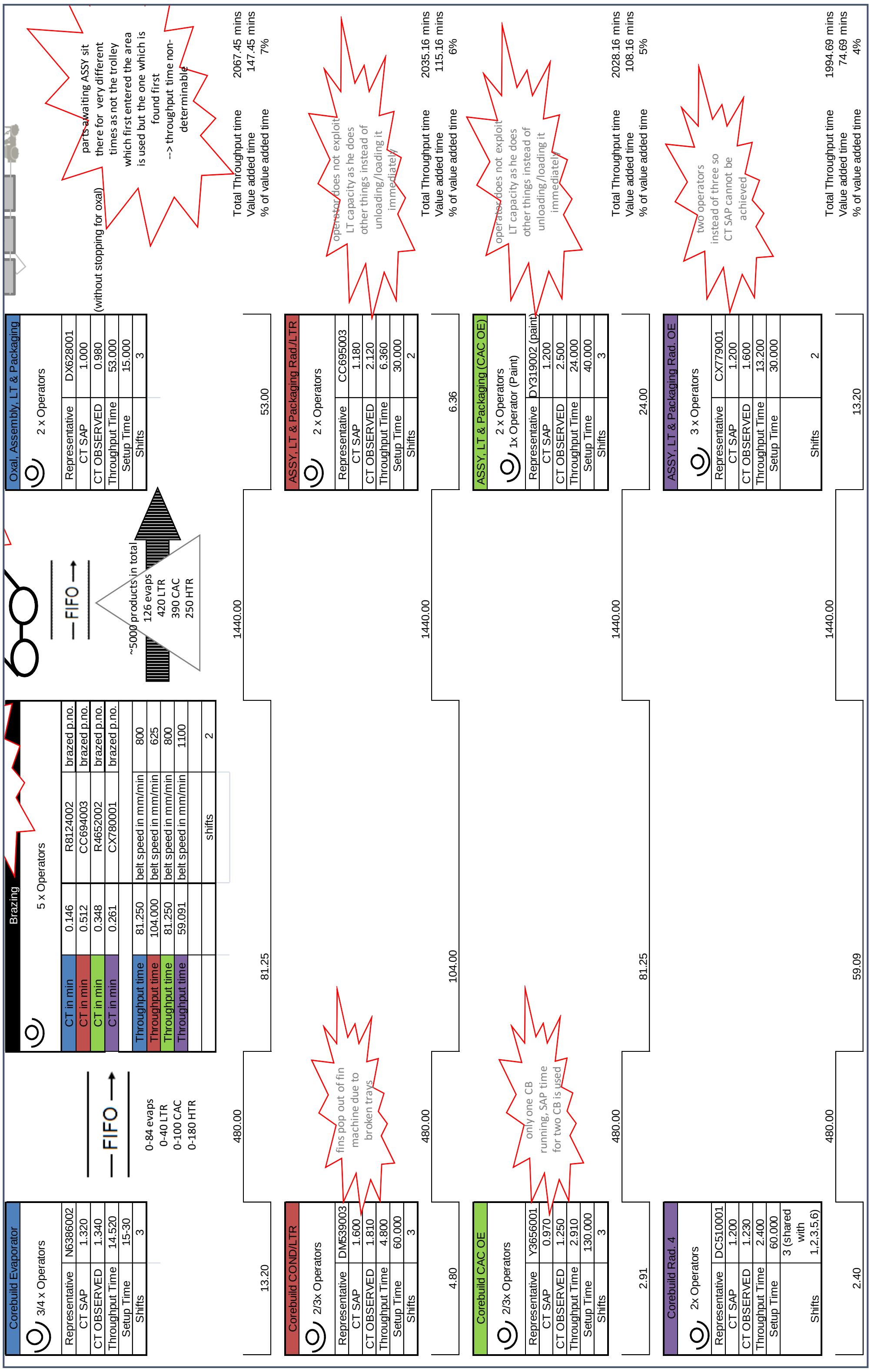


Figure 1: Snapshot of Value stream map for the EC plant

The brazed parts wait for assembly in the FIFO lanes at the assembly line and in the post brazing FIFO lanes next to the furnace. The post brazing FIFO lanes are not used as FIFO lanes and new brazed parts are stored randomly where the operators find some space. Therefore, in the value stream map the symbol for push material transfer and the triangle for inventory is indicated next to the FIFO lane and a Kaizen flash underlines this issue. The numbers of the parts that are selected for the VSM are counted at the point in time when the line walk was done. In addition, it is indicated that there are about 5000 parts in total awaiting assembly. The furnace/brazing throughput and cycle times are taken from the SAP system. There is no Kaizen flash as there was no problem observed. The inventory in the FIFO lanes in front of the furnace inlet was inserted as a range because the products were brazed during the observation. The maximum numbers of products were low because at that point in time when the observation for the specific product was done the furnace was brazing those parts. This does not necessarily have to be the case at every point in time as there are three different brazing families and sometimes products had to wait until their brazing programme is used again.

Oxal, Assembly, Leak Testing, & Packaging	
2 x Operators	
CT Observed	0.980
Throughput Time	53000
Setup Time	15000
Shift	3

Figure 2: Evaporator line cycle times

The VSM shows that the radiator assembly line had a manning shortage since there were two operators working instead of three operators, which would be the ideal situation. The charge air cooler (CAC) and low temperature radiator (LTR) lines had no breakdowns and or manning shortage but the leak testers were not fully utilised to their full capacity, and the leak testing process is the bottleneck of the lines. Instead of immediately unloading and loading, the leak testers, were performing other activities when they finished one product. For instance, the operator at the CAC line would pack the products in the cardboard box for shipping and at the same time the operator of the crimping station would stop working as the First-In-First Out (FIFO) lane in-between crimping and leak testing is full. After more than five minutes the crimping operator would then unload the leak tester that was waiting to be unloaded. Poor utilisation of the bottleneck station results in inability to achieve the SAP cycle time and tose deficiencies were marked with a Kaizen flash.

4.2 Value Analysis

VSM highlighted Kaizen flashes for which detailed value analysis would be undertaken. Value analysis was then conducted by assessing the activities in terms of value-adding, non-value adding essentials and non-value adding waste. Value-adding is a change of form or character of the product and things that the customer is willing to pay for. Non-Value Adding Essentials is waste that cannot be avoided that lean seeks to eliminate. Non-Value Adding Waste is waste that does not add value and can be avoided [8].

It was noted that in some instances, the core building production line could not keep pace with SAP cycle time but the deviations were not as high as at assembly line. The CAC core building line has one core builder instead of two possible core build stations and this led to higher cycle time. TYhe observed cycle time at the condenser/LTR line was found to be higher than the SAP time. One reason is that the trays that were filled with airways from the fin machine were broken fins popped out of the trays and the operator had to collect and insert them manually. These issues had been highlighed as Kaizen flashes in the VSM. A key challenge was determining the duration for which brazed cores wait for assembly since the operators took readily accessible coreswithout following the FIFO rule.

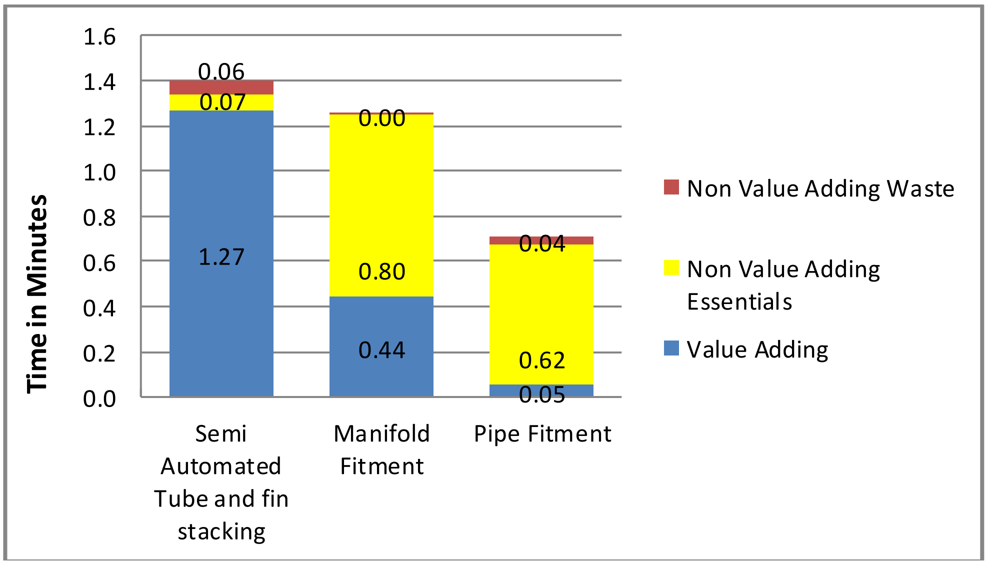


Figure 3: Core Building Value Analysis

Figure 3 displays value analysis per operation on the core build line of evaporators to classify value adding, non-value adding waste and non-value adding essentials.

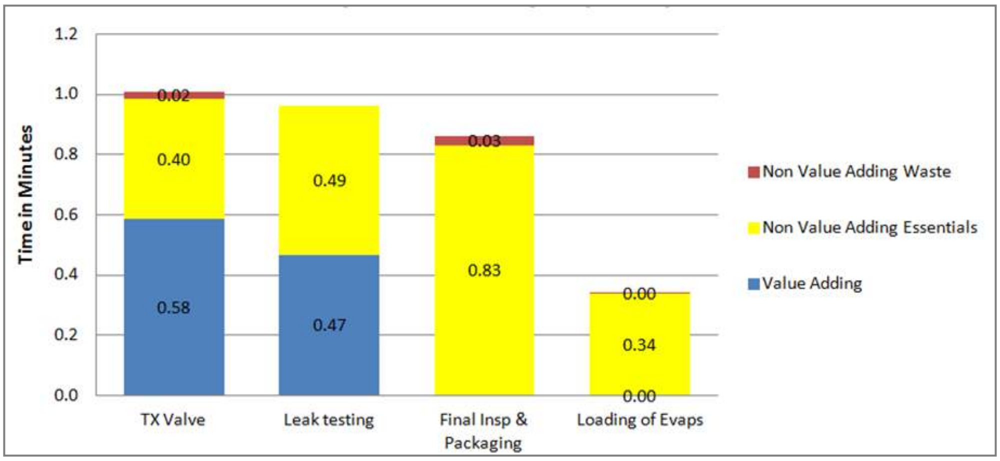


Figure 4: Assembly Value Analysis

Figure 4 displays value analysis per operation on the assembly line of evaporators. It also presents the different elements on the Evap assembly lines and classifies the activities into value adding, non-value adding waste and non-value adding essentials.

4.3 Furnace Downtime

The furnace is a costly station and has high-energy consumption, and thus it is crucial to have a detailed record of its downtime. The furnace is never switched off entirely; nevertheless the furnace consumes more energy when running than in standby mode. Table 1 shows the downtime in minutes for a six month period from January to June. The downtime due to no parts decreased from January to May but then increases slightly in May and June. It is revealed that high maintenance cost that takes up about 13% of the total operating cost, and this could be reduced by good planning.

Table 1: Extract from Furnace Downtime Report July 2016

Furnace Downtime (minutes) Due To Number of Cores					
Jan	Feb	Mar	Apr	May	Jun
2550	1170	260	0	120	190

4.4 Work in Progress Awaiting Assembly

The brazed parts were found in the FIFO lanes next to the furnace, at the lines and at other areas that are supposed to be used for other purposes. Figure 5 shows the cores in the FIFO lanes in the background and even more cores in the area next to them. In addition, the area marked in red is often used to store brazed parts.

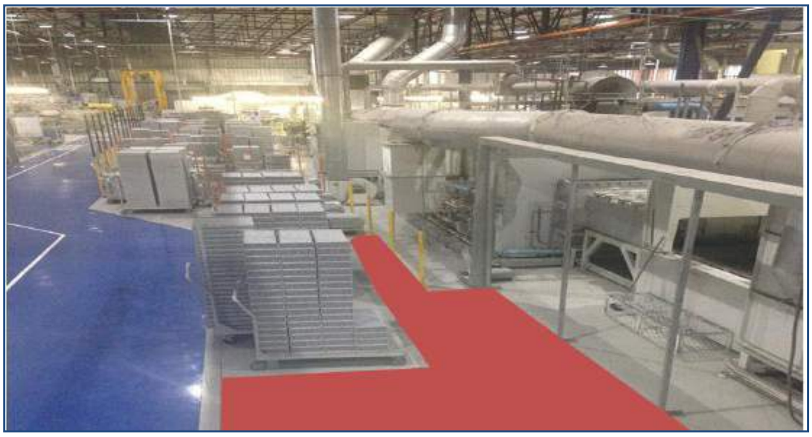


Figure 5: Inventory next to the furnace

In order to find out how many heated parts are awaiting to assembled, function in SAP was used to identify all the part numbers, quantities and value that are scanned after brazing. The evaluation of the inventory using SAP and manual counting done for CW 29 and 30 revealed that there were 6 348 heated cores/parts on average. This includes parts that are already in the processing at the assembly lines and assembled parts that are packed into a box but not scanned yet. The data was filtered by MRP controllers in order to see which parts account for the highest proportion, and as the pie chart in Figure 6 shows, the number of CACs is almost half of the total quantity, followed by radiators that come from the core building lines. The percentages of evaporators and condensers, LTRs and radiators from the other core building lines are quite small compared to CACs and OE radiators. Therefore, the majority of work in progress was being processed on the CAC lines and the second most parts are parts that are supposed to go to the Rad. AM or the Rad. OE line.

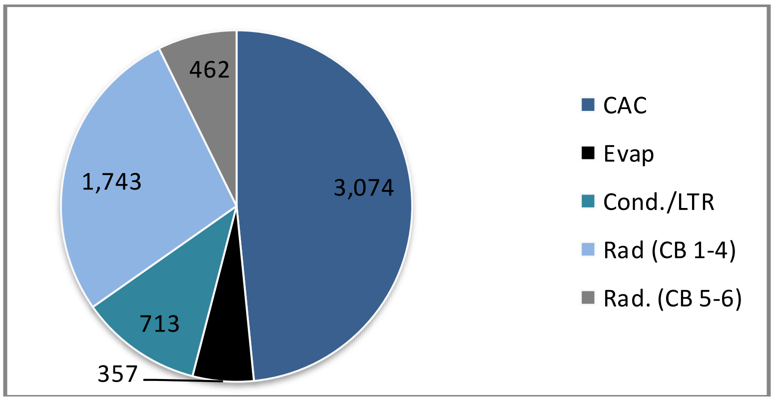


Figure 6: Summary of Inventory Level for brazed parts awaiting assembly

Overall, the average of the manual core count and the result of the SAP analysis is 5500 parts of work in progress between brazing and assembly. The remaining question was to ascertain the number of parts should ideally be the WIP inventory. About 5500 parts is more than two daily production quantities, volumes for July divided by 23 working days in July.

5. DISCUSSION OF CONTINUOUS IMPROVEMENT INITIATIVES

Balancing charts were drawn and they revealed that there is no heavy imbalance in the plant although there was notable inventory behind the furnace. As the balancing chart is based on SAP times and includes scrap, defects, downtime and changeovers it must be the case that assembly section does not achieve their targets even if there is no documented downtime, defect or scrap. This could possibly be due to several reasons such as that the targets were too high which means the cycle times in SAP times are too low. However, this was very unlikely as the SAP times are the result of a MTM study done according to a video of the actual process. Moreover, the employees responsible for those time studies are trained. Alternatively, the lines run with fewer operators than planned, which can happen because of absenteeism or when the plant has to run on short time due to low volumes. Either the operators did not run the lines properly thereby failing to achieve their targets. This is exactly what the VSM revealed, the operators at assembly did not fully utilise the bottleneck workstations and therefore could not produce according to the SAP times. In addition, some lines were less manned yet the balancing chart assumes that the lines run with maximum number of heads. In addition, the VSM revealed that there are two planning points in the EC plant instead of one. Those two plans sometimes differ, and as a consequence, it is impossible to use the FIFO lanes after the furnace as FIFO lanes because often products that got brazed after another product are assembled first.

5.1 Improvement Opportunities for Training and Removal of Second Planning Point

A major problem that the value stream analysis revealed was that the assembly lines often could not achieve targets even if there are no defects or downtime. This is often due to operators not full-time utilizing the capacity of the bottleneck workstations. In order to improve situation, a workshop for the operators was proposed. The workshop would raise operator awareness of the existence of bottlenecks and the importance of utilising them all the time. The workshop should emphasise that a chain is only as strong as a weakest link and the criticality of teamwork. In case one operator at the bottleneck workstation has difficulties to cope, the other operators should help. After the theoretical part, the lessons learnt should be practised on the shop floor. It was antciapted that the assembly lines would be able to process all the brazed material after operator training , and thus it would be possible to introduce dedicated FIFO lanes after brazing. Moreover, when the FIFO lanes are used properly, the second planning point will not be necessary anymore.

The training of operators and the removal of the second planning point allows the reduction of work in progress so that a saving of costs of carrying inventory is possible. Besides, lead times will be shorter and the FIFO lanes at the furnace will be clearer with the result that it will not be necessary to search for the right trolley for assembly anymore. In addition, training can possibly increase the motivation of the operators. However, there are cost inplications since the operators are not working during the training sessions. An estimate of costs from the Training and Development department was R8 410 for the preparation and three rollouts of the training as there are three shifts of operators. Still, there would be a saving of R99 000 per year, presuming that a workshop is performed once a year.

5.2 Continuous Improvement from Economical Evaluation of the VSM

Compared to the current situation, the workshop and the elimination of the second planning point do help to reduce the level of inventory and therefore also the cost of carrying inventory. The costs of carrying inventory comprise of cost for the physical space occupied, the opportunity cost because of the capital lockup, handling costs and the cost for deterioration and obsolescence [9]. In order to compare the annual cost of carrying inventory in the EC plant with and without the implementation of the proposal the focus lies on the cost for the space and for the tied up capital. This is because there is one worker per shift responsible for the internal material handling. Deterioration was also not taken into account. This is because products are made to order and even if the products spend more time awaiting brazing, there is no chance that they are not taken by the customer.

Table 2: Required Space for Post Braze FIFO lanes

Required Space For Post Braze Fifo Lanes								
Assembly line	Min. qty on trolley	Cycle time CB	Produced qty in one CB shift	required FIFO lane bays	provided FIFO bays at line	provided post braze bays	Cycle time brazing	Time to Braze Qty (hours)
CAC OE	30	1.50	293.20	10	9	6	0.28	1.37
CAC AM	30	4.00	109.95	4	4	6	0.18	0.33
Rad.	42	2.56	171.93	5	8	6	0.42	1.21
LTR	45	1.60	274.88	7	2	6*	0.51	2.35
Rad. AM	87	1.98	222.57	3	6	6	0.47	1.73
Evap	42	1.71	257.19	7	7	-	0.15	0.62
Cond.	45	1.62	271.48	6	6	-	0.26	1.16
Rad. OE	50	1.20	366.50	8	6	6	0.26	1.59

Table 2 shows the required space for post-braze FIFO lanes, and currently, the brazed parts occupy about 183 square meters in the plant. This space includes the area of the post brazing FIFO lanes next to the furnace and the FIFO lanes at the assembly lines. Furthermore, the area between the post-braze FIFO lanes and the steps of the control panel of the furnace is included as CACs and radiators are stored. In the course of the manual core count, brazed cores were seen in different places, but most of the time the previously mentioned area was covered with cores. This yields inventory costs of R96 782 per year including rent, insurance and overheads. Assuming that there are 5 500 brazed parts on average, which is the average quantity of the manual count and the observation using SAP, and an average value of R246 per part, this results in R94 710 of opportunity costs. This is presuming a rate of imputed interest of seven percent per year, which is the rate used for calculations.

After the workshop and the removal of the second planning point the inventory level will be lower. At an average there would be one sixth of the daily production quantity of core building. This is because the daily production quantity is built at core building during three shifts. The furnace is able to braze this quantity during two shifts whereas assembly processes this quantity during three shifts. Using the gross cycle time minus the additional time for changeovers of the balancing chart for the period from august to December, this results in an average daily production quantity of 2 587 parts. One sixth of this quantity times the average value of a brazed core of R246 results in the average value of R106 069. This results in opportunity costs for the tied up capital of R7 422 per year.

Concerning the space, it was possible to use the provided post brazing FIFO lanes and the assembly FIFO lanes so that the extra space that is currently occupied by cores is not required anymore. This is because they were designed to cater for the brazed parts. This proposal can save R107 650 of costs for carrying inventory per year and Table 3 shows the calculation of inventory costs. Consequently, the required space will reduce to 129 meters square, which results in annual costs for the space of R68 006. Therefore, proposal could save R107 650 of costs for carrying inventory per year.

Table 3: Inventory Carrying Costs Before and After Implementation of Proposal 1

	CURRENT	AFTER WORKSHOP/ ONE PLANNING POINT
average qty awaiting assembly	5500	431
average value per unit (ZAR)	246	246
imputed interest	7%	7%
opportunity cost of capital per year (ZAR)	94 710.00	7 421.82
required space (sqm)	183.30	128.80
cost per squaremeter per year (ZAR)	528.00	528.00
cost for space (ZAR)	96 782.40	68 006.40
cost for carrying inventory post braze	191 492.40	75 428.22
average qty awaiting brazing	431	431
average value per unit (ZAR)	201.00	201.00
imputed interest	7%	7%
opportunity cost of capital per year (ZAR)	6 064.17	6 064.17
required space (sqm)	61.60	61.60
cost per squaremeter per year (ZAR)	528.00	528.00
cost for space (ZAR)	32 524.80	32 524.80
cost for carrying inventory pre braze	38 588.97	38 588.97
annnual cost for carrying inventory pre and post braze	**191 492.40**	**75 428.22**

5.3 Continuous improvement opportunities on shift models

As the furnace is running two shifts and assembly and core building is running three shifts there is always WIP building up in front of the furnace when it is not running. On the other side, the inventory level is increasing behind the furnace when it is running and decreasing when it is not running [10]. The average number of parts awaiting brazing must be one sixth of the daily production quantity sincecore building is producing the daily production quantity during three shifts and the furnace brazes that quantity within two shifts. Moreover, there is one sixth of the daily production waiting brazing at the other end of the furnace on average, since assembly consumes the number of parts in three shifts that the furnace brazes during two shifts. In order to reduce this level of inventory it would be an option to run core building and assembly on a two-shift model as well. If all three sections were running in two shifts, the inventory level could be reduced even more. In addition, lead times would decrease, provided that all cores are brazed and assembled within the same day. Besides, running the EC plant only on two-day shifts would save the night shift allowance for the operators.

5.4 Roadmap for Reduction of Waste

Three proposals to improve the situation in the EC plant are developed. The first embraced a workshop for the operators at assembly, which should enable them to meet their targets for the hourly production quantities. In addition, it would be possible to use the post brazing FIFO lanes as dedicated FIFO lanes again and the second planning point will not be necessary. The introduction of this proposal was highly recommended as it generates a saving due to the reduction of inventory and does not entail disadvantages, except for the fact that the high inventory level currently works as a huge safety buffer. After the implementation of this proposal there remains an imbalance due to the different shift models of the three production sections, two further proposals were developed. The other proposal was on adjusting the shift models to a two-shift model for the entire plant. In order to achieve this, high investments would be required and it would not pay off, as the potential saving of inventory costs is small. This would only be a short-term improvement, as the furnace would have to run three shifts again in the following year when demand is anticipated to increase.

The other option was to create a slower brazing profile, so that the furnace runs three shifts, as CB and assembly. Nevertheless, the economical evaluation revealed that this option would not generate a saving and would fall in line with the organisational strategy of increasing volumes in the near future. Nevertheless, even if the adjustment of the shift models did not make good economic sense, the implementation of the first proposal allowed a good saving. Table 2 summarises the arguments that speak for and against the proposals and costs estimates with anticipated savings.

Table 4: Summary of Proposals

Summary of Proposals				
	Status Quo	Training + One Planning point	Two Shifts	Three Shifts
basis for comparison of savings and costs				
cost of carrying inventory for post braze WIP	ZAR 191 492 per year (space and opportunity costs)	ZAR 75 428 per year (space and opportunity costs)	probably less than ZAR 75 428 per year	probably less than ZAR 75 428 per year
cost of carrying inventory for pre braze WIP	ZAR 38 589 per year (space and opportunity costs)	ZAR 38 589 per year (space and opportunity costs)	probably less than ZAR 38 589 per year	probably less than ZAR 38 589 per year
other costs and required investments	-	• Workshop (once a year) ZAR 8 410	• tools and LT jigs: more than ZAR 1.350 million •space for additional machines	• labour ZAR 610 580 per year • energy ZAR 826 000 per year • brazing jigs, • brazing trials, research
savings	-	ZAR 107 654 per year	• no saving because of high investment for tools etc.	• no saving because of costs for energy and labour and the investment in brazing jigs
(+)	-	• reduction of inventory • no more searching for brazed material • reduction of lead times	• further reduction of inventory •further reduction of lead times for most parts	• removal of constraint of brazing families • further reduction of inventory
(-)	• chaotic, products are hard to find in piles of material	• imbalance of the 3-2-3 shift model	• **high investment in tools/ LT jigs or entire lines** • expensive HV CB and ASSY lines are only utilized for two shifts • **not future oriented as furnace must probably run three shifts next year**	• **feasibility** • high investment • higher costs for energy and labour • **furnace runs out of capacity when volumes increase as expected**

6. CONCLUSION

Utilising lean manufacturing as a culture that creates and sustain long-term commitment is a challenge to most organisations. Value stream mapping is an excellent starting point for impementing continuous imrpovement initiatives in organisations. Removing the second planning point in the EC plant and the training of the operators allows the reduction of inventory, as it aims to reduce all non-value adding activities or waste. The balancing chart was converted into a template so that new parts and volumes as well as changes in shift patterns or in terms of the scrap, defect or downtime rates could be readily monitored.

7. REFERENCES

[1] **Ohno, T.** 1988. *Toyota production system: beyond large-scale production*. crc Press.

[2] **Chen, J. C., Li, Y. & Shady, B. D.** 2010. From value stream mapping toward a lean/sigma continuous improvement process: an industrial case study. *International Journal of Production Research*, 48 (4): 1069-1086.

[3] **Santamaria, E. M.** 2013. Implementation selected tools of Lean Manufacturing. 19-21 (Accessed 17/08/2016)

[4] **Lasnier, G.** 2007. Le lean-manufacturing (systèm de production à haute performance) dans les industries travaillant en juste-à-temps avec flux régulés par takt-time (rythme de la consommation du client). *La Revue des Sciences de Gestion : Direction et Gestion*, 42 (223): 99.

[5] **Soliman, M. H. A.** 2017. A Comprehensive Review of Manufacturing Wastes: Toyota Production System Lean Principles.

[6] **Chaple, A. P. & Narkhede, B. E.** 2017. Value stream mapping in a discrete manufacturing: A case study. *International Journal of Supply Chain Management*, 6 (1): 55-67.

[7] **Karim, A. & Arif-Uz-Zaman, K. 2013.** A methodology for effective implementation of lean strategies and its performance evaluation in manufacturing organizations. *Business Process Management Journal*, 19 (1): 169-196.

[8] **Gu, J., Zhang, G. & Li, K. W.** 2015. Efficient aircraft spare parts inventory management under demand uncertainty. *Journal of air transport management*, 42: 101-109.

[9] **Soliman, M. H. A.** 2017. A Comprehensive Review of Manufacturing Wastes: Toyota Production System Lean Principles.

[10] **Chen, L. & Xu, J.** 2015. Optimization of soft beverage inventory managementin practice for SMEs: A case study of JN Ltd. In China.

[11] The Centre for Management & Business Development The 8 Wastes of Lean Management Available: www.c4mbd.com (Accessed 16/08/2016).
[12] **Lam, N. T., Toi, L. M., Tuyen, V. T. T. & Hien, D. N**. 2016. Lean Line Balancing for an Electronics Assembly Line. *Procedia CIRP*, 40 (Supplement C): 437-442.

ROOT CAUSE ANALYSIS FOR REDUCTION OF WASTE ON RUBBER CABLES

M. Dewa[1*]

[1]Department of Industrial Engineering
Durban University of Technology, Country
mendond@dut.ac.za

ABSTRACT

In view to meet rising demand of infrastructure needs, rubber electrical cable manufacturing plants should be equipped with latest state of the art modern machinery to offer quality products that are specifically designed to meet electrical and physical specifications and standards. The case-in-point plant experienced a sudden significant increase in excess length scrap and this scenario necessitated an investigation to ascertain the root cause of waste on rubber cables. The methodology that was employed embraced determination of the cable process and material flow and then development of a cause and effect diagram. The cause and effect diagram highlighted that there were numerous causes for the waste on rubber cables, with human error as the greatest source of waste. Additionally, concerning measurement, improper calibration resulted in huge variance in length from machine to machine. It was recommended that machine length counters should be calibrated to a single reference machine and use of two main length counters. An internal printing line was also created to eliminate the rewind line phase. It was also imperative to educate the operators about the correct standard operating procedures and methods to be followed.

[*]Corresponding author

1. INTRODUCTION

The scope of globalization has amplified, posing intense pressure on manufacturing organisations to strive for continuous improvement in their products and processes to achieve world class performance [1]. It is against the backdrop of increasing globalization that rubber electrical cable manufacturing plants should be equipped with latest state of the art modern machinery to offer quality products that are specifically designed to meet electrical and physical specifications and standards. It is a pre-requisite that specifications are designed to assure that utilities receive cables that meet their performance expectations. [2]. The case-in-point plant experienced a sudden significant, questionable and unexplainable increase in excess length scrap and this scenario necessitated an investigation to ascertain the root cause of waste on rubber cables. It was therefore decided that an investigation be done to determine the reason for the increase in excess length scrap. The objective of this paper is to determine the root cause of the excess lengths, identify possible scrap reduction areas and develop a methodology for optimising the process flow and reduce or eliminate scraps in a rubber electrical cable manufacturing context.

2. LITERATURE REVIEW

2.1 Cable manufacturing process

Copper wire is drawn to the proper size in a two-step drawing process, annealed and coated with polyethylene insulation during the first phase of cable manufacturing process before being tested and passed on to the next manufacturing phase. They may also be tinned to avoid corrosion. Figure 1 shows the second manufacturing phase is characterised by twisting the insulated copper wire into wire pairs and combining the twisted wire pairs using a strander to form cables. The laid up cable is then sheathed with a layer of rubber or other insulating material, after which the cable is again tested for electrical resistance. It is vital to measure wire and insulation diameter and testing for electrical properties such as capacitance and resistance before the reels move to the next manufacturing operation [3].

Figure 1: Twisting and stranding steps in cable manufacturing process

It is worth noting that the production of the conductive core is characterised by cold deformation processes, such as stranding and compaction whereby the wires are subjected to tangential forces emanating from the friction between the wires (twisting and pooling forces). The use of the compacting die also results in compressive forces and these forces influence the metallurgical state and overall mechanical and electrical performance of the conductor. The optimisation of lay length, the number of wires in each layer, number of layers and twisting direction parameters may yield substantial reduction of copper use, while concurrently maintaining the desired properties of the cable [4].

2.2 Insulating materials

The insulating materials for electrical cables include thermoplastics and thermosetting materials [5]. A thermoplastic is a material such as PVC and nylon, which during cable manufacturing, when processed by an extruder can be repeatedly made to flow under the application of heat. On the other hand, a thermoset, such as rubbers (neoprene, hypalon) and polyester resin cannot be made to flow under the application of heat. Consequently, production is generally fast when using thermoplastics and is slower when using thermosetting materials.

A rubber material is capable of recovering quickly and forcibly from large deformations, and can be modified to a state in which it is essentially insoluble, yet can swell in boiling solvents such as benzene. When free of dilutants, a rubber in its modified state can retract to less than 1.5 times its original length in less than one minute after being stretched at room temperature to twice its length [6]. The properties of rubber used for insulations include its resistance to abrasion, reaction with acids, alcohols and other hydrocarbons. The properties also include rubber sustenance in water, oil, sun, flame and nuclear radiation [7]. The selection of insulation material influences the pulling tension, tactile feel as well as the degree of stretch before breaking, which is encountered during installation, processing, and flexing [8].

2.3 Defects and reduction approaches

A defect is a component that does not meet the internal or external customer specifications thereby resulting in reworks and rejects that are costly for production processes [9]. Defects are generated from poor production processes as an outcome of human error or machine problem and reworking consumes more time and hence increases the cost of the finished product. Defect assessment is characterised by identifying the defect; defining the root causes and related significances; cause classifications; identifying the reasons of root causes; and instituting corrective actions to reduce the probability that a problem will recur [10]. An organization should also train its operators to autonomously perform routine checks, inspections, cleaning, lubrication and adjustments. The operators could perform tasks that prevent deterioration of production equipment and improve its performance. It is also vital to inculcate a team-working culture, a team that identifies problems and suggest ideas for elimination of the defects, and help to abate barriers that are inherent in the traditional manufacturing [11].

3. METHODOLOGY

The methodology that was adopted for root cause analysis for reduction of waste on rubber cables comprises of 5 steps that include:

Step 1: Determine the cable process flow and the material flow
This woould aid to map out the path that the cable goes through from raw material to a finished product. Once the production path is clear a systematic approach can be created to track the cable through each stage.

Step 2: Create a cause and effect diagram
The cause and effect diagram aids to determine the root cause of the rubber excess lengths and the contributing factors. Using the cause and effect diagram the key factors that can be controlled during the study will be noticed (e.g. ensuring the correct procedure is followed).The idea is to eliminate controllable varying factors so that the actual underlining cause that may be unnoticed can be picked up.

Step 3: Create an action plan and communicate with the necessary people
The action plan should clearly state what the focus of the study is and the procedure to be followed. It is also crucial that production is clear of the intention of the study as they will be directly involved in the project.

Step 4: Track the selected order of the 16mm^2 rubber cable
There needs to be clear and constant contact between production and the study members so that the accuracy of the outcome is high and there are no stages that are missed.

Step 5: Analyse the results of the study and determine the improvement areas
Using the information gathered when the cable was tracked from stage to stage, the results will highlight the issue and show the way forward to eliminate the problem.

4. ANALYSIS AND RESULTS

For every product each stage of production is planned accordingly taking into consideration the necessary allowances (i.e. start-up allowances) so that the final cable is of the correct length and design. According to the specifications for the 16mm^2 rubber cable, to achieve the final cable length of 400m, a factor of 8% needs to be incorporated into the core lengths to accommodate for the lay length. Figure 2 shows the rubber cable process flow and the cable chosen to be tracked was the 16mm^2 3 core + 1 core 16mm^2 black pilot cable.

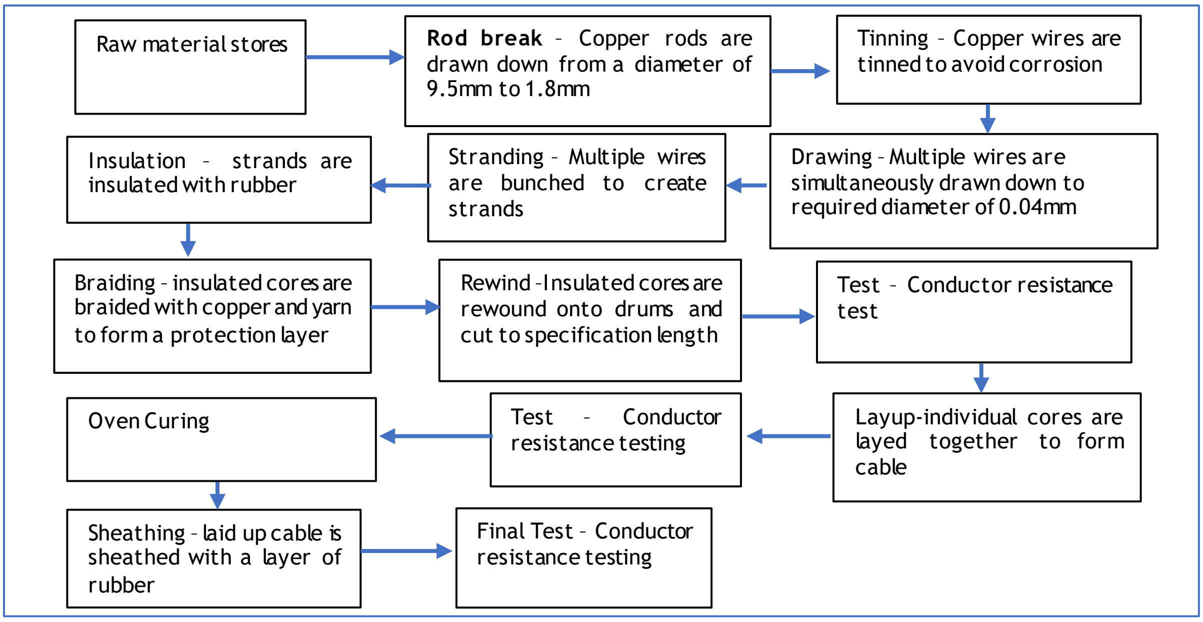


Figure 2: Rubber cable process flow

The recommended problem solving tools include the cause-effect (Ishikawa), why-why analysis, questioning technique, scatter diagrams, and failure mode effect and criticality analysis [12]. The cause-effect diagram was used in this paper for the reduction of waste on rubber cables. Cause-effect diagrams are a strong tool for the identification of different potential causes of a problem and used as a guide to allocate resources for problem solving.

4.1 Problem areas

The main departments involved in the investigation of excess and short lengths on the final cable lengths included Yield Department, Continuous Improvement, Rubber Testing Department, and Rubber shop factory (Production). Table 1 shows the key areas of focus for ensuring that each stage was monitored and tracked to avoid losing information as well as the actions that should be executed to ensure reduced excess waste for the 16mm^2 cable that was described in Figure 2.

Table 1: Summary of key focus areas and actions to be taken

No.	Task	Action
1	MES	>Ensure all the correct orders have been selected, procedures executed correctly and scanning at each phase done.
2	Insulation 9C1	>Make 1 set of 4 splices for all colours >Record lengths
3	Test	>Verify lengths via conductor resistance testing.
4	Rewind 10C1/2	>Check measure lengths
5	Braiders	>Braid cable drum to drum (no length changes)
6	Lay-up 10B7	>Verify lay lengths and diameters >Record length.
7	Test	>Verify lengths via conductor resistance testing.
8	Sheathing 9E1/2	>Record length >Verify with previous process
9	Test	>Bulk test and verify length via conductor resistance

10	6D5 Printing	a) Check if all counters correspond.
		b) Drum required lengths and verify lengths via conductor testing
		c) Record excess
		d) Rewind to verify print legends distant

4.2 Data Collection

This section focuses on the actual data that was collected from each stage of production from stranding to printing. The actual order that was being tracked was 1 600m of final cable length which was to be delivered in 4 drums of 400m each. This implied that 1 728m of each core was required at layup. However due to production and planning purposes the orders were combined accordingly to decrease the amount of set-up changes and this is an unavoidable factor.

Stranding

It was found that various strand lengths are produced and stored in the Kanban area. Some strands were found to be stored in the correct area while some were incorrectly stored. It was also noted that the fork-lift drivers do not necessarily check the drum lengths before transporting them to rubber shop factory and instead take the first available drums to rubber shop factory. As a result, the incorrect drums are used for the incorrect orders, creating unnecessary scrap as excess lengths. Strand lengths are made according to order specifications and for the $16mm^2$ cable being tracked, the required strands for insulation of 20 splices of 432m strand was 8670m + 30m start-up allowance. Only the lengths are recorded on the reel cards and stored at the Kanban area along with other drums for different order specifications. If the incorrect length is taken to insulation then there will be excess lengths and scrap created other than the start-up allowance.

Insulation – 9C1

The core lengths were recorded and as shown in Table 2, there are 4 cores required from insulation therefore the input strand length is divided into splice lengths of 432m each.

Table 2: Recorded input and output data for insulation of cores

Input	Output	Recorded Data	Specification conductor mass	Actual conductor mass
4 Drums of 8670m strand	Red insulated core: 20 x 432m	Red (P3D130) = 8670m (± 15m balance)	144g/m	Variation from 137g/m - 139g/m
		3 kg scrap		
	Blue Insulated core: 20 x 432m	Blue (P3D086) = 8843m (± 122m balance)		
	Yellow Insulated core: 20 x 432m	Yellow (P3D089) = 8843m (± 100m balance)		
	Black Pilot Insulated core: 20 x 432m	Black (P3D075) = 8843m (± 80m balance)		

The recorded data shows that the incorrect input length (8 843m) was used which resulted in unaccounted scrap of (± 122m) which could have been avoided. Since the balance strand length was more than the allocated ± 30m start up allowance it was scrapped as excess length.

Rewinding

Each insulated core from 9C1 is rewound and cut to the length of 1 728m which is 4 splices. The recorded data is shown in Table 3.

Table 3: Difference in input and output length from rewinding

Input length (from 9C1)	Output Length (from 10C2)	Difference
4x432m = 1 728m	1740m	12m

The difference of 12m is scrapped as excess length. This implies that the 9C1 and 10C1 length counters do not match.

Braiding

Drum to drum run occurs as insulated cores are braided with copper and yarn to form a protection layer and there is no change in lengths.

Lay-up

The recorded data for input and output length is shown in Table 4.

Table 4: Input and output length for lay-up

Input length	Output Length	Specification lay length
4x432m = 1 728m	1600m	114m

Samples of the laid up cable were taken from the running end of each drum so that the lay length could be measured. When measured, the variation in lay lengths taken ranged between 116mm – 135mm. The operators were using the incorrect lay lengths when laying up the cable. This is done when one of the core lengths is short. The lay length is increased so that the final lay-up length is the required 1 600m. By doing this there is a balance of core lengths left for 3 cores while one core is completely used. The balance lengths of the cores is scrapped as excess length.

Test

The conductor resistance test is done using the lay-up length (1600m). This is done to ensure that the electrical properties required is correct. If the actual results are higher than the specification, more than the standard amount of material required was used. This is a loss in terms of yield because extra material has been "consumed" by the product. On the other hand, if the actual results are lower than the specification, less than the standard amount of material required was used. This is a gain in terms of yield because less material has been "consumed" by the product which means material was saved. A 5% variation between the specification resistance and the actual resistance is allowed. Table 5 shows the results for conductor resistance test.

Table 5: Results for conductor resistance test

Specification	Actual Results	Difference	Percent Variation
1.33 Ω/km	1.26 Ω/km	0.07 Ω/km	5,26
	1.26 Ω/km	0.07 Ω/km	5,26
	1.25 Ω/km	0.08 Ω/km	6,02
	1.28 Ω/km	0.05 Ω/km	3,76

These results demonstrate that more than standard amount of material was used, leadings to a loss in terms of yield.

Sheathing

The recorded data for input and output length is shown in Table 6.

Table 6: Input and output length for lay-up

Input length (P3D169)	Output Length (P3D165)	Difference in length
4x432m = 1 600m	1633m	33m

The clock at 9E2 over-reads by 20m for every 1000m, and the collected data also showed that the length counters on 9E2 and 10B7 were not reading the same length.

Test

Table 7 shows the results for testing of the final cable length. The testing of the final cable length is essential as it is the final test done to ensure that the electrical cable being delivered is according to the required specifications and is safe. The conductor resistance test is done using the final length of 1 633m from 9E2. This was to determine if there was a difference between the results obtained from the test done after lay-up. Since 5% variation is allowed between the specification resistance and the actual resistance, the results demonstrate that less than the standard amount of material required was used.

Table 7: Results for testing of the 1633m cable length

Specification	Actual Results	Difference	Percent Variation
1.33 Ω/km	1.28 Ω/km	0.05 Ω/km	3,76
	1.29 Ω/km	0.04 Ω/km	3,01
	1.27 Ω/km	0.06 Ω/km	4,51
	1.26 Ω/km	0.07 Ω/km	5,26

Of the 1 600m final cable, 3 final drums of 400m were tested. This was done to also determine if there was any difference between the results obtained using the 1 633m length from 9E2. The results are recorded Table 8 below:

Table 8: Results for testing of the 400m cable length

Specification	Drum 1		Drum 2		Drum 3	
	Actual Results	Percent Variation	Actual Results	Percent Variation	Actual Results	Percent Variation
1.33 Ω/km	1.25 Ω/km	4,51	1.25 Ω/km	4,51	1.25 Ω/km	4,51
	1.26 Ω/km	3,76	1.26 Ω/km	3,76	1.25 Ω/km	4,51
	1.27 Ω/km	4,51	1.28 Ω/km	3,76	1.27 Ω/km	4,51
	1.26 Ω/km	3,76	1.27 Ω/km	4,51	1.26 Ω/km	3,76

Printing

The clock gives the same length as 9E2. This implies that the 3D5 length counter is also over-reading 20m for every 1 000m. Any leftover cable is scrapped as excess length.

4.3 Results

The collected data revealed some problems such as incorrect drums being used for the incorrect orders during stranding, incorrect input length during core insulation, and mismatch of length counters during rewinding, which resulted in excess lengths and scrap. It was this crucial to conduct a root cause analysis for reduction of waste on rubber cables. Figure 3 shows a cause and effect diagram which highlights the main problem areas which were the main contributors to the excess lengths.

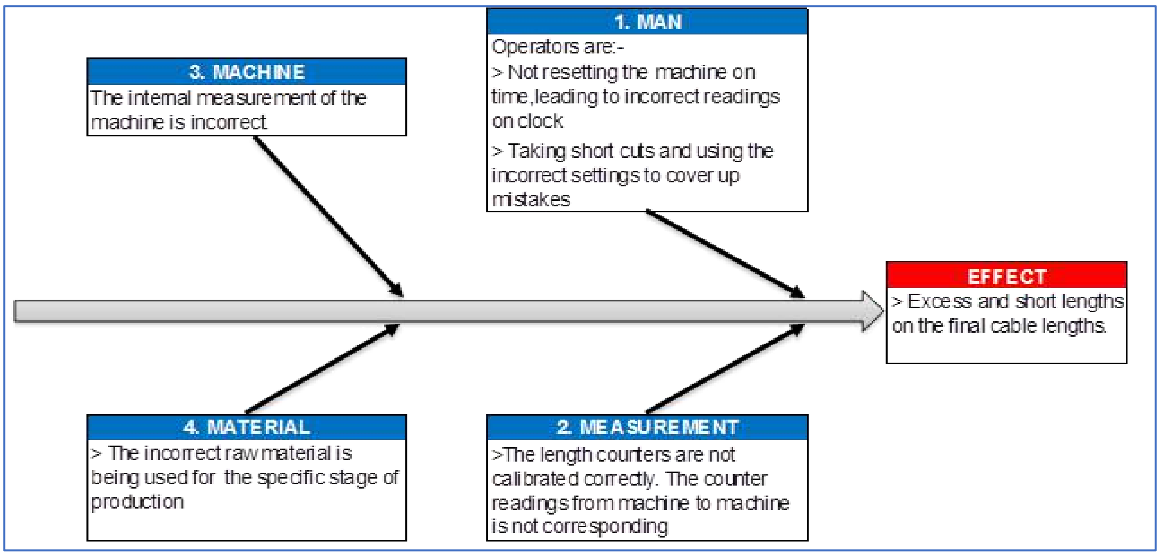


Figure 3: Cause and effect diagram for excess and short cable lengths

The cause and effect diagram highlighted that there are numerous causes for the excess problem, especially the 16mm^2 rubber cables but the main factors that were focused on during the study include:

- Man – The greatest deal of variance arises because of human error which occur when the operators are recording the correct information, resetting the machine on time, and loading the orders onto the machine so that the correct settings and standards are used.
- Machine – The machine settings and functioning were also to be checked to ascertain whether the machine was running at the correct speed with the correct order, whether the inputs were recorded correctly (i.e. raw materials, drum numbers), whether the correct standards were being used, and whether the parameters of the machine were correct.
- Measurement – The variance in lengths from machine to machine is high and implies that the length measuring clocks are not reading the same length and may not be calibrated correctly. So the focus was on checking if the calibration factor on the machine is correct, ensuring that all the machines involved have been properly calibrated, and finding out at which stage the greatest difference in length difference lies.
- Material – Material had to be checked to ascertain whether poor quality material used, or whether incorrect raw material was being used for the specific stage of production
- Method – The methods had to be checked if there was any incorrect procedure being followed and shortcuts were taken to do tasks instead of following Standard Operating Procedure (SOP)
- Environment - The working environment conditions were to be checked for optimality and also to check if 5'S was being practiced.

These issues then lead to the incorrect method being used. The left over length at any machine is scrapped as excess lengths and this is avoidable because the incorrect input values are used (Material, Machine, Man) and the measurement between machines is not corresponding. Therefore, a single length counter cannot be trusted as the base measurement.

5. RECOMMENDATIONS

The following recommendation were made to the production department in order to possibly reduce or eliminate the problem:

- *Calibrate the machine length counters to a single reference machine* -This will create a definite reference point that all the machines are calibrated to. The chosen machine is the printing machine 3D5 as it is the final length counter before despatch. This will also eliminate the variance from machine to machine because instead of having individual clocks that are set independently, all the clocks are set in relation to one point.
- *Use 2 main length counters* - Main length counters should be installed at the beginning of the process and at the final stage of the process. This will create two key spots that ensure that the input length is correct and the final output length is correct. This will help decrease the excess lengths at each machine because there are only two definite length counters.
- *Educate the operators about the correct SOP and methods to be followed* - Train the operators on quality so that they are aware of the implications of their actions. It is vital to retrain them so that they follow the correct procedures. It is also crucial for management to promote honesty instead of fear on the factory floor so that problems can be fixed at the source instead of shortcuts being taken, creating larger unnecessary problems. Operators should maintain basic equipment conditions (cleaning, lubrication, bolting); maintain operating conditions (proper operation and visual inspection); discover deterioration, mainly through visual inspection and early identification of signs of abnormalities during operation; enhance skills such as equipment operation,
- *Recalibrate all the machine clocks* - This is a long-term solution and it will ensure that all the length counters are correct. It will also decrease the variance from machine to machine by ensuring that they are all correct.
- *Create an internal printing line on 9E2* - This will combine the sheathing and printing phase and save the company time and money.
- *Eliminate the rewind line phase* - The insulated cores should go directly to the braiders instead of to the rewind line then to the braiders. This will eliminate the length variance of one phase as the lengths are not cut at the rewind lines.
- *Team-working culture* - It is important to inculcate a team-working culture, a team help to break down the barriers that are inherent in the traditional approach and to identify problems.

6. CONCLUSION

Cause and effect diagrams can be deployed for reduction of waste on rubber cables. Errors and mistakes during production need to be solved and rectified at the source so that defects are not carried over to the subsequent stage of production. It was found that the initial problem was not major but when it is not solved it leads to more problems. If the initial input length and output length of the first stage is accurate then there should be no excess lengths left over at any of the subsequent stages of production other than the start-up allowances which are accounted for in production planning. It is vital to implement continuous improvement methodology in order to sustain an organisation in this highly competitive environment. It is imperative to ensure that the clocks are calibrated correctly, and that the operators use the correct procedures and methods, as well as promoting clear communication and honesty among employees so that the root cause of problems can be found and resolved.

REFERENCES

[1] **Paolucci, M. and Sacile, R**. 2016. Agent-based manufacturing and control systems: new agile manufacturing solutions for achieving peak performance. Crc Press.

[2] **Mendelsohn, A., Ramachandran, R. and Richardson, B**. 2008. Importance of quality compounds for utility cable reliability. In Rural Electric Power Conference, IEEE pp. C2-C2.

[3] **Randive, A.J., 2010**. Scrap reduction model: By combination of DMAIC and design of experiments. The University of Texas at El Paso.

[4] **Napieralska-Juszczak, E. and Napieralski, P.** 2017. Optimisation of the manufacturing process of a stranded cable. In Environment and Electrical Engineering and 2017 IEEE Industrial and Commercial Power Systems Europe (EEEIC/I&CPS Europe), IEEE International Conference, pp. 1-6.

[5] **Powers, W.F.** 1994. The basics of power cable. IEEE transactions on industry applications, 30(3), pp.506-509.

[6] **Rosato, D.V. and Rosato, M.G**. 2012. Injection molding handbook. Springer Science & Business Media.

[7] **American Insulated Wire Corporation**. A Guide to Wire and Cable Construction, Pawtucket, Rhode Island, p18.

[8] **Anixter Inc**. 2008. Thermoplastic vs. Thermoset Wire and Cable Materials, Wire Wisdom.

[9] **Santos, J., Wysk, R. A. And Torres, J. M**. 2014. Improving Production with Lean Thinking. John Wiley & Sons.

[10] **Elpidio, R. and Teresa, M.D.** 2013. Lean maintenance model to reduce scraps and WIP in manufacturing system: case study in power cable factory. WSEAS TRANSACTIONS on SYSTEM, 12(12), pp.650-666.

[11] **Romano, E. L. P. I. D. I. O., Murino, T., Asta, F. E. L. I. C. E., & Costagliola, P. I. E. R. V. I. N. C. E. N. Z. O.** (2013). Lean Maintenance model to reduce scraps and WIP in manufacturing system: case study in power cables factory. WSEAS Transactions on Systems, 12(12), 650-666.

[12] **Dewa, M., Naicker, A. and Singh, R**., 2013. Root cause analysis for reduction of waste on bottle filling and crowning operations. SAIIE25 Proceedings, 9th-11th of July.

CAN THE MACHINE!

D. Hartmann[1*], T. Hattingh[2] & N.R. Mokoena[3]

[123]School of Mechanical, Industrial and Aeronautical Engineering
University of the Witwatersrand, South Africa
[1]dieter.hartmann@gmail.com, [2]teresa.hattingh@wits.ac.za

ABSTRACT

Informal waste collection is a sophisticated network of waste collectors, scavengers, sorters and transporters. This system sources waste to recycling companies in exchange for money. Millions of people make a living as waste pickers. Their work contributes positively toward the economy, the environment and public health. Informal waste collection is well described in many countries, however, little information on the commercial and spatial structure exists for South African operations.

This study explored the operations of informal waste collection in Bethlehem in the Free State and Johannesburg in Gauteng, with 20 and 40 informal participants respectively. Data were collected through interviews, group discussions and participant and direct observations.

Informal waste pickers collect, sort, transport and sell recyclables to registered and privately-owned companies in exchange for money. The most common materials collected are white paper, plastic bottles and aluminium cans. Informal waste collectors in Bethlehem and Johannesburg collect their recyclables from the streets, dumpsites, landfill sites, shops, rubbish bins and municipal containers. Most informal waste pickers are male. Waste pickers use trolleys and wheeled bins as transport. Informal waste collectors usually collect recyclables at proximity to where they live and trade their recyclables at the nearest formal recycling company.

Informal workers in Johannesburg work longer hours a day compared to those in Bethlehem, but those in Bethlehem earn more money due to lower competition.

*Corresponding author

1. INTRODUCTION

1.1 Background to the research problem

The informal sector in the waste management and waste recycling industry is a sophisticated network of collectors, transporters, sorters, merchants and managers. This system supplies waste to formal recyclers in exchange for money. It plays a vital role in the waste management and recycling industry. However, there is little knowledge or information available on how this system works especially when it comes to the supply chain network and the remuneration model used.

This study had to consider the legal framework within which the industry operates, whilst establishing roles and responsibilities in the network, processes that are followed, material interchanges between actors in the network, the methods used for material transport and the flow of money in the system.

This paper is perhaps an ironic nod to the conference theme – realising that much as industry can move towards an Internet of Things centred knowledge economy – there are some industries for which the integration of technological means is probably not likely in the medium term.

1.2 Motivation

Informal waste collection and recycling plays an important role in the solid waste management industry as it contributes positively to the economy, the environment and, public health and well-being. Millions of people across the world make a living from selling waste material for money.

Additionally, the informal sector plays an important role in keeping the streets clean, saving municipalities, in countries like South Africa, millions, reducing the volume of waste in landfills, creating jobs and therefore reducing the level of unemployment.

Informal waste collection mainly occurs in developing countries due to their low level of economic development [1]. “Poor wages and low prices for products and services create viable profit margins from collecting and selling secondary raw materials. If alternative employment opportunities and associated wages were higher, scavenging would be less financially attractive” [1].

Literature contains information on how this system works in many developing countries but there is little information available on the material exchange network and the remuneration model used in South Africa. This research project aims at tapping into areas of informal waste collection and recycling where there is limited knowledge available. The purpose of this study is to find out how the system works in two different places in South Africa: Johannesburg (a city) and in Bethlehem-Bohlokong (a town). This will allow for a better understanding to be gained on the difference between the material exchange network and remuneration model in urban and rural South Africa.

2. RESEARCH FOCUS

2.1 Research question

How does informal waste collection and recycling work in selected South African locations?

2.2 Research objectives

The research question will be answered by addressing the following objectives:

1. Identify the actors involved in the informal waste recycling industry and their responsibilities.
2. Identify all the processes involved in the informal waste collection and recycling industry.
3. Investigate the structure of the waste material exchange network or supply chain network.
4. Investigate the type of waste materials collected by informal waste pickers/scavengers.
5. Investigate areas where informal waste pickers/scavengers collect waste.
6. Investigate where informal depots are located in relations to the sources and customers of waste.
7. Investigate the remuneration model used in this industry.

3. LITERATURE REVIEW

3.1 Waste management

Americans alone produce over 389 million tons of waste a year [2]. This number is by far more than any other nation in the world [2]. “Solid waste management is a challenge faced by African cities today. Most cities are unable to manage the increase in volume of solid waste due to urbanisation and population growth” [3]. “The generation of high volumes of waste material by the economically active consumer population, especially in the urban areas fuelled by urbanisation, has led to the emergence of the so-called waste picker generation” [3].

“Waste management is the collection, transportation and disposal of garbage, sewage and other waste products” [4]. Ultimately, good, and green, waste management aims to treat waste products to be converted into valuable resources. Waste disposal options range from dead end options such as landfills and incineration to recovery and recycling options, which include gasification, decomposition, and reuse.

3.1.1 The relationship between the formal and the informal sector in solid waste management

In many countries, waste management is approached formally and informally. The informal waste collection and recycling industry in India, Egypt [5, 6], Mexico [7, 8], Jordan [9] and Ethiopia [3] were investigated. Considerable urban waste picking entrepreneurship was identified [10].

From reviewing literature on the countries above, it was discovered that the type of materials collected by informal workers was generally the same, the waste material exchange network or supply chains have similarities, with the major difference being related to scale.

In countries like Ethiopia, formal approach is the sole responsibility of the government [3], handling solid waste in two ways: through the door-to-door waste collection and secondly through a container system into which residents are required to carry and dump waste [3]. Collected waste is disposed of on landfill sites. The formal system is complemented by informal recovery and recycling system. This is where individuals, families or small enterprises source waste, which gets sold to recyclers.

"The informal sector is characterised by small scale, labour-intensive, largely unregulated and unregistered, low-technology manufacturing or provision of services" [1]. Unlike the formal sector enterprises or entrepreneurs, the informal sector receives low pay, are untaxed, and they do not have trading licenses and are not included in any government insurance schemes or social welfare [1].

An investigation in Addis Ababa, Ethiopia on solid waste management showed that informal actors play a vital role in solid waste management most specifically in recycling, an area where the government has no capacity of its own [3]. Solid waste generation in Addis Ababa public spaces, households and institutions is estimated to be around 0.36 kg per capita each day, most of the waste generated is organic [3]. The government solid waste collection system collects and disposes wastes like vegetables, bones, wood, and leaves and miscellaneous other waste while informal actors collect paper, glass, metals, plastics and clothes to resell or recycle [3].

3.1.2 General flow of waste in the solid waste recycling industry

In the waste recycling industry, waste flows from the source to the recyclers. As mentioned in the section above some countries do not have any formal recycling service and all recycling activities are undertaken by informal actors. For countries where there are formal recycling services, informal actors supply formal recyclers with waste material. Figure 1 shows the flow of waste from generators to recyclers.

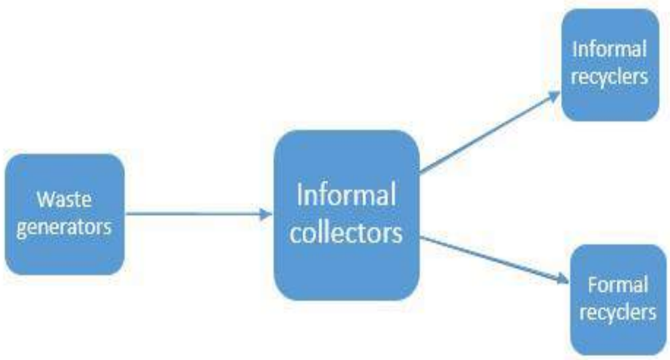


Figure 1: Flow of waste materials

3.2 Informal waste recovery and recycling

The informal sector is the unregistered, unregulated and unorganised sector in the solid waste management system which simply means that there are no barriers to entry. Actors in the informal sector collect waste materials and trade it for recycling. This sector seems to exist in both urban and rural areas but the quantity of waste materials recovered in rural areas is much smaller.

A South African study on waste picking revealed a difference in the income generated by informal workers across cities although no correlation between the size of the cities and the income generated [11]. An interplay however exists between the variation in the amount exchanged for the waste collected, the availability of waste and the competition for the waste [11].

3.2.1 Processes undertaken in the informal waste collection and recycling industry

Informal actors in the waste management industry unlike formal actors see an opportunity to make money by trading waste material for recycling instead of throwing it away. These actors believe in adding value to waste materials others have discarded and have considered useless. Many different steps and processes are undertaken to add value to waste material which had been discarded. These steps are: [1]

- Collection
- Sorting
- Accumulation
- Pre-processing
- Small manufacture and craftsmanship
- Trading

3.2.2 Informal waste collection and recycling in developing countries

Table 1 shows the average population size and waste generated, the percentage of waste collected and the waste properly disposed for developed and developing countries. The amount of waste generated was measured in kg/capita/day.

Table 1: Waste generation for developed and developing nations (adapted from [12])

Countries	Population size	Waste generation	Waste collection	Proper disposal
Developed (High-Income)	1.0 Billion	1.3 kg/capita/day	100%	100%
Developing (Middle-Income)	3.0 Billion	0.77 kg/capita/day	60%	30%
Developing (Low-Income)	2.4 Billion	0.59 kg/capita/day	40%	5%

Table 1 shows that in developed countries waste management systems are very effective as all the waste is collected and is properly disposed, therefore leaving no room for informal recycling. In developing countries, insufficient collection, inappropriate disposal in open dumps and uncontrolled street collection points are a major issue [7]. Thus Informal waste recycling "is widespread throughout urban areas of the developing world and it is reported that up to 2% of the population in Asia and Latin American cities depend on waste picking to earn their livelihood" [1].

Informal waste recyclers can be categorised into at least four types.

3.2.2.1 Itinerant waste buyers

These individuals purchase pre-sorted waste material for accumulation and batched resale into the formal recycling economy. This approach is particularly common in China with highly disciplined domestic waste sorting behaviours [1].

3.2.2.2 Street waste picking

This practice is characterised by picking from mixed waste found on the streets and/or inside bins. This method is commonly used in countries like South Africa, Jordan, and Ethiopia.

3.2.2.3 Municipal waste collection crew

With this type of informal waste recycling, recyclables are recovered from municipal pickup trucks. Access to material is either free or secured through payments or bribes to employees in the formal waste management economy. This method of recovering waste material is common in developing countries like Colombia, Philippines, Mexico and Thailand [1].

3.2.2.4 Waste picking from dumps

Recyclables are collected from dumpsites before they are covered or treated with chemicals. Most waste collectors live near the dumpsite. This is common in cities like Rio de Janeiro, Mexico City, Manila, Bangalore, Guadelajara, Dar es Salaam, Cape Town, Guatemala City and many others [1].

Figure 2 shows the network and interaction of actors in the informal waste system in Addis Ababa. The complex interwoven network is typical for a complex social network.

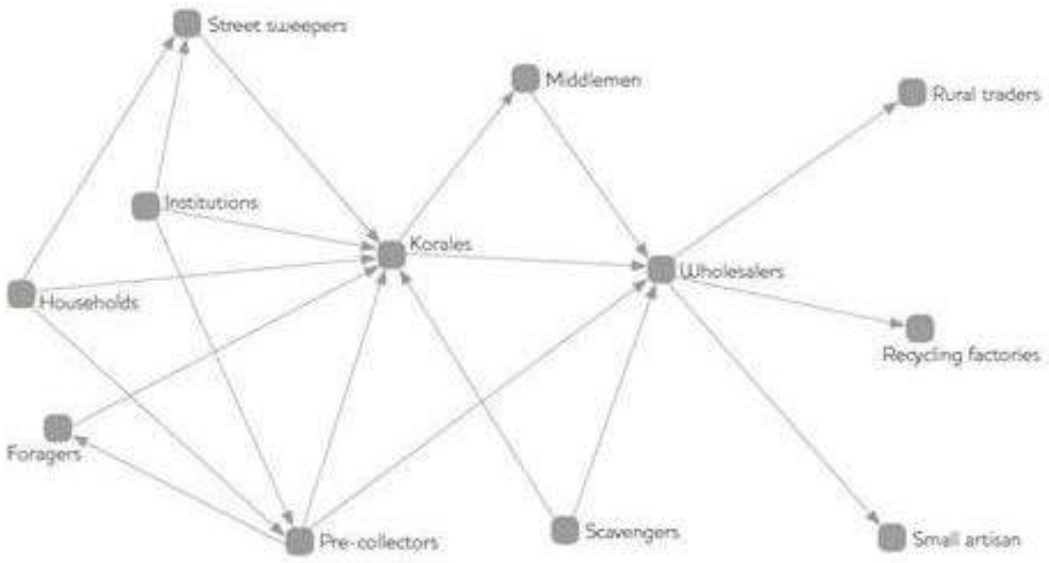


Figure 2: The waste system in Addis Ababa [3]

Secondary materials that have been collected by informal workers are often traded locally with end users such as artisans and craftsmen. As shown in Figure 2, there are intermediate dealers such as brokers, junk shops, wholesalers and intermediate processors between waste pickers and the end-users [1]. Activities carried out by these intermediate dealers can be informal or formal.

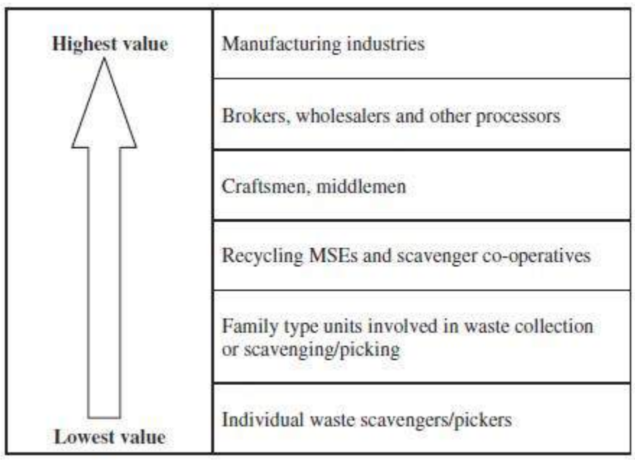


Figure 3: The recycling network hierarchy [1]

Figure 3 shows a recycling network hierarchy. “The higher a secondary raw material is traded, the greater the added value it possesses” [1]. Informal recyclers are at the base of the secondary material trade hierarchy [1] which means that they make small amounts of money from trading waste materials. Individual waste pickers or scavengers have limited capacity when it comes to processing and storing waste materials, they are the most vulnerable and are easily exploited by those at higher levels in the trade hierarchy [1].

3.3 The current situation in South Africa.

South Africa is an upper middle-income country producing about 1.7 kg/capita/day in urban areas and 0.35 kg/capita/day in rural areas [11]. It is estimated that about 62 000 people in South Africa make a living as waste pickers [13]. Waste pickers in this country collect, sort and sell recyclables and reusable materials such as cardboard, paper, plastic, metal and glass, and they earn on average an income ranging from R 50 to R 150 per day [12]. A non-profit organisation based in Durban called Asiye eTafuleni estimated that an average income generated by waste pickers in Durban ranges between R2,400 and R3,360 a month. 80 – 90 % of waste collected by waste pickers gets recycled, and it is estimated that each picker diverts up to 24 tons of waste a year [6]. Major product prices include cardboard at about 80c/kg, plastic at R2/kg, newspaper at 40c/kg and light steel at 70c/kg [14].

Informal waste collection and recycling exists in both the urban and rural areas in South Africa. Plans are to slowly formalise the informal sector in urban areas whereas most recycling activities in rural areas are carried out by private companies [8]. Informal waste collection seems to be more difficult in rural areas due to the erratic availability of trucks from recycling companies. This means that informal waste collectors need to have areas where their recyclables are stored, typically outside landfill sites [8].

4. RESEARCH METHOD

A qualitative descriptive field study [15] was carried out in two places; in Bethlehem (a town) in the Free State and in Johannesburg (a city) in Gauteng.

4.1 Data Collection

Unstructured or semi-structured techniques were used for the collection of data. The study was carried out through individual interviews, group discussions and direct participant observations [16]. Informal waste collectors and recyclers are very busy individuals and those methods were considered the most effective for this research project [15]. Data collection tools such as surveys and questionnaires were not used to avoid issues around texture and nuance, and also to compensate for low literacy levels [17].

During the time the field research was conducted, all actors who participated in the study were asked verbally for permission to participate in the study before interviews and group discussions were held, and before observations were made. Participants were also briefed on what the project is about, the purpose of it and also the importance of them taking part in the study. In the section below, a brief explanation on how the methods were used is given.

4.1.1 Participant Observation & Direct Observation

These methods were chosen as it allowed the researcher to be integrated in the activities carried out by informal waste collectors and recyclers, which made it easier to gain insight. During participants'- and direct - observation, photographs were taken using a mobile phone camera. Participants had to grant the researcher permission to take the photographs. Where photography was declined, notes were taken on paper.
The observation protocol [18] evaluated:

- The physical conditions of the area where the participant was working;
- The activities carried out by participants;
- The type of equipment participants used as they carried out their work;
- The type of recyclables participants collected and/or sold;
- The way they interact with other actors in the informal recycling industry;
- Points of sale of collected recyclables.

4.1.2 4.1.2 Individual interviews and group discussions

During interviews and group discussions, conversations were audio recorded using a mobile phone sometimes the information shared was written down on paper. Participants were only audio recorded if they consented. For those who did not consent, information was logged on paper.

5. OBSERVATIONS

From Bethlehem, a total of 20 waste pickers participated in the study of whom 60 % were male. All pickers in Bethlehem participated in collection, sorting and selling however only 50 % of pickers transported material.
In Johannesburg, a sample of 40 randomly chosen participants showed only a single female waste picker. Of this sample, all pickers participated in sorting, transporting and selling. 96 % of participants also collected material, the balance bought collected material from other recyclers. A minority (7 %) also weighed, priced and bought material. This minority was the highest earning, even having vehicles to transport.

As waste pickers collect recyclables, some put the recyclables inside black plastic bags while some in polypropylene bags referred to as 'tough bags'. Figure 4 shows the polypropylene bag and black plastic bags used by waste pickers/scavengers.

Figure 4: Tough bags and black refuse bags

After collecting the recyclables, waste pickers/scavengers separated the recyclables into categories. For transport, some waste pickers carried the black plastic bags over their shoulders while some used trolleys as shown in Figure 5. In Bethlehem, the local recycling company offers free transportation services for those who have more than four 'tough bags' full of recyclables.

A

B

C

Figure 5: A) modified trolleys, B) storage areas, C) unmodified dollies

5.1 Areas where participants collect their recyclables

Waste pickers have preferences when it came to the areas where they collect waste materials. Table 2 shows the different areas where recyclables were collected by participants for the two locations.

Table 2: Location choices for waste collection

		Percentage of pickers who make use of locations	
	Location	**Bethlehem**	**Johannesburg**
1	Streets	65	100
2	Dump Sites	60	60
3	Households	60	N/A
4	Landfills	20	95
5	Shops	25	12.5
6	Rubbish Bins	55	95
7	Municipal Containers	55	32.5

Some of the waste pickers/scavengers prefer collecting recyclables from the streets, whilst others prefer scavenging in dumpsites found mostly in open spaces in their township or collecting from households where waste pickers arrange for communities to put certain materials aside and not to throw them away. Some prefer scavenging for recyclables at a landfill site located a couple of kilometres away from the township. Some pick from rubbish bins located in front of stores, restaurants, cafés and in public spaces such as parks and marketplaces. Some scavenge in municipal containers located in open spaces in the township, where residents dump their waste. Municipal waste collectors with pickup trucks only collect waste once a week, so if it happens that rubbish bins get full, residents usually empty their bins into those containers. Lastly some waste pickers/scavengers collect recyclables from shops. They arrange with shop owners to put aside certain waste materials for them to collect later.

In Bethlehem, the picking domains are strongly zonal, with pickers either concentrated in the town, or in the nearby township, with little or no overlap. In Johannesburg, waste pickers/scavengers who participated in the study collected their recyclables in Braamfontein, Hillbrow, Newtown and the central business district (CBD).

5.2 Daily working times

The average picker in Bethlehem worked an observed 6.5 hours per day, ranging between 4 and 12 hours, whilst in Johannesburg, pickers worked an average of eight hours per day, ranging from 5 to 11 hours.

5.3 Trade frequency

Pickers differed in the frequency of trading. Pickers sold their goods either daily or every two, three or five days. Notably, the income normalised per day, was highest for those trading daily, and lowest for those trading weekly. The average income for the sample was R 202.58 per day in Bethlehem and R 128 in Johannesburg.

Table 3: Average daily income

Trade every .. Days	Income per day (Bethlehem)	Income per day (Johannesburg)
1	R 351.87	R 99
2	R 112.5	R 266
3	R 108.33	R 78
5	R 95	R 87

5.4 Informal depots

An informal depot is a place where recyclables are stored after collection. This research showed that participants who collect recyclables from the landfill site do not need a storage area for their recyclables since they trade daily, simply contacting recyclers from the formal sector to fetch the recyclables. Daily traders did not need storage space unless some accumulation was needed to reach minimum trading batch sizes. For those who traded on a two, three or five-day cycle, storage space was necessary. It was found that some participants stored the recyclables in their homes while others did so in the streets where they sleep.

Figure 6 shows an open area in Marshall and Nugget street Johannesburg, where some of the participant store their recyclables.

Figure 6: Informal depot between Marshall and Nugget streets, Johannesburg.

The participants generally collect recyclables at close proximity to where they live. For Bethlehem, those who live in the township (Bohlokong) collect recyclables from nearby tuckshops, households, dumpsites and/or municipal containers. Those who live in town (Bethlehem) usually collect recyclables around town.
In Johannesburg, pickers who stay in Hillbrow or Braamfontein usually collect recyclables in the area where they stay. The only challenged faced by these informal workers is walking long distances to trade their recyclables. In both locations, most participants are homeless and typically sleep on the streets.

5.5 Other observations

In no case was the use of any protective clothing or equipment observed. It was notable, particularly in Bethlehem, that up to 50 % of the participants were children.
Despite the potential benefit of forming syndicates, the level of organisation amongst waste pickers was very low. This is seen as a potential area for optimisation, as a great deal of work duplication takes place between pickers.

6. RESULTS

6.1 Identified actors

After processing the data that was collected, Figure 7 and Figure 8 shows the actors that were identified in Bethlehem and in Johannesburg respectively.

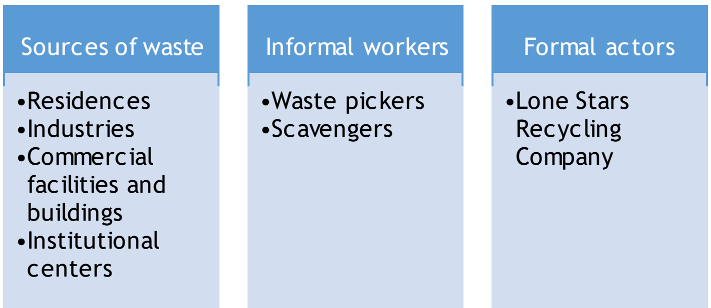


Figure 7: Actors in Bethlehem

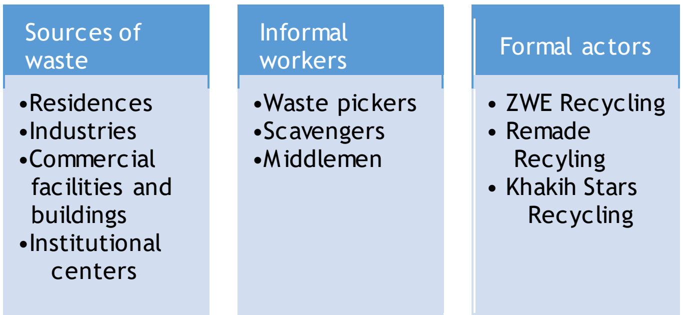


Figure 8: Actors in Johannesburg

6.2 Process Map

Processes or activities carried out by the identified actors are mapped in Figure 9.

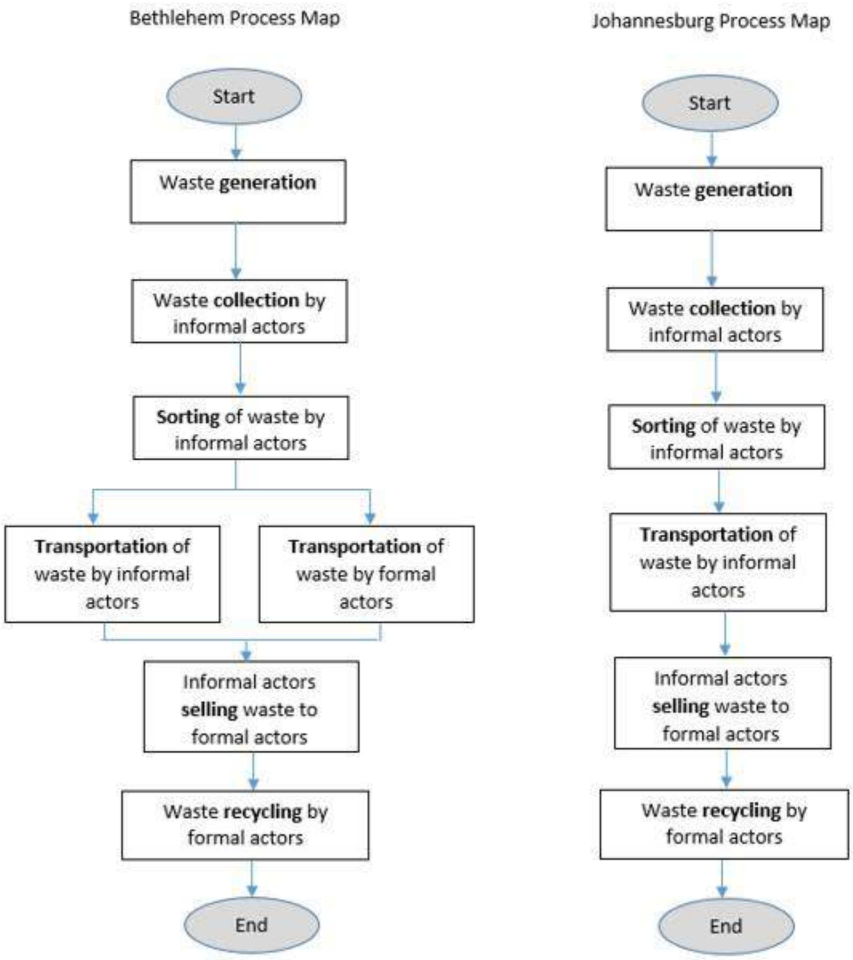


Figure 9: Waste process maps

6.3 Waste material exchange network

Figure 10 and Figure 11 show the flow of waste materials amongst the different actors in the waste collection and recycling industry, with the arrows pointing in the direction in which the waste is flowing.

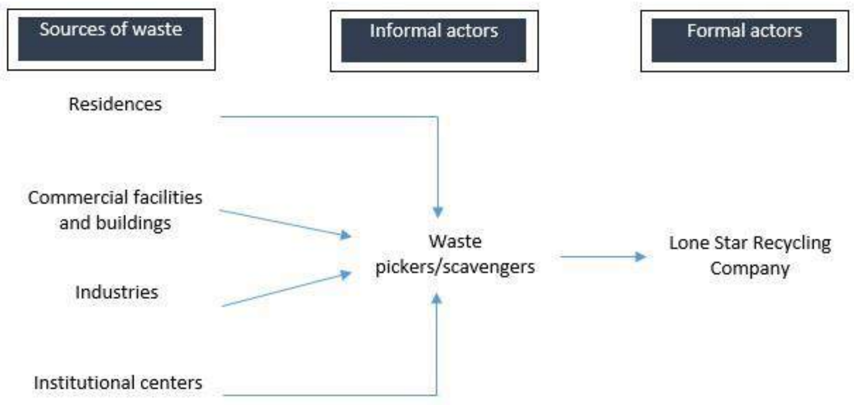


Figure 10: Waste material flows in Bethlehem

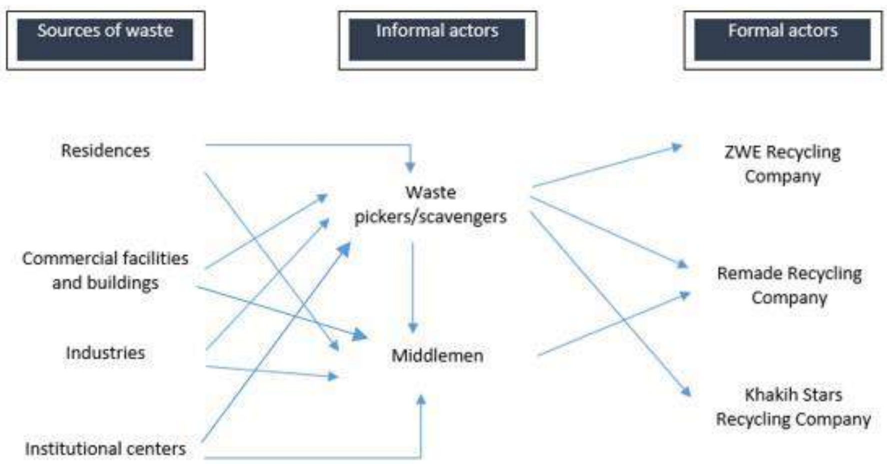


Figure 11: Waste material flows in Johannesburg

6.4. The types of recyclables collected and traded by informal waste collectors Informal waste collection isbroad, and although there are specialised operators who will collect only a single category of waste, this is uncommon, and most collectors will collect a variety of products, which includes plastic, paper, metal, glass and other wastes, which includes industrial waste.

Table 4 shows the categories of collected waste materials.

Table 4: Products collected

Code	Description	Products
PET	Polyethylene	Cold drink bottles
HDPE	High Density Polyethylene	Milk, margarine, cleaning material containers
Make PLSM (LDPE)	Plastic - Mix	Shopping bags, bread bags and frozen vegetable plastic bags
K4	Used cardboard	Cardboard
FN	Flat News	Newspaper
CMW	Common Mixed Waste	Magazines
HL 1	Heavy Letter 1	White paper

PP	Polypropylene	Paint, ice cream and yoghurt containers
UBC	Used beverage cans	Aluminium cans
GL	Glass	Glass bottles and jars
IMW[3]	Industrial Mixed Waste	Cereal boxes

7. DISCUSSION AND CONCLUSION

This study found that there is an unexpectedly low level of integration of the informal waste picking industry [19]. The expected segregation of roles, between pickers, transporters, sorters and traders was not found, with almost all waste pickers performing all these roles, which is inefficient, slow and reduces the earning potential of each participant in the system.
Not all pickers make use of trolleys which has the potential to significantly raise the earning potential of a waste picker. Given that waste picking is not only of benefit to the pickers, but indeed to society, the simplified provision of trolleys to pickers would be beneficial and assistance in this regard, perhaps from commercial donors, would be of universal benefit.

The forming of operational syndicates would significantly improve the effectiveness of the informal waste collection system. This would also allow for more direct and speedier integration into the formal recycling system, allowing for greater cash flows, as accumulation of waste is no longer required.

The working conditions, stigma and earning potential in the industry is regrettable. The use of vulnerable people, the risk of personal harm and injury, and social ostracisation may require assistance from formal structures to protect people in this field.

The working hours found in this study were surprisingly low, particularly considering the distances all pickers had to cover. The sources of waste materials, and the types of materials picked, were however as expected.
The role of the formal economy is considerable. Pricing, and therefore earning potential is dictated by prices that these agents are willing to pay for waste.

8. REFERENCES

[1] **Wilson, D.C., Velis, C., and Cheeseman, C.** 2006. Role of informal sector recycling in waste management in developing countries, *Habitat international*. 30(4), pp 797-808.
[2] **Van Haaren, R., Themelis, N., and Goldstein, N.** 2010. The state of garbage in america, *BioCycle*. 51(10), pp 16-23.
[3] **Teshager Alemu, K.** 2017. Formal and informal actors in addis ababa's solid waste management system.
[4] **Demirarslan, K.O. and Demirarslan, D.** Chapter methods and recommendations on evaluation of furniture wastes from ecological aspects.
[5] **Gerdes, P. and Gunsilius, E.** 2010. *The waste experts: Enabling conditions for informal sector integration in solid waste management: Lessons learned from brazil, egypt and india*, GTZ.
[6] **El-Dabh, B.** 2012. *President's controversial waste collection programme becomes institutionalised*, in *Daily News Egypt*.
[7] **Medina, M.** 2005. Serving the unserved: Informal refuse collection in mexico, *Waste management & research*. 23(5), pp 390-397.
[8] **Schneider, P., Sembera, J., and Silva, R.** 2017. *The role of the informal sector in a rurbanised environment* in *Solid waste management in rural areas*. InTech.
[9] **Aljaradin, M., Persson, K.M., and Sood, E.** 2015. The role of informal sector in waste management, a case study; tafila-jordan, *Resources and Environment*. 5(1), pp 9-14.
[10] **Rogerson, C.M.** 2001. The waste sector and informal entrepreneurship in developing world cities, *Urban Forum*. Springer. pp 247-259.
[11] **Schenck, R. and Blaauw, P.F.** 2011. The work and lives of street waste pickers in pretoria—a case study of recycling in south africa's urban informal economy, *Urban Forum*. Springer. pp 411.
[12] **Viljoen, J.M.M.** 2014. *Economic and social aspects of street waste pickers in south africa*, University of Johannesburg.
[13] **Makhasana, S., Sidoni, N., Rule, S., Richards, R., Dlulane, G., and Sibeko, L.** 2012. *Report on the determination of the extent and role of waste picking in south africa*.
[14] **Farber, T.** 2017. *Meet the bin scavengers saving sa r750-million a year*, in *Times Live*, tiso blackstar group (Pty) Ltd.

[3] Mixed Industrial Waste was the only product category which was collected in only one of the locations, namely Johannesburg.

[15] Yin, R.K. 1987. *Case study research: Design and methods*, Sage Publications, Inc.
[16] Kvale, S. and Brinkmann, S. 2009. *Interviews : Learning the craft of qualitative research interviewing*, Sage.
[17] Phillips, E. and Pugh, D. 2010. *How to get a phd: A handbook for students and their supervisors*, McGraw-Hill Education (UK).
[18] Rossman, G.B. and Sharon, F.R. 2003. *Learning in the field: An introduction to qualitative research*, Sage.
[19] Medina, M. 2008. The informal recycling sector in developing countries: Organizing waste pickers to enhance their impact.

USE OF MULTIPLE ACTIVITY CHARTS FOR PROCESS IMPROVEMENT IN RUBBER WIRE CABLE MANUFACTURING

M. Dewa[1*]

[1]Department of Industrial Engineering
Durban University of Technology, Country
mendond@dut.ac.za

ABSTRACT

Waste reduction becomes an increasingly dominant topic for companies that strive to survive in the modern manufacturing world. This paper applies multiple activity charts for process improvement in rubber cable manufacturing at a case study company. The process is first mapped to outline the basic manufacturing process which included mixing, milling, straining and cooling. Tasks were then categorised as value-adding task, non-value adding but necessary, and non-value added. The non-value adding waiting was noted where the mixer operator waited for the miller operators to complete rolling a batch before mixing a new batch. Two scenario, where mixer waits for miller and where miller waits for mixer, were analysed through the use multiple activity charts. The analysis revealed that the current method employed was unbalanced since more time was wasted in waiting since the mixing time was longer than milling time. It was thus proposed that the mixer operator continuously works while the miller milled the batches as they became available. The mixer waiting time was eliminated as the mixer was continuously loaded, leading to productivity improvement. It was also recommended that the technique of job rotation should be employed due the disparity in workloads for mixing and milling operators.

*Corresponding author

1. INTRODUCTION

Waste reduction becomes an increasingly dominant topic for companies that strive to survive in the modern manufacturing world. One of the most significant challenges that is faced by manufacturing organisations is the continuous process improvement that entails new insights about the behavior of processes in order to understand their potential for optimization or improvement [1]. Multiple activity charts are used as tools for process improvement to describe activities of more than one subject, recorded on a common time scale to show their interrelationship. The objective is to rearrange activities, organise teamwork, minimise maintenance time, reduction of change-over time and idle time [2]. The case-in-point cable manufacturer produces its own rubber compounds which is used for insulation and sheathing. Rubber is the only raw input material that is produced on site and other materials such as copper for the cable making process are outsourced. Rubber is produced in batches and there are various compounds with different properties that are used for different types of cables. The problem that was being faced by the case-in-point cable manufacturer is that it was failing to achieve its target of 60 batches per shift and it was argued from the operators' perceptive that the target was unattainable. Therefore, the aim of the study was to determine if the target could be achieved and to find ways to increase the output by reducing or eliminating waste.

2. LITERATURE REVIEW

2.1 Types of waste

The principle of lean manufacturing is grounded on the identification of value, elimination of waste, and generation of flow of value to the customer. Waste is described as any process activity that does not add value to the customer [3]. The expenditure of resources for any objective that does not create value for the end customer is considered to be wasteful, and thus should be targeted for elimination [4].

The seven most common types of waste include:

- Transport - This activity relates to transport of materials, which is often indispensable, but does not create value. It is worth noting that that there are more chances of delays or an increase of damage and loss when a product is transported more often.
- Motion - These are unnecessary movements by people or equipment moving or walking more than is necessary when performing activities. Excessive travel between workstations and machine movements cause stress to employees and machines and can cost time and money, as well. Movements that are done more than is required may also cause damage and injury to the operator [5].
- Waiting — This form of waste happens when people, equipment or products waits to be processed, and it does not add any value to the customer. Operators may stopped waiting for parts, machines, or other workmates.
- Defects - These are errors that occur during processing and may require re-work or additional work. Defects are the most visible of the seven types of waste, yet are not easy to catc h before they reach the consumers. Defects retard production and increase lead time. Since every defective part needs a replacement or repair, which wastes materials and time, in terms of quality, defects tend to cost more and thus, waste of defects should be reduced or eliminated. There should be systematic and valid control methods in place so that defects can be identified before the product is unnecessarily processed, causing an unnecessary cost to the company [4].
- Inventory - this type of waste comprise of raw materials, work in process, or finished goods in the shop-floor, warehouse and other areas of a production plant. Inventory must be stored, and this requires space, packaging and transportation. Inventory has a possibility of being damaged during transport. Waste of inventory tends to hide other production waste such as machine breakdown, poor scheduling, quality problem, transportation time of raw material, poor vendor delivery times, lengthy set up time and line imbalance. Thus, minimum inventory is encouraged since it exposes all other types of waste [6].
- Over-processing – It occurs when a particular process step does not add value to the product and more work is done on a process than is required.
- Overproduction - It is created by producing more products than are required by the customer. Overproduction is related to inventory waste and producing in excess consumes an organisation's resources in advance, resulting in extra inventory. [5].

2.2 Process Improvement

Process improvement is the proactive task that is characterised by identification, analysis and improvement of existing business processes within an organization for optimization and to meet standards of quality [7]. There are different approaches to process improvement such as lean manufacturing, business process re-engineering and benchmarking, and a systematic methodology is followed for each, of which the process can either be complemented with sub-processes, modified or eliminated for the ultimate goal of improvement. It is an ongoing practice of continuous improvement and should be expedited with the analysis of tangible evidence of improvement from performance measurement. Process improvement is characterised by initially understanding the purpose of a process, workflow or activity and then analysing to identify major problems and how they can be eliminated. It also embraces comprehension of the standards that must be achieved and activities should be monitored to ensure that performance targets are met [8].

Improving operation procedures is a simply method that make optimum use of the raw materials that are used in the production processes. Standard operating procedures reduce system variation and thus are crucial in ensuring consistent quality, and they support training and reduce risk [9].

2.3 Multiple Activity Charts

Multiple activity charts are used to show the interrelationships of individuals in teams of workers, or the relationships between workers and equipment. They are used to illustrate parallel activities and time relationships between two or more resources and are useful where the interactions between workers, equipment and materials repeat in periodic cycles. The charts can aid in identifying bottlenecks and idling resources by showing the utilisation profile of each resource in parallel bars. They can also provide a valuable tool for monitoring progress in critical situations where a detailed understanding of the workflow is needed [10].

3. METHODOLOGY

The methodology that was adopted is chacterised by 5 steps that include the following:

Step 1: Communication with all concerned
In order for the project to be successful, the co-operation of the operators and the team leaders is required. This is because they have authority and control over the tasks to be improved. If the operators and relevant supervision have the correct understanding of the study then they will co-operate accordingly instead of being against the study. This means that there will be fewer human error factors that would occur, the study would show the true potential and substantial improvements can be made at the case-in-point plant.

Step 2: Understanding the machine processes
There are 3 ways to identify waste by classing each process step under the following categories:

- Value-Added task: The actual process that transforms the input material into the desired output.
- Non-Value Added task but is necessary, e.g. transportation of WIP from process to process.
- Non-Value Added task but is unnecessary, e.g. waiting for the process 2 to be complete the first batch before process 1 starts the next batch, tasks can be completed simultaneously so that there is a continuous flow of production.

Therefore is essential to understand the process so that the correct improvement and waste areas can be identified.

Step 3: Study of the process and data collection
A production and time study was conducted for a month, with the aim to monitor and study the process in order to collect a reliable amount of data that would be used to calculate a standard time to produce one complete batch. The necessary trials were done in order to determine the actual output, cycle time and improvement areas at the case-in-point plant.

Step 4: Summarise and analyse the study results and outcome
The study was analysed in stages so that each stage was monitored and improved accordingly. Creating point improvements was avoided as it may not influence the process as a whole and would not necessarily cause an improvement in the ability to reach the target or increase overall output.

Step 5: Presentation of results and recommendations to production
The results and improvement areas that were found were presented to production department so as to aid in meeting targets, improving efficiency and daily output.

4. PROCESS ANALYSIS

4.1 Basic stages in rubber production

The basic stages at the case-in-point plant are:

- **Mixing -** This is where the necessary compounds are mixed. The rubber output batches re made up of the following compound groups that are added to the mixer accordingly: - Polymers, Fillers, Chemical additives, Accelerators. The operator loads the mixer which independently mixes the compound. It also automatically opens at specific set temperatures to allow for the necessary compounds to be added. Once the batch has been mixed for the necessary time it drops onto a conveyor which leads to the next process. Although they do not necessarily work with set times for mixing but instead they work with specific temperatures.

- **Milling -** This operation is when the mixed batch is blended/milled between rollers to further mix and blend the rubber compound. There are specific blending times and temperatures to be set in order to ensure proper blending. It is then cut into strips and feed to the next stage via a conveyor system.
- **Straining -** This in when the rubber compound is fed into a strainer which strains the rubber (to remove contamination) and then extrudes the rubber which is cut into 2 strips.
- **The cooling system - This** is when the rubber strips pass through a water trough and then onto a series of conveyors which pass by fans which cool the rubber down. This is done so that the rubber cools down and does not stick to each other when packed onto the pallet.

The case-in-point plant has the target of producing 60 batches per shift and has not been meeting this target. It produces various types of rubber that have different properties, i.e. outer/inner sheathing, outer/inner insulation; heavy-duty cables; light duty cables; and other rubber product.

4.2 Pareto Analysis

Pareto analysis is a formal technique that can be used by Industrial Engineers to identify key areas to derive better benefits after some course of action. Figure 1 shows the Pareto results for percentage monthly demand for different types of rubber by the plant.

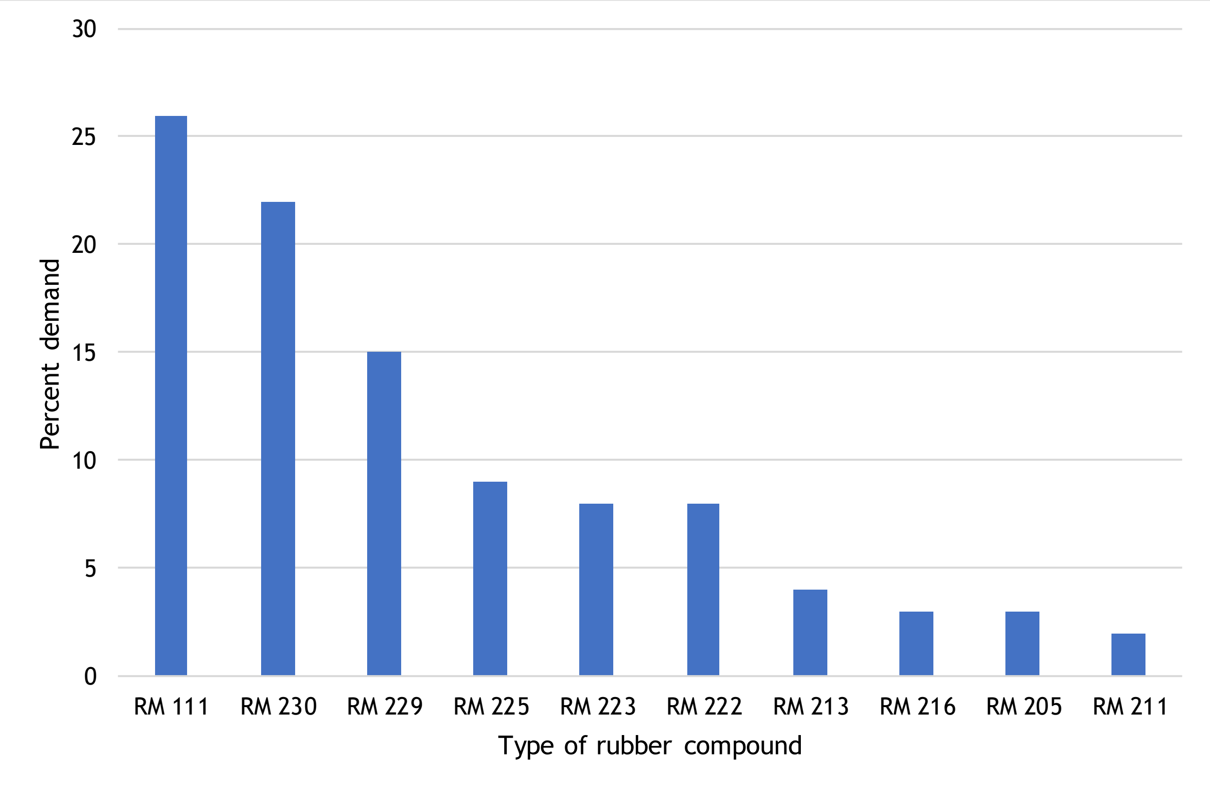


Figure 1: Pareto results for demand of different types of rubber

Pareto analysis revealed that high voltage cable (RM 111), extra heavy duty CR sheathing material (RM 230) and heavy-duty CR sheathing material (RM 229) were the most utilised rubber compounds, and thus the study focused on RM 111.

4.3 Multiple Activity Charts

Table 1 shows the time study results for RM111 which is characterised by 4 elements that include loading compound, first mixing, loading accelerator and second mixing.

Table 1: Time study results for RM111

El. No.	Element	Total basic time	No. of elements	Average basic time	Allowance	Std time
1	Load Compound	15.54	10	1.55	25	1.94
2	First mix	20.52	10	2.05	0	2.05

3	Load Accelerator	2.98	10	0.30	25	0.37
4	Second mix	22.40	10	2.24	0	2.24

4.3.1 *Multiple Activity Chart 1*

Figure 2 shows the current method employed for the high voltage cable (RM 111). The mixer operator waits for the batch to be milled before mixing a new batch. The chart shows that in 1 hour, 6 batches (as per target) can be produced, given that there are no disturbances. With cycle times longer than that of RM 111, in an hour, less than 6 batches can be made and this is below the hourly target. The mixer waits for 21 minutes in a 60 minute time frame. The mixer operator idles for 38.5 minutes in a 60 minute time frame. The miller and the miller operator wait for 39 minutes in a 60 minute time frame.

Figure 2: Multiple Activity Chart if mixer waits for miller

It was noted that the mixer operator waits for the miller operators to complete rolling a batch before mixing a new batch. Time is wasted waiting because the mixing time is longer than the milling time. Therefore batch B can be mixed while batch A is milled and the miller should complete batch A before batch B is dropped. This will create continuous flow of batches and eliminate idle time.

4.3.2 *Multiple Activity Chart 2*

Figure 3 shows an improved state where the mixer operator continuously works and the miller mills batches as it is dropped. In one hour 8 batches of RM 111 can be made considering that there are no disturbances. This is 2 batches more than the target. With cycle times longer than that of RM 111, in an hour 6 batches can be made. The mixer waiting time is eliminated as the mixer is continuously loaded. The mixer operator idle time is reduced to 29.5 minutes. The idle time is reduced by 9 minutes. This is only a 23.4% improvement but is still positive because the number of batches produced in an hour increased from 6 to 8.

The miller and miller operator idle time decreased to 27.5 minutes. The idle time decreased by 11.5 minutes (29.5% improvement). Table 2 shows the machine and labour capacity utilisation of the before-and-after scenario implementation of the method study. When scaled up, according to chart 1, 48 batches, of RM 111, can be produced in an 8-hour period where a single material is continuously mixed. According to chart 2, 64 batches, of RM 111, can be produced in an 8 hour period where a single material is continuously mixed. This is 16 batches more which is a 33% improvement.

Table 2: Machine and Labour capacity utilisation

	Before state	Improved state
Mixer	65%	100%
Mixer Operator	65%	51%
Miller	35%	54%
Miller operator	35%	54%

Figure 3: Improved state for multiple activity chart

These results demonstrated that the target could be achieved and method study can be deployed to increase the output by reducing or eliminating idling time.

4.4 Areas for concern

Kaizen is an excellent strategy that can be adopted in Industrial Engineering to create continuous improvement by adopting small, ongoing positive changes to reap major improvements. It was noted that there was also room for improvement concerning the standard operating procedures (SOPs) that were not followed correctly, thereby resulting in increasing cycle time increasing. The mixing cycle time increased because the ramp was not functioning properly. The ramp was sitting at a single position (960) instead of fluctuating in position (1100 - 920) to create the necessary pressure to mix the batch and increase the temperature. Figure 4 shows the fluctuation in cycle times for RM 111 and it was crucial to understand the reason behind the fluctuation of the cycle times. The cause was poor preventive maintenance frequency since it was noted that there was a build-up of dust in the mixer ramp causing a blockage. The ramp was supposed to be cleaned/dusted every time the mixer is loaded to ensure the build-up is avoided. At cycle 4, the time to reach temperature had increased by approximately 0.5 min. From cycle 4 to 10, the time to reach temperature continued to increase. At cycle 9, a technician from the maintenance department was called to check the mixer ramp that was not functioning correctly to create the necessary pressure. At cycle 10 the ramp was given a quick clean before being loaded. It is noticed that the time to reach temperature dropped but not down to the usual time taken. From cycle 12 onwards the ramp was cleaned after loading, the dust on, in and around the mixer was brushed down. The time to reach temperature dropped down to the “normal time” usually taken. The time taken to reach optimal operating temperature decreased after the interventions and the ramp operated without problems.

The second area of concern was time lost due to low morale since it was noted that the operators took longer breaks than the allowed time. For a final stage to stop exactly at or few minutes before break, the operators have to stop mixing at approximately 10 minutes. A comparison was made between Number of batches made considering extra break times taken versus Number of batches that can be made if standard times are followed. The base information for the times for the two scenarios are:

Number of batches made considering extra break times taken:

- Cycle time taken 10 minutes
- Actual Break times taken = 60 minutes lunch, 20 minutes tea time (x2)
- Meeting Time = 30 minutes
- Wash up Time = 30 minutes
- Set up time = 30 minutes

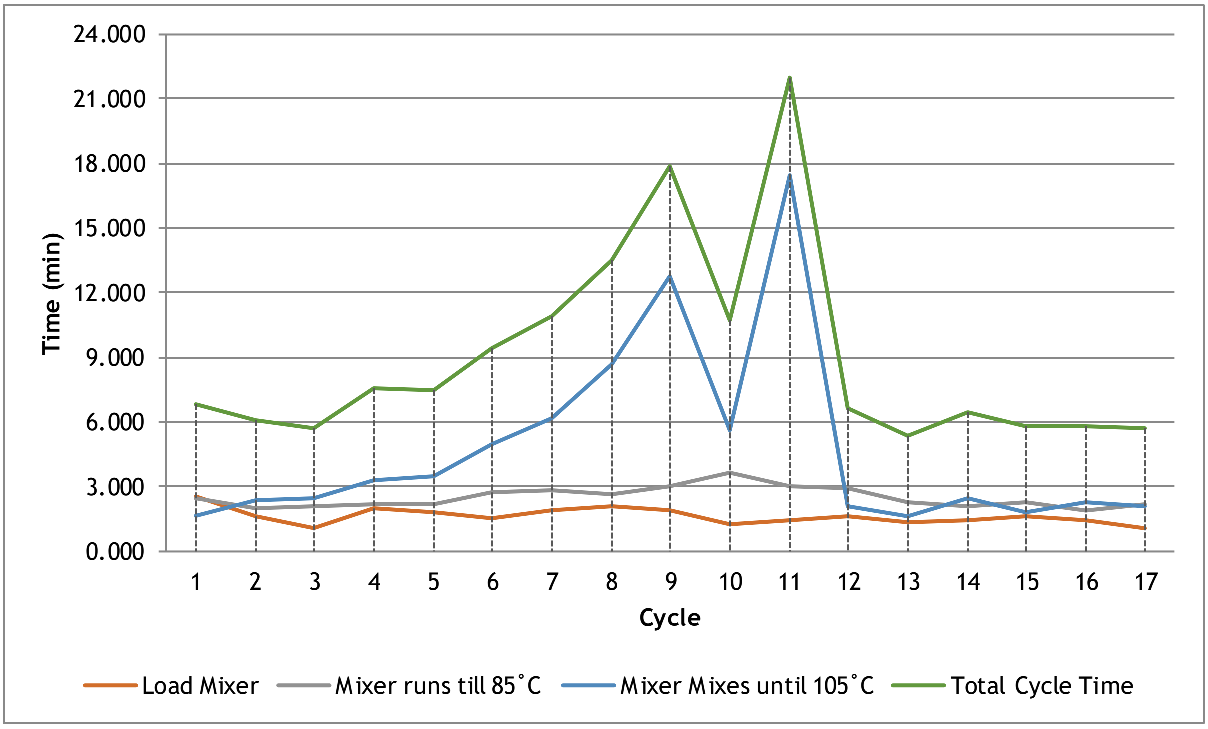


Figure 4: Fluctuation of cycle time during mixing

Number of batches that can be made if standard times are followed:

- Cycle time taken 10 minutes
- Break time = 45 minutes lunch, 10 minutes tea time (x2)
- Meeting Time = 15 minutes
- Wash up Time = 15 minutes
- Set up time = 30 minutes

An output trial was conducted to determine the output of the strainer which was unknown.
The trial was conducted at 6 and 5 RPM, the results were as follows

- 8.42 kg/min at 6 RPM
- 6.9 kg/min at 5 RPM

Using the output trial results the strainer cycle time was calculated as follows:

Table 3: Output of the strainer

Material	Mass (kg)	Time (min)	
		6 rpm	5 rpm
RM 111	79.79	9.476	11.564
RM 229	71.99	8.550	10.433
RM 230	70.53	8.376	10.222
RM 235	80.60	9.572	11.681

Using the study results, the average batches to be produced when using standard times, was determined and is indicated in Table 3.

Table 4: Average batches to be produced when using standard times

Material	Standard Time (min)			Batches Produced	
	Mixing	Mill	Strainer	Per Hour	Per Day
RM 229	6.270	3.880	8.550	7	140
RM 230	4.000	3.840	8.376	7	143
RM 111	6.610	5.540	9.476	6	127

From the results it was concluded that the strainer is the bottleneck process at the rubber production plant. The operators stop mixing batches when the strainer pot is full and start mixing again when the pot is almost empty. The operators should rather monitor the pot level and mix accordingly instead of waiting for the strainer pot to empty. The operators were abuse lunch times. They left early for lunch and arrive late from lunch. The operators would also take breaks at different times. The strainer operator could take a later lunch so that rubber in pot is not left to get cold. The mixing and mill operator can take lunch when the pot is full.

5. RECOMMENDATIONS

The following is suggested based on the results and analysis to improve the Case-in-point plant performance:

- Continuously mix batches so that the milling operation waits for the next batch instead of the mixer standing until the milling operation is completed.
- Ensure that the standard operating procedures are followed correctly so that batches can be continuously and consistently mixed.
- Discipline operators so that they do not take advantage of break times and stick to the stipulated times instead of leaving early for breaks and arriving late from breaks.
- The operators can take their breaks at different times so that the output is maximised and there will be a continuous flow of operations.
- The technique of job rotation should be employed due the disparity in workloads for mixing and milling operators.

6. CONCLUSION

In today's business environment, the expenditure of resources for any objective that does not create value for the customer is wasteful, and thus should be targeted for elimination. With constant pressures to cut costs and innovate, process improvement can be deployed to identify and eliminate major problems and embraces comprehension of the standards that must be achieved and activities should be monitored to ensure that performance targets are met. Multiple activity charts can provide a valuable tool for monitoring progress in critical situations where a detailed understanding of the workflow is needed. SOPs must make bottom-line economic sense, especially if an organisation invest the time and energy to develop and implement effective SOPs. The study revealed that the target of 60 batches a shift is achievable under ideal situations. It also revealed that there is excessive time wasted in areas that are unnecessary and the operators need to follow the correct methods and operating procedures to avoid time wastage.

REFERENCES

[1] **Stojanovic, N., Dinic, M. and Stojanovic, L.** 2015. Big data process analytics for continuous process improvement in manufacturing. In *Big Data (Big Data), 2015 IEEE International Conference. IEEE.* pp. 1398-1407.

[2] **Kanawaty, G.** 1992. *Introduction to work study*. International Labour Organization.

[3] **Womack, J.P.; Jones, D.T.** Lean Thinking: Banish Waste and Create Wealth in Your Corporation; Simon & Schuster: New York, NY, USA, 1996; pp. 305–311.

[4] **Melton, T.** 2005. The benefits of lean manufacturing: what lean thinking has to offer the process industries. *Chemical engineering research and design*, *83*(6), pp.662-673.

[5] **Simboli, A., Taddeo, R. and Morgante, A.** 2014. Value and Wastes in Manufacturing. An Overview and a New Perspective Based on Eco-Efficiency. *Administrative Sciences*, *4*(3), pp.173-191.

[6] **Arshad, M.R.** 2007. *A Case Study on Manufacturing Waste Reduction Through Just in Time*, Doctoral dissertation, University of Malaysia Pahang.

[7] **Stein, S.J.** 2017. *The EQ leader: instilling passion, creating shared goals, and building meaningful organizations through emotional intelligence*. John Wiley & Sons.

[8] **Chang, J.F.** 2016. Business process management systems: strategy and implementation. Auerbach Publications.

[9] **Hunt, G.E.** 1990. Waste reduction techniques: An overview. Pollution Prevention Review, 13, p.26.

[10] **Heap, J.** 2015. Multiple Activity Charts. *Wiley Encyclopedia of Management*, pp.1-1.

TAYLOR SERIES METHOD FOR THE ANALYSIS OF NON-LINEAR AND DYNAMIC ENERGY CONSUMPTION MODEL UNDER A FLEXIBLE INVENTORY MODEL

B.B. Benga[1], T.B. Tengen[2] & A. Alugongo[3]

[123]Faculty of Engineering and Technology,
Department of Industrial and Mechanical Engineering,
Vaal University of Technology
[1]bengaebouele@yahoo.fr, [2*]thomas@vut.ac.za, [3*]alfayoa@vut.ac.za

ABSTRACT

Existing literature shows that operational inefficiency is one important cause of energy consumption problem encountered within most manufacturing plants. Energy consumption can be perceived as an important key performance indicator because it is mostly influenced by the production rate, which in turn is linked to the lot-size and lead-time. Machines and processes within a plant produce goods at variable rates. When speeds vary, slow rates typically result in dropped profits while faster speeds affect quality control. Producing more, earlier, or faster than required by the next process is waste. It causes inventory to accommodate the excess and frequent reprocessing. This is why it is important for operating speeds to remain consistent with the inventory model's parameters, and it can be done by carefully handling the random behavior of lot-size and lead-time. In this research paper, Taylor Series Method is used to analyze, approximate the dynamic property of energy consumption around the average lead-time or lot-size, which are two important components of a hybrid inventory ordering policy. The results revealed that it is possible to achieve good approximation of energy consumption in the neighborhood of important average point (decision parameter of the inventory model).

[1] The author was enrolled for a DTECH (Mechanical) degree in the Department of mechanical Engineering, University of Vaal, South Africa
*Corresponding author

1. INTRODUCTION AND BACKGROUND

Existing literature shows that operational inefficiency is one important cause of energy consumption problem encountered within most manufacturing plants. Energy consumption can be perceived as an important key performance indicator because it is mostly influenced by the production rate, which in turn is linked to the lot-size and lead-time. Note that both lead-time and lot-size (order quantity) are two important input of every inventory policy [1] [2] [3] [4]. Many inventory-ordering policies have been proposed to address important issues in most manufacturing plants and unfortunately none of them has effectively addressed the existing energy consumption problems. These problems consist of carefully manipulating variable lot-size or orders quantities in order to minimize the inventory cost under a set of appropriate constraints. The present state of inventory costs within most plants is a proof that such issue still needs to be addressed [4]. The concept of inventory control and management is not new in organizations [5] [6] [7] [8]. For instance, the economic order quantity model was developed to help organizations with the problem of determining the optimal quantity to order [9]. In order to improve inventory control in complex and dynamical environment, a probabilistic model was developed [9]. Recently, a hybrid inventory model was developed to enhance inventory control [10]. Although cost-effective, the aforementioned hybrid inventory model presented some drawback, which is an indication for improvement on the modeling approach.

In general, the state of a hybrid inventory model can be described by the values of continuous variables and discrete mode. Information can be seen on references [11] [12], for an introduction of hybrid systems. In the same order of idea, a linear combination technique was used to hybridize both continuous *(r,Q)* and periodic *(R,S)* inventory model. Although the results of implementing such hybrid system were found useful in reducing the total inventory cost, it should be observed that linear equations cannot predict and control effectively the stochastic manufacturing plant performance behavior. At this stage, a new science is therefore needed because the old one is just insufficient. A non-linear combination technique should be used to hybridize both continuous *(r,Q)* and periodic *(R,S)* inventory model. It is then necessary to build a robust hybrid inventory system and test it to see whether it can be used to deal with the issue of efficiency within a manufacturing plant. In this research paper, the proposed hybrid inventory model is seen as the one that both flow and jump. Thus, an appropriate framework meant to study Such hybrid inventory behavior is needed. Note that a hybrid inventory model that flow and jump can be described by a differential equation. An optimization approach that exploits the randomness of the production rate under specific circumstance (down time machine) and studies its effects on the hybrid inventory level is proposed. More attention is finally devoted to formulating accurate model for lot-size with regard to variation in energy consumption, which is perceived as key performance indicator.

One of the things energy consumption change can make happen is the high cost of implementing the proposed hybrid inventory model. Given the fact that the cost of implementing a hybrid inventory model within a manufacturing plant is subjected to variations, energy consumption may be seen as an important factor that prevent such cost to remain as low as possible. This high hybrid inventory cost is mostly due to random production rate. Unfortunately, random production rate mostly impacted by downtime machine is not well monitored. Note that production rate is function of the lead-time and lot-size. Note that both lead-time and lot size are mostly random in most manufacturing plants. Further, lead-time which is an important input of the hybrid inventory model is highly affected by delay (downtime resulting either from material hardness or over production). Here, the lead-time is totally dependent on the lot-size and production rate. Meaning that for a specified lot-size, the lead time will increase if the speed at which items are produced decrease. The aforementioned relationships should be analyzed with enough care because they indirectly lead to energy cost. Hence, it is also required to study the impact of both production rate and lot-size on energy consumption. The analysis of the aforementioned relationship is then justified. Note that a proposed method meant to analyze the relationship between energy consumption and inventory is organized as follows: Theoretical foundation of the model; Properties of both hybrid and energy model; Empirical part of the models (model estimation and empirical study); Simulation study; results.

2. METHODOLOGY

2.1 Case study company background

A case study company is a manufacturing firm around the Moungo region (Cameroun) which produces and sells more than 2 chemical products all over the country. There, products are distributed via company owned distribution. More frequently, the distribution center orders products from the factory, stores them at its facility, and waits for customers (agents or retailer), who then sell the products to their own customers. One of the products that are distributed and that represent the majority of annual sale is analyzed. The current inventory policy being used is periodic review *(R, S)*. At the beginning of each week, the inventory manager of the distribution center reviews weekly sale forecast together with the space available in the distribution center and decides on the quantity to order by using his own experience. Further, the inventory manager orders enough products to prevent stock-out problem. For this study, the company would like to further examine other possibilities such as hybrid inventory review systems that is made of combination of both periodic *(R, S) and* continuous *(r, Q)* review model. In addition, the company also would like to better understand the impact of different factors (random production rate, demand) on energy consumption level. Primary data over a year period were then collected and analyzed in order to fully define and understand the problem (see figure 1).

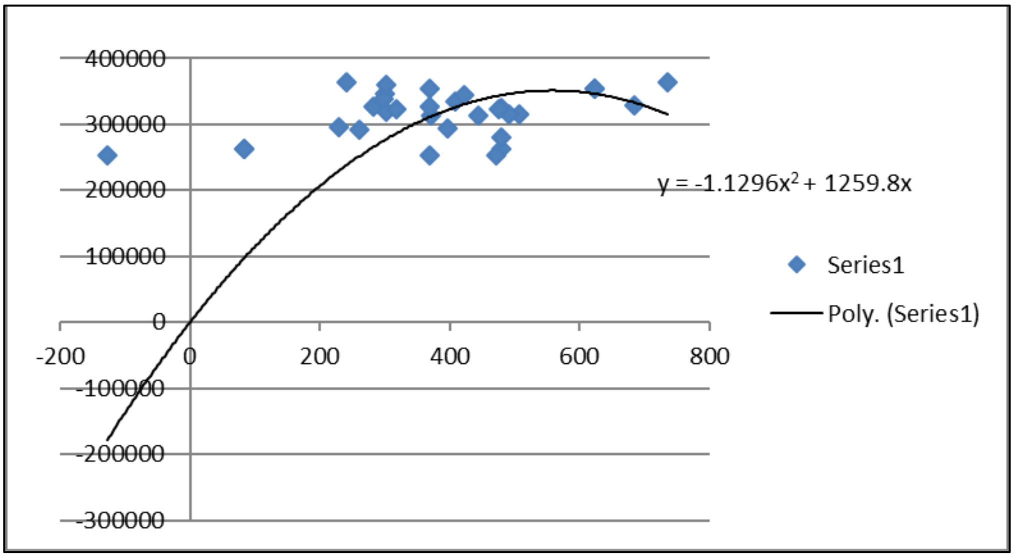


Figure 1: Energy consumption data versus lead-time

As can be seen from figure (1), it is observed a polynomial energy function of degree 2. This is called a quadratic energy function with a lading coefficient of -1.12. It is also observed a concave form (increase and decrease) of energy consumption. Such pattern of energy consumption may come with many disadvantages that are not fully addressed in most research papers. For instance, the polynomial pattern of energy could be most influenced by the random production rate, which in turn is related to lead-time and the lot-size based. Thus, it is then developed an optimal energy inventory model in which energy is derived from energy source. A conceptual approach in which such energy model is derived from past data has been dealt with by many authors [13]. However, most of these models have not been fully developed within a plant, taking into consideration the impact of a implementing a flexible and hybrid inventory ordering policy. This was a good reason to analyze the cost implication of energy consumption under hybrid inventory model that combines the feature of both a periodic and continuous inventory models. The dynamics of such hybrid inventory model is fully described in previous research papers [10]. For a practical demonstration of such hybrid inventory system, read the next paragraphs.

2.2 Hybrid inventory review problem

Genaraly, the annual hybrid cost of ordering, holding inventory and incurring backorder as described by Rossetti [14] [10] [20], provides a good starting point.

$$TC(r,R,Q) = k * N(t)_{hybrid} + h * \frac{1}{T} * \int_0^T I(t)_{hybrid}\,dt + j * \frac{1}{T}\int_0^T B(t)_{hybrid}\,dt \qquad (1)$$

Where *TC* is the total cost per unit time of implementing a hybrid *(r, R,Q)* inventory review policy, *r* is the re-order point, *Q* is the lot-size, *R* is the review period, *k* is the order preparation cost per order, *N(t)* is the number of replenishment orders made per unit time, *h* is the holding cost for the item in units per time, *T* is the period or cycle time, $I(t)_{hybrid}$ is the inventory on hand, *j* is the backordering cost for the item in unit per time, and $B(t)_{hybrid}$ is the inventory backordered. It should be noted that the number of replenishment per time $N(t)_{hybrid}$ is dependent on the demand rate and the inventory on hand. $I(t)_{hybrid}$ is dependent on depletion rate.

Number of replenishment $N(t)_{hybrid}$

The features needed to describe Periodic *(R, S)* and Continuous *(r, Q)* review combination system is used to determine the number of replenishment over the period. This hybrid system may depend on the relative numbers of replenishment performed under both review systems. Let $N(t)_{con.h}$ be the number of replenishment performed per period under a Continuous *(r, Q)* review system and $N(t)_{per.h}$ the number of replenishment performed per period under a Periodic *(R, S)* review system needed for hybrid model at a specified capacity utilization level *U*. Represented as a percentage, the capacity utilization is the extent to which the productive capacity of a firm is being used in generation of goods and services. It is then possible to formulate the equation for the total number of replenishment in performing both Periodic *(R, S)* and Continuous *(r, Q)* review system as follows [14] [10].

$$N(t)_{Con_h} * T_{Cont_h} + N(t)_{Per_h} * T_{Per_h} = T * U \quad (2)$$

$$N(t)_{Con_h} * Q_{Cont_h} + N(t)_{Per_h} * Q_{Per_h} = D_{Total} \quad (3)$$

$$N(t)_{Cont_h} + N(t)_{Per_h} = N(t)_{Hybrid} \quad (4)$$

Where U is the capacity utilisation during the period in percentage, $T*U$ is the total time (number of weeks) of operation per period, $T_{cont.h}$ is the replenishment cycle time under a continuous *(r, Q)* review system in weeks/year, $T_{per.h}$ is the replenishment cycle time under a periodic *(R, S)* review system in week/year, $Q_{con.h}$ is order quantity under a continuous *(r,Q)* review system needed for a hybrid model, $Q_{per.h}$ is the order quantity under a periodic *(R,S)* review system needed for a hybrid model, $N(t)_{hybrid}$ is the total hybrid inventory replenishment.

In this case, the cycle time (time/cycle) under a periodic *(R, S)* inventory system can therefore be expressed by

$$T_{\text{Prev}} = \frac{(Q_p)}{\left(\frac{\partial D}{\partial t}\right)} \quad (5) \qquad T_{con} = \frac{(Q_c)}{\left(\frac{\partial D}{\partial t}\right)} \quad (6)$$

Where Q_p or (Q_{per}) is the order quantity following periodic review system. Further, the cycle time under a continuous *(r, Q)* review system has Q_c or (Q_{cont}), *which* is the order quantity.

Inventory held $I(t)_{hybrid}$

From a mathematical perspective, the total hybrid inventory level may be described by three components: the normal consumption inventory, drop inventory and lead time safety stock inventory. The first one, called "the normal consumption inventory may be described as the inventory that is depleted from the initial stock. A possible model is then.

$$I(t)_{A3} = \int_0^T (I_0 - \mu(t))dt I(t)_{A3} = \int_0^T \left(I_0 - \left(\frac{\partial D}{\partial t} * t \right) \right) dt \quad (7)$$

Where, $I(t)_{A3}$ represents the instantaneous inventory before the re-order (jump process) takes place with, t ($t > 0$) is the time that elapse between the initial time (beginning of operations) to the re-order time, I_0 is the initial inventory, $\mu(t)$ is the degradation/utilization.

The second type of inventory, called "drop inventory" is the one needed to deal with stock-out events when the on-hand-inventory drops below a specified point before the review date R as a result of external factors such marketing and promotion. Note that the quantity needed to cover the expected need (demand) at this particular time can automatically be placed. It implies that there should be restocking times t_r and associated order amounts Q_r. This drop inventory performed through the spontaneous replenishment process may be described by.

$$I(t)_{B3} = \frac{1}{T} \int_0^T (\delta(t - t_r) * Q(t_r)) dt \quad (8)$$

Where δ is the Dirac delta function or step function, t_r is the time at which the jump process is observed and Q_r is the order quantity at time t_r.

The third inventory known as safety stock inventory is the one required for dealing with uncertainty during restocking or replenishment. This safety stock inventory is given by

$$I(t)_{C3} = \left(S_{hybrid} - \left(\frac{\partial D}{\partial t} \right) * T_L(t)_{hybrid} \right) \quad (9)$$

Where S_{hybrid} is the target hybrid inventory level, $T_L(t)_{hybrid}$ is the consumption time or protection demand interval. Hence the instantaneous hybrid inventory held

$$I(t)_3 = I(t)_{A3} + I(t)_{B3} + I(t)_{c3} \quad (10)$$

Inventory backordered $B(t)_{hybrid}$

It should be noticed that the hybrid backorder inventory can be computed by applying the same principle as for the Periodic *(R, S)* and Continuous *(r, Q)* inventory review systems. Therefore, the general expression of the instantaneous backorder inventory per time can be given by.

$$B(t)_{Hybrid} = \int_0^T (B_{bo.hybrid} * T_{B.hybrid})dt \quad (11)$$

Where $B_{bo.Hybrid} = \frac{1}{period} \int_0^T \left[\frac{N_{bo.hybrid}}{N(t)_{Hybrid}} \right] dt$, (12)

Where $B_{bo.hybrid}$ is the hybrid backorder rate, $T_{b.hybridt}$ is the hybrid time length of the backorder, $N_{bo.hybrid}$ is the practical number of back order per time and $N(t)_{hybrid}$ is the ideal number of order per period under a hybrid review policy, $T_{bo.hybrid}$ (t) is the time length for the hybrid backorder during a period. Then, $T_{bo.hybrid}$ (t) can also be given by

$$T_{bo.hybrid.}(t) = MLT(t)_{BO\,hybrid} + PROD(t)_{BO_{hybrid}} \quad (13)$$

Where $MLT(t)_{BO.hybrid}$ is lead time for backorder, $PROD(t)_{BO.hybrid}$ is production time for backorder. The value of the decision variables *r, R, S* and *Q* can then be obtained by solving appropriate differential equations.

Energy used is perceived as constraint for most manufacturing plant performances. Since it was demonstrated that the implementation of a hybrid inventory model can bring more value in most systems, the question now is to understand its implication on other performances. Note that energy consumption's behaviour and structure within most manufacturing plant can be analysed differently. In this case, one had to figure out what the true dynamic of energy is with respect to inventory. How to put it in way that most managers could understand what the true dynamic of energy is? So, part of the development undertaken in the next sections is dealt with that question.

2.3 Properties of the energy model

It is almost impossible to determine the energy consumption pattern of electric equipment without a good understanding of the power theory. So, the term power theory of electrical installation can be understood as the state of knowledge on their power properties. In that sense, it is a set of true statements, interpretations, definitions and equations describing these properties [15]. Such equation of power in given as follows [16] [17]:

$$P = \frac{1}{T} \int_0^T p(t)dt = V_{RMS} * I_{RMS} \therefore P = S \cos\varphi \quad (14)$$

Where V_{RMS}, I_{RMS} are the effective value of the varying voltage and current, then *S* is the apparent power, $\cos\varphi$ is the phase difference between the current and voltage drop across the load.

2.4 Empirical part of the model (model estimation and empirical study)

For instance, energy consumption that results from implementing a hybrid inventory model within a manufacturing plant hang in the balance, it is uncertain. This may be due to the random nature of production rate, which in turn is linked to the lot-size and lead-time. Hence it is important for production rate to remain consistent with the inventory model's parameters, and it can be done by carefully handling the random behavior of lot-size and lead-time. In this research paper, mass balance (inventory level)for unsteady state process is dealt with in order to lower or keep energy consumption as much steady as possible. So, in term of interpretation, this energy consumption is then developed as follows:

$$E(t) = \sum P * M_{LT} = \sum_{i=1}^{\infty} P * \frac{1}{Th_r} * Q(t) \qquad (15)$$

Where Th_r is the theoretical processing rate (speed for actual product output), P is the electrical power consumed by the equipment of the production line, $M_{Lt}(t)$ is the manufacturing lead time, and Q is the lot-size of implementing a hybrid inventory model. Equation (15) was inserted into the energy balance equation that result from producing item, which is meant to represent what could had happened to the plant by looking at the difference between the new and old value of energy. A possible model form of energy balance in this case is then given by

$$\frac{dE(t)}{dt} = \frac{d\left(P * \frac{1}{Th_r} * Q(t)\right)}{dt} = P * \frac{1}{Th_r} * \frac{dQ(t)}{dt} \qquad (16)$$

Note that equation (16) is seen as a non-linear differential equation. From the aforementioned equation, the basic idea is to make a change of variables in order to come up with a different form of the lot-size by reshaping it. In general, the temporal variation of energy consumption is defined by

$$\frac{dE(t)}{dt} = \frac{E(t_2) - E(t_1)}{t_2 - t_1} \qquad (17)$$

Where dE/dt is the variation in energy consumption, $\boldsymbol{E_2}$ is the energy consumed during the second period $\boldsymbol{t_2}$, $\boldsymbol{E_1}$ is the energy consumed during the first period $\boldsymbol{t_1}$. Note that t_2 and t_1 describes the fluctuating lead time. Equation (16) is equated to equation (17) in order to have a clear expression of the lot-size (Order quantity Q) pattern. This led to the following lot-size balance for the lot-size in the plant:

$$P * \frac{1}{Th_r} * \frac{dQ(t)}{dt} = \frac{Th_r}{P} * \left[\frac{E(t_2 - E(t_1))}{t_2 - t_1}\right] \qquad (18)$$

Observe that the proposed hybrid inventory model is governed by a continuous and discrete dynamics. In other word, the hybridization takes place because of the dynamical switching process from continuous to discrete. Of the many basic modeling approaches for hybrid system, an explicit approach is used to describe and control the evolution of the lot-size Q. This explicit approach is described by cautious equations that lead to a system of differential equations.

$$\frac{dQ(t)}{dt} = \rho * E(t_2) - \rho * E(t_1) \qquad (19)$$

The lot-size as described by equation (19) is subject to change that depends only on initial state (before) and final state of the energy consumption, but not on the manner used to realize the change in energy consumption. One could then think about the manner used to realize such change from the initial to the final state of energy consumption. Equation (6) becomes:

$$E(t_2) - E(t_1) = \frac{1}{\rho} * \frac{dQ}{dt} * (t_2 - t_1) \qquad (20)$$

Where $E(t_2)$ is the new value of energy consumption under a continuous inventory model, $E(t_1)$ is the old value of energy consumption under the hybrid inventory model, φ is the production rate, t_2-t_1 is the change in time. It may be that there is evidence that the instantaneous rate of change at time t is actually a function of previous time. This is referred to as a delay differential equation. Equation (20) logically leads to a continuous change in energy.

$$\frac{dE}{dt} = \frac{1}{Th_r} * dQ * (E_2 - E_1) = \alpha * E \tag{21}$$

Equation (21) is the energy balance equation and represents the increase in total energy as result of change in time by one unit. It shows that the new value of energy consumption is dependent of the old value of energy consumption and the effect of variation of the lot-size. In other word, energy balance is the relationship between new value of energy consumption and old value of energy consumed. In this case, the difference between *$E(t_2)$ and $E(t_1)$* shows up as the change in energy consumed by the plant.

2.5 Solving the mathematical problem

Equation (21) can be rearranged by approximation. Of course, additional terms can be added that modify the evolution of the change in energy consumption. A possible approximation of such change in energy consumption over time is obtained using the Taylor series method as follows.

$$E(t) = \alpha_1 * E_T^{\ 1} + \alpha_2 * E_T^{\ 2} + \alpha_3 * E_T^{\ 3} + \ldots\ldots\ldots \alpha_n * E_T^{\ n} \tag{22}$$

The parameter α meant for the flux rate can then be set as follows:

$$\alpha = C^{te} \therefore \frac{dE_T}{dt} = \left[\alpha_1 * E_T^{\ 0} + 2 * \alpha_2 * E_T^{\ 1} + 3 * \alpha_3 * E_T^{\ 2} + \ldots\ldots\right] * \frac{\partial E_T}{\partial t} \tag{23}$$

$$\alpha \neq C^{te}$$

$$\frac{dE_T}{dt} = \left[\frac{d\alpha_1}{dt} * E_T^{\ 1} + \alpha_1 * \frac{dE_T^{\ 1}}{dt}\right] + \left[\frac{d\alpha_2}{dt} * E_T^{\ 2} + 2 * \alpha_2 * E_T * \frac{dE_T}{dt}\right] + \left[\frac{d\alpha_3}{dt} * E_T^{\ 3} + 3 * \alpha_3 * E_T^{\ 2} * \frac{dE_T}{dt}\right] + \ldots\ldots\ldots \tag{24}$$

Important input data and parameters of such hybrid inventory model were collected from archival record. important parameters needed for this study are represented as follow: The ordering cost *k=R125 per order*; The holding cost *h=R0.511 per unit per time*; The back order cost *j=R201per unit per time*; The service level was set to 90%

3. SIMULATION OUTPUT RESULTS

In this section, a simulation/experiment was performed on the constructed mathematical models. Note that Engineering Equation Solver (EES) is the software used to run the energy models. The simulation output results are consolidated in data reports and presented by graphs. The graphs obtained from simulation studies are useful in that they can raise questions, which in turn stimulate further investigation.

3.1 First output result

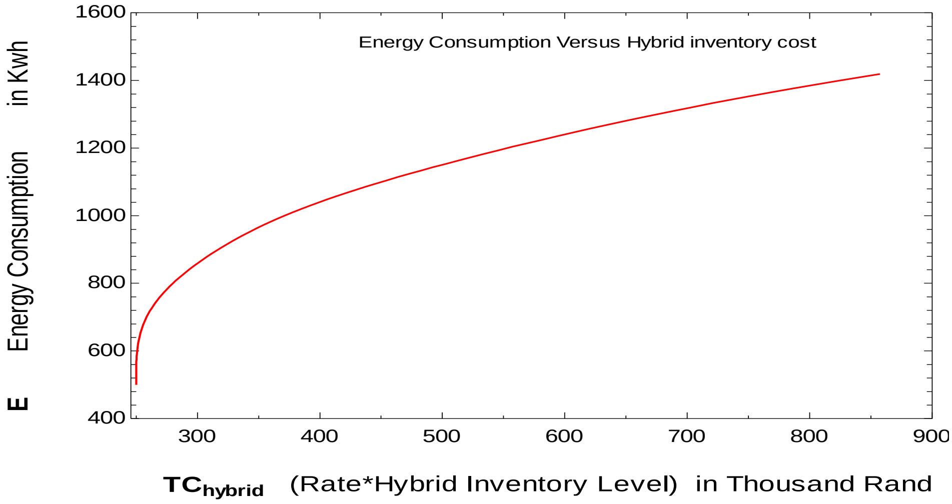


Figure 2: Energy consumption versus hybrid inventory cost

It is noticed an unusual correlation between energy consumption and hybrid inventory cost. Pointing this fact out under this section is perceived as the starting point of an effective analysis. It is observed a particular path of energy consumption, which of course does not, tends to infinity really fast but rather is subject to stabilization. One shouldn't worry about such type of energy behavior for at least many reasons. Notice on the graph 1 that as the hybrid inventory cost increases, the energy consumption also increases. It is clear that such energy consumption function is an increasing function. Since an increasing function is observed (a one to one relationship), it is destined to have a unique inverse. It is also observed on figure (1) that an increase in the value of energy consumption is really considerable between two point of the inventory cost (300 and 400 thousand rand). After the hybrid inventory cost that equate 400 thousand rand, as the hybrid inventory cost gets larger and larger, the increasing evolution of energy consumption start slowing down. The aforementioned relationship between energy consumption and hybrid inventory happened for an important reason. The presence of the logarithmic growth of energy consumption may come from the fact that the production rate increases in a -fixed percentage. Such event must compels one to understand what could happen if the production rate does not increase at a fixed percentage. The product that is manufactured or processed is made up of some type of material, which of course may impact the speed. A reasonable answer could be obtained from signs. For instance, a visible sign of plant ineffectiveness may be perceived as the impact of production rate change. Hence it is crucial to handle the speed at which the products are produced.

A logarithmic growth of energy consumption like that shown in figure (1) is often not well understood by the decision makers. In two important points (300 to 400), the amount of energy consumed has significantly increased. The logical consequence of such energy consumption pattern is that it may get out of hand over time and affect the hybrid inventory cost. Such logarithmic growth of energy consumption is then justified by the fact that demand and lead-time fluctuation have significant implication on production rate, which in turn is linked to energy consumption. In this particular case, the lead-time may be mostly impacted by the downtime machine. Problems like that become easy to be recognized, because of the large number of activities in the production process [18]. For example extended lead-time (made up of value and non-value activities) can lead to poor throughput rate, which in turn has a direct impact on the economic performance/effectiveness. Mostly, the value activities are lot-size related. Knowing the link that exists between lot-size and lead-time, it is crucial to focus more on value activities. This is a clear indication that a proper management of production rate through the lot-size and lead-time management is capital. Otherwise, it may have negative implications on energy consumption in the long run (see figure 3). Although this research paper is able to provide some insights, it is acknowledged a limitation. For instance, it is used various adjustments of the model parameters and forcing within the range of the uncertainties without emphasizing on the right number of replications. However, much could have been done if an appropriate number of replications were done. Future research should include model calibration with multiple and appropriate replications.

3.2 Second output result

In this section, it is used a 3Dimensional graph scatter plot to see how the production rate relate to the hybrid inventory policy that is implemented and the energy consumption. This three Dimensional graph is used to explore three important aspects: the strength; direction; and non-linearity of the relationship between the three aforementioned variables.

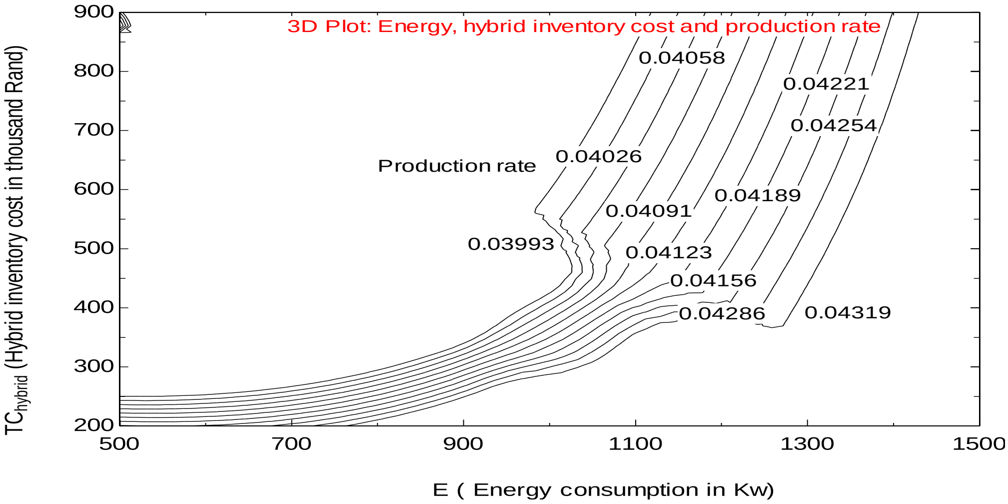


Figure 3: Impact of production rate on energy consumption and hybrid inventory cost

As can be seen from figure (2) it is observed a system with a production that seem to keep a constant path until an unexpected even happen. Most goods have mechanical properties that must be taken into consideration during the manufacturing process. The amount of time required to produce one unit of good may then be set accordingly because it can reflect on the system flow rate. The same level of activity indicates that such system is moving on an easier pace (move at a particular pace). That means, it is working much efficiently. However, as the production rate changes suddenly, it might seriously impact on the energy consumption and hybrid inventory cost. Studies for the use of energy in production showed that different production rate can generate different levels of energy consumption [19] [20]. Manufacturing firms have recognized this trend and are now trying to implement corporate strategies so as to improve energy efficiency [19].

Recall that figure (2) portrayed a hybrid inventory cost that tend to follow an unusual path with respect to energy consumption, but is permutated by the change in the production rate. Further, figure (2) showed that the pattern of energy consumption versus a hybrid inventory model is totally affected by the speed at which the items are produced. The aforementioned path is unstable as result over time. Launching a production with low speed resulted in reduced energy consumed while the hybrid inventory cost increased. However, having a production line that performs at a high speed compelled the production line to have high energy consumption while the hybrid inventory cost was low. Future research should then emphasize on the balance between energy consumption and production output.

3.3 Third output result

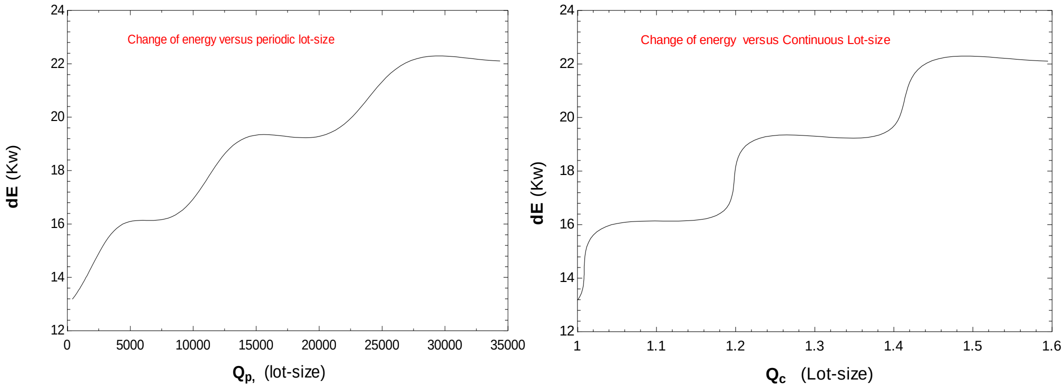


Figure 4: impact of the lot-size on the change of energy

It is presented a graph describing how the level of change of energy had increased with respect to the lot size. The step function in this case is found useful in representing the data as the nature of the change of energy is not naturally continuous. The energy function is increasing on the interval (0,500), constant on the interval (800,1000) and increasing again on the interval (1000, 1100). This function appeared to be a step function, which is one kind of piecewise function defined by constant value over each part. Note that a conclusion that is drawn from the graph portrays that step graph can describe both change of energy and discrete nature of the change due to sometime the constant movement of the lot size.

4. MODELS VERIFICATION AND VALIDITY

The independent models are run such that artificial data set (figure 2) be generated and be checked against the real data that were seen on figure (1). In other word, it is checked whether the same pattern of energy consumption versus time emerged after running the models. As can be seen from real data (figure 1) and simulation output result (figure 2), the same pattern of energy consumption versus Hybrid inventory cost almost emerged. Note that such hybrid inventory cost model is function of time. In this case, there is a relation between energy consumption and hybrid inventory cost, which in turn is related to time. Thus, energy consumption is related to time. Such equality is a transitive relation, which of course compelled one to say with a certain degree of confidence that the hypothesis/theory was validated. It should be further noticed that some results that confirm the validity and verification of the proposed models are displayed on figure 2 and 3.

5. CONCLUSION

Energy consumption is an interesting property of most manufacturing plants and items produced might affect the occurrence of energy consumption. Managers should be aware of the risk factors for higher energy consumption, which may be proven useful in identifying items produced at high risk for energy consumption peak. This can help one to minimize the risk of higher energy consumption through the avoidance of all the unnecessary activities that may cause different type of breakdown machine (or stoppage), which in turn are tied to the production rate. In other word, everything has to be done at an easy pace. That means it is happening efficiently. A system with the right vision will have higher efficiency than the others.

6. REFERENCES

[1] Kotler, P. 2002, Marketing Management: The Millenium Editio*n,* 2nd Ed, New Delhi: Prendice Hill Of India

[2] Kumar. J & Pradeep, K.S. 2013. Survey Of Demand Fluctuation Of Dairy Products At Raipur Dugdh Sangh, Chhattisgarh, India For Selection Of Appropriate Forecasting, *Uncertain Supply Chain Management*, 1 (2013) 133–144

[3] Chalotra, V, 2013, Inventory Management And Small Firms Growth: An Analytical Study In Supply Chain. Vision, *Sage Journal*, 17(3), 213-222

[4] Isaksson, O.H.D. & Seifert, R.W, 2014, Inventory Leannes And Financial Performance Of Firms. Production Planning & Control: *The Management Of Operations Journal*, 25(12), 999-1014

[5] Anichebe, N.A & Agu,O.A. (2013), Effect Of Inventory Management On Organizational Effectiveness, *Information And Knowledge Management*, 3 (8), 92-100

[6] Mohamad, S.J & Suraidi, N.N & Abd.S.N & Suhaimi,R..D. 2016, A Study On Relationship Between Inventory Management And Company Performance: A Case Study Of Textile Chain Store, *Journal Of Advanced Management Science*, 4(4), 299-204

[7] Mpawanya, F, 2005, Inventory Management As A Determinant For Improvement Of Customer Service*, Master Thesis, Pp 23-24*

[8] Sietienei, E & Memba, F. 2015, The Effect Of Inventory On Profitability Of Cement Manufacturing Companies In Kenya: A Case Study Of Listed Cement Manufacturing Companies In Kenya, *International Journal Of Management And Commerce Innovations*, 3(2), 111-119

[9] Harris, F. 1913. How Many Parts To Make At Once, *Factory, The Magazine Of Management*, 38(6), 947–950.

[10] Benga,E.B.B. 2015. Optimizing The Inventory Ordering Policies In Supply Chain Management, *Thesis Submitted In Fulfilment Of The Requirements For The Degree Magister Technology In Industrial Engineering . 52-54-16*

[11] Lunze,J & Lammabhi-Lagrigue, F. 2012. Even Based Control With Communication, *22 Osilar Conference On Signal & System And Computers*

[12] Perona, M & Saccani, N & Zanoni, S . 2009. Combining Make-To-Order And Make-To-Stock Inventory Policies, *An Empirical Application To A Manufacturing Sme Marco*

[13] Neshat, N & Amin-Naseri, M.R. 2014, "Energy Models: Methods And Characteristics" Journal Of Energy In Southerrn Africa, 25(4), 101-110

[14] Momeni, Z & Azizi, A. 2018. Current Order And Inventory Models In Manufacturing Environments: Areview From 2008 T0 2018, *International Journal Of Mechanical And Production Engineering Research And Development,* 8(1) 1087-1096

[15] Seitbmetz, C.P. 1997. Theory Of Alternating Current Phenomena, New York

[16] Czamecki, J.S.1997. Power And Compensation In Circuits With Periodic Voltage And Currents, *Outline Of The History Of Power Theory Development*, 3(2), Pp 37-46

[17] Schulze, S & Seuring,S Ewering,C, 2012. Applying Activity-Based Costing In A Supply Chain Environment, *Int. J. Production Economics*, 135(1), Pp. 716-725,

[18] Kumar,R. & Sundar, R & Balaji, A, 2014. A Review On Lean Manufacturing Implementation Techniques, *Procedia Engineering*, 97(1), Pp. 1875-1885.

[19] Fysikopoulos, A & Alexios, P & Georgios, S & George, C, 2013. Energy Efficiency Of Manufacturing Processes: A Critical Review, *46 CIRP Conference on Manufacturing Systems*.

[20] ROSSETTI, M.D. 2010. Simulation Modeling & Arena. 1st ed. Hoboken: John Wiley & Son,Inc

[21] Liang, Y. 1997. The Development Of An Intelligent Inventory System, *Thesis, Usur*

PROPOSE A CONCEPTUAL PROGNOSTIC MODEL BASED ON THE CONDITION OF RAILWAY VEHICLE MOVING COMPONENTS

J. Coetzee[1*] & L. Pretorius[2]

[1]Research Group for Systems Energy and Innovation
University of Pretoria, South Africa
bing.elek@gmail.com

[2]Research Group for Systems Energy and Innovation
University of Pretoria, South Africa
pretoriussystems@gmail.com

ABSTRACT

The aim of all companies is to increase their profitability and it is no different for the railway environment. Possible steps toward obtaining this goal includes reducing costs, increasing productivity and efficiency. One focus area within the railway technical environment that can facilitate in obtaining these goals is maintenance. Even though conventional strategies such as corrective and preventive maintenance have proven to be lifesaving in some cases, it does not always adhere to the cost saving and process efficiency drives by companies. A new strategy of predictive maintenance is being investigated which is based on forecasting of system degradation through a prognostic process by using condition monitoring data. The objective of this paper is to propose a prognostic model which will be used for predicting the most cost and time effective maintenance schedule for a railway wagon based on the condition of its bogies, wheels, bearings and brake blocks.

*Corresponding author

1. INTRODUCTION AND RESEARCH QUESTIONS

For any freight moving company, transporting goods from A to B at the highest profit margin and at the lowest risk is the name of the game. Railway companies are looking towards the maintenance department to assist in obtaining these goals. They must ensure that the locomotives, wagons and infrastructure are in the optimal operating condition in order to deliver on the commitments made to the customer.

From a strategic viewpoint, is there a way to support the maintenance department in obtaining the company's set goals? The researcher identified focus areas and questions, when answered will provide the maintenance department with a model that will empower them to optimally identify system failure with minimum downtime and maximum cost saving as a reward.

- What is the current maintenance strategy applied by the company?
- What is predictive maintenance for railway rolling stock?
- What are the advantages to the company when pursuing this strategy?
- What instruments and methodologies are available to assist in developing a predictive model?

Currently a lot of time, effort and money are being spent on maintenance strategies where equipment is only being repaired after catastrophic failures or where scheduled maintenance is applied. These two strategies have a negative impact on the cost saving and improved safety objectives of the company due to high spare part and labor costs as well as possible train derailments. A new perspective on the current maintenance strategies is needed in order to obtain the required objectives. One such method is to revisit the maintenance strategy. Recently a new movement has developed of applying a predictive maintenance strategy started. Predictive maintenance is being used to predict the time of a potential system or component failure by using the system's condition monitoring data. This leads to only performing maintenance actions when it is needed which leads to less downtime and reduced spare part costs.

Modern-day railway companies also see advancements in technology. These advancements include condition monitoring systems. Condition monitoring systems are being used to monitor the real-time condition of rolling stock as well as the infrastructure. Examples of rolling stock and infrastructure condition monitoring systems are:

- Wheel bearing monitoring systems
- Wheel profile monitoring systems
- Skew bogie detection systems
- Brake block condition monitoring
- Ultrasonic broken rail detection systems
- Systems that measures the long stresses in the rail

Currently, the data generated by these systems are being used to generate stop train alarms which could lead to a prevented derailment and in some cases for trending purposes. One example where trending led to discovering an inherent fault within the wheel bearing monitoring system; the system reported the same temperature values for all wheels on the left side of the train. This is impossible in real life since the bearings are not the same age and did not travel the same distances. The fault was reported to the manufacturer and was rectified.

This example confirms the value of condition monitoring data and it is natural to take the next step to apply modeling techniques on the data to predict equipment failures and allow the opportunity to take preventative actions.

The aim of this paper is to encourage the implementation of a predictive maintenance strategy as well as choosing a prognostic model best suited for the available data to ensure accurate predictions to support cost-saving initiatives and safety improvements in the railway environment.

2. LITERATURE SURVEY

2.1 Maintenance

If you have two islands with one called technology and the other called profitability you will have to connect the two with a bridge called maintenance. Choosing an optimal maintenance strategy could lead to reduced costs as well as increased productivity and efficiency without compromising on safety.

In today's day and age, equipment is designed to run at optimal levels and in supercritical environments. Catastrophic failures in these cases usually lead to major cost implications whether it is the replacement of critical components or production downtime. In the case of the railways, a catastrophic failure might be the derailment of a train. When a train derails inside a tunnel it could cause the line to be closed for a week or even more and the loss of revenue has a huge impact on the profits of the company. It is for this reason that maintenance is becoming more vital than ever before. Good maintenance leads to more reliable systems and an increased safety environment.

Maintenance is defined by [1] as follows; "*Actions necessary for retaining or restoring a piece of equipment, machine, or system to the specified operable condition to achieve its maximum useful life*."

Maintenance can broadly be classified into two categories namely planned and unplanned. These two categories can be further divided into three major maintenance strategies [2]. They are; *Corrective, Preventative and Predictive Maintenance*.

Figure 1 below gives a general indication of where the different types of maintenance strategies fit into the lifecycle of a component or system.

When new equipment is introduced into an operating environment it initially has a high failure rate that decreases with time usually caused by factors like poor design and incorrect installations. Corrective maintenance is usually associated with the infant period. During the useful life period, the equipment displays a constant failure rate with random failures occurring. During this period the preventative maintenance strategy is being followed. As the equipment nears its end-of-life or wear-out phase it displays an increased failure rate and is caused by amongst other things material wear and fatigue. The predictive maintenance becomes the preferred strategy during this period [3].

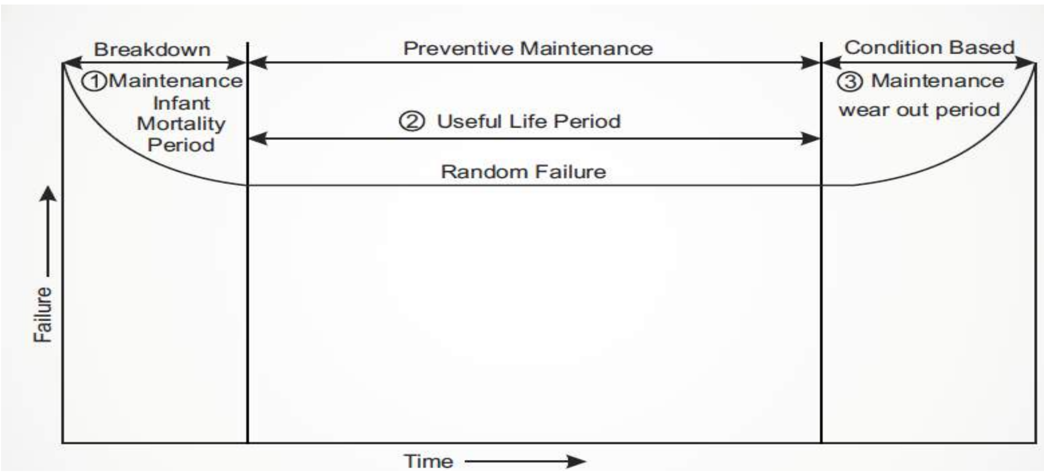


Figure 1: Bathtub curve [3]

2.1.1 Corrective Maintenance

Categorised under unplanned maintenance, Corrective Maintenance also known in some circles as the "run-to-breakdown" strategy takes place after equipment/component failure or breakdown with the aim of restoring the asset to its intended operating condition. This can be realised by replacing the complete asset or by repairing sub-components of the asset. The corrective maintenance strategy entails that no scheduled actions should be taken to maintain any equipment or part thereof before it breaks down. Like everything in life, this strategy has advantages and disadvantages associated with it [4] and are listed below:

Advantages

- There are lower short-term costs associated with it in the sense that no actions need to be performed before a failure occurs. This also cuts down on travel time and costs.
- Real-time alarming is usually the method used to trigger awareness of the failure or breakdown and this drastically minimises maintenance planning.
- There are no complicated processes related to this strategy, reaction is only required when a problem occurs.
- In some non-safety critical cases, the cost involved in only maintaining after a failure is much less than preventative maintenance actions.

Disadvantages

- Since the condition of the equipment is not monitored, a failure can be very unpredictable.
- Unexpected failures are very often associated with the unavailability of spares and this leads to increased downtimes.
- The lifetime of the equipment drastically decreases due to the fact that the equipment is not kept at its optimal working condition.

- Catastrophic failures can have high costs associated with it but it can also carry serious safety risks and poor customer satisfaction with it.

One example from the railway infrastructure perspective is when a break in the rail occurs. In this case, corrective maintenance includes closing the line and repairing a piece of track or even clearing and fixing derailed rolling stock. In all scenarios, it is evident that it involves huge cost implications due to operating downtime and possible loss of life.

In conclusion, this is often the most expensive option but if a company wants to pursue the corrective maintenance strategy they should ensure that they have a highly skilled maintenance team with immediate access to all spares required to repair the faulty equipment and keep the downtime to a minimum. It can be deduced by evaluating the above-mentioned reasoning that an alternative maintenance strategy should be investigated that can more comprehensively support the company’s cost-saving initiatives.

2.1.2 Preventative Maintenance

Moving from the unplanned to planned maintenance category, Preventative Maintenance aims to prevent catastrophic failures, safety violations and extended downtimes [2]. This can be realised by replacing worn components before they fail. This maintenance strategy is also called Scheduled Maintenance and understandably so. In order to know when to replace a component before a failure occurs, it is necessary to plan for it. There might be a servicing period suggested by the manufacturer of a specific component. An example of this might be that the bearing on a railway wagon must be reconditioned every 2 years, this implies that the maintenance department must keep track of that wagon and set up a schedule for it to be serviced before two years has passed. This strategy has as an outcome improved efficiency of equipment, extended equipment life and it minimises unexpected failures.

Advantages of Preventative Maintenance [5] includes the following:

- The downtime caused by equipment breakdowns can now be significantly reduced
- Since the equipment is now running close to its optimal running condition, the life expectancy of the equipment is increased
- Overtime due to major repairs caused by catastrophic failures can now be reduced
- It improves the safety conditions for the personnel working in the vicinity of the equipment

Disadvantages of Preventative Maintenance [6] includes the following:

- It can become costly due to frequent traveling to and from a site as supposed to just waiting for components to start failing
- The risk of over maintenance is imminent, components might not be assessed as frequent as planned
- The workforce will increase with this strategy due to the regular equipment checks instead of a dedicated team only in the case where a failure occurred
- Increased costs due to the frequent replacement of expensive parts before their end of life

In the railway environment, the preventative maintenance strategy is implemented by means of a monthly scheduled task for cleaning and servicing the trackside condition monitoring systems. The advantage is that the condition monitoring systems have a high availability and it can be relied upon to give stop train alarms that can prevent derailments. The disadvantage, on the other hand, is on remote lines that stretch across the country, the traveling distance between the depots and the different measurement sites is vast and that has a huge impact on the traveling and overtime budget of the company. The average derailment cost amounts to approximately R20 million.

Summarising this strategy; the occurrence of catastrophic failures is drastically reduced but as a result, it could lead to unnecessary maintenance. Preventative maintenance in some cases are deemed to have the highest maintenance costs involved but has reduced downtime as a countermeasure. Once again, when evaluating the preventative maintenance strategy information and taking into account the valuable data that is available from the condition monitoring systems, it is evident that an alternative maintenance strategy must be investigated that fully support the company’s cost-saving objectives.

2.1.3 Predictive Maintenance

In recent times a great deal of equipment gets delivered with some sort of condition monitoring system accompanying it. With the Internet of Things (IoT) and storing all data on the “Cloud”, it becomes easy to keep an eye on all assets. All these technological advancements gave birth to Predictive Maintenance, sometimes called condition-based maintenance. This strategy uses the condition monitoring data obtained from the equipment and combines it with analysed historical trends [2] also taking into account the system degradation. By doing this, a predictive model can be developed and used to continuously monitor the health of the system and predict equipment failures or breakdowns before they happen. As with the previous maintenance strategies, predictive maintenance also has inherent advantages and disadvantages [7].

Advantages:

- The operational lifetime of the equipment increases due to the fact that it is constantly monitored
- The downtime caused by catastrophic failures can now be eliminated by taking preemptive corrective actions which in turn lead to an increase in system availability
- Costs related to labor and parts gets reduced due to the fact that maintenance actions are only actioned when required and not per regular scheduling
- This strategy allows the company to sweat its assets and as a result increases profitability

Disadvantages

- Upfront capital investment is necessary to procure condition monitoring equipment that is capable to extract diagnostic data from functioning equipment
- It is necessary to invest in skilled personnel and constant staff training
- The savings potential after the upfront investments made to establish this strategy is not always seen by management

Figure 2 depicts a predictive maintenance strategy. Predicting a possible equipment failure is just one gear within the maintenance gearbox. Prediction is being driven by determining the components and factors that could cause equipment failures and in turn, it drives the action to be taken to avoid any failures and major costs implications.

Figure 2: A model for the predictive maintenance strategy [14]

Predictive maintenance makes use of predictive analytics. Predictive analytics is a field of statistics that extract information from data and use it to predict trends. It makes use of techniques like data mining and machine learning to achieve that goal. In principle, predictive analytics captures relationships between predictor parameters and data from past occurrences and through algorithms predict the unknown outcome. In the case of predictive maintenance, it is predicting the time to failure of a system or component.

In closing, predictive maintenance enables the company's maintenance department to be proactive and make maintenance decisions based on the condition of their equipment instead of regularly scheduled maintenance and running the risk of over maintenance. Hitachi makes the following statement; "*Predictive maintenance could reduce cost by 10 to 40 percent and reduce downtime by 50 percent*." [16] In the rail environment this could mean that instead of bringing in a wagon for a routine two-year service that can on average amount to R24 000 per wagon, it might still be able to run for an additional year. For an asset base of 90 000 wagons; presume that on average 20 000 wagons are scheduled for maintenance in a particular year. If through applying a predictive maintenance strategy, 8000 wagons was predicted to only be serviced the next year, it would lead to a saving of R192 million. Any company will be inclined to enforce a maintenance strategy that could have such a cost-saving implication.

2.2 Prognostics

Predicting a type of failure and the time at which it might happen with a certain confidence level forms a critical part of the predictive maintenance strategy. The term generally used to describe the time between the evaluation starting time and the time of failure is Remaining Useful Life (RUL) [8]. This can be confirmed in Figure 3 below. Time (t_{now}) is the evaluation starting time and the red arrow identifies the time of a component failure. Remaining Useful Life estimation is central to the research field of prognostics and health management. It is estimated from data obtained from the condition monitoring systems. Factors that have an influence on the RUL of a system includes its system health and operating conditions.

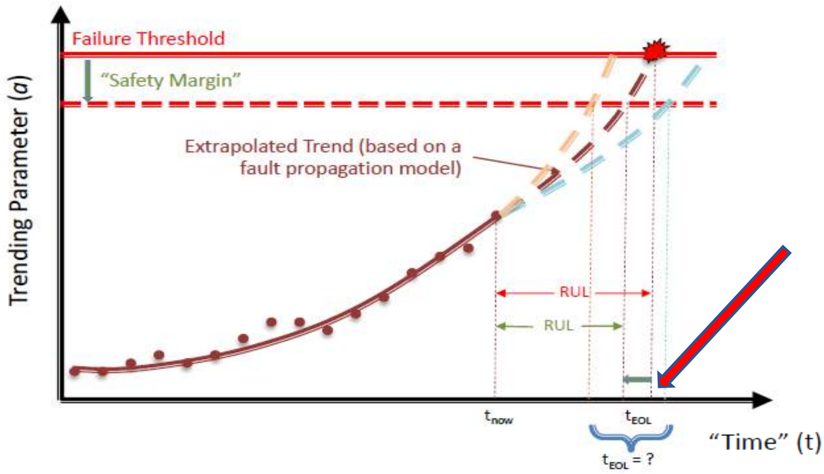


Figure 3: Graphical representation of Remaining Useful Life

The prediction of an outcome in the healthcare fraternity is referred to as prognosis. Prognosis is defined as: "*Prognosis (Greek: πρόγνωσις "fore-knowing, foreseeing") is a medical term for predicting the likely or expected development of a disease, including whether the signs and symptoms will improve or worsen (and how quickly) or remain stable over time; expectations of quality of life, such as the ability to carry out daily activities; the potential for complications and associated health issues; and the likelihood of survival (including life expectancy).*" [9].

Prognosis was adopted into the engineering field and customised to form the discipline known as Prognostics. Prognostics is defined as: "*Prognostics is an engineering discipline focused on predicting the time at which a system or a component will no longer perform its intended function.*" [8].

Prognostics is a field that has made a positive impact on the system life cycle management of complex systems and components [10]. Some benefits related to this discipline include the following:

- It establishes a change in maintenance strategies, from corrective to predictive
- It reduces the spares inventory, a just-in-time strategy can now be followed
- It leads to reduced downtimes by the prevention of catastrophic failures
- It does not only assist in sweating the assets but it can also monitor the deterioration rate and alarm on an asset that must be prematurely maintained

From the above, it is evident that prognostics is an extremely valuable tool for someone that needs to initialise a maintenance action. In order to fully realise its full potential, its output should be the input into a system health management program.

The approaches to building models in the prognostics field can be classified into three groups namely; data-driven, model-based and hybrid approaches [8].

3. INITIAL CASE APPLICATION FOR PROGNOSTIC MODELS

3.1 Model-based approach

Model-based also called Physics of failure-based prognostics is a frequently used approach. This approach involves physical models of the system or components under investigation. The physical model can either be developed by the researcher from first principles or obtained from the manufacturer of the system or component where the degradation and failure modes are described mathematically. The developer of the model-based approach needs to have a thorough understanding of the system and its operating conditions which includes workloads and the environment [10].

One example of a model-based prognostic approach is the growth of a crack through a metal plate. The rate of damage growth can be expressed using the Paris model (Paris & Erdogan, 1963) as $\frac{da}{dN} = C(\Delta K)^m, \Delta K = \Delta\sigma\sqrt{\pi a}$ where a is the crack size, N is the number of cycles, m and C are damage model parameters, ΔK is the range of the stress intensity factor, and Δσ is the stress range [11].

This methodology is known for its accuracy and precision but it depends on model fidelity. The advantage of this approach is that it is easily validated but with the limitation of being time-consuming and costly [10].

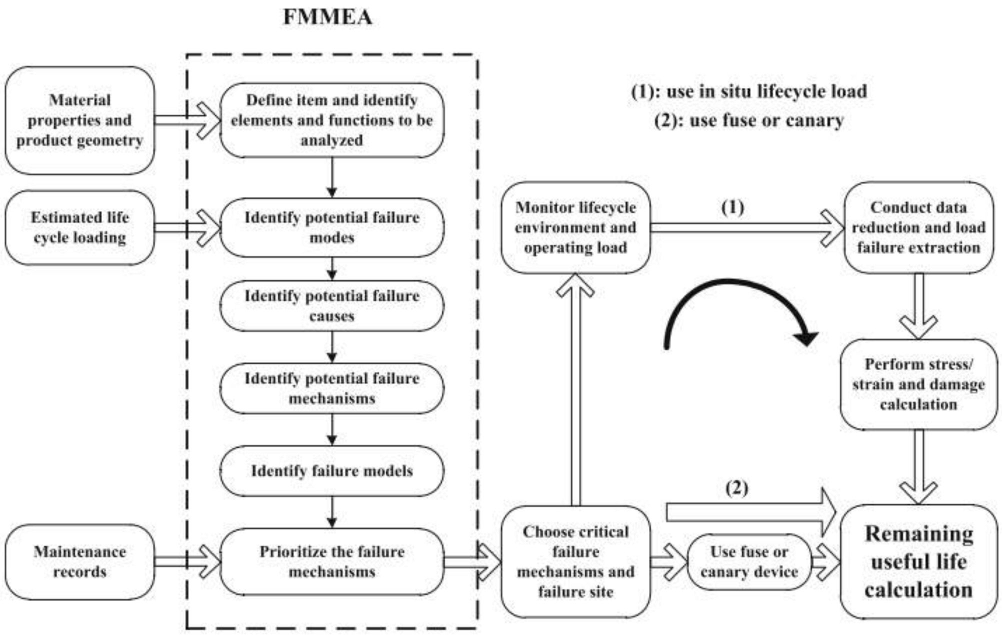


Figure 4: Model-based prognostics methodology [10]

3.2 Data-driven approach

As the name states, the data-driven approach requires having a large amount of multivariate historical system or component data for normal and degradation scenarios. This approach is widely used when difficulty exists in obtaining the system or component degradation model. This methodology is mainly used for systems or components which displays gradual degradation monitored by various sensors. Figure 5 below describes the data-driven prognostics methodology most commonly followed [12]. One reason for making this approach a preferred one is the ease with which it can be implemented and the speed of deployment. It makes use of condition data such as speed, temperature, volts, etc. The researcher needs no or little knowledge of the system to be able to develop a model. Machine learning and the data-driven approach goes hand in hand. Machine learning software and techniques are widely available with built in applications which ensure the ease of development and solving of data-driven algorithms. There are however factors that can cause difficulties for the data-driven approach and they are [10]:

- Overfitting as a result while training the algorithm
- It can be computationally intensive due to large datasets
- Run to failure data might be missing if it is a new system or component

There are however techniques available within the machine learning environment to cater for the challenges posed above. Thus taking into account the quick and cost-effective deployment and manageability of the difficulties, the data-driven approach still remains a preferred methodology.

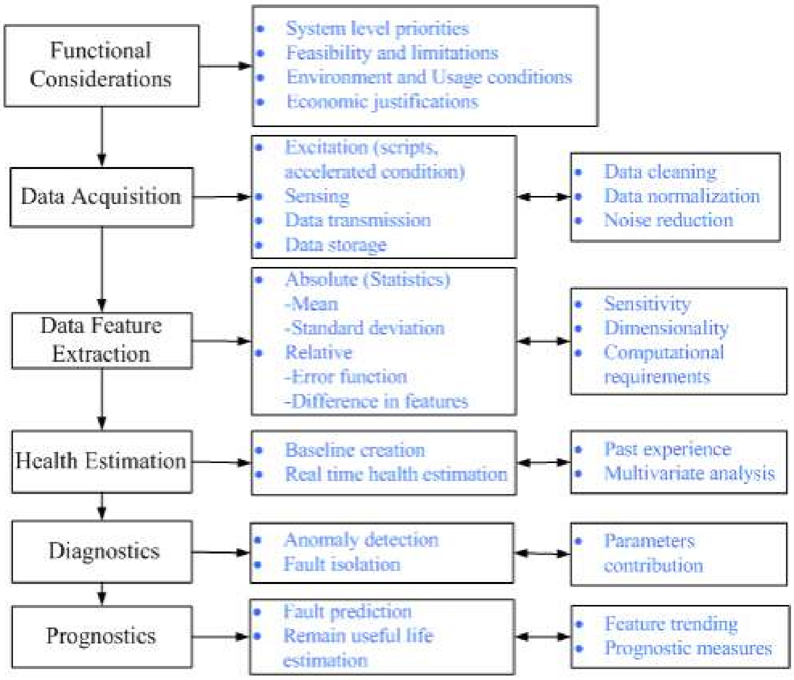


Figure 5: Data-driven prognostics methodology [12]

3.3 Hybrid approach

Both the model-based and data-driven approaches have their advantages, but combining them can even have more benefits. The approach when combining the two approaches is called the Hybrid or Fusion approach. The outcome of all approaches is to estimate the remaining useful life (RUL) of a particular system or component. The combination or fusion of the model-based and data-driven approaches can either be before the RUL estimation (pre-estimate) or after the initial RUL estimation where the individual approach results can be combined to estimate a final RUL (post-estimate).

The Kalman and Particle filters are frequently used for the implementation of this methodology. An example for illustration purposes; a guidance system calculates the exact point of impact but through years of experience the officer knows that by adjusting the cannon's aim with 1 degree to the left will assure a hit.

Figure 6 below describes the data-driven prognostics methodology most commonly followed [10].

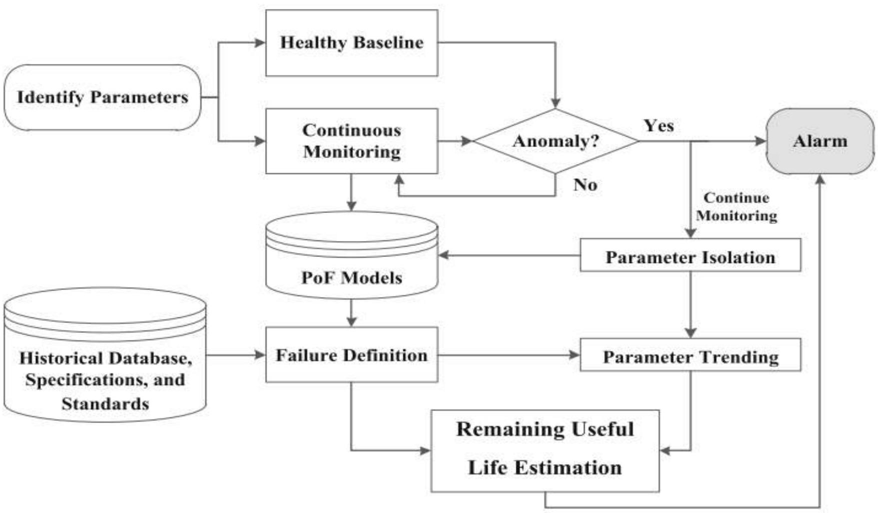


Figure 6: Hybrid approach prognostics methodology [10]

3.4 Failure Modes, Mechanisms and Effects Analysis (FMMEA)

In order to establish an accurate prediction for the remaining useful life (RUL) of an asset, it is necessary to know exactly what parameters has an influence on the condition thereof. The well-known method; Failure Modes, Mechanisms and Effects Analysis (FMMEA) will be utilised to assist in determining the prediction parameters of the system. Identifying possible failures and how it impacts the condition of an asset is the reason why this method is being used to determine the prediction parameters. It is necessary to know exactly what causes the failure in the asset and how it becomes visible [13]. To fully understand what caused the failure, it is essential to know the failure mode and the failure mechanism. Knowing this can assist in selecting the correct sensor for detecting and monitoring the particular condition. That sensor may already exist and positioned correctly or through experience, the operator can suggest a new sensor and or position. By not having the correct condition sensor in the correct position may cause the selection of prediction parameters to be unfitting. In turn, not having the correct prediction parameters will cause the selected model to provide inaccurate prediction results that could have cost implications caused by safety impacts or unnecessary maintenance actions. Figure 7 stipulates the steps to be taken during the FMMEA methodology.

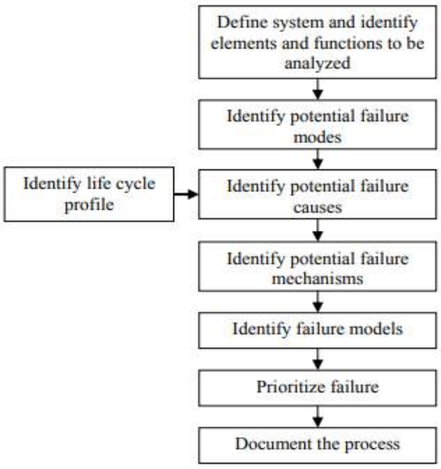


Figure 7: FMMEA Methodology

The FMMEA methodology was applied to a railway wagon and the result is shown in Table 1. The elements were deduced from identifying potential failures on a railway wagon. It is evident from the results gathered that condition monitoring systems are needed to monitor each wheel of the wagon, all wheel bearings of the wagon and the wagon's bogies. These three elements will become the prediction parameters for the prediction model.

Table 1: Result of the FMMEA methodology applied to a railway wagon

Element	Potential Failure Mode	Potential Failure Cause	Potential Failure Mechanism	Mechanism Type
Bearing	Axle turns off	Loss of lubrication	Lubrication did not recover original properties	Overheat
Wheel	Wheel loses contact with the rail	Tread wear exceeded the limit	Poor rail condition	Wear out
Wheel	Wheel loses contact with the rail	Flange height exceeded the limit	Wheel passed allowed kilometers	Wear out
Bogie	Derailment	Excessive lateral forces	Wheels grind inside of the rail	Wear out
Element	**Severity**	**Occurrence**	**Risk**	
Bearing	Catastrophic	Occasional	Very High	
Wheel	Catastrophic	Reasonably Probable	Very High	
Wheel	Catastrophic	Reasonably Probable	Very High	

Bogie	Catastrophic	Reasonably Probable	Very High	

Figure 8 is an example of a seized bearing, a stop train alarm was generated using condition data. The train was stopped in time and a derailment was prevented. The sweating of assets is not the only advantage of the predictive maintenance strategy, in this case, it would have given an early warning to service or replace the bearing before catastrophic failure.

Figure 8: Picture of a seized bearing

3.5 Initial condition data

The measurement system's physical sensor outputs get processed and stored in a central database. The researcher obtained raw data from the central database that addresses the identified prediction parameters. This data will be fed into the chosen prognostic model approach. These are the wheel bearings, the wagon's wheels and the wagon's bogies. The data displayed graphically below is for a period of one year. The Y-axis depicts the relevant condition measurements. For the wagon wheel bearings they are; degrees Celsius (absolute temperature) and decibel (acoustic signature). For the wagon wheels they are; millimeters (tread wear and flange height). For the wagon bogie it is; tons (gauge spreading force). The X-axis depicts the number of measurements taken for the specific wagon during the selected time period per measurement system.

3.5.1 Wheel bearings

The first prediction parameter is the wheel bearings of the wagon. There are two measurement systems that were used to obtain condition data for the bearings. From the first system (Hot Bearing Detector System - HBD) the real-time absolute temperature can be obtained. The data for the past year for a specific wagon is shown in Figure 9. From the second system (Acoustic Bearing Evaluator - ABE) the real-time acoustic signature of the bearings can be obtained. A graphical representation of the past year's ABE data for a specific wagon is shown in Figure 10.

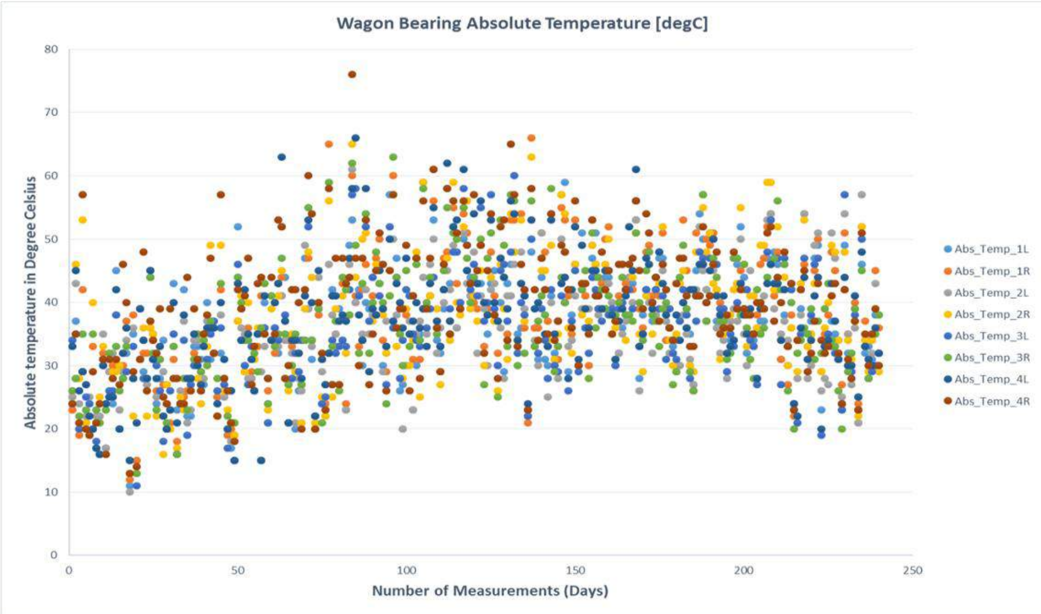


Figure 9: Raw data showing the absolute temperature for the wagon's bearings

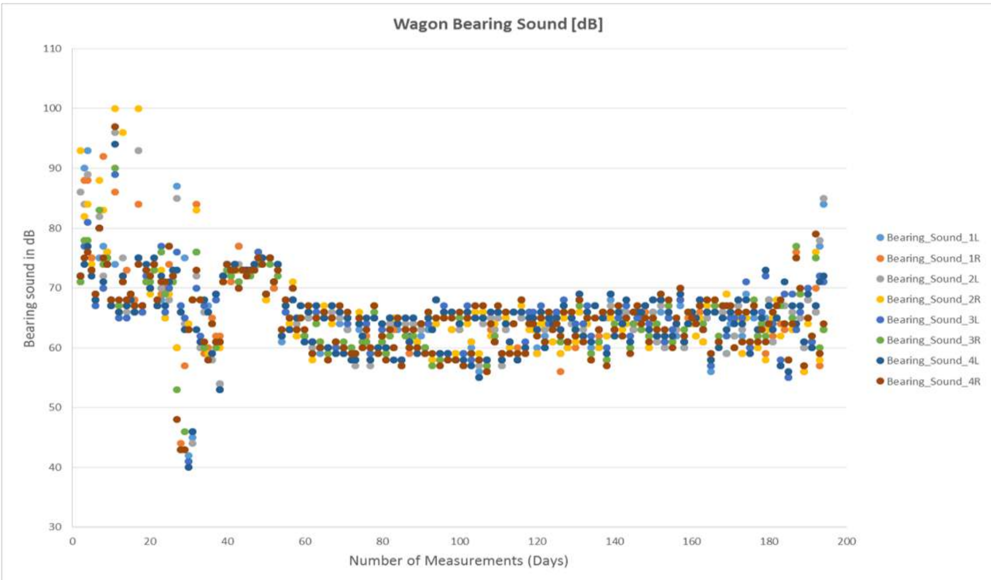


Figure 10: Raw data showing the acoustic signature for the wagon's bearings

3.5.2 Wheels

The second prediction parameter is the wheels of the wagon. The measurement system used to obtain the condition data for the wheels is called the Wheel Profile Monitoring System (WPMS). The wheel tread wear data for the past year for a specific wagon is shown in Figure 11. This graph displays two distinct clusters of measurements. This wagon went in for service and new wheels were put on the wagon. The measurements on the most left shows more wear (before service). The flange height data for the past year for a specific wagon is shown in Figure 12. These graphs clearly display the degradation of the wagon wheels.

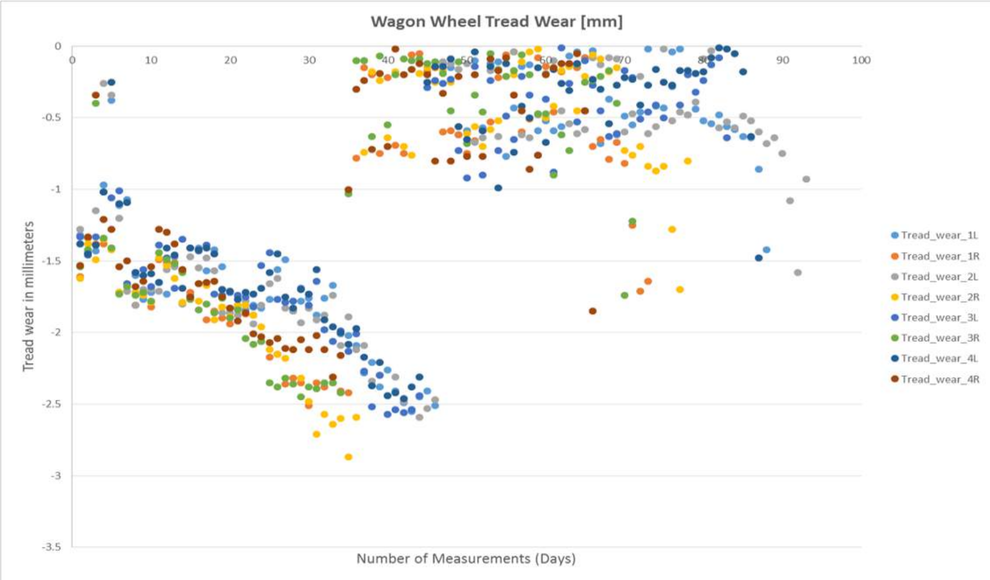


Figure 11: Raw data showing the tread wear for the wagon's wheels

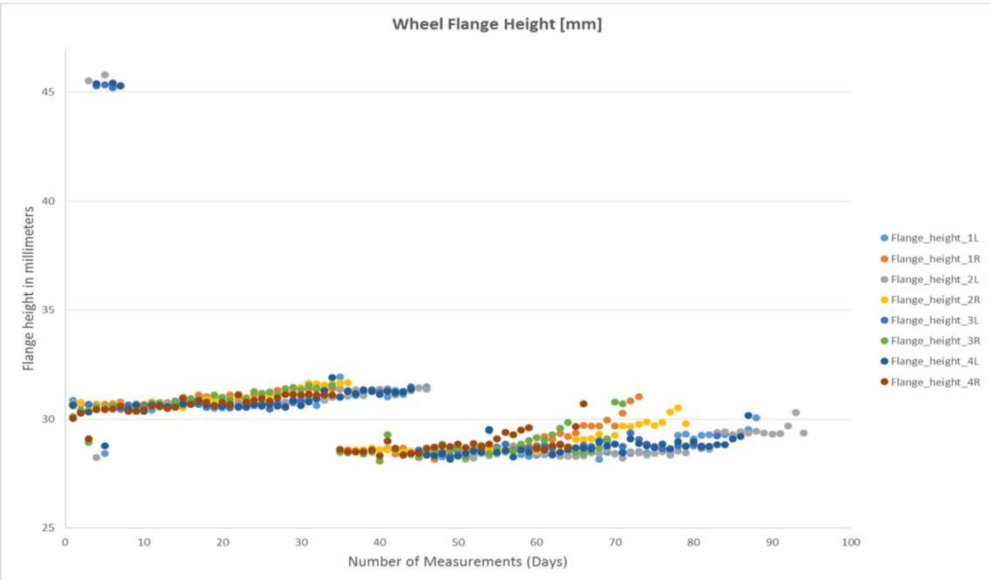


Figure 12: Raw data showing the flange height for the wagon's wheels

3.5.3 Bogies

The third prediction parameter is the wagon's bogies. The measurement system used to obtain the condition data for the bogie is called the Skew Bogie Detector System (SBD). The lateral force (gauge spreading force) that a skew bogie applies to the rail is measured and the data for the past year for a specific wagon is shown in Figure 13.

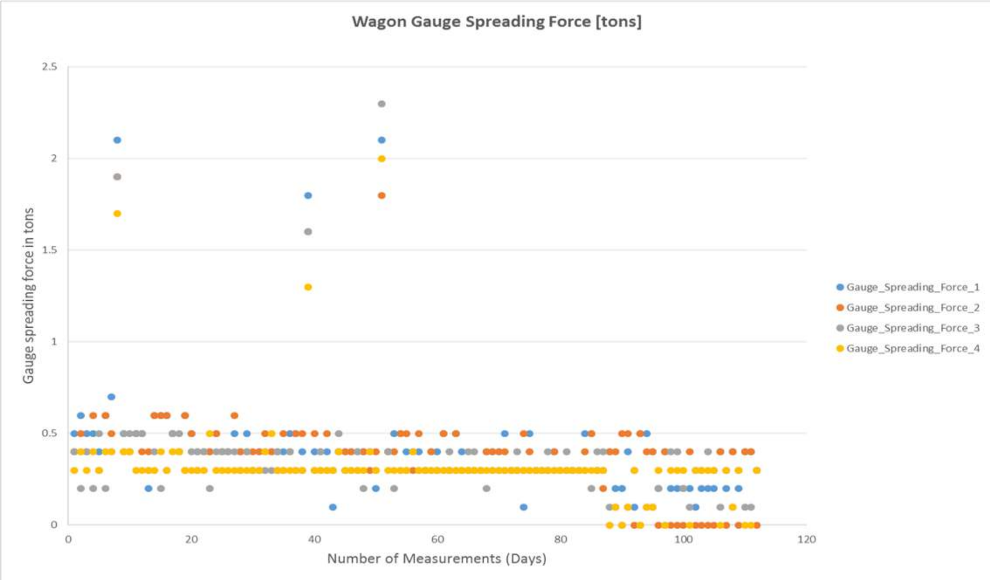


Figure 13: Raw data showing the gauge spreading force for the wagon's bogies

3.6 Brake blocks

The feasibility of adding the brake blocks as an additional prediction parameter to the research will form part of future work. The brake block top thickness data for the past year for a specific wagon is shown in Figure 14. The brake block bottom thickness data for the past year for a specific wagon is shown in Figure 15.

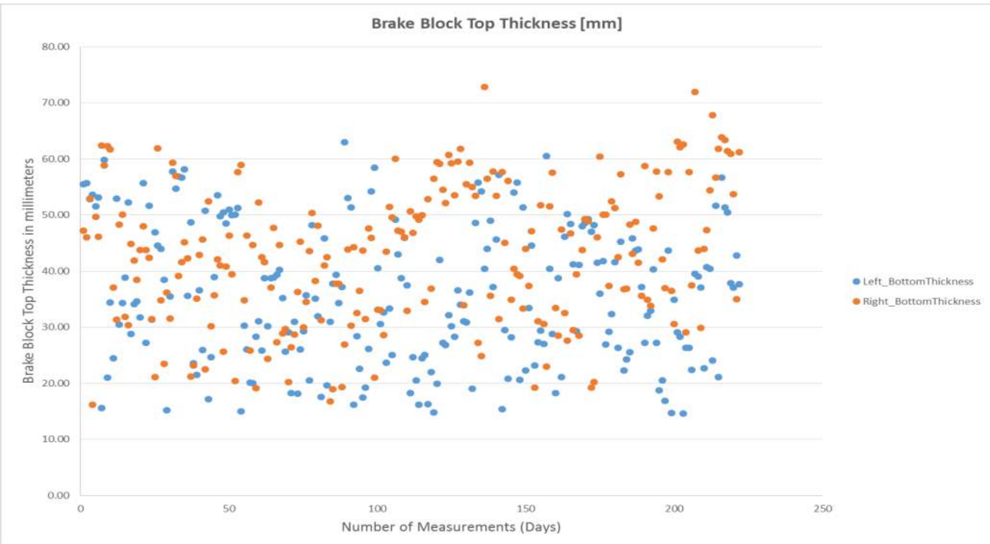


Figure 14: Raw data showing the top thickness of the wagon's brake block

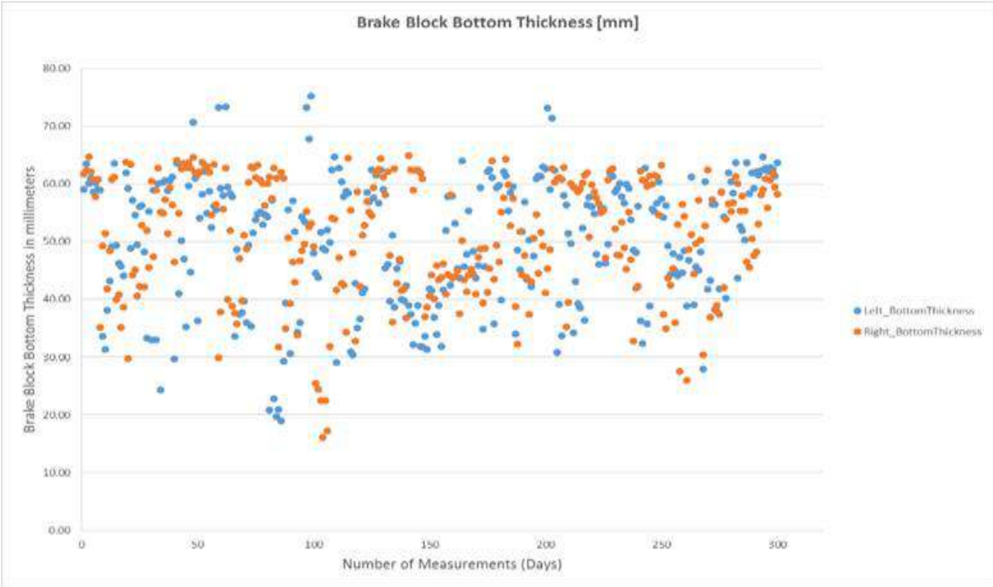


Figure 15: Raw data showing the bottom thickness of the wagon's brake block

4. SUMMARY AND FUTURE WORK

Maintenance is an integral part of society, whether it is to paint your house or to keep expensive railway equipment in optimum working condition. There are various maintenance strategies available to follow. Careful consideration must be given when a maintenance strategy is being chosen. The skill level of technicians, spare part inventory and availability, cost-saving initiatives and safety considerations are the major influencers on the decision. Predictive maintenance is a relatively new concept but is making huge strides forward. It utilises predictive analytic techniques to combine predictor parameters with historical data to predict when equipment failure could occur. "If it's not broken don't fix it." It not only demonstrates huge cost savings, decreased downtimes and improved safety conditions but it has an add-on advantage of being able to more effectively manage staff. Predictive maintenance based on a railway wagon could allow it to run for a longer time than the scheduled two years or it can detect a premature failure and initiate a corrective action in order to prevent a derailment.

Prognostics serves as the perfect tool for estimating the time to failure for a system or component in question. The time to failure is more often referred to as remaining useful life (RUL). There are various suggested approaches that can be followed to arrive at the estimated RUL. These include; model-based, data-driven and the hybrid approaches. Model-based relies heavily on knowledge of the system, data-driven relies on a huge amount of data in the absence of physical knowledge of the system and the hybrid approach is a combination of the two. Combining the model-based and data-driven approaches make for a more accurate estimation of the RUL. These approaches require predictor parameters and that can be determined by following the Failure Modes, Mechanisms and Effects Analysis (FMMEA) methodology. The researcher followed the methodology for a railway wagon and identified the parameters to be; wheel bearings, wheels and the bogie.

The researcher obtained four years' worth of condition data relevant to the predictor parameters. A decision was made by the researcher to follow the data-driven approach but at the end add knowledge of the systems in order to establish a more accurate RUL estimation.

The following topics will be addressed for future work:

- Analysis of the data
    - Cleaning the data
    - Perform feature extraction
    - Determine whether there are missing failure data or not
    - Determine whether the data are censored or uncensored
- The categorisation of the predictor parameters
- Investigate survival analysis
- Investigate the possible use of the Cox proportional hazards regression model
- Investigate the principal component analysis
- Determine requirements for inputs into the machine learning process

This model will be implemented in the railway environment to support the cost-saving initiatives of the company.

REFERENCES

[1] BusinessDictionary.com. (2018). Maintenance. [online] Available at: http://www.businessdictionary.com/definition/maintenance.html [Accessed 10 Jan. 2018]

[2] En.wikipedia.org. (2018). Maintenance (technical). [online] Available at: https://en.wikipedia.org/wiki/Maintenance_(technical) [Accessed 10 Jan. 2018]

[3] Khan, M. and Siddiqui, N. (2017). Maintenance Engineering and Management. India: New Age International (P) Ltd pp. 1-10.

[4] Martins, L. (2018). Advantages and disadvantages of Corrective Maintenance Infraspeak Blog. [online] Infraspeak. Available at: http://home.infraspeak.com/advantages-and-disadvantages-corrective-maintenance/ [Accessed 21 May 2018]

[5] Services, A. (2018). The Benefits of Preventive Maintenance - Able Services. [online] Able Services. Available at: http://ableserve.com/issue-1/the-benefits-of-preventive-maintenance/ [Accessed 21 May 2018]

[6] ProTech Property Solutions. (2018). Advantages and Disadvantages of Preventative Maintenance - ProTech Property Solutions. [online] Available at: http://www.protechpropertysolutions.co.uk/advantages-and-disadvantages-of-preventative-maintenance/ [Accessed 21 May 2018]

[7] Preventive-predictive-maintenance.blogspot.co.za. (2018). Predictive Maintenance Implementation: Advantages and Disadvantages. [online] Available at: http://preventive-predictive-maintenance.blogspot.co.za/2010/01/predictive-maintenance-implementation.html [Accessed 21 May 2018]

[8] En.wikipedia.org. (2017). Prognostics. [online] Available at: https://en.wikipedia.org/wiki/Prognostics [Accessed 13 Dec. 2017]

[9] En.wikipedia.org. (2017). Prognosis. [online] Available at: https://en.wikipedia.org/wiki/Prognosis [Accessed 13 Dec. 2017]

[10] Elattar, H., Elminir, H. and Riad, A. (2016). Prognostics: a literature review. Complex & Intelligent Systems, pp.125-154

[11] An, D., Choi J. and Kim, H. (2012) A tutorial for model-based prognostics algorithms based on matlab code. Annual Conference of the Prognostics and Health Management Society 2012, pp. 1-9

[12] Pecht, M. and Kumar, S. (2008) Data Analysis Approach for System Reliability, Diagnostics and Prognostics. Pan Pacific microelectronics symposium, Kauai, Hawaii, USA.

[13] Mathew, S., Alam, M. and Pecht, M. (2011). Identification of Failure Mechanisms to Enhance Prognostic Outcomes. Journal of Failure Analysis and Prevention, [online] 12(1), pp.66-73

[14] Dunn, S. (2018). Big Data, Predictive Analytics and Maintenance - Assetivity. [online] Assetivity.com.au. Available at: https://www.assetivity.com.au/article/maintenance-management/big-data-predictive-analytics-and-maintenance.html [Accessed 24 Jan. 2018]

[15] Goebel, K., Saxena, A., Daigle, M., Celaya, J., Clements, L. and Scott, M. (2012) Introduction to Prognostics. First European Conference of the Prognostics and Health Management Society 2012

[16] Hitachiconsulting.com. (2018). [online] Available at: https://www.hitachiconsulting.com/documents/solutions/hitachi-predictive-maintenance/hitachi-predictive-maintenance-overview-so.pdf [Accessed 14 Jun. 2018]

LEAN STARTUP AND ITS APPLICATION IN BUSINESS AND OPERATIONS STRATEGY IN SOUTH AFRICA

N.M. Kola[1] & B. Emwanu[2]
[1]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
nadeemkola@gmail.com

[2]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
Bruno.Emwanu@wits.ac.za

ABSTRACT

Lean startup is a relatively new concept initially developed to help startups grow their business in a more effective and scientific way. This methodology uses data to help inform decisions such as entering new markets in a fast-changing business environment with several uncertainties. The methodology contrasts with the traditional approach by supporting experimentation over preconceived assumptions and iterative development over linear methods to evaluate new business ventures.

This study evaluates the extent to which lean startup principles have been applied across industry in South Africa and how these have been used in conjunction with the corporate and operations strategy. While Lean startup methodology has been applied in other parts of the world, its applicability and extent of application within the South African context is not evident. Prevalence of Lean startup methodology is the desirable state whereas paucity would be a concern.

The results of the study show that a number of participants used the methodology to help develop the business and operations strategy. Participants also typically saw value in the concepts for application across industry as long as the right structures are in place to help support the initiative.

[1] Corresponding author

1. INTRODUCTION

1.1 Background

Traditionally, the role of business strategy, among other things, is to help mitigate risk by having a clear roadmap for the company. Companies typically spend time and effort developing financial models to forecast the impact of the strategy and to have a strategy that is as robust as possible. Using the traditional approach, a business project typically undergoes in-depth research and analysis to back-up the idea it is using. While there is merit in the traditional approach that has been tried and tested over decades, the world has changed. Today with much lower barriers to entry, competitors are able to bring products to market quicker, hence execution is taking place alongside strategy [1]. Businesses need to be able to test ideas first hand with actual customers before building scaled end products, and in so doing businesses can establish early in the process if they should pursue the idea in question, change it, or abandon it. In building the strategy alongside the execution (i.e. building the product), the risk of failure can be greatly reduced. Similarly, when looked at from an operations perspective, having a well-developed strategy helps to reduce the risk of failure or production changes which in turn can mean massive cost implications.

One can find many examples of companies in Silicon Valley that have risen from garage based startups to literally the biggest corporations in the world, all in a relatively short space of time. During this time other organisations have not done as well and in certain circumstances, had to close. Many of these Silicon Valley startups were built using an iterative process that allowed the organisation to design their strategy while validating the assumptions in parallel. This methodology known as the “lean startup” methodology encourages experimentation rather than detailed planning, customer feedback in place of assumptions and iterative development in place of scaled up development [2].

While the lean startup methodology has been implemented by many startups across the world, the level of research in academia has been limited. Further, while the lean startup method has been implemented and documented in “First World” countries, very little published research was found showing its applicability in the South African context.

Thus, this research attempts to determine the extent to which the lean startup methodology has been applied in selected firms in South Africa, as well as understand the extent in which the lean startup methodology would assist in informing the development of the operations strategy by gathering actual customer data to inform the implementation of the business strategy.

This paper has the following objectives:

1. To establish the extent to which the lean startup concepts have been applied in selected South African firms.
2. To determine which of the lean startup concepts were applied.
3. To establish whether the lean startup concepts have been applied with the intention of helping develop the business and operations strategy through the top-down development approach.
4. To determine what the success or failures were in implementing the lean startup concepts.

The proposition is that South African companies will typically be aware of using concepts like agile for systems development but will typically not have applied lean startup concepts, especially in a strategy function.

2. LITERATATURE REVIEW

The literature review aims to provide perspective on both business strategy and the lean startup methodology as these two areas of knowledge are fundamental to the research.

2.1 Strategy

Various academics have defined strategy over the years, for example defines strategy as, “the direction and scope of an organisation over the long term: which achieves advantage for the organisation through its configuration of resources within a changing environment, to meet the needs of markets and to fulfil stakeholder expectations” [3]. On the other hand, Porter [4] defined strategy at a high level as, "The essence of strategy is choosing to perform activities differently than rivals do”. Wit and Meyer [5], defined strategy based on earlier works by Porter [6] as, “The search for a favourable competitive position in an industry, the fundamental arena in which competition occurs. Competitive strategy aims to establish a profitable and sustainable position against the forces that determine industry competition”.

A universal definition of business strategy has been elusive. While strategy has been defined differently over time, Hax and Majluf [7], collated the various definitions in a unified study that summarises strategy as, “a multidimensional concept that embraces all the critical activities of the firm, providing it with a sense of unity, direction, and purpose, as well as facilitating the necessary changes induced by its environment." For this research the definition from Hax and Majluf [7] will be used as it incorporates aspects of the other definitions.

While the definition of business strategy is varied, the way in which strategy is developed also has various interpretations. This has led to different schools of thought on how a strategy is achieved. Because of the different interpretations of strategy development, two schools of thought have emerged namely, prescriptive and emergent strategy [8]. These schools of thought have differing views on the three core areas of strategy which are analysis, strategic development and implementation [9]. In the prescriptive school of thought the three areas are done in a sequential way, where in the emergent school of thought the three areas are interrelated and there is no clear distinction between development and implementation [8]. This study adopts the position of the prescriptive school of thought of strategy development.

Bradley et al. [10] provides an example of a strategy development framework shown in Figure 1 below. One key factor to note from this framework is that the feedback loop is only included at the end of the process and not at each step of the process as described in the customer development process [11].

Figure 1: Strategy Building Blocks [10]

2.2 The Lean Startup

The advent of the Lean Startup movement is attributed to Blank [11] who introduced the concept of customer development where the customer was involved from the initiation of the idea. Through this process the feedback from the customer is used to ensure that the final product is in line with what customers want. Blank [11], contrasts the differences between the traditional product development as shown Figure 2 below, with the newer customer development cycle as shown in Figure 3 below.

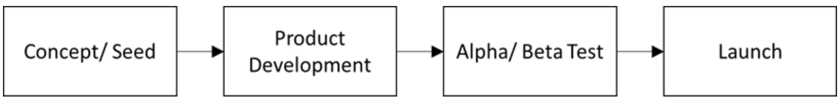


Figure 2: The Product Development Diagram (traditional method) [11]

Due to the iterative nature of the cycle, shown in Figure 3, the concept of "Customer development", helps to manage risks by involving customers throughout the process. A key to this process is the fact that is it iterative and not linear as in the traditional product development process.

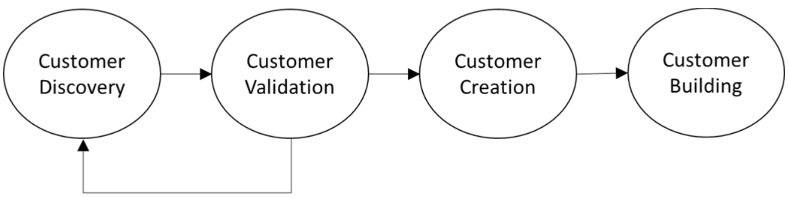


Figure 3: The Customer Development Cycle [11]

While Blank [11] is associated with pioneering the lean startup movement, the name (lean startup) was actually coined by Ries, a student of Blank [12]. The lean startup process is based on the concepts made famous by Lean Manufacturing and the Toyota Production System [13].

The lean startup method is built on including customers from the very start, avoiding waste and building a minimum viable product as quickly as possible and iterating based on the feedback received [15]. Ries [13] goes on to redefine a startup as, "human institution designed to create new products and services under conditions of extreme uncertainty".

The key principles to the lean startup concept developed by Ries [13] are summarised in Table 1 below.

Table 1: Key lean startup principles [13]

Principle	Description
Entrepreneurs are everywhere	Using Ries's definition of entrepreneurship, the applicability of these concepts can be to anyone working in any organisation, big or small.
Entrepreneurship is management	Since a startup is an organisation, the associated processes need to be appropriate for the environment with extreme uncertainty.
Validated Learning	Validation of the startup experiments should be run. The results of which can be used to iterate on the vision and idea.
Build-Measure-Learn	Fundamental to the process is to get the product to market, test it with customers and then use this feedback to determine whether to pivot or persevere.
Innovation accounting	To get a successful outcome, the process needs to be closely measured. This includes setting KPI's and holding people accountable to deliver.

The lean startup methodology helps new ventures develop products that customers want, far quicker and cheaper than traditional methods [14], the cost of getting a product to market has decreased drastically. Although the lean startup concepts were initially designed for startups, these concepts can be extended to an established organisation as well. Ries [13] expands the concepts of entrepreneurship and Lean Startup to not only include startups but to expand the concepts to include all organisations building or launching a new product in environments of extreme uncertainty.

2.3 Conclusion

A theoretical framework is developed by incorporating the lean startup methodology as defined by Ries [13]. The framework is illustrated in Figure 4 below. One of the research objectives is to establish the relationship between the lean startup methodology and business strategy. The theoretical framework will be useful for investigating this relationship. The relationship is shown in the theoretical framework by illustrating the lean startup method as a link between the business and operations strategy. Each of the pillars is then broken down into key concepts that are integral to the lean startup method. The four pillars and these concepts are adapted from the key principles as developed by Ries [13].

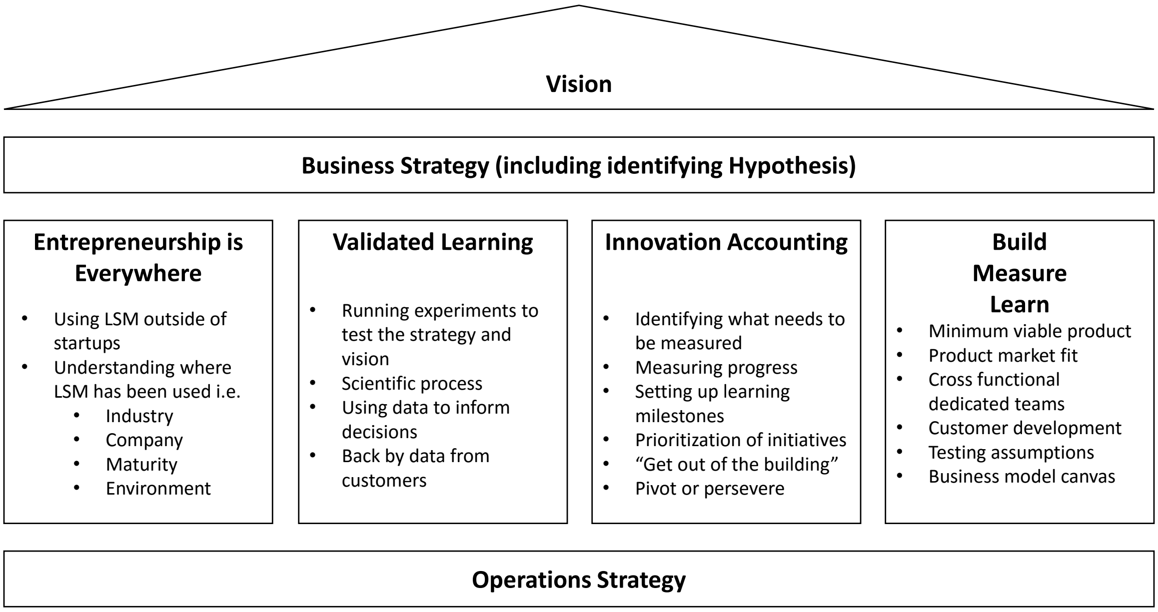


Figure 4: Theoretical Framework

Having reviewed the relevant literature and developing the theoretical framework to guide this study, the chapters that follow will look at how data was collected, what information was obtained and how it was interpreted, leading to the findings of this study. The next chapter explains the methodology used to collect the data.

3. METHODOLOGY

3.1 Research Approach

A qualitative research approach was used. A number of qualitative methods were considered for this exploratory study [16] and it was decided that interviews would be used as it allows for interpersonal contact which enables one to probe issues deeper and follow on certain responses from the interviewee. This is useful in evaluating the way in which strategy development changes and how these changes affect the implementation. This method also allows for one to obtain a large amount of information in a relatively short space of time which is ideal due to the tight time constraints of this research.

The interviews could be conducted in a number of different ways [17], however, for the purposes of this study the semi structured interview was considered to be the most suitable. Due to the exploratory nature of this research, semi-structured interviews help to keep the questions consistent while at the same time being able to extract extra information over and above what the questions directly asked. An interview guide was developed before the actual interviews are conducted. Some of the questions were open ended to allow for participants to express their views adequately with the interview guide ensuring that the conversion is kept on topic. The interviews were recorded and later transcribed and coded. The coding process allowed the data collected to be segmented into various themes which are then analyzed [18].

3.2 Analysis of Data

Once the data were collected and transcribed, the data were coded to identify themes in line with the theoretical framework. Leedy and Omron [19] have developed a step wise method to ensure reliability of the analysis. This includes applying the initial coding to a sample of the data to test if the codes developed can capture all the insights from the data. This process would help inform the final code list thus ensuring both validity and reliability of the coding. Another important way to ensure reliability is to use direct quotes for each code in order to improve the reliability of the analysis [19].

The analysis has been adapted from the multiple case study analysis methodology developed by Yin [20]. Each individual project that the participant discusses is tested against the theoretical framework to identify the extent of application. The interview was split into two sections. The first portion of the interview was used to determine the participant's understanding of lean startup as a concept. The second portion of the interview was where the participants discussed the project examples. Each project was individually analyzed against the theoretical framework to understand the extent to which the methodology has been applied. This process was adapted from the "within" case study analysis [20]. The projects were then analyzed together to determine if there were any themes that could be identified across the various projects.

4. RESULTS

A total of nine interviews were scheduled but only eight interviews were eventually conducted. The participants all had some type of consulting experience but with various backgrounds. While all the participants had strategy and consulting experience, majority of the participants had experience in international consultancies, followed by niche consultancies and only one participant being in an internal strategy role. The profiles of the various people interviewed is summarised in Table 2. The participants were all senior management or partners in their respective roles.

Table 2: Interviewee Profile Information

Coded Name of Participant	Role	Industry Focus	Years of Experience	Areas worked in
A	Internal Strategy	Telecoms	7 Years	South Africa Rest of Africa
B	International Consultancy	Financial Services	16 Years	Europe, Brazil, SA
C	International Consultancy	Financial Services	23 Years	Europe, USA, SA
D	International Consultancy	Financial Services Retail	20 Years	SA, Rest of Africa
E	Niche Consultancy	Across Industry	20 Years	SA
F	Niche Consultancy	Mining Telecoms Financial Services	20 Years	SA
G	International Consultancy	Financial Services Telecoms Rail & Freight	10 Years	SA Middle East
H	International Consultancy	Financial Services	20 Years	Europe SA

By asking each of the participants to explain their understanding of lean startup, the understanding of the concepts were assessed. This question was also used as an indicator to understand the extent to which the concepts were applied.

Figure 5 below shows the number of people that knew compared to those that didn't know what the concept meant. A specific question was asked on the interpretation and understanding of lean startup. The question was asked upfront without too much of an explanation to get a non-biased view of the concepts.

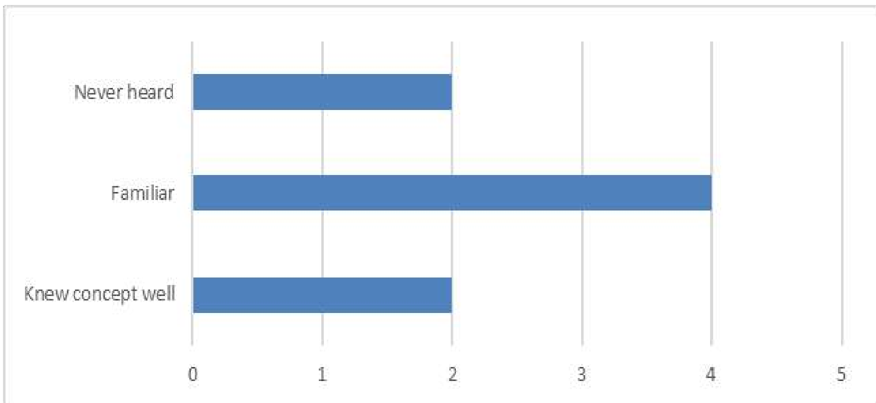


Figure 5: Familiarity of LSM

The second portion of the interview was focused on the interviewee talking through various projects that they were involved in. In total participants walked through eighteen projects across the eight interviews. It must be noted that due to the interviews being semi-structured, this portion was driven by the interviewee with the interviewer probing only when needed.

The tables below are summaries of the detailed analysis that was done for each project. Table 3 below shows how each project scored against the 4 elements of the theoretical framework. The sc oring was based on a qualitative assessment of the application of the concept against what was defined in the literature. The scoring using the 3-point approach where “1” meant the concept was not applied, “2” meant the concept was applied partially and “3” meant the concept was applied in totality.

Table 3: Project Scoring Summary

	Description of project type	**Validated Learning**	**Innovation Accounting**	**Build Measure Learn**	**Strategy Applicability**
Project 1	Operating Model Change	2	2	1	1
Project 2	Video Streaming	2	1	3	1
Project 3	Base stations	2	1	2	1
Project 4	Global Commercial Bank	2	3	2	3
Project 5	Grow Financial Services	2	2	2	3
Project 6	Automotive Reorganisation	1	2	2	1
Project 7	Automotive Car Launch	2	1	2	2
Project 8	New Channel Launch	1	2	2	2
Project 9	Retail Business	2	1	1	1
Project 10	Bank	3	1	2	2
Project 11	Trade Union Investment	1	1	1	1
Project 12	Cosmetic Manufacturer	1	1	2	2
Project 13	Strategy Consulting	1	1	2	2
Project 14	In house Consulting Capability	1	1	1	1
Project 15	Property Management	1	2	2	3
Project 16	Port Company	2	1	2	3
Project 17	Engaging Clients better	2	1	1	1
Project 18	Mass savings product	2	1	1	1

From the results, a small proportion of the projects applied the concepts in full, with majority of the projects applying the concept partially or not at all.

Each project from the various participants was analysed to determine the extent to which the lean startup methodology and its various components were applied. Figure **6** below shows the extent of the application of the various lean startup concepts. From the chart, the application of the concepts varied with innovation accounting being the least used concept as it was not used at all in 67% of the projects, followed by strategy applicability which was not used in 50%, validated learning which was not used in 39% and build measure learn which was not used in 33% of the projects.

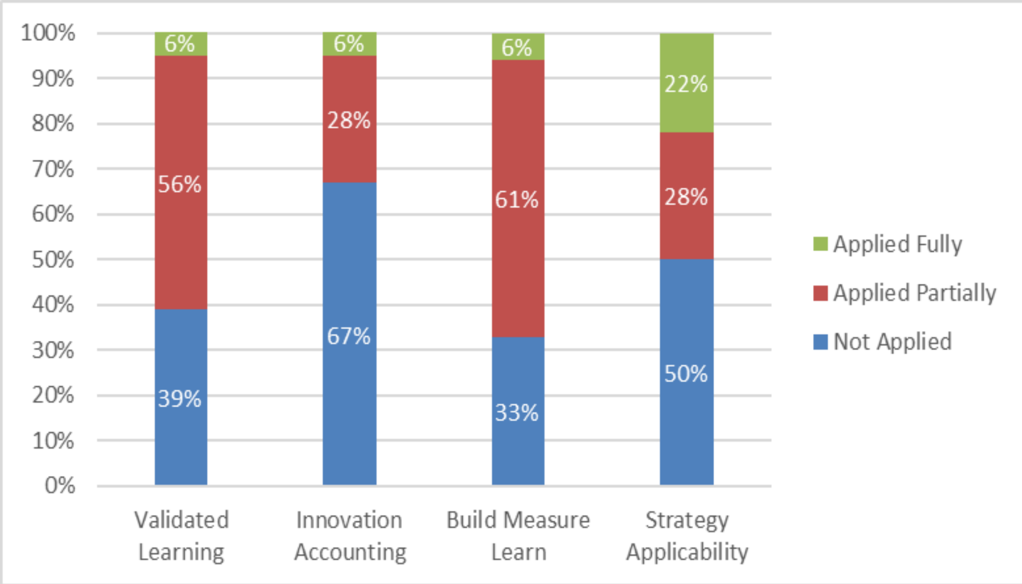


Figure 6: The extent of the concepts applied across the project studies

Validated learning, innovation accounting and build measure learn were all applied fully in 6% of the projects with strategy applicability applied in 22% of projects. Build measure learn and validated learning were the concepts most frequently applied partially, being applied 61% and 56% respectively. Innovation accounting and strategy applicability were both found to be applied partially in 28% of the projects.

A scoring model was then set up for all projects where the concepts that were either partially or full applied were counted. Using this model, a heat map of the various concepts was identified as shown in Table 4: Heat Map of the Concepts per industry for the selected firms.

Table 4: Heat Map of the Concepts per industry for the selected firms

	Number of Projects	Validated Learning	Innovation Accounting	Build Measure Learn	Strategy Applicability
Financial Services	7	71%	43%	57%	57%
Automotive	2	50%	50%	100%	50%
Retail	1	100%	0%	0%	0%
Manufacturing	1	0%	0%	100%	100%
Professional Services	2	0%	0%	50%	50%
Rail and Freight	1	100%	0%	100%	100%
Telecoms/ Media	4	75%	50%	75%	25%

From the projects that were discussed, a large majority of the projects were based in financial services. With only one project in each of retail, manufacturing and rail. Looking at each of the industry specific projects, the rail and freight project was the one with most of the concepts fully applied however, as noted, there is only one project in this category.

Each of the projects discussed by the interviewees was evaluated against the four concepts to determine if certain participants discussed the concepts more than others, as shown in Figure 7 below.

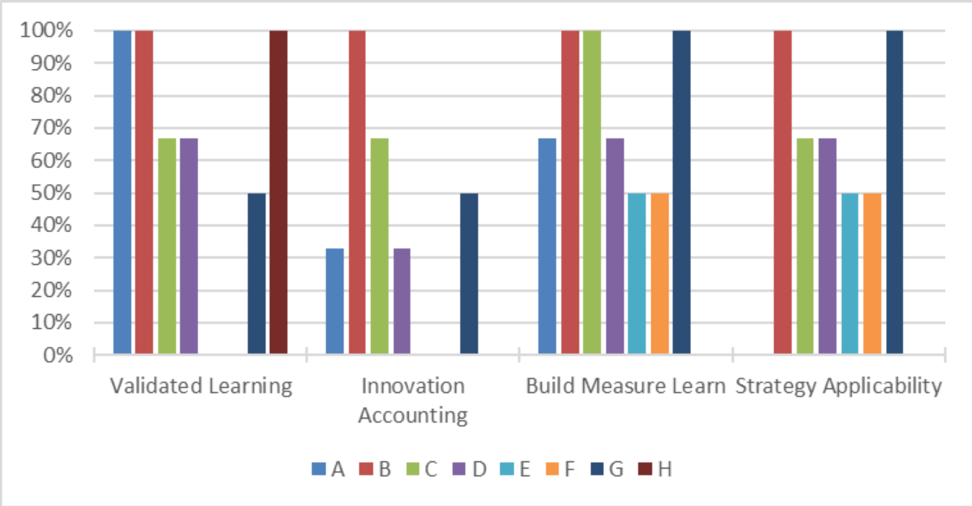


Figure 7: Concept Application per interviewee

The projects discussed by participant B, C, D and G applied the lean startup concepts to various degrees in at least one of the projects they discussed. Where participants B, C and G also showed a strong understanding of the lean startup concepts while participant D applied the concept even though an understanding of LSM was not shown. On the other hand, participant F was not familiar with LSM and thus it showed, as the projects discussed did not apply the lean startup concepts. There were however participants E and H who were familiar with the concepts but did not apply the concept to a large extent.

The various customer engagement models namely B2B, B2C and hybrid were analysed against the LSM concepts as shown in Figure 8 below.

Validated learning was most commonly used in the B2C model, while innovation accounting was most commonly used in the B2B model. For both build measure learn and strategy applicability, these were most common used in the hybrid customer engagement model.

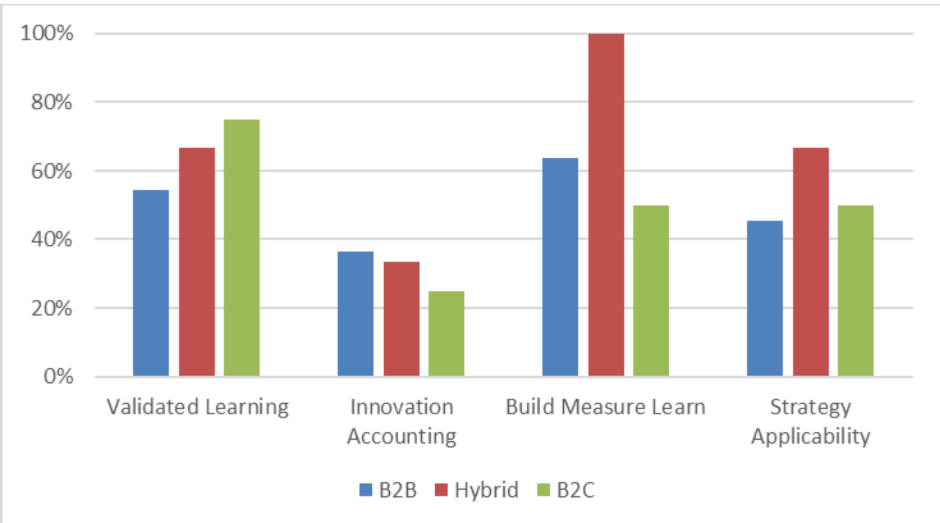


Figure 8: Concept Application per Customer Engagement Model

5. DISCUSSION

The purpose of the research was to better understand the extent to which lean startup principles have been applied in selected firms across industry in South Africa and how these have been used in developing the corporate and operations strategy. To carry this out professionals with extensive strategy consulting experience were interviewed to answer the research question. The research question sought to establish the extent to which the lean startup concepts were applied to the selected projects and the extent to which it influences strategy.

To help answer the research question, a proposition to be investigated was posited and four research objectives were defined. The proposition of the study was that the professionals interviewed may have been familiar with the concepts, but they did not necessarily apply the concepts. From the results obtained this proposition was supported, as some of the professionals were familiar with the concepts but no project had all the concepts applied in totality. This suggests that knowledge of LSM alone is not enough to ensure that the LSM concepts will be applied in totality and thus other considerations also need to be considered. Some of these are suggested below such as LSM being a relatively new concept outside of the technology environments.

Of the four research objectives defined, three were met and one was inconclusive. The objectives are expanded upon below.

5.1 Objective 1: To establish the extent to which the lean startup concepts have been applied in selected South African firms.

It was found that while most of the participants interviewed were familiar with lean startup as a concept, none of the participants applied the concept in totality. This is despite the observation that the number of participants who were familiar with the lean startup concepts was greater compared to research that was done earlier the concept was presumably still new. The reason for these participants, who were familiar with the lean startup concept, not applying the concept in totality could be attributed to several factors including the type of industry they focussed on, the way the South African corporate environment operates, or it could have been due to the LSM concepts being still relatively new outside of technology environments. On this point it should be noted that the cases presented in the literature on application of LSM are predominantly in technology environments, hence there may be unforeseen differences in application outside of these environments.

From the research it was found that heavy industries typically did not apply the concepts except for rail and freight. The lack of application of LSM in heavy industries could be attributed to the nature of the industry being more traditional while the exception of rail and freight could be attributed to the participant who discussed the project, being familiar with the LSM concepts. Another industry which did not apply the concepts was retail. From the literature it was found that the South African retail industry generally does not do as much experimentation but rather follows international trends, suggesting that the retail industry in South Africa is cautious and somewhat conservative and may not be as receptive to LSM concepts that require experimentation. On the other hand, it was noted that the participant involved with the retail project did not know about LSM which might have influenced the outcome. Both considerations might therefore have influenced the outcome from the retail industry. It was found that participants that were familiar with lean startup tended to apply the concepts more, even though they did not understand all the concepts in detail or were not aware that they were applying the concept. It was also found that participants with international experience or working for international organisations were more familiar with the concepts. This outcome, associating familiarity of concepts with international experience, was expected since the literature documents most of the successful cases of application of LSM as located outside of South Africa, particularly in developed countries.

5.2 Objective 2: To determine which of the lean startup concepts were applied.

The results showed that both validated learning and build measure learn concepts were applied partially with more-or-less similar frequency while, innovation accounting was applied to a lesser extent. When looking at the full application of the LSM concepts, the frequency of application was much less with the three LSM concepts (namely Validated Learning, Innovation Accounting and Build Measure Learn) being applied 6% each. From the results it was also found that participants who were not familiar with the concept would not apply the concepts in totality but might unknowingly apply the concepts partially. This suggests that while aspects of the LSM concepts have been applied on various projects analysed, the full application of the concepts was much fewer. The partial application of LSM, rather than the full application of the concepts, can be attributed to LSM being made up of several more traditional concepts which participants are more familiar with, and therefore perhaps more comfortable with.

5.3 Objective 3: To establish whether the lean startup concepts have been applied with the intention of helping develop the business and operations strategy through the top-down development approach.

Participants used the lean startup concepts to help develop the business or operations strategy (50% of projects discussed). When the projects were further analysed it was found that even participants not familiar with the concepts still unknowingly used some of the concepts to help develop strategy. This can be attributed to the participants being familiar with elements of LSM and using those elements to help develop the business and operations strategy. This suggests that using LSM and its concepts to help develop the business and operations strategy would be possible as confirmed by the projects that used it. On the other hand, participants that did not apply the concept could be attributed to lack of understanding of LSM or not viewing LSM as a tool to help develop strategy. This outcome serves as an interesting prospect that LSM, if well managed and better understood, has the potential of becoming an important tool when developing the business and operations strategy.

5.4 Objective 4: To determine what the success or failures were in implementing the lean startup concepts.

Since none of the projects analysed applied all the lean startup concepts in totality it was difficult to find from the partially applied LSM what the successes and failures were in applying the concepts at the planning stage. An alternative approach would be to follow the process even for partially applied LSM projects into implementation. The participants interviewed however, were strategy consultants who typically do not enter implementation of projects and hence would not be aware of what the long-term outcomes of the projects were. Thus, this objective of the research was not achieved.

6. CONCLUSION

The research question sought to determine the extent to which the lean startup methodology and/or its principles have been applied in selected South African organisations and how this influences the development of the business and operations strategy, using the top down development approach. From the results obtained, through addressing the research objectives, it was found that the LSM methodology was used to varying degrees depending on the type of project as well as the background of the professional being interviewed. Since the professionals interviewed had extensive experience in strategy development using the top down method, they were able to relate their use of lean startup methodology to help influence the strategy.

While the projects discussed were not necessarily labelled as lean startup or follow the LSM methodology explicitly, they showed signs of applying the concepts with the extent of application varying by project. Interestingly, even though there was variation of application by project, some features of LSM became apparent, for example it was found that validated learning and “build measure learn” were used more frequently than innovation accounting.

Overall, the application of the LSM concepts helped in developing the business and operations strategy although this was not done in all projects. From the participants it was found that lean startup is a concept that can used across projects if there is proper support from management. Among limitations mentioned by participants as hindering application of LSM in South African industries was the making of decisions too quickly without evaluating the concepts thoroughly. LSM avoids this quick fix approach as by its nature LSM is methodological, being quantitatively based and using rigorous hypothesis testing to help develop a business model

This work pioneers a study on the application of lean startup across industries in South Africa thus making a contribution. Although exploratory in nature, it provides a useful foundation for researchers wishing to carry the work forward. A framework has been provided and useful findings documented that other researchers can build upon. Further research should consider the limitations to application of LSM in South African industries and provide suggestions on how to overcome these.

REFERENCES

[1] C. Gagliardi, “Strategy made simple in a digital world,” *PWC*, 2015. [Online]. Available: http://pwc.blogs.com/ceoinsights/2015/09/strategy-made-simple-in-a-digital-world.html. [Accessed: 23-Sep-2016].

[2] S. Blank, “Why the Lean Start-Up Changes Everything,” *Harvard Business Review*, 2013. [Online]. Available: https://hbr.org/2013/05/why-the-lean-start-up-changes-everything. [Accessed: 25-Jul-2016].

[3] G. Johnson, K. Scholes, and R. Whittington, “Exploring corporate strategy : text and cases,” in *P 10*, FT Prentice Hall, 2008, p. 878.

[4] M. E. Porter, *On competition*. Harvard Business School Pub, 2008.

[5] B. de. Wit and R. Meyer, *Strategy : process, content, context*. Cengage Learning, 2010.

[6] M. E. Porter, *Competitive advantage : creating and sustaining superior performance*. Free Press, 1985.

[7] A. C. Hax and N. S. Majluf, “STRATEGY AND THE STRATEGY FORMATION PROCESS,” *Sloan Sch. Manag.*, no. August, 1986.

[8] R. L. Lynch, *Strategic management*. Pearson, 2012.

[9] R. A. Jamil and S. T. H. Shah, “Critical Exploration of Prescriptive and Emergent approaches to Strategic management: A review paper,” *Int. J. Information, Bus. Manag.*, vol. 7, no. 3, 2015.

[10] C. Bradley, A. Dawson, and A. Montard, “Mastering the building blocks of strategy | McKinsey & Company,” *Mckinsey Quarterly*, 2013. [Online]. Available: http://www.mckinsey.com/business-functions/strategy-and-corporate-finance/our-insights/mastering-the-building-blocks-of-strategy.

[11] S. Blank, *The Four Steps to the Epiphany*, Second Edi. K&S Ranch, 2009.

[12] S. Virani, “The History Of Lean Startup,” 2016. [Online]. Available: http://www.salimvirani.com/the-history-of-leanstartup-and-how-to-make-sense-of-it-all/.

[13] E. Ries, *The lean startup : how today’s entrepreneurs use continuous innovation to create radically successful businesses*. 2011.

[14] A. Heryey, “Moore’s Law is Probably Dead – Future Crunch – Medium,” *Medium*, 2015. [Online]. Available: https://medium.com/future-crunch/moore-s-law-isn-t-the-only-technology-law-it-may-not-even-be-the-most-important-1964581dc0ab#.uo9fmr8y6. [Accessed: 26-Feb-2017].

[15] E. Chen, “Business strategy meets lean startup: Lessons from agile development,” *TechBeacon*, 2015. [Online]. Available: https://techbeacon.com/business-strategy-meets-lean-startup-taking-lessons-agile. [Accessed: 22-Feb-2017].

[16] J. Yates and T. Leggett, “Qualitative Research: An Introduction,” *Radiol. Technol. (RADIOL TECHNOL)*, vol. 88(2), pp. 225–231, 2016.

[17] S. Kvale, *InterViews: An Introduction to Qualitative Research Interviews*. 1996.

[18] J. W. Creswell and J. W. Creswell, *Qualitative inquiry & research design : choosing among five approaches*. SAGE Publications, 2013.

[19] P. D. Leedy and J. E. Ormrod, *Practical research : planning and design*. 2015.

[20] R. K. Yin, *Case study research : design and methods*. 2013.

APPLICATION OF MANUFACTURING STRATEGY PRINCIPLES IN ADAPTING CRAFT BREWERS FOR SUSTAINABLE GROWTH IN THE SOUTH AFRICAN BEER INDUSTRY

G. Ramorula[1] & B. Emwanu[2]
[1]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
gt_rams@yahoo.com

[2]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
Bruno.Emwanu@wits.ac.za

ABSTRACT

The beer industry is known for having few big producers who dominate and control the market such that it is considered saturated and mature with high barriers to entry. Nevertheless, in-roads into the beer market by much smaller players, namely craft brewers are observed, that are apparently growing and managing to retain market share. Being SMMEs, the ability of craft brewers to achieve this position points to their agility and resilience, however it is not clear how this is achieved through the manufacturing function expected to meet consumer requirements. This raises the question whether craft brewers use manufacturing strategies at all and if these align with consumer needs and company long term plans.

This study investigates the craft beer industry in South Africa by interviewing five craft beer producers and surveying over four-hundred consumers to understand expectations of consumers and practices of craft brewers regarding manufacturing strategy.

Results indicate that application of manufacturing strategy principles is limited and does not seem purposive to improve the manufacturing process and sustain performance but seems to be survival moves. This suggests long-term risk to stability of craft breweries in South Africa and the need to adapt manufacturing strategy principles to improve this.

[1] Corresponding author

1 INTRODUCTION

There are many economic and brand benefits that can be realised through the consideration of product differentiation, specifically in the fast moving consumer goods sector [1]. This is especially true in the beer industry where there has been a noticeable change in the types of beer that are preferred by the consumers in markets that were previously thought to be saturated and mature.

The growth of craft beer in recent years is largely attributable to consumer changing tastes and earning potential of craft breweries. These craft breweries are perceived to be producing a more diverse, higher-quality beer than the traditional mass-produced products that are manufactured by the commercial brewing giants of the world such as ABInBev, Heineken and others [3]. Craft brewing continues to take market share away from traditional non-craft brewers in recent years [4].

The beer business has consistently shown that it is a high volume, low margins industry, a characteristic of the FMCG sector that it belongs to [5]. This presents a high cost barrier to entry and requires economies of scale to overcome. Competing on high volume and low margins is currently not widely available to Craft Brewers as they are confined to producing small batches of products and sell to a much localised market. This however, may not be a perceived weakness as it is the attribute that allows craft brewers to diversify their portfolio and produce beers that are very specific in style and have a broad offering in contrast to the highly inflexible mass production of large scale breweries.

Nevertheless, the sustainability of craft brewers' growth is questionable given that little or no strategies seem to be in place to survive and remain sustainable. Being SMEs, standing up against the more resourced and hostile competitors and holding their own requires agility and resilience as the big players react to the threat. There is need to investigate the growth and impact of craft brewers in the highly competitive FMCG environment, what significance this has for corporate brewers as this will determine the level of reaction craft brewers can expected, and the implication for craft brewers using manufacturing strategy to achieve sustainability.

2 LITERATURE REVIEW

2.1 Craft brewing in South Africa

The history of craft brewing and its origins is varied across the world but appears to have evolved as a result of the amalgamation of corporate brewers that resulted in what can be described as a standard offering and diminishing variety of traditional beer taste and the birth of home brewing by passionate consumers [6]. The quality of these beers improved significantly over time and this was followed by their popularity from those who were seeking more from their drink than just a beer [7]. In South Africa where the apartheid regime prevented the majority of the population to legally consume alcohol until the dawn of democracy, the illegal brewing and trading of alcohol remained suppressed and thus no major inroads into beer brewing were made [4]. This gave rise to the large contingency of "shebeens" across the country where traditional alcohol could be made and sold. These however were never made part of the formal sector. In the formal sector, South African Breweries (SAB) continued to dominate the brewing industry with very little competition.

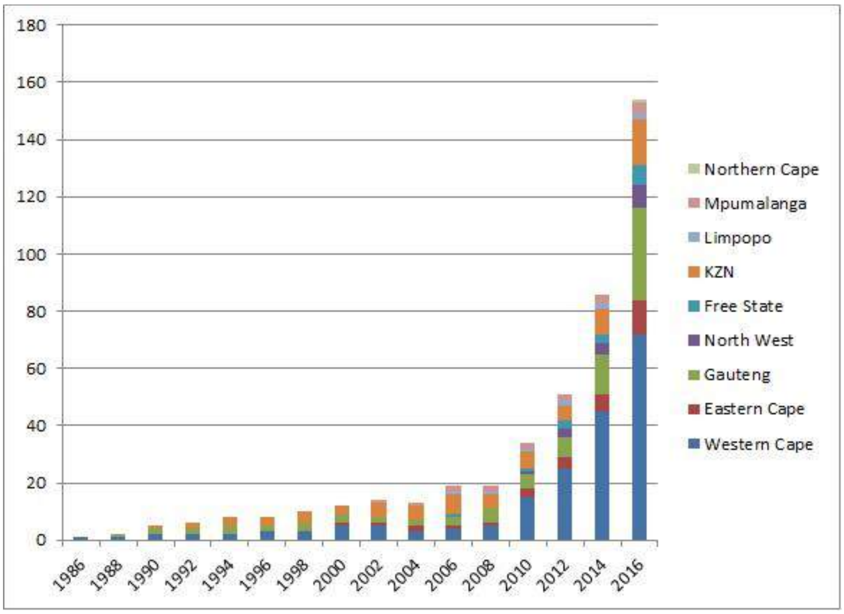


Figure 2-1: Growth of craft breweries in South Africa over the years [6].

Figure 2-1 clearly shows that the majority of the craft breweries built are in the Western Cape province of South Africa followed by the Gauteng Province. It also shows an exponential growth in the number of craft brewery operations registered in South Africa.

In recent years, the total volume of beer has increased by 1.5% compared to the overall growth in total liquor consumption of 2.0% per year and considering that the population growth in South Africa has grown at a rate of over 2.0% per annum, this means that the per capita consumption of liquor has declined overall [8].

The South African craft beer scene growing from a relatively small base compared with other markets reported an estimated 30 per cent growth in 2015, followed by an accelerated 35 per cent growth in 2016 and estimates of up to 18 million litres of craft beer by the end of 2017 [2]. Given the high rate of growth in craft beer, one can infer that craft beers are gaining market share in South Africa from the conventional beer market where large corporates like ABInbev and Heineken are operating.

2.2 Manufacturing Strategy

Manufacturing strategy is summarised as the total set of decisions that shape the long term capabilities of an organisation's operation and how they contribute to the overall business strategy [9]. Quantifiable strategic objectives that indicate the "end state" that the organisation hopes to achieve by addressing what it deems to be the opportunities available in the market and the challenges in achieving its objectives are put together. When these have been put together some strategic intents are laid down to drive a path of achieving these objectives through work programmes and control measures which are put in place to monitor progress against timelines [10]. There must therefore be an alignment between the overall business strategy and the more specific manufacturing strategy. To achieve this level of alignment, the manufacturing strategy's main objectives must not only be based on developing capabilities in the areas such as cost, quality, reliability, and flexibility, but must have an overarching purpose of realising long term increasing market share and profitability [11].

Manufacturing strategy undertakes to look at developing capabilities in fulfilling performance objectives for manufacturing that may include one or more of the following objectives, namely, quality, cost, flexibility and reliability of the organisation.

2.3 Manufacturing strategy frameworks

2.3.1 The Framework

A framework consists of the selected theories that underlines the thinking with regards to how one understands the topic which in this case is manufacturing strategy, as well as the concepts and definitions from that theory that are relevant to the topic [13]. The most common manufacturing strategy framework found by the researcher consists of how strategy is made and the constituents of a manufacturing strategy as shown in Figure 2-2.

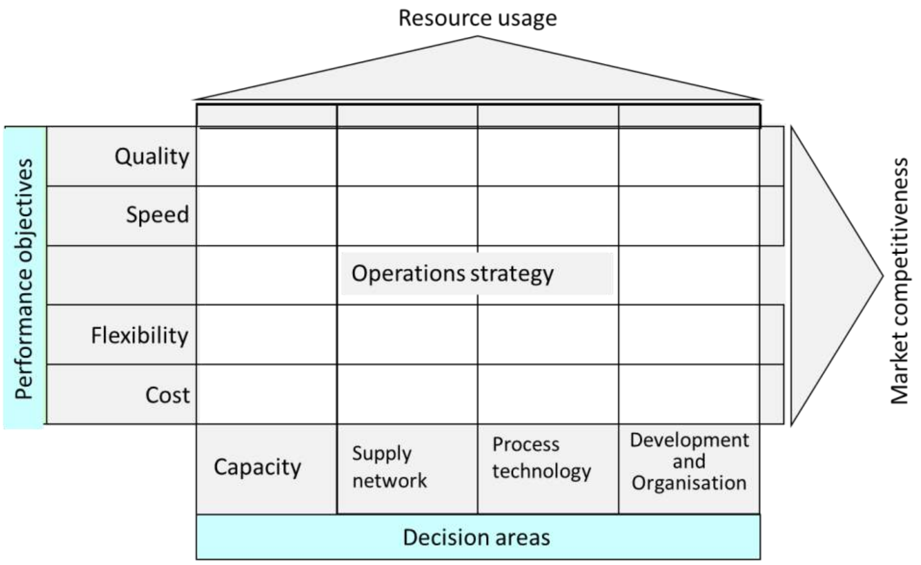


Figure 2-2: Manufacturing strategy theoretical framework [9].

The concepts and constituents of the manufacturing strategy as presented in Figure 2-2 are explained in the following subsections. Evidence of manufacturing strategy will be presented when a firm deliberately has a total pattern of decisions on all the interlinked constructs to improve operations and sustainability.

2.3.2 Strategic Decision Areas

2.3.2.1 **Brewery Capacities**

2.3.2.1.1 LOCATION OF THE MANUFACTURING FIRM

Breweries are demand driven operations where large proportions of the volume of beer they produce are distributed as perishable goods. As a result, the distribution costs of a brewery tend to account for a large proportion of the costs of running such a business. In order to decrease the costs associated with distribution, it is important for the manufacturing plants to be located in population centres. These decisions are made early on in the planning phase of building a brewery to define the ideal location. Locations of demand driven operations are chosen primarily in proximity to labour, transportation and technology while meeting procurement/distribution requirements of the firm [14]. Manufacturing firms also need to consider the proximity to the supply of water, electricity, waste streams and supply of raw materials which may affect the quality and costs of the operations and products. The location of a brewery in general forms part of a strategic decision that forms part of the overall manufacturing strategy [15].

2.3.2.1.2 CAPACITY OF THE MANUFACTURING FIRM

Capacity decisions for any brewery are highly interconnected with the quantity of beer it is able to produce and the flexibility it is able to do. This is determined by the equipment and human capital that is currently employed by the firm [12]. In a conventional manufacturing operation, important capacity decisions include how to deal with cyclical demand that is inherent in the beverage sector, whether to add capacity in anticipation of forecasted market conditions and how to use capacity to improve the economies of scale of the organisation [16]. This is to be investigated at craft brewery level.

2.3.2.2 **Process Technology**

A brewery like any manufacturing system is an input-output system in which manufacturing resources such as raw materials and energy are transformed into products. One way by which breweries adapt to the environmental dynamics is the engagement in manufacturing process technology. Employing the correct technology in manufacturing can result in reduced operations cost, improved quality of products and flexibility [17]. There is a requirement for manufacturing firms to adapt production processes to the latest technological trends and developments and adapt to environmental dynamics including short product life cycles, growing product complexity, and rapid advances in technologies by continuously engaging in manufacturing process innovation [18]. There have been major developments in the industry in recent times which include refrigeration and pasteurization to preserve the beer quality, with today's advances having to do with brewery equipment and quality control. Most large commercial breweries are built with stainless steel material of construction for equipment in direct contact with the beer and have advanced heating and cooling systems. In a brewing environment, the employment of process technology will enhance other manufacturing strategy objectives such as cost, quality and reliability. The extent of sophistication of the process technology largely depends on the ability of the firm to invest capital to realise the benefits. Such resources are generally not available to firms on a small scale such as craft brewers in contrast with larger corporations that can invest heavily in process technology to automate tasks and realise benefits associated with the investment. As equipment becomes more specialized, it also becomes more expensive. Moreover, upgrades may be necessary to maintain competitive advantage; however, a reasonable return on investment may be more difficult to achieve for the smaller and less resourceful craft brewers [19].

2.3.2.3 **Organisational Structure**

The degree of decentralization and "divisionalization" of the organisation affects the allocation of decision-making responsibilities among department managers, whereas the relative use of cost and profit centres within a business unit affects the allocation of decision-making responsibilities among business unit managers and functional manufacturing managers [16]. All the factors that can impact on the designing of the organisational structure should be well analysed and decisions made as part of strategy formulations to make them fit for purpose for that particular organisation [20]. McMillan [21] states that there is a strong link between organisational structure and the size of the organisation's facilities, process technologies employed, operating environment and culture. It follows that the organisation structure that is in place in craft breweries should be informed by variables it affects and compliments as part of its overall manufacturing strategy.

2.3.2.4 **Vertical integration considerations**

According to Perry [22], "a firm can be described as vertically integrated if it encompasses two single output production process in which the entire output of the upstream process is employed as part or all of the quantity of one intermediate input into the downstream process; or the entire quantity of one intermediate input into the downstream process is obtained from part or all of the output of the upstream process". The decision to vertically integrate involves the replacement of a supply network function over which the brewery managers have limited control internally [16]. In doing so, the respective function then forms part of the responsibility of the managers in the firm. In the context of craft brewing, vertical integration may sometimes not be considered due to resource constraint but there may be a few areas available where the output of the craft brewery is partially employed in the downstream distribution or supply chain network that is managed by the craft brewery organisation. In a study conducted in the United Kingdom, over 80% of the craft brewers sold all their beer within a 65 kilometre radius of the brewery [23]. In the same study, it was also found that around one in five craft breweries had either acquired or leased at least one pub for the sale of beer produced from their craft brewery. It appears therefore that craft brewers are open to the acquisitions of pubs or restaurants where their beer is sold. When considered as part of the overall manufacturing strategy, the decision to vertically integrate may be an important one for craft brewers to remain competitive and other avenues of vertical integration may be a possibility to explore.

2.3.3 The theoretical framework applied to breweries

The theoretical framework presented shows the elements of manufacturing strategy to be considered in a manufacturing firm. It is however not plausible for all of these elements to be made priority for the success of craft breweries given the limited resources available to implement. The framework can be used by manufacturing firms to assist in quickly determining which objectives are to be prioritised to improve operations. Figure 2-3 shows how the conceptual framework is applied in the commercial brewing industry to identify immediate objectives for improved brewing based on the successes of companies such as South African Breweries.

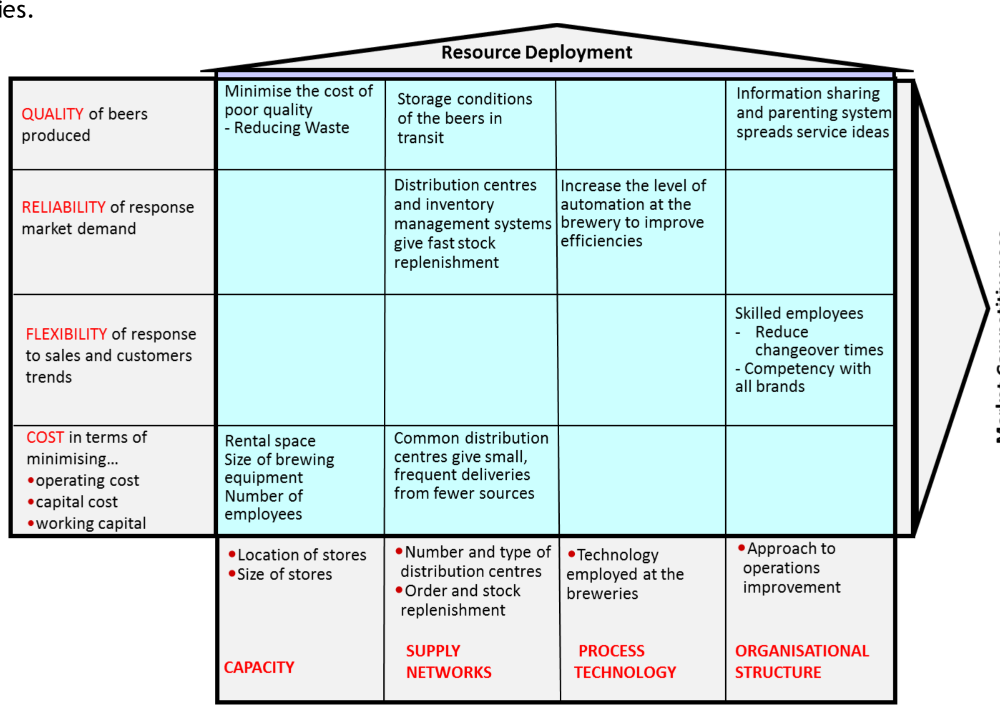


Figure 2-3: Manufacturing strategy framework applied to breweries [9].

2.3.3.1 Resource deployment

Manufacturing strategy is executed at different levels of an organisation and depending on the size of that organisation, this can involve one person or several people. In the case of craft breweries, it is expected that this execution will be carried out by relatively few people. Using the 'resource-based' view of strategy helps decision makers in manufacturing strategy identify resources that may lead to sustainable competitive advantages in the areas of quality, cost, flexibility and reliability and how those resources may either be acquired or developed internally [24].

2.3.3.2 Market Competitiveness

The pursuit of market competitiveness involves selection of resources to develop within the firm and decision how to accumulate them. Firms have to decide how to deploy their unique resources in order to fully realize their potential rents [25]. The consideration of the different deployment alternatives requires firms to fully exploit the potential rents associated with their valuable assets.

3 RESEARCH METHODOLOGY

3.1 Research methodology/paradigm

The chosen method of study includes descriptive and quantitative approach to answer the exploratory research question. The literature shows that the fundamental principle of mixed methods research is that multiple kinds of data can be collected with different strategies and methods in ways that reflect complementary strengths and non-overlapping weaknesses, allowing a mixed methods study to provide insights not possible when only qualitative or quantitative data are collected [28]. The strength of the qualitative method component in this particular case being that craft brewers may tend to have their story that have a similar timeline which would not necessarily be elicited by quantitative method component. Also, to understand in-depth whether craft brewing has had an impact on commercial brewing would require some understating of several indicating variables that can be objectively measured by quantitative analysis.

3.2 Research methodology design

The study intends to complete a survey using a semi-structured questionnaire on consumers that indulge in the experience of craft brewing and consumption of craft beers. A separate survey targeting craft brewery owners or managers will be used investigate a large variety of factors and relationships and how they influence each other, or in a scenario where no basic fundamentals exist to show which factors and relationships are more important such as the process of manufacturing strategy [29].

3.3 Population and sampling

A minimum of 400 respondents to the survey were targeted as individuals who indulge in beers. These individuals were selected conveniently (see elaboration below) in Gauteng, Western Cape and Kwa-Zulu Natal cities and formed part of the sample that represented the population. The population itself comprised of South African alcohol consumers over the age of 18. It is suggested in the literature that where the population exceeds 1 million individuals, a sample of over 380 will be sufficient to draw reliable conclusions [30].

At least five (5) craft breweries owners or managers were approached to complete the survey. The required permission from the said owners was sourced accordingly for participation.

Convenience sampling was used because questionnaires were distributed to consumers that were found at a specific location. Sampling units were selected by virtue of being at a specific place at a specific time where the surveyor had been located. The surveyor was during the course of data collection placed at different premises of alcohol sale and consumption areas where the target population was most likely to come. Of the total sample obtained, 259 (62.6%) were male and 155 (37.4%) female, fairly similar to Brooks [31] findings in USA. 179 (43.2%) respondents were in Johannesburg, 118 (28.5%) in Cape Town and the balance of 117 respondents (28.3%) from Pretoria, Durban, Vereeniging and Rustenburg. Similarly, convenience sampling was used for the selection of craft brewery owners selected to participate in the study. They are selected due to the geographic location of being close to Gauteng Province which is within reach of the conveyor. This method has some shortcomings in that not every South African alcohol consumer has an equal chance of being selected for participation. The rationale and advantage of such an approach is that it will save time and other resources that are not available for the research to be completed. Consequently, inference to the entire population using the results of this study must be cautiously done.

3.4 Ethics clearance

Ethical clearance that guides how participant information is to be protected and how consent from participants will be sort has been obtained from the University of the Witwatersrand, Johannesburg. A copy of the ethical clearance is available at the school.

4 RESULTS

4.1 Consumer results

24.9% (or 103 respondents) thought that craft beer was “very expensive” while 72.0% thought that craft beer was “expensive” compared to other beers. Only 2.9% (or 12 respondents) thought that craft beer was comparable with other beers when it comes to price per unit.

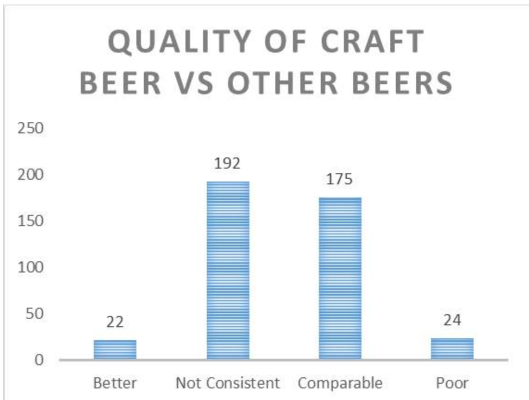


Figure 4-1 shows the perceived quality of the craft beers and price per unit compared to other beers.

In terms of accessibility 48% of the respondents had poor access to craft beer and a further 43% believe that craft beer is generally inaccessible. The remainder of the respondents had generally good (5%) and “ok” access to craft beers. On the issue of quality of craft beers, 46.4% (192 of the respondents) thought that the quality of craft beers is not consistent compared with other beers, while 5.8% (24 of the respondents) believe craft beers have a poorer quality in comparison to other beers. The balance of 47.6% (197 respondents) believe that the quality is comparable with that of other beers.

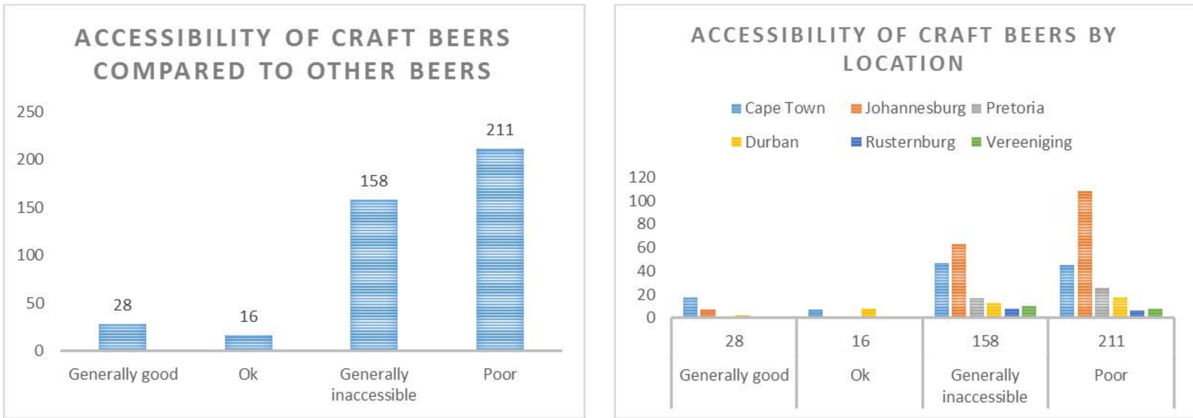


Figure 4-2 shows accessibility of craft beers in the market as reported by consumers.

4.2 Craft breweries capacity and capabilities assessment

The capacities of the responding breweries were studied and the results were such that Craft Brewery A at current production levels of 250 000 litres of beer per month had massive capacity compared with the rest of the other breweries.

The other breweries had much less capacities with current production levels ranging from 9600 litres per month to 32 000 litres per month. The capacity of a brewery is determined by the size of each batch that can be produced by the brewery and complemented by the processing equipment and the quantities of human capital operating the equipment that is currently in the employ of the craft brewery. Out of the five breweries, only Craft Brewery E was running at full capacity compared with the design capacity. Craft Brewery D was running at 40% of design capacity, Craft Brewery B at 80% capacity and Craft Brewery C at 76% capacity.

Of the five craft breweries three of them had been in operations for 3 years, namely Craft Brewery B, C and E. Craft Brewery A has been in operation for one year and Craft Brewery D for only half a year. It must be noted that these numbers represent the time since the first saleable product was commissioned and that planning of the operations started years earlier for all the craft breweries.

4.3 Determination of size and location

The pattern of decisions regarding the choice of equipment sizing to determine the overall size of the manufacturing facility were studied. All the craft breweries that responded had a variety of reasons that were considered when determining the final size of the manufacturing plant and its location.

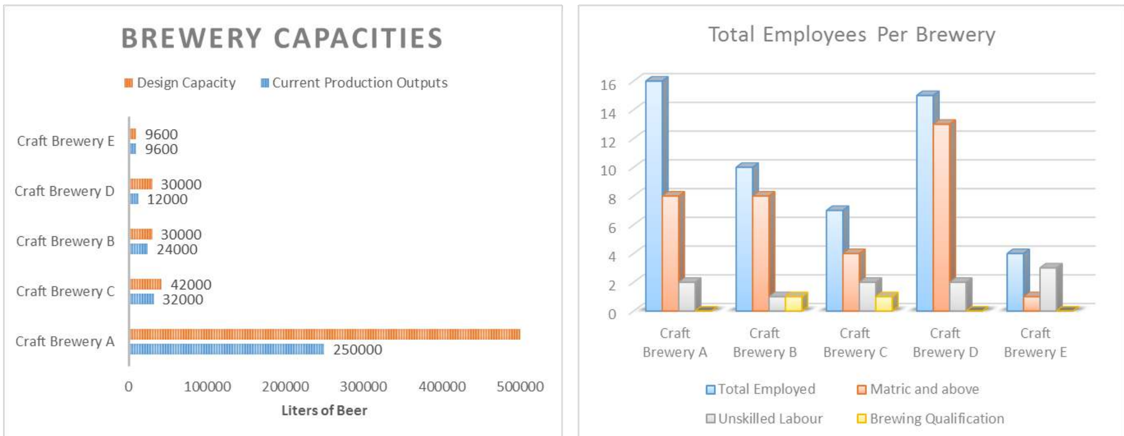


Figure 4-3 shows capacities of each brewery and number of employees at each brewery and their respective skill level that responded to the survey.

The smaller firms in the case of Craft Brewery B, C and E were largely influenced by affordability of the equipment that was required and the affordability of the space that was to be used as the location of the plant. The larger firms in Craft Brewery A and D were largely influenced by the market size they wanted to reach. In terms of staff, Craft Brewery A employed the most employees at 16 followed by Craft Brewery D at 15. Craft

Brewery B brewery employed 10 personnel; there were 7 employees at Craft Brewery C and a total of only 4 employees at Craft Brewery E brewery. Only Craft Brewery B and C had a person with previous brewing experience as well as brewing qualifications with previous careers at a large commercial brewery.

4.4 Process technology decisions

The type of process technology employed at each facilities was recorded as follows:
Craft Brewery A: The brewery is mostly automated at approximately 80% automation of all activities that go into brewing the beer. The plant is divided into the brewhouse facilities that largely handles hot product from the cooking process. This is followed by the cellars area where the actual fermentation and maturation of the beer takes place. The filtration and bottling plant then follow. The supporting facilities such as the refrigeration plant, water treatment and heating plant are all fully automated. The cleaning component of all the product handling equipment are fully automated.

Craft Brewery B and Craft Brewery D: Both are about 50% automated with the rest of the brewing activities completed manually by the brewing operators. In both cases cleaning of the beer handling facilities is manually undertaken by the operator. To this extent, some consistency of achieving the same amount of cleanliness is reduced. The supporting facilities such as the heating plant, refrigeration plant and water treatment are all handled automatically.

Craft Brewery C and Craft Brewery E: At these two breweries, there are more manually completed brewing activities than automated ones. Only activities that cannot be performed by an operator are automated such as the heating plant and refrigeration plant. The cleaning of the plant is manually completed by the operators. Most of the beer transfers between different stages of the process are also manually handled.

4.5 Vertical integration consideration

Some of the breweries that responded were found to have considered vertical integration to some extent in making decisions regarding risk, product quality, the cost structure of the brewery and the degree of focus of the organisation.

Craft Brewery E: Has had no thoughts of vertically integrating any part of the supply chain given the brewery's relatively small size in terms of quantities of beer produced. Raw materials such as malted barley and hops are largely procured through SAB Ltd. Distribution is outsourced to a company that specialises in this field. There is no on premise sales licence for beer to be drank on site and the location of the brewery is in an industrial area.
Craft Brewery C: Its major raw materials are sourced from SAB Ltd and packaging materials are sourced from Console Pty Ltd and Coleus Pty Ltd. Approximately 30% of the beer made out of this brewery is sold at a restaurant which is a sister company of the brewery. Also, about half of the beer made from the brewery is contract brewing. This means that customers in essence rent out the facility to produce their own beer which they will market, distribute and sell completely separately from the brewery.

Craft Brewery B: Similar to Craft Brewery C, its major raw materials are procured from SAB Ltd and packaging materials are also sourced from large corporations such as Coleus and Console. A large quantity of the beer produced here is sold in kegs which are re-usable. In the past, the brewery has tried to distribute internally but found it practically impossible to distribute reliably as demand increased. The distribution was later outsourced. A small percentage of the beer is sold through the beer garden that is on the same premise as the brewery. The location of the brewery is ideal for such an arrangement. The brewery is currently considering opening sales outlets in and around the local area that will exclusively sell products from the brewery. This requires large capital investments and is currently on hold.

Craft Brewery D: The decision was deliberately made to have international option for the supply of brewing raw materials like malted barley and hops. This gives the brewery a differentiation option as well. The brewery has invested resources in developing an in-house restaurant and beer garden where currently most of the beer is consumed. This decisions means that beer does not have to be packaged and thus saves cost accordingly. The brewery has no intention to do contract brewing at this stage as this would take away focus from developing its own brands.

Craft Brewery A: Given issues previously experienced by the company regarding quality of supply of the raw materials, a decision was made to source raw materials from outside South Africa. Negotiations resulted in the ability to match prices for raw materials procured outside to be similar to raw materials in South Africa. The brewery was built to a relatively big size on the basis that it will do mostly contract brewing in addition to in-house brands. While the intention was not to own the distribution part of the operation, it was decided very early on to partner with a sister company in this regard. As a result there is better control of costs involved with the distribution part of the operation. Currently over half of beer produced here is brewed under contract brewing and the rest constitute in-house brands.

4.6 Quality assurance systems

The study investigated the quality systems that have been put in place at the responding craft breweries and found as follows:

Craft Brewery A and D: Consideration for quality is extensive as proven by investment in a quality analysis laboratory on site. A laboratory in breweries is required to be provided and maintained at the plant site with the necessary equipment and supplies for conducting quality control testing. Automated instrumentation have been installed in the process to track quality continuously. Every batch that is packaged is tasted by experienced tasters to analyse how consumers may react to the product. An extension to this is further analysis of the product in trade when the life of the beer is expected to deteriorate with time. Craft Brewery A site is a certified ISO 14001 site which is a standard for operation like craft breweries and allows for regular health and safety audits to be completed on site. Craft Brewery D is planning to be ISO certified site in the near future and thus putting all the systems in place that are required.

Craft Brewery B, C and E: There is no on site laboratory facility as this is not considered a requirement. Basic instrumentation are used to analyse some quality parameters however these are not continuous monitoring instruments. Tasting of every batch of beer packaged is completed on site before it is dispatched. The sites are not accredited with any quality certification authority and thus no regular site inspections are done to test systems that have been put in place for conformance to quality. In the specific case of Craft Brewery D and E, there is no follow up with quality in trade as the general belief is that the beer will be consumed in time before its "best before" time elapses.

All sites claim to have some form of quality system in place where analyses are performed at a predetermined rate and results are recorded. This however can only be verified by a valid accreditation by a relevant authority.

5 DISCUSSION OF RESULTS

5.1 Quality of the craft beers

The survey asked respondents about their impression of craft beer quality in which 52.2% of the response was that the quality of the craft beers tasted or drank by them were not consistent or were poor. While no literature could be found to compare these results, this is a very low satisfaction rate with quality by consumers of craft beers. It is worth pointing out that there are costs associated with failure to meet the requirements and expectations of the consumer which could include the consumer declining to use or consume the product again and in the case of craft beer this could mean the consumer opts to try alternative beers.

With regard to quality assurance systems, only Craft Brewery A had all the systems in place for the management of quality in response to the survey questionnaire. The other four breweries that responded had a mix of some quality assurance systems in place but not all. This may potentially leave a gap where quality issues are not identified and corrected. All these issues that are not resolved will in time filter through to the consumer and be reflected in the consumer's perception of the quality of craft beer.

Only Craft Brewery A and D currently have an on-site laboratory for this analysis which forms the foundation of understanding every batch of beer that is produced. The other breweries having no laboratory facilities means they are not able to have an in-depth understanding of the products that they make. The finding of this study is that there is a clear requirement for a set of decisions to improve the quality front of beer production at the craft breweries.

5.2 Unit cost of craft beer

To control the external cost influencing factors that are key inputs to the craft brewery relates to the management of key stakeholders such as suppliers and vendors. The competitive approach suggests that the development and usage of multiple sources for most or all key materials inputs to the brewing of beer will result in lower negotiated prices as suppliers are made to bid against one another. To this regard, the five responding craft breweries have indicated that they are largely dependent on SAB Ltd for all their malt and hops purchases. This is contradictory to the pattern of being competitive, as this is a sole supplier and requires a new set of decisions to be made regarding the sourcing of key materials for the manufacturing plant. Only Craft Brewery A indicated a move away from being dependent on a sole supplier for these two key raw materials to the brewing process. This decision by Craft Brewery A was primarily made for differentiation and to strengthen the supply of these materials but not necessarily to reduce cost.

Other patterns of decision that have a direct bearing on the price that the consumer will eventually pay for craft beer are internal to the craft brewery and include but is not limited to the quantity of labour employed to make a unit of beer and improvement on the cost of quality. Decision patterns that are related to these parameters must be made based on a clear understanding of how these impact the total cost of manufacturing.

5.3 Accessibility of craft beer

There are many variables that affect the accessibility of craft beers in the market and the single most important one of them is the size of the facility built to produce sufficient quantities of beer as per the demand from the market. Under the circumstances of uncertain fluctuation of market demand, the degree of matching between production capacity and customers demand can affect the cost and the efficiency of making beer available to the market.

The primary influencing factor for accessibility of beer to consumers is how the beer is packaged and taken to markets [26]. The majority of the craft beer is served in kegs and therefore not available in convenient packs made available to the consumers through conventional market channels such as the 330ml bottles and cans. Craft Brewery A was found to be the only brewery that sold these convenience packed beers and thus able to take their beers into convenience stores and other places where consumers can buy and take the beers to any place they prefer. The other craft breweries largely sold their beers in kegs which limited the reach of the market to places where alcohol is consumed in-premise.

5.4 Vertical integration considerations

The response by the craft breweries owners indicated that some thoughts towards vertical integration had been entertained. Given the relatively small sizes of the craft breweries, the power to consider backward integration on the raw material side is close to zero. This means that all the craft breweries in South Africa have to rely on SAB Ltd owned Malting Plant to buy malt from or else face the increased costs of importing malt from market overseas, which will significantly increase the costs of manufacturing.

Respondents to the research questionnaire reveal several challenges. Craft Brewery A and Craft Brewery C in particular in their response to the relationship with raw material suppliers, pointed out that the quality received from the only supplier in South Africa is inferior at best and poses a risk to the quality of their products. While Craft Brewery A was then able to move on and negotiate prices of raw materials out of South Africa, Craft Brewery C, similarly to other breweries that responded, did not have the resources to do so.

5.5 Current strategy shortfalls

Performance objectives that are directly related to cost and quality proved to be the area where most decisions were deliberately made. Most of the craft breweries studied delay investment in expensive equipment that are customary in the brewing industry at the early stages of their operations. Obtaining adequate access to capital has been proven to be one of the biggest hurdles to starting a new business and growing facilities that are already in production [27]. Financial constraints is an issue for flexible solutions to all challenges, ranging from product development problems to production problems [28]. In setting up their organisation structures, the craft breweries must develop an understanding of the requirements of the operations and job designs that will focus on understanding the roles of all resources including personnel. An optimisation of all these resources can be done but must cover all the responsibilities that are to be performed by personnel in order to deliver against the manufacturing strategy. Brewing beer requires not only very specific equipment but also very specific human resource skills to understand all the inputs and outputs of the manufacturing plant.

For Craft Brewery A, B, C, D and E:

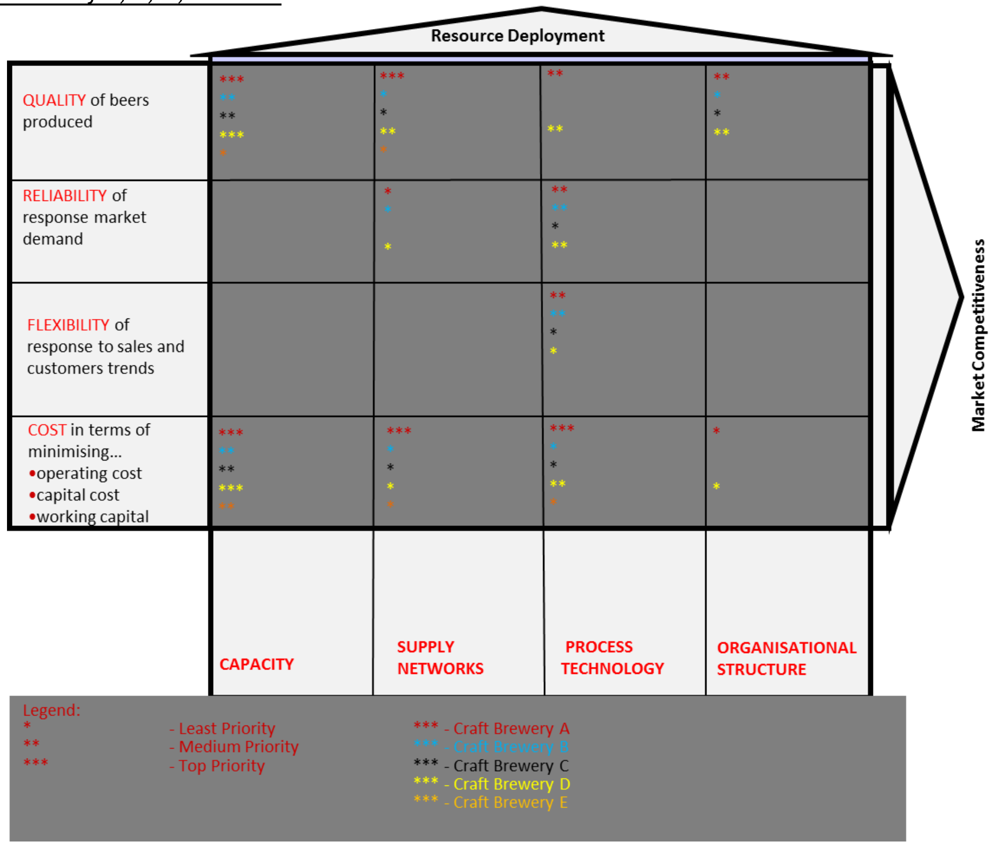


Figure 5-1 Combined Craft Breweries fit to the manufacturing strategy conceptual framework

For Craft Brewery A:

- Cost of brewing capacity – Given that this is the largest craft brewery, there has been significant investment on the capability to achieve economies of scale.
- In-house quality assurance – Craft Brewery A has an onsite laboratory equipped with the relevant brewing equipment to analyse and monitor quality.
- Product quality in transit – Craft Brewery A has developed a networking relationship with a distributor from a sister company that can be managed and monitored to keep the integrity of the beer while in transit.
- Process technology deployment – Craft Brewery A employs process technology on site and up to 80% of the brewery has been fully automated.

For Craft Brewery B:

- Cost of brewing capacity – The capacity of the brewery is approximately 32000 litres per month which is large compared to others. Significant investments into acquiring the large equipment were made as part of the decisions regarding size incurring some investment costs for capacities.
- In-house quality assurance – The required facilities have not been put in place such as an in house laboratory and all the quality analysing instrumentation that comes with it.
- Process technology deployment – At 50% approximate level of automation of all the tasks of the production, Craft Brewery B has invested in some process technology.

For Craft Brewery C and D

- Cost of brewing capacity – The equipment in the brewery was bought mostly second hand and thus capacities were limited to the existing size of that equipment. Decisions on plant capacities were thus largely based on the cost of procuring the equipment instead of designing fit for purpose equipment as per the market demands that were perceived.

- In-house quality assurance – The craft brewery owner is the main driver of quality with no major investment in quality monitoring process technology employed. Sensory analysis of quality is the main way of final check of batches of beer before they are shipped of site.

- Process technology deployment – A large proportion of tasks at this brewery are completed manually by operators. There has not been any deliberate intervention of automating sections of the brewery for cost, quality and reliability of beer production.

For Craft Brewery D:

- Cost of brewing capacity – Craft Brewery D is running at 40% of design capacity of approximately 30 000 liters per month output. The investment into the size of the brewery has been made already and now the work will be to increase production based on the market demand.

- In-house quality assurance – Accreditation is a commonly used quality management method and the focus from Craft Brewery D has been to comply with outlined process.

- Process technology deployment – To insure batch uniformity and integrity of product through each of the process, Craft Brewery D is equipped with several in process instrumentation to measure quality.

For Craft Brewery E:

- Cost of brewing capacity – The largest investment made towards the manufacturing operations of Craft Brewery E is the capacity for production.

- In-house quality assurance – The brewery has not invested in laboratory equipment to brewing analysis for monitoring quality.

- Product quality in transit – There are no systems in place to check for quality-in-transit. The products from this brewery do not have an extensive reach and therefore are not kept in transit for long.

- Process technology deployment – Production tasks are mostly completed manually by operators in the plant. There is no automation for the key tasks in relation to cost, quality and reliability of beer production.

6 CONCLUSION

There are various craft brewery attributes that seem to affect whether some or most of the decisions that are made in relation to cost, quality, reliability and flexibility are adequately dealt with in the operation of a craft brewery. Out of the demographic variables that were included in the study, size of the plant had stronger relation with all the decisions that are manufacturing strategy related.

In conclusion, the craft breweries management must understand that a manufacturing strategy must be comprehensive in the sense that it should not only guide long terms decisions that must be made for sustainability but must also provide guidelines for business managers to address the many facets of manufacturing decision making given the constantly changing environment of the brewing industry.

Given that craft breweries are in essence SMMEs identified by the government as crucial for the development of the economy, this study will provide guidance to prospective craft brewers and those that are already established in the craft brewing industry by providing some principles on what elements of strategy should be in hand during this phase of market growth.

7 REFERENCES

[1] Daud, M. A. (2013). A differentiation strategy for FMCG companies on ingredient branding. *International journal of engineering and management sciences*, 83 - 86.
[2] Palmer, K. (2017). *There's a craft beer revolution*. Cape Town: Beyond Luxury Media.
[3] Carpenter, A. J. (2015). *Craft beer & consumer behaviour*. Denver: Not Specified.
[4] Banda, F., Matumba, C., & Modliwa, P. (2015). *Research on Competition and Regulation: Competition, barriers to entry and inclusive growth*. Johannesburg: University of Johannesburg.
[5] Malhotra, S. (2014). A Study on Marketing Fast Moving Consumer Goods (FMCG). *INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH & DEVELOPMENT*, 1-3. Retrieved from http://www.ijird.com/index.php/ijird/article/viewFile/45586/36923
[6] brewmistress.co.za. (2016). *Statistics for the South African craft beer industry*.
[7] Warner, A. G. (2010). The evolution of the American brewing industry. *Journal of business case studies*, 31 - 46.
[8] Holtzkampf, E. (2014). *Liquor consumption patterns in South Africa*. Cape Town: Wineland media.
[9] Slack, N., & Lewis, M. (2011). Operation Strategy. Pearson Education.
[10] Mburu, S. N., & Thuo, J. K. (2015). Understanding the concept of strategic intent. *International Refereed Journal of Engineering and Science (IRJES)*, 58-59.

[11] Ulusoy, G. (2003). *COMPETITIVE MANUFACTURING STRATEGIES FOR THE MANUFACTURING INDUSTRIES IN TURKEY*. Istanbul, Turkey: Sabancı University.
[12] Julka, N., Baines, T., Tjahjono, B., Vitanov, V., & Lendermann, P. (2007). A review of multi-factor capacity expansion models for manufacturing plants: Search for a holistic decision process. *International Journal of Production Economics*, 607 - 621.
[13] Grant, C., & Osanloo, A. (2014). UNDERSTANDING, SELECTING, AND INTEGRATING A THEORETICAL FRAMEWORK IN DISSERTATION RESEARCH:. *ADMINISTRATIVE ISSUES JOURNAL*, 12 - 26.
[14] Henderson, J. R., & McNamara, K. T. (2000). The Location of Food Manufacturing Plant Investments in Corn Belt Counties. *Journal ofAgricultural and Resource Economics*, 680-697.
[15] Chen, L., Olhager, J., & Tang, O. (2014). Manufacturing facility location and sustainability: A literature review and research agenda. *International Journal of Production Economics*, 154-163.
[16] Fine, C. H., & Hax, A. C. (1984). *Designing a manufacturing strategy*. Michigan: Sloan School of Management, MIT.
[17] Schrettle, S. (2013). *Managing manufacturing process innovation – New manufacturing technology adoption as a dynamic capability*. Bamberg: University of St.Gallen.
[18] Lee, B. (1996). *The justification and monitoring of advanced manufacturing technology*. Management Accounting Research,.
[19] Pontinen, J. (2011). *Industry Overview: The Brewing Industry*.
[20] Fadeyi, O. (2015). *Review of Organizational Strategy and Structure*. Ota: International conference on African Development Issues.
[21] McMillan, E. (2001). *Considering organisation structure and design from a complexity paradigm perspective*. Milton Keynes: Open University.
[22] Perry, M. K. (1989). *Handbook of Industrial Organization*. Elsevier B.V.
[23] Cabras, I. (2016). *British beer: The report on the 2016 members' survey of the society of independent brewers*. Newcastle: The society of independent brewers.
[24] Almeida-Costa, L., & Dierickx, I. (1998). *The strategic deployment of resources*. Fontainebleau, France.: INSEAD.
[25] Kanagal, N. (2003). Role of Relationship Marketing in Competitive Marketing Strategy. *Journal of Management and Marketing Research*, 1 - 17.
[26] Dai, J., & Zhou, C. (2008). *Beer Distribution in China*. Atlanta: Georgia Institute of Technology.
[27] Kerr, W. R., & Nanda, R. (2009). *Financing constraints & entrepreneurship*. Harvard Business School.
[28] Dawson, C. (2002). A user-friendly guide to mastering research. *Practical Research Methods*, 14-22.
[29] Fidel, R. (1984). The Case Study Method. *A case study*, 1-9.
[30] Sekaran, U. (2003). *Research Methods for Business – A skill building approach*. Hoboken: John Wiley & Sons, Inc.
[31] Brooks, J. R. (2011). *Craft Beer Uncapped*. Advantage Publishing Ltd.

A RECONFIGURABLE ROBOTIC END EFFECTOR FOR MACHINING AND PART HANDLING – MACHINE TOOL SIMULATION

C.E. Reddy[1], J. Padayachee[2] & G. Bright[3]

[1,2,3] Department of Mechanical Engineering
University of KwaZulu Natal, South Africa
[1]212502866@stu.ukzn.ac.za, [2]PadayacheeJ@ukzn.ac.za, [3]Brightg@ukzn.ac.za

ABSTRACT

Robot machining forms a significant portion of Industry 4.0 in Smart Factories. Industrial robots require an end effector to perform tasks. The purpose of this study is to research and develop a reconfigurable robotic end effector for machining and part handling. Such a device would eliminate the need of multiple robots for part handling and lengthy end effector changes. This research will involve the complete mechanical, conceptualization, detailed design, manufacturing and testing of this new end effector. The reconfigurable platform requires modular gripper architecture and compact machine tool system. Flexible tendon cable-driven grippers are versatile, compact and offer a large degree of compliance. Vibrations at the tool-tip results in reduced surface quality while instability can decrease spindle life. The spindle will be designed to reduce chatter and enhance dynamic stability. The concept utilizes a flexible gripper system in conjunction with a compact, lightweight milling cutter capable of machining non-ferrous metals. Analytical tools such as Kinematic analysis to determine gripper-finger dynamics, Vibration analysis of the spindle-tool to determine the frequency response function, Stability Lobes to determine chatter-free cutting parameters and Finite Element Analysis to determine strength properties of the mechanical design; will be utilized to enhance and validate the concept.

[1] The author Is enrolled for an MSc Eng (Mechanical) degree in the Department of Mechanical Engineering, University of KwaZulu-Natal
[2] The author is a Senior Lecturer in the Department of Mechanical Engineering, University of KwaZulu-Natal
[3] The author is a Professor in the Department of Mechanical Engineering, University of KwaZulu-Natal

1. INTRODUCTION

South African manufacturers find it difficult to remain competitive due to the high cost of imported CNC machines among other reasons. Robot based machining provides an alternative to expensive, bulky CNC machines. Both serial and parallel machines are currently being used for robot based machining, however there are some drawbacks to using these robots for machining.

The market is in need of a compact two-in-one solution to machining and part handling. An end effector that is able to retrofit to most serial robots and parallel robots will offer a greater level of flexibility in manufacturing. Commercially available part handling systems and machine tools are specifically designed for certain precise operations. A dedicated machine tool is used for drilling, milling, and boring. Industrial grippers are each typically designed to grip objects of similar geometry, this means that multiple grippers are needed for different parts. Most small scale manufacturers cannot afford the initial investment and often do not require the level of precision of industrial machine tools and grippers.

Robot based machining applications have received recent attention from the scientific community as a low-cost alternative to CNC machines [1],[2],[3],[4]. Research is being conducted on parallel kinematic robot architectures that will enable robust machining capabilities. Nonetheless, additional research must be conducted on the design of suitable end effectors for machining applications. Robots are highly flexible devices that can be adapted to a wide variety of manufacturing applications. This flexibility is a key enabler of manufacturing competitiveness. This research will involve the research and development of a reconfigurable robotic end effector. Presently, there are end effectors on the market that are capable of machining, however, the same end effector is not capable of part handling. In order to enable part handling the robot would have to undergo a time-consuming change of the end effector, which would increase robot downtime and reduce productivity. This research aims to address the challenge of developing an end effector that is capable of both machining and subsequent part handling. Such an end effector would eliminate the need for a second robot to unload parts after machining. It would also eliminate the need for lengthy end-effector changes. The research involves the complete mechanical, conceptualization, detailed design, manufacturing and testing of this new, innovative end effector. To improve the performance of the end effector, it is designed with sensors such as accelerometers and position sensors that will enable process parameters to be optimized for high part quality.

This paper presents a guide to reconfigurable end effector development. Key insights in the design and development process are highlighted. The paper also presents a novel reconfigurable end effector design. Engineering analyses are performed on the end effector – machine tool design.

2. LITERATURE STUDY

There are limited attempts to merge both cutting and gripping into a single system. Rahman et al [5] had attempted to design a robot-farmer for gripping and cutting crops. The design incorporates a cutter and a two finger robotic manipulator. The cutter subassembly consists of the cutting blade, blade holder, and gearbox mechanism that allows the cutter to linearly extend and retract into and out of the holder. The robotic arm subassembly consists of four parts. The proposed model cuts crop similar to a human farmer. The two finger gripper is actuated by a DC servo motor.

2.1 Part handling systems: Reconfigurable grippers

According to Yeung [6], a gripper must have multiple fingers with reconfigurable positions to be capable of handling various shaped objects. Furthermore, each of the fingers must have movable joints so that the finger is able to conform to the shape of the object. The design of a reconfigurable gripper should be simple, light-weight and incorporate position correction. Simplistic designs reduce cost and increase reliability while maintaining a minimalistic design with fewer parts and mechanisms; this lowers the manufacturing cost and chance of failure. The robot payload, moments generated by the end effector, and inertia due to the change in velocity of the gripper fingers all determine the size of components. Position-correction removes the need for accurate object location and pre-planning grasps.

Much research has been conducted on the development of reconfigurable grippers. Different examples of reconfigurable grippers will be reviewed in order to establish a state of the art in reconfigurable gripping systems. Zhang et al [7] proposed a compensatory grasping design for a parallel jaw gripper. The research aimed to reposition the part to the required orientation by rotation during grasping. The design utilized four tips, two on each of the parallel jaws of the gripper. Ease of reconfiguration to handle different industrial parts is another advantage of compensatory grasping.

Riedel et al [8] developed an adjustable gripper as a reconfigurable robot with a parallel structure. The work proposed a robot system which combined an adjustable gripper with a parallel manipulator. The operation consisted of four individual phases. The main idea was to generate a closed-loop kinematic chain which is formed by the object and the robot’s fingers.

Molfino et al [9] proposed a low-cost reconfigurable gripper for assembly and disassembly tasks. The gripper mechanism was required to grasp both prismatic and cylindrical shape families. Two parallel fingers generate planar grasp while the third finger is fixed and allows cylindrical geometries to be grasped. The gripper has three degrees of freedom. The parallel fingers are actuated by pneumatic cylinders. Ziesmer [10] worked on a reconfigurable end effector for in-hand manipulation without finger gaiting or regrasping. The proposed gripping mechanism made use of the concept of a part-finger interface. The key to this concept exists when the interfaces are modelled as joints, this means that the end effector and part form a closed loop mechanism. As a result, the part-finger interface allows reconfigurability between hard finger contacts and planar contacts. Actuation of the gripper mechanism is provided by two 2-position linear actuators.

Dottore et al [11] proposed a tendon-driven modular continuum arm with soft reconfigurable gripper. The reconfigurable gripper is underactuated, it consisted of four soft fingers linked to an independent actuation module. The gripper fingers were mounted on parallel discs at 180° intervals, i.e. two fingers per disc. The discs were concentric about a centre axis, a geared DC motor was used to rotate the discs. The fingers were actuated by a tendon-driven mechanism, the cables were actuated by geared DC motors.

Makris et al [12] presented a reconfigurable gripper for dextrous manipulation in flexible assembly. The design was intended to result in a high-speed multi-fingered gripper. The gripper consisted of three fingers with a total of eight degrees of freedom. The gripper had a wide range of grasping modes, the centripetal grasp, the parallel grasp, and the encompassing and pinching grasp. Each of the fingers were actuated by a motor, the fingers were directly coupled to the output shaft of the motor. The fingertips were also motor driven at the tendons. Furthermore, the application of reconfigurability in grasping technologies was investigated. The aim of their research was to create a gripper that is capable of reconfiguration in order to achieve a wide range of grasping modes. In this approach the gripper can take one of two grasping orientations, centripetal or parallel. Each orientation allows the gripper to grasp objects of varying geometries.

Cipra et al [13] investigated a reconfigurable and foldable hexapod robot inspired by Origami. The HexaMorph is unlike most modular robotic systems with tree or open-chain architecture as it does not have a central body. It has a closed-loop form which permits the robot to seamlessly envelope itself. Since there is no interconnection of modules, the robot reconfigures its orientation by actuating servo motors. The two main methods of reconfigurabilty for motion are self-deploying and locomotive squirming.

Reconfigurable gripping technology has influenced the manufacturing industry due to the multipurpose impact of these grippers. New methods of reconfigurabilty are being understood as researchers attempt to increase productivity and efficiency in manipulation operations. Reconfigurable grippers that are able to grasp a wide variety of geometries are valuable in the automotive and aerospace industries. Figure 1 graphically illustrates the different end effectors.

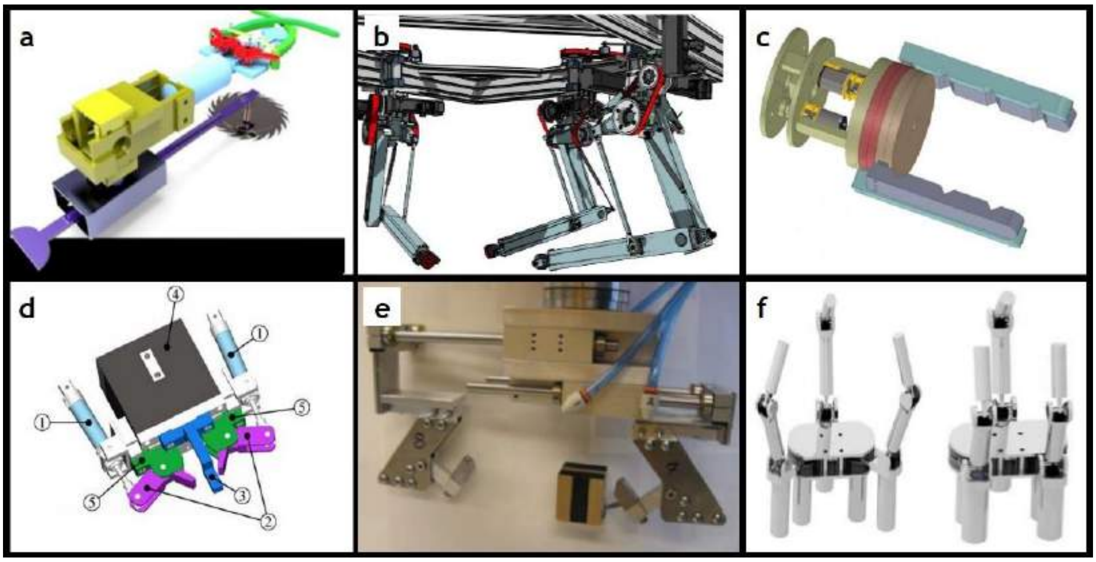


Figure 1: Different end effectors from literature study (a) Robot-farmer [5] (b) Parallel reconfigurable gripper [8] (c) SIMBA Soft Reconfigurable Gripper [11] (d) Multipurpose SPI3 gripper design [9] (e) Reconfigurable gripper prototype [10] (f) Reconfigurable gripper for dextrous manipulation [12]

2.2 Machine tool: Machining end effectors

Bologa et al [14] conducted an investigation using serial industrial robots in CNC milling processes. Point-to-point movements are easily attained by serial robots. Milling processes require accurate control of the entire path, every point on the trajectory must be reached with high accuracy. These intermediate points are calculated using linear and circular interpolation algorithms. Two key issues with utilizing serial robots for machining processes are the difficulty of path control and the low rigidity of the robotic structure.

Jie and Shu-Hui [15] performed a study on the design method for robotic drill end effector. The proposed end effector was developed to be used in conjunction with a six axis serial manipulator. A disadvantage of the serial robot relates to high compliance, this causes perturbations and deformations in machining applications. Drilling processes generate thrust forces which will cause vibration due to the serial robot joint configuration. A solution to this problem was suggested by altering the end effector to serial robot mounting platform. The clamping force should be greater than the thrust force in order to stabilize the system. A key issue in the drilling operation is the lack of accuracy in the serial robot.

Karpiel and Petko [16] proposed a mechatronic design of a parallel manipulator for milling. A novel 3-RRPRR, parallel manipulator capable of three translational degrees of freedom resulted in a high payload capacity, large workspace and increased accelerations. The manipulator design amalgamates the advantages of the Steward-Gough platform and Delta. The end effector platform has an attachment point for the mounting of milling spindles. The maximum achievable milling force is greater than 100 N, this is sufficient to machine aluminium. The design employs direct actuation to the platform, this increases the payload and overall stiffness of the robot. A key issue with parallel manipulators and machining is the complex control that is required.

3. SYNTHESIS OF A RECONFIGURABLE END EFFECTOR

The final design of the end effector is largely constrained by the method of reconfiguration that is taken forward. A brief system overview of the solution is presented in Figure 2. In order to accomplish the design target of a dual system, the method of reconfiguration is of critical importance.

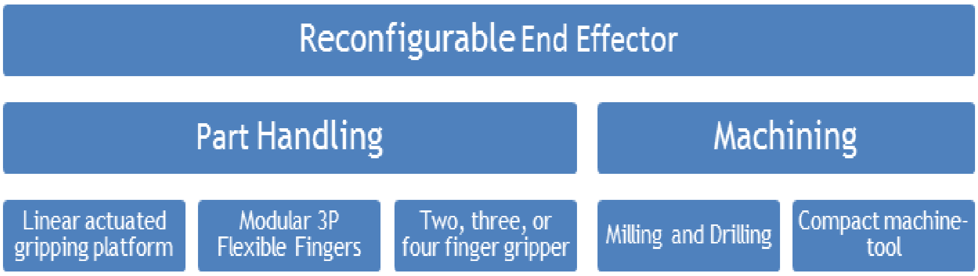


Figure 2: System overview

The proposed design incorporates both part handling and machining features on one end effector. Part handling is accomplished via a reconfigurable gripper. The gripper consists of a main actuation platform that reaches down toward the object. The platform can hold two, three and four gripper finger modules. Hence, a wide variety of part geometries can easily be manipulated. The finger modules consist of a multi-link flexible joint tendon actuated finger architecture. The addition of a flexible rubber joint allows for added compliance and dexterity of the gripper. These fingers are driven from their foot-points; hence they are compact. The linear actuated gripping platform creates the reconfigurabilty required to give the machine two poses, i.e. the grasping pose (Figure 3a and 3b) and the machining pose (Figure 3c and 3d).

Machining is realized by the compact, high-speed spindle. The spindle is designed to machine woods, plastics and non-ferrous metals such as aluminium, copper and bronze. The compact, brushless motor is coupled to the spindle via a pulley-belt drive system. The spindle is designed for high speed milling. Angular contact bearings are stacked in the front and back of the spindle shaft in the configuration. The machine-tool is designed to work with ER11 metric collets, the collet holder can clamp a maximum end mill of 7 mm diameter. The end effector can be mounted onto serial robots, parallel robots and three axis machines. The design of the unit permits easy adaption to different mounting configurations. Figure 3 shows a rendering of the final design; all subcomponents can clearly be seen.

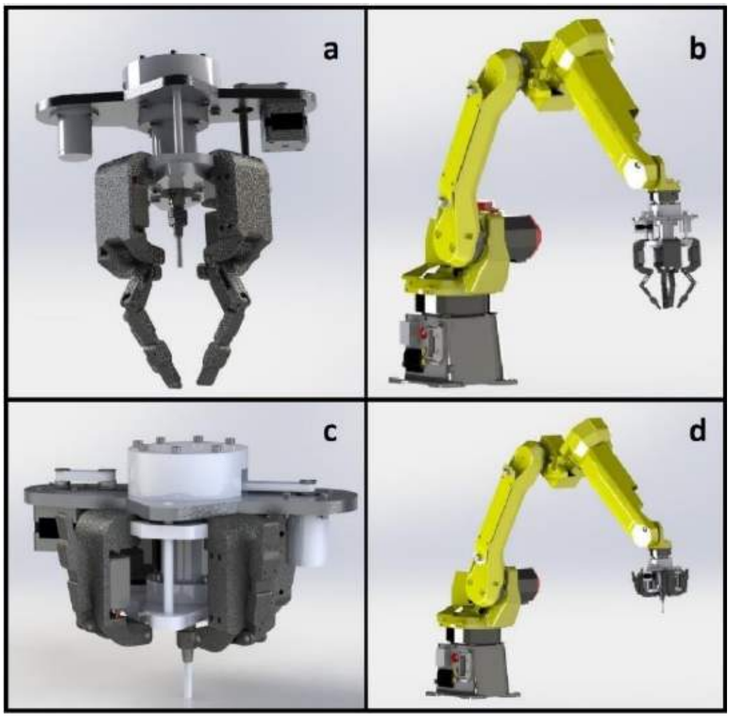


Figure 3: CAD model of reconfigurable end effector

The machine tool design comprises of a high speed brushless motor, machine spindle and pulley-belt drive coupling. In order to understand the principles required to successfully design the system, Madea's [17] expert spindle design system was used in conjunction with other theories. Cutting conditions such as: Tool geometry, workpiece material data, cutting speed, spindle speed, depth of cut and width of cut were initialised. Thereafter, the cutting requirements are computed based on the cutting condition data. These requirements are torque, power and speed. Based on the calculated cutting requirements, the power transmission system is selected and bearing span optimization can be conducted. The method is presented in Figure 4.

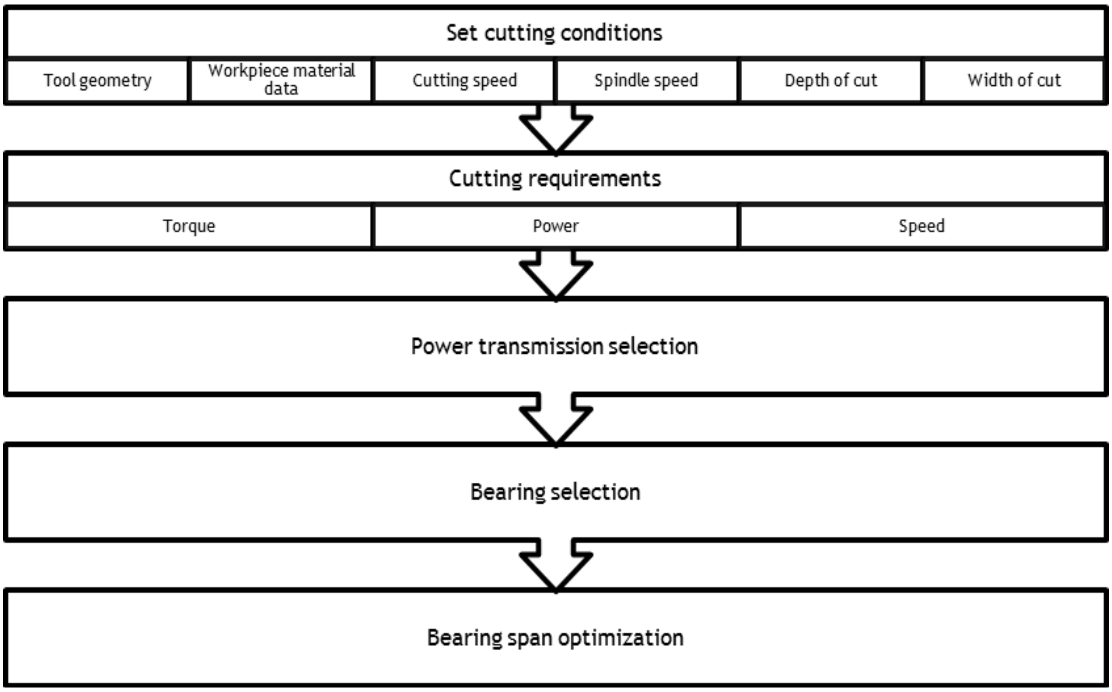


Figure 4: Spindle design methodology

The mechatronic system overview in Figure 5 illustrates the overall system layout. The Arduino DUE development board will be used to control the motors. The stepper motor which linearly actuates the gripping platform –

provides the gripping pose when fully extended and provides the machining pose when fully retracted. Each gripper finger module (maximum of four) will be tendon driven with the actuation of a micro servo motor. The high speed brushless motor will be switched on/off and speed controlled by the ESC and Arduino board.

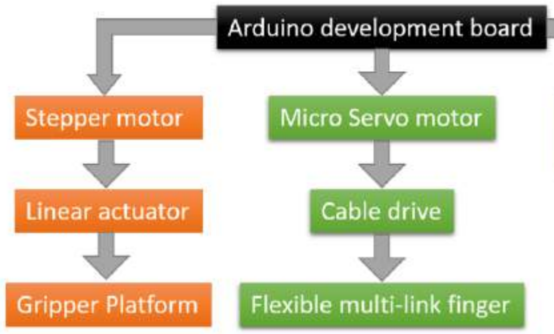


Figure 5: Mechatronic system overview

4. ANALYTICAL MODELLING OF END MILLING FORCES

In order to proceed with machine-tool design, the cutting forces are needed. The cutting forces constrain the spindle design as motor power and torque are strongly related to cutting forces. Altintas [18] developed a mechanistic cutting model to determine the forces in the end milling process. Tangential, radial, and axial forces acting on a differential flute element with height dz are given by the following equations:

$$dF_{t,j}(\varphi,z) = [K_{tc}h_j(\varphi(z)) + K_{te}]dz \quad (4\text{-}1)$$

$$dF_{r,j}(\varphi,z) = [K_{rc}h_j(\varphi(z)) + K_{re}]dz \quad (4\text{-}2)$$

$$dF_{a,j}(\varphi,z) = [K_{ac}h_j(\varphi(z)) + K_{ae}]dz \quad (4\text{-}3)$$

The elemental forces are resolved into feed, normal, and axial directions:

$$dF_{x,j}(\varphi_j(z)) = -dF_{t,j}\cos(\varphi_j(z)) - dF_{r,j}\sin(\varphi_j(z)) \quad (4\text{-}4)$$

$$dF_{y,j}(\varphi_j(z)) = dF_{t,j}\sin(\varphi_j(z)) - dF_{r,j}\cos(\varphi_j(z)) \quad (4\text{-}5)$$

$$dF_{z,j}(\varphi_j(z)) = dF_{a,j} \quad (4\text{-}6)$$

These forces were used to simulate the cutting forces generated in the end milling of Aluminium 6061-T6. Table 1 indicated the assumptions that were made based on the work material and cutter geometry:

Table 1: Cutting force simulation assumptions

Variable	*Symbol*	*Assumption*
Tangential cutting force coefficient	K_{tc}	805 MPa
Radial cutting force coefficient	K_{rc}	418 MPa
Axial cutting force coefficient	K_{ac}	227 MPa
Tangential edge force coefficient	K_{te}	6 N/mm
Radial edge force coefficient	K_{re}	6 N/mm
Axial edge force coefficient	K_{ae}	1 N/mm
Start immersion angle	φ_{st}	85°
Exit immersion angle	φ_{ex}	100°
Number of flutes	N	2
Helix angle	β	30°
Edge contact length	a	2 mm
Feed rate	c	0.1 mm/tooth
Cutter diameter	d	6 mm

To compute the total cutting force, the cutting flute is segmented into a finite number of differential elements along the cutting flute curve. The total cutting force components acting on the flute at an instant are obtained by numerically integrating the force components acting on a differential element [19]. The tangential, radial, and axial cutting forces were solved using equations 4-1, 4-2, and 4-3 and the constants tabulated in Table 1. These forces were integrated over an axial depth of cut to produce the three elemental cutting forces using equations 4-4, 4-5, and 4-6. The computation was done in MATLAB. The red curve (top most) represents the normal cutting force, the green curve (middle) represents the axial cutting force. The blue curve (lower most) represents the feed cutting force. All three cutting forces are plotted againt the immersion angle of the tool. The three normalised cutting forces are plotted against immersion angles in Figure 6.

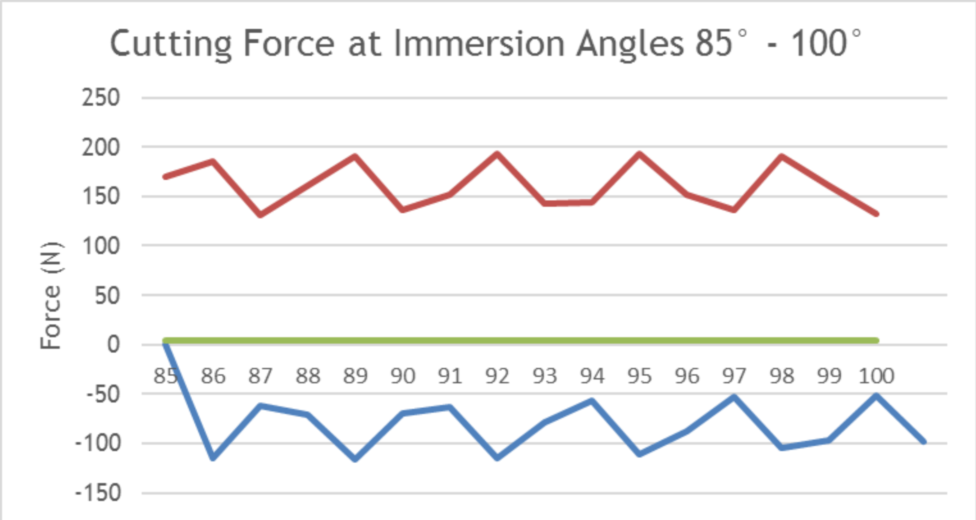


Figure 6: Simulated cutting forces

5. RESULTS

Static and dynamic finite element analyses were conducted to validate the spindle shaft design.

5.1 STATIC ANALYSIS OF THE SPINDLE SHAFT

Metal cutting is a precise process; machine tools are often used for prolonged periods. As such, machine tool failure as a result of fatigue fracture of the spindle shaft is frequently observed. There are however, cases of machine failure as a result of the spindles large deformation self-excited vibration which is caused by the cutting force. Therefore, the static design of the milling spindle unit is mainly related with the static stiffness of the spindle which is referred to as spindle stiffness. The spindle stiffness relates closely to load capacity and vibration resistance, which is a crucial performance index of the spindle unit. Spindle stiffness includes both the axial and bending stiffness. Normally, bending stiffness is more significant than axial stiffness. This is due to the large lateral loading that occurs from the vertical milling process [20].

A static structural analysis (NX10, SOL101, Linear) was conducted to determine the Von-mises stress and maximum deflection of the shaft. A 3D tetrahedral mesh CTETRA10 with 4 mm element size was used, the simulated material is AISI 4340 alloy steel. Figure 7c indicates load conditions of normalised forces at the tool-tip (F_x, F_y and F_z). Figure 7d shows the non translational constraint on the radial surfaces where the bearings are in contact with the shaft. Figures 7a and 7b indicate the Von-mises stress and maximum deflection respectively. The largest deflection occured at the tool-tip while maximum stress occured at the second step of the shaft (collet holder – front bearing collar).

Furthermore, simulations were conducted for different mesh materials. This was done to verify the material choice of AISI 4340 alloy steel. This steel did not generate the most significant safety factor but is chosen due to its ease of machinability as compared to AISI 1005. Table 2 gives the results of these simualtions.

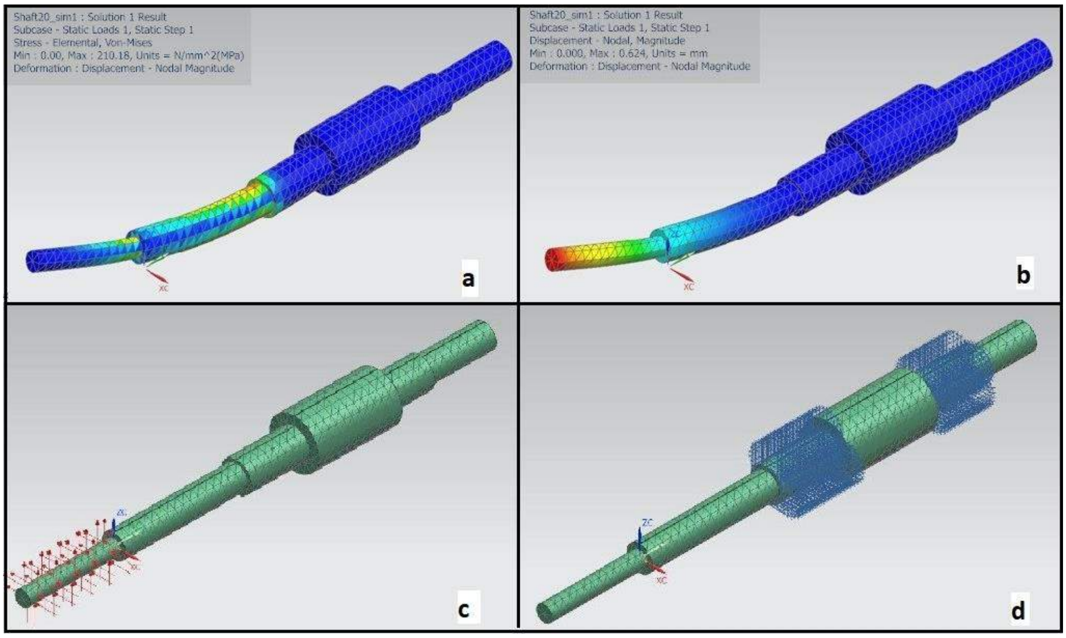


Figure 7: FEA simulation results, loads and constraints

Table 2: Static structural analysis results for different shaft materials

Material	**Deflection (mm)**	**Elemental Stress (MPa)**	**Yield Strength (MPa)**	**Safety Factor**
AISI 4340	0.624	210.18	710	3.378
AISI 1005	0.603	212.10	1034	4.875
AISI 310 SS	0.622	209.25	520	2.485
AISI 410 SS	0.549	211.09	415	1.966
Aluminium 6061-T6	1.738	208.24	276	1.325

5.2 DYNAMIC ANALYSES OF THE SPINDLE SHAFT

Vibrations are definited in metal cutting operations. These vibrations can change the relative position of workpiece and milling cutter to affect the machining accuracy. It also accelerates the wear of the milling cutter which further influences machining accuracy. Research [21],[22],[23] shows that the processing quality depends on the vibration produced by the machine. High speed machine tools with high accuracy can be significantly influenced by vibrations. Therefore, the modal analysis of the spindle is the primary issue of dynamic characteristics [20].

The natural frequencies of the spindle shaft were determined. Figure 8 displays the graphical results of the twelve different mode shapes. Non-translational constraints were applied to the radial surfaces where the bearings are located. The frequency and nodal displacements are tabulated in Table 3.

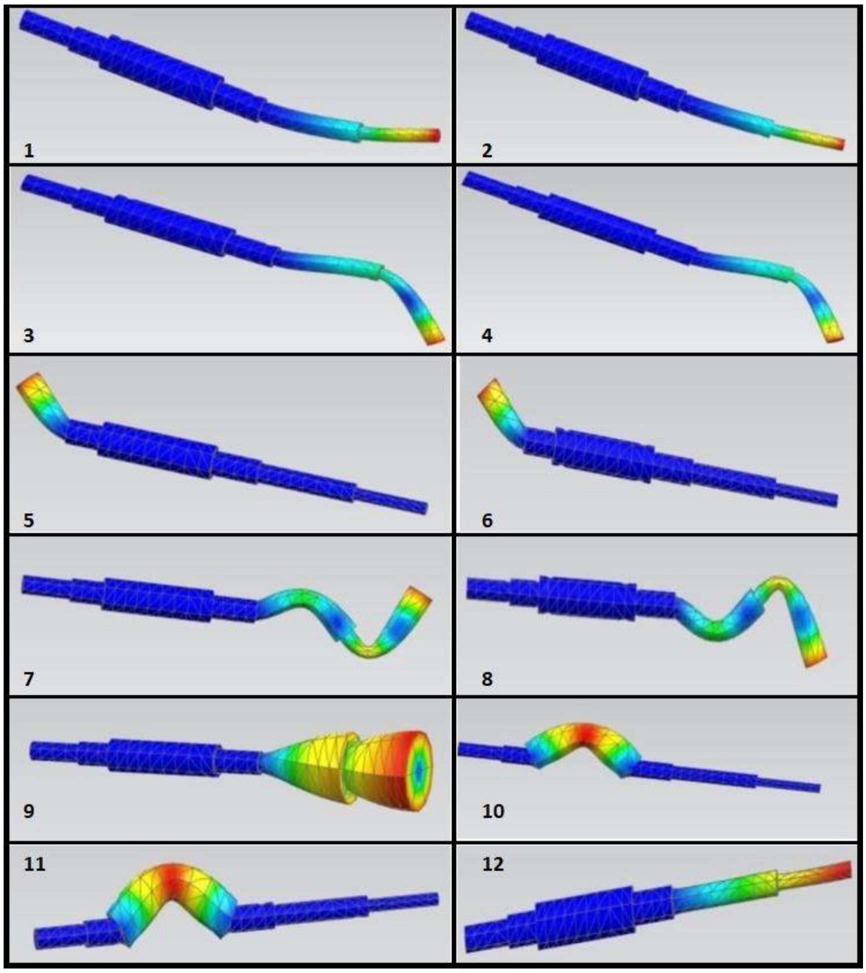


Figure 8: Modal frequency simulation of spindle shaft

Table 3: Simulated mode frequencies and displacements

Mode	Frequency (Hz)	Nodal displacement (mm)
1	1480	19.29
2	1481.35	19.29
3	5152.09	21.57
4	5155.58	21.57
5	10002.6	20.97
6	10042.9	20.98
7	14639.7	18.32
8	14657.7	18.29
9	15350.9	14.66
10	17082	6.209
11	17168.6	6.225
12	21498.2	12.28

6. FUTURE WORK AND DISCUSSION

A considerable amount of developement is still required to validate the system design. The stability model for end milling process will ensure that the correct speed and depth of cut are selected for chatter free machining. The avoidance of chatter will allow decreased vibration of the machine-tool hence, insuring a good surface finish. A prerequisite to calculating the chatter stability lobes involves obtaining the transfer functions of the machine tool system and evaluating the dynamic cutting coefficients for a specific cutter, workpiece material, and radial immersion of cut [18]. Once these are known, the following algorithm will be used to calculate the stability lobes:

1. Select a chatter frequency from transfer functions around a dominant mode
2. Solve the eigenvalue equation xy

$$\boldsymbol{a_0\lambda^2 + a_1\lambda + 1 = 0} \quad (6\text{-}1)$$

3. Calculate the critical depth of cut using equation xx

$$\boldsymbol{a_{lim} = -\frac{2\pi\lambda_R}{NK_t}(1+\kappa^2)} \quad (6\text{-}2)$$

4. Calculate the spindle speed for each stability lobe $k = 0,1,2 \ldots$

$$\boldsymbol{T = \frac{1}{\omega_c}(\epsilon + 2k\pi)} \quad (6\text{-}3)$$

$$\boldsymbol{n = \frac{60}{NT}} \quad (6\text{-}4)$$

5. Repeat the procedure by scanning the chatter frequencies around all dominant modes of the structure evident on the transfer functions.

A dynamic response analysis will be conducted to determine the operating frequencies of the spindle while it is in operation. These operational frequencies can then be compared to the natural frequencies of the spindle. If it is found that the operating frequencies are unrelated to the natural frequencies, the design is safe. Harmonic response analysis will therefore be conducted to validate safe operation of the machine tool.

Bearing stiffness should be incorporated into the dynamic model of the spindle system. For a high speed spindle unit and machine tool structure, bearings with rolling balls create non negligable effects. In these components, the interface between rolling balls and ball grooves are critical points affecting the mechanical characteristics such as the static or dynamic stiffness. The accurate approach in simulation is to factor-in the rolling elements within the ball grooves of the linear component. The rolling interface will determine whether the simulation results approach the real characteristics of the system.

The mechanical design and components of control have been completed, structural and dynamic finite element analyses have been conducted. Key insight include the role of chatter prediction and prevention on a machine-tool design, and spindle design and stiffness optimization.

7. CONCLUSION

The research project presents an innovative reconfiguration technology that is currently not exploited by the manufacturing industry. Literature reveals that research into reconfigurable grippers is on the increase with the onset of Industry 4.0. The design permits part handling and machining with the reconfigurable platform. Research into the simultaneous force loading due the material removal process has been conducted. These oblique cutting forces have been identified using the Mechanistic oblique cutting model. Engineering design principles and analyses such as FEA and FEM have been applied to the machine tool. The stiffness of the spindle is acceptable but not ideal, it is for this reason that future work on stability will allow the design to yield a higher stiffness. The natural mode frequencies of the spindle have been obtained. Future work on the dynamic response of the spindle to the metal removal operation will determine if the system is safe from a dynamic point of view [24].

8. REFERENCES

[1] **J. Bruninga, B.D., M.A. Dittricha, H.-S. Parkb** 2016. "Simulation based Planning of Machining Processes with Industrial Robots," in *16th Machining Innovations Conference for Aerospace Industry - MIC 2016*, Garbsen, Germany, pp. 16 - 24.

[2] **U. Schneider, J.R.D.P., M. Drust, A. Verl** 2013. "Position Control of an Industrial Robot Using an Optical Measurement System for Machining Purposes," in *11th International Conference on Manufacturing Research (ICMR2013)*, Cranfield University, UK, pp. 307 - 312.

[3] **U. Schneider, J.R.D.P., M. Drust, A. Verl** 2014. "Combining holistic programming with kinematic parameter optimisation for robot machining," in *45th International Symposium on Robotics (ISR 2014)*, Munich, German, pp. 524 - 530.

[4] **F. Domroes, M.R., B. Kuhlenktter** 2014. "Towards Autonomous Robot Machining," in *45th International Symposium on Robotics (ISR 2014)*, Munich, Germany, pp. 3 - 6.

[5] **Kevin Tai, A.-R.E.-S., Mohammadali Shahriari , Mohammad Biglarbegian and Shohel Mahmud** 2016. "State of the Art Robotic Grippers and Applications," *MDPI Robotics*, pp 1 - 20.

[6] **Yeung, B.H.B.** 2000. "Development of a 6 DOF Reconfigurable Gripper for Flexible Fixtureless Assembly " Master of Applied Science Master Dissertation, Graduate Department of Mechanicd and Industnd Engineering University of Toronto Toronto, Canada.

[7] **Tao Zhang, G.S., Ken Goldberg** 1999. "Compensatory Grasping with the Parallel Jaw Gripper," *4th Workshop on Algorithmic Foundations of Robotics*,

[8] **Martin Riedel, M.N., Mathias Hüsing, Burkhard Corves** 2008. "An adjustable gripper as a reconfigurable robot with a parallel structure," in *Second InternationalWorkshop on Fundamental Issues and Future Research Directions for Parallel Mechanisms and Manipulators* Montpellier, France.

[9] **Rezia Molfino, R.P.R.a.M.Z.** 2006. "A LOW-COST RECONFIGURABLE GRIPPER FOR ASSEMBLY AND DISASSEMBLY TASKS IN WHITE INDUSTRY," Department of Mechanics and Machine Design University of Genova, Genova, Italy.

[10] **Ziesmer, J.A.** 2009. "Reconfigurable End Effector Allowing For In-Hand Manipulation Without Finger Gaiting Or Regrasping," Degree of Master of Science, Faculty of the Graduate School, Marquette University, Milwaukee, Wisconsin, USA.

[11] **Anand Kumar Mishra, E.D.D., Ali Sadeghi, Alessio Mondini and Barbara Mazzolai** 2017. "SIMBA: Tendon-Driven Modular Continuum Arm with Soft Reconfigurable Gripper," *Frontiers in Robotics and AI* 4(4),

[12] **Jason Spiliotopoulos, G.M., Sotiris Makris** 2018. "A Reconfigurable Gripper for Dexterous Manipulation in Flexible Assembly," *MDPI Robotics*, 3(4),

[13] **Wei Gao, K.H., Jasjeet S. Seehra, Karthik Ramani, and Raymond J. Cipra** 2014. "HexaMorph: A Reconfigurable and Foldable Hexapod Robot Inspired by Origami,"presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, Chicago, IL, USA, 14-18 September 2014,

[14] **Radu-Eugen Breaz, O.B., Anca Lucia Chicea and Sever Gabriel Racz** 2015. "Using Serial Industrial Robots In CNC Milling Processes," University of Sibiu, Engineering Faculty, ResearchGateOctober 2015 2015, Available: https://www.researchgate.net/publication/291336878_Using_Serial_Industrial_Robots_in_CNC_Milling_Procesess, Accessed on: 17 January 2018.

[15] **Liang Jie, J.S.-H.** 2016. "The Study of Design Method for Robotic Drill End Effector," in *13th Global Congress on Manufacturing and Management, GCMM 2016*, China, 174: Elsevier Ltd, pp. 206-210.

[16] **Maciej Petko, G.K.** 2005. "Mechatronic Design of a Parallel Manipulator for Milling," in *International Conference on Advanced Intelligent Mechatronics*, Monterey, California, USA: IEEE.

[17] **Maeda, O., Cao, Y., and Altintas, Y.** 2005. *Expert spindle design system*. pp 537-548.

[18] **Altintas, Y.** 2000. *Manufacturing automation: metal cutting mechanics, machine tool vibrations, and CNC design*. The Pitt Building, Trumpington Street, Cambridge, United Kingdom: Cambridge University Press.

[19] **Wu, B., Yan, X., Luo, M., and Gao, G.** 2013. "Cutting force prediction for circular end milling process," *Chinese Journal of Aeronautics*, 26(4), pp 1057-1063.

[20] **Liu, D., Zhang, H., Tao, Z., and Su, Y.** 2011. *Finite Element Analysis of High-Speed Motorized Spindle Based on ANSYS*.

[21] **David, C., Sagris, D., Stergianni, E., Tsiafis, C., and Tsiafis, I.** 2018. "Experimental Analysis of the Effect of Vibration Phenomena on Workpiece Topomorphy Due to Cutter Runout in End-Milling Process †," *Machines*, 6(3), p. 27.

[22] **Thomas, M., Beauchamp, Y., Youssef, A.Y., and Masounave, J.** 1996. "Effect of tool vibrations on surface roughness during lathe dry turning process," *Computers & Industrial Engineering*, 31(3), pp 637-644.

[23] **Chen, C.-C., Liu, N.-M., Chiang, K.-T., and Chen, H.-L.** 2012. "Experimental investigation of tool vibration and surface roughness in the precision end-milling process using the singular spectrum analysis," *The International Journal of Advanced Manufacturing Technology*, 63(5), pp 797-815.

[24] **Jui P. Hung, Y.L.L., Tzuo L. Luo and Hsi H. Hsiao** 2012. "Prediction of the Dynamic Characteristics of a Milling Machine Using the Integrated Model of Machine Frame and Spindle Unit," *World Academy of Science, Engineering and Technology International Journal of Mechanical and Mechatronics Engineering*, 6(7),

A MODEL FOR THE RELATIONSHIP BETWEEN SUSTAINABILITY AND OPERATIONS STRATEGY PERFORMANCE OF A WATER UTILITY IN SOUTH AFRICA

M.F. Lushozi[1] & B. Emwanu[2]
[1]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
mthunzi.lushozi@gmail.com

[2]School of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
Bruno.Emwanu@wits.ac.za

ABSTRACT

Water is a scarce resource in South Africa and is recognised as a priority issue in the National Development Plan, one of whose key milestones is to ensure that all South Africans have access to clean running water in their homes. The increased expectation for clean water has come with an increased water demand but with little or no expansion of capacity to match the increased demand. The country is currently faced with an aging infrastructure being used to meet increased demand which threatens the sustainability of water utilities. This study considers the sustainability of South African water utility infrastructure, defined by maintenance and capital expansion, and the role of operations strategy defined by five generic operations strategy performance objectives. A framework was developed to investigate the relationship between water utility sustainability in South Africa and operations strategy, and the mediating role of trade-offs in this relationship. A questionnaire was used to collect data from key operatives in several water utilities and data analysed for results. It is expected that results will support a model that can be used to guide South African water utilities towards sustainability by proposing the most appropriate key performance objectives to pursue and indicating expected trade-offs.

[1]The author was enrolled for a MSc. Eng. (Industrial) degree in the School of Mechanical, Industrial and Aeronautical Engineering of University of Witwatersrand
[2]The author lectures in the School of Mechanical, Industrial and Aeronautical Engineering of University of Witwatersrand

1 INTRODUCTION

Infrastructure is the backbone of a successful and competitive economy which requires among other things sustainable water infrastructure [1]. However, aging infrastructure is a general concern for all utilities not only in South Africa but around the globe. For instance, in the United States, it is reported that the Leaders of the utilities in all regions of the country unanimously selected aging infrastructure as their top concern, including pipelines, tunnels, dams, pumping, storage and treatment facilities among the water utility infrastructure [2].In European countries such as the United Kingdom, water utility's infrastructure aging is resulting in billions of litres of water lost every day [3]. African countries such as Botswana have embarked on projects to rehabilitate aging infrastructure as one of the focus areas for their Water Utility Corporations [4]. Furthermore, South African water utilities are also on the same predicament of the aging water infrastructure. However, in South Africa, it can be argued that the challenges related to utility's infrastructure are to a large extent attributable to how post-apartheid transition was managed. The introduction of apartheid Native Land Act No. 27 of 1913 resulted in the deprivation of millions of South Africans living in informal settlements and rural areas of access to basic services such as clean water, sanitation, and electricity [5]. Post 1994 election a new constitution was introduced to the country, within which section 27 of the Bill of Rights of the Constitution declared that everyone has the right to have access to sufficient food and water. Therefore, based on this mandate the government started to extend the basic services (such as water and electricity) to a larger population of the country as per this section of the constitution. However, these basic services were distributed without significantly increasing the infrastructure and at the expense of existing infrastructure maintenance [6]. Moreover, Mugabi *et* al [7] are of the view that inappropriate strategic management practices in Public water utilities in developing countries result in the enormous challenges they face in meeting the water demands of their growing urban populations. Adopting an operations strategy that seeks to achieve effective asset utilization can assist the utilities to deliver on the agreed service level [2]. Service levels of organisations may be considered in terms of performance objectives. According to Slack and Lewis [8], there are five performance objectives that organisations need to focus on namely quality, speed, dependability, flexibility, and cost. Therefore, this study seeks to examine the current state of South African water utility infrastructure with respect to the five performance objectives (cost, dependability, flexibility, quality, and speed) and their trade-offs, and how this relates to its sustainability. Consequently, it seeks to develop an operations strategy model to be applied in South African water utilities. Figure 1 is the hypothesised model for the study. This model hypothesises that operations strategy performance (independent variable) has a significant relationship with water utility sustainability (dependent variable) and trade-offs (mediating variable).

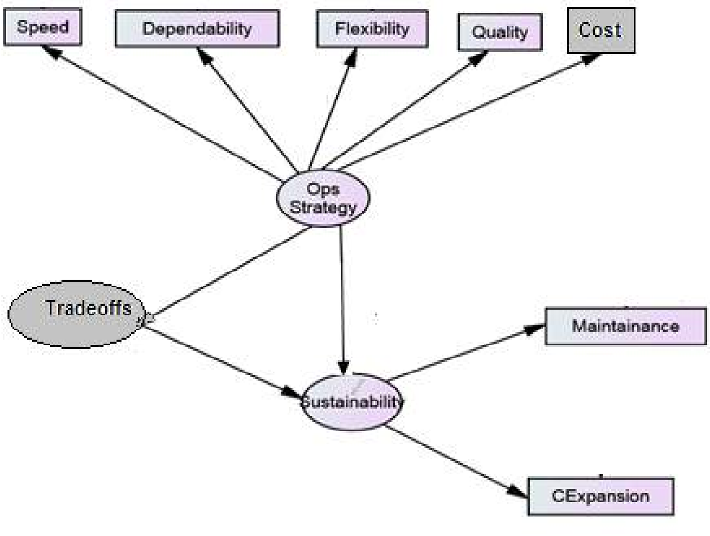


Figure 1: Hypothesised model

The rationale for the model is that trade-offs are considered important for operations strategy performance following the literature (section 2) that one should choose among performance objectives. This raises the question whether trade-offs are also practiced in water utilities and subsequently have influence on operations strategy performance and sustainability of a water utility. Given their importance in the literature, it was anticipated that they have an influence on water utility as well and therefore needed to be investigated for a water utility. A study of these parameters and their relationship was undertaken.

The parameters were measured differently. Operations strategy performance was measured by using the operations strategy performance objectives (cost, speed, flexibility, dependability, and quality). On the other hand, Water utility sustainability was measured by current maintenance status and capital expansion status. Having considered trade-offs as likely to have an effect on operations strategy performance and subsequently to also have an effect on sustainability of the water utility, hypotheses were constructed. Consequently, the model had the following three hypotheses:

- **H1**: operation strategy performance has a significant and direct effect on water utility sustainability.
- **H2**: operation strategy performance has a significant and direct effect on trade-offs.
- **H3:** trade-offs has a significant and direct effect on water utility sustainability.

The three hypotheses were developed to be used to construct and test a Water utility's strategy operations performance model. They were tested using multiple-regression and the final model was constructed using AMOS. The rest of this paper is divided into five sections. Section 2 reviews the relevant literature for theory and key concepts used in the study; Section 3 is the research methodology that was adopted for the study; Section 4 presents the results; Section 5 discusses the results and proposes the final model; and Section 6 gives the conclusion and recommendations for further research.

2 LITERATURE REVIEW

This section reviews the literature highlighting key concepts relevant to this study including operations strategy, trade-offs and sustainability. In doing so, the concepts are categorised into three key variables namely independent, mediating and dependent variables that may be possible to relate to each other conceptually. The rationale for categorising them that way is given in the text. The section first considers operations strategy, what it means and its importance for organisations in order to achieve competitive advantage. In addition, the components of operations strategy are highlighted describing the nature of impact they can have on the organisation. Next, the theory of operations strategy is considered specifically highlighting the important role trade-offs play in organisations. Lastly, the section looks at sustainability in utilities, what it means, its importance for organisations, and how it may be perceived for water utilities.

2.1 Operations strategy performance

The success of an organisation is dependent on having an effective and clearly defined operations strategy [9]. The operations strategy is normally formulated based on the competitive advantages of the organisation [10]. Different scholars have defined operations strategy differently. For instance Skinner [11] describes it as a set of manufacturing policies designed to maximize performance to meet corporate strategy; alternatively, it is a total pattern of decisions that shape the long-term performance of the organisation [8]. On the other hand, Walters [12] argues that operations strategy is more than a set of decisions and suggest that it is more about long-term goals, policies, culture and resources, and decisions that relate to operations. According to Slack and Lewis [8], there are five generic operations performance objectives and four operations decision areas. The five operations performance objectives are cost, quality, speed, dependability, and flexibility; while the four decision areas are capacity, supply network, process technology, and development and organisation. An effective operations strategy should maintain and improve the competitive advantages of the organisation based on their corporate operations resources capabilities [10].

In manufacturing better quality and less variance are invaluable. Higher quality infers lower costs and increased productivity; subsequently a greater market share and a better competitive position for the organisation [13]. Slack and Lewis [8] define quality as the consistent conformance of a product or service to customers' expectations. Standardization and careful attention are vital in assisting organisations to maintain a high and consistent quality level which ultimately leads to the profitability of the organisation [14]. On his manufacturing strategy model, Hill [15] suggested that performance objectives can either be order qualifiers or order winners. A water utility is a good example of an organisation whose product requires a certain level of quality for one to get an order in accordance with South African National Standards (SANS) 241.
Cost can be described as an absolute term and measures of the number of resources used by the organisation's operations to produce the product [16]. The cost for water utility would involve mainly operations cost, plant maintenance, and capital expansion.

Flexibility is the ability of operations to easily change product mix and production volumes in response to the market demands [17]. Hallgren [19] argues that a characteristic that differentiates flexibility from other dimensions of operational performance is that it is a measure of potential as opposed to actual performance. The ability to change water supply volumes with demand would constitute flexibility in a water industry.

Amoako-Gyampah & Acquaah [18] maintain that organisations with reliable and on-time deliveries can anticipate greater customer satisfaction that can potentially lead to increased sales growth and market share.

Dependability in the manufacturing fraternity measure the ability of an organisation to keep delivery promises made to the customer [8]. It is an ability of the organisation to do things in time for their customers to receive their product exactly when they are needed, or at least when they were promised [20]. Customers normally gauge if the product required was delivered on time or not [12]. He suggests that dependability is both internal and external. According to Awwad *et* al [2], external dependability is usually regarded by customers as a good thing because being late with delivery of the product can be a considerable frustration to them. Dependability can be a determining factor sometimes that is used by clients to determine whether suppliers have their contracts renewed [19].

It is therefore clear that the external effects of this performance objective are to increase the odds of customers returning with more business [21]. According to Russell1 & Millar [22], the internal dependability has an influence on cost at least in three ways. Internal dependability can result in the utility saving time by saving money directly or giving the utility stability which allows it to advance its efficiencies. Highly dependable systems can assist increase speed performance [19].

Speed indicates the elapsed time between the beginning of an operations process and its end such as the time taken to fulfill the customer requests [8]. According to Jeena [23], improved speed assist operations influence the time taken for goods and services to be delivered to the customer She is of the opinion that externally, speed is a significant facet of customer service, while internal; it decreases throughput time and reduces risks by delaying the commitment of resources. Therefore, in water utility, speed would be associated with the time taken by treatment plants to clean up water and pump it to strategic reservoirs.

In the initial model, it is anticipated that the operations strategy performance which is defined by the five operations objectives namely cost, quality, speed, dependability, and flexibility directly influences sustainability of the water utility but is not influenced in turn by sustainability. On the other hand, operations strategy performance may also influence sustainability of the utility depending on the trade-offs position. Operations strategy performance is therefore considered an independent variable in the initial model as it affects other variables directly or indirectly but is not affected by them.

2.2 Theory of trade-offs

Effective operations strategy requires an awareness of trade-offs, whereby management is able to consider the different variables required to achieve set objectives and have the ability to trade one against the other, to optimise the chances of success [8]. According to Skinner [22], a trade-off scenario is one in which a high, market-leading performance in one competitive criterion (e.g. flexibility) is achieved at the expense of lower levels of performance in a different capability (e.g. speed). Filippini *et* al [16] maintain that two types of performance are in trade-off if the achievement of high values in one type of performance means low values in another. Therefore, a trade-off situation is characterized by high levels of difference between the two types of performance [24]. According to Skinner [11], it is paramount to recognise the variables that can be controlled and then to focus on them. Skinner [25] argues that manufacturing plants should have specific and concise tasks with regard to their products, technologies, and markets. He suggested three different dimensions of focus; Product focus: producing a narrow mix of products; market focus: serving a carefully and narrowly defined market segment or niche and process focus: focusing on a certain type of production technology. Clark [26] suggests that organisations that attempt to excel in everything without developing competitive priorities will end up second best to organisations that focus theirs. The focused factory theory [14] is based on the premise that no single organization can do equally well in all things. Applying this perspective to water utilities, this suggests that it is not critical to be perfect in all aspects of water delivery from source to the consumer; rather by focusing on few manageable objectives water utilities should be able to improve their service and efficiency in the long term. Subsequent to Skinner's work, [22] and Hill [15] continued to emphasize the importance of operations strategy and expand on it.

In the initial model, it is anticipated that whereas trade-offs have an influence on sustainability of water utilities they in turn are influenced by the strategy performance objectives, so that being in between they have a mediating role between performance objective and sustainability. Trade-offs are therefore considered a mediating variable in the initial model.

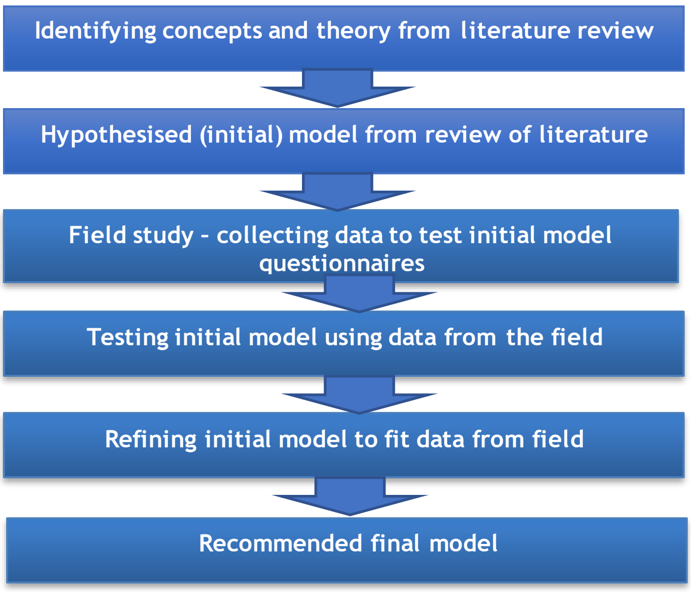
2.3 Utility Sustainability - Maintenance and Capital Expansion

At the World Summit on Sustainable Development that was held in 2002 at Johannesburg, South Africa, the then President Nelson Mandela said: 'Among the many things that I learned as president was the centrality of water in the social, political and economic affairs of the country, the continent, and the world' [27]. On that note, the importance of sustainability of water utilities cannot be overemphasized. According to the Environmental Protection Agency [28], utilities that efficiently integrate sustainability considerations into their operations strategies can expect to accomplish numerous benefits such as minimizing costs, maximising investment results, improvement of the ability to analyse a range of alternatives, and ensuring that financial and revenue strategies are sufficient to finance, operate, maintain, and replace critical infrastructure throughout its operational life. According to Water and Sanitation [29], the research indicates that the water demand in South Africa will exceed supply by 2025 if the current water resources are not supplemented. They warn that the sustainability of the water sector as a whole is also under threat due to the poorly maintained and often ill-equipped infrastructure. The primary objective of maintenance is to slow down asset deterioration and extend the period before rehabilitation and renewal is required [30]. He maintains that normally plants apply a mix of preventive and breakdown maintenance. The International Water Associations [31] also assert that plant maintenance seeks to enhance the effectiveness and sustainability of capital investments and existing infrastructure for water utilities. The Department of Water Affairs [32] believes that Asset Management is fundamental in the maximization of the value of an asset over its lifecycle; thereby ensuring that the water utilities derive the most benefit from its investment. This comprises of constructing, operating, repairing or replacing assets at the optimum time to ensure system reliability at the lowest cost and least impact to the utility.

In the initial model, it is anticipated that sustainability of the water utility is a variable that is directly influenced by and therefore dependent on both operations strategy performance and the trade-offs selected.

3 RESEARCH METHODOLOGY

Research can be progressed using a number of research strategies, such as quantitative, qualitative or mixed-methods approach [33]. This study was conducted using a quantitative method. The strategy was to develop a conceptual framework on a step by step basis beginning with locating the applicable theory to developing an initial framework that gets refined progressively to a more realistic framework that fits the data obtained from the field. Diagram 1 illustrates the steps that were followed to conduct this study.



The first step was to identify relevant concepts and theory from the literature so as to develop an initial hypothesised model. When this was done, an instrument was developed to collect relevant data from practitioners in the field, basing on information from literature and considering the relationships investigated. Data collection was done using a questionnaire to obtain primary data from South African water utilities. The data were used to test whether the framework fits the data. Where it did not, adjustments were made to the model that led to the final conceptual model.

3.1 Sample characteristics

The target population was water utilities' employees that could read and write. The questionnaires were issued to 70 people excluding those that were issued for pilot testing of the questionnaire. Only 62 respondents of the 70 responded. Out of the 62 respondents, only 50 were completed in full for analysis and 12 were not completed fully. Demographic characteristics of the sample are presented in Table 1. The main aim of the questionnaire was to collect data for the study in order to test the hypothesised model.

Table 1: Sample characteristics

Demographic	Sub-groups	Percentage %
Gender	Female	36
	Male	64
Department	Strategic Asset Management	14
	Capital Projects	42
	Operations and Maintenance	30

	Project Controls	6
	Other	8
Positions	Engineers	11
	Sectional Managers	7
	Project Managers	10
	Area Managers	3
	Operations Managers	4
	Maintenance Managers	3
	Others	12
Age	20-25	2
	26-30	12
	31-35	24
	36-40	20
	41-45	20
	46-50	4
	50 and above	18
Qualifications	Grade 12	0
	Certificate	4
	National Diploma	14
	Degree	30
	Post Graduate	52

3.2 Measuring instrument

A questionnaire was used to collect data for the study. It was composed of closed-ended questions developed specifically for this study. The questionnaires were personally administered to the respondents.

3.3 Statistical analysis

The statistical analysis was carried out by means of the SPSS program. In order to ensure validity, each construct was tested to determine the percentage of variance explained by a single factor. Confirmatory factor analysis (CFA) was conducted to establish the structure of the relationship between the studied variables from the sample responses. Descriptive statistics including Cronbach alpha coefficients were determined to assess the reliability of the measuring instruments. This study used Pearson product-moment correlation to identify the relationships between the variables. The study has three hypotheses as indicated in the introduction. Therefore, this was used to test significance among the variables. The effect size in the case of multiple regressions was given by the formula Steyn [33] $f^2 = R^2/1-R^2$. The following parameters 0.02 (small effect size), 0.25 (medium effect size) and 0.4 (large effect size) were set for practical significance of f^2 [34].

4 RESULTS

The result section presents the findings on correlation analysis where the correlation between the variables was conducted. The correlation significances and effect sizes were tested. Due to the small sample size, the confirmatory factor analysis was conducted to test the validity. Furthermore, AMOS 20 was used to conduct structural modeling in order to test the hypothesised model. Lastly, the hypotheses were tested using the multi-regression. The results for the above-mentioned test are presented in the sections below.

4.1 Confirmatory Factor Analysis and Reliability Test

In order to ensure validity, each construct was tested to determine the percentage of variance explained by a single factor. Confirmatory factor analysis (CFA) was conducted to establish the structure of the relationship between the studied variables from the sample responses. Furthermore, the internal consistency reliability was assessed by the use of Cronbach's alpha. Field [35] suggests that Cronbach's alpha level of 0.7 is generally a suitable cut-off point for reliability test; however, other scholars such as Ruperto *et* al [36] also take values as low as 0.6. Therefore, the cut-off for this study was 0.6 and all the values below 0.6 were dropped from the further analysis and from the model.

Table 2: Confirmatory Factor Analysis results and Internal Consistency Reliability

Component	Total	% of Variance	Cronbach Alpha
Speed	2.055	41.099	.609
Dependability	1.653	33.057	.473
Flexibility	2.028	50.693	.662
Quality	2.198	43.969	.673
Cost	2.702	54.047	.781
Trade-off	2.548	50.960	.753
Maintenance	2.793	55.860	.796
Capital Expansion	3.361	67.223	.870

Inspection of Table 2 suggests that the total variance explained for in almost all the items are above 40%; therefore the factor structure is confirmed as Stacciarini and Pace[37] on their study recommended variance explained as low as 34%. Table 2 also indicates that the Cronbach's alpha values that were obtained ranged from 0.47 to 0.78. All the Cronbach's alphas were above the 0.6 cut-off except the one of dependability which was 0.5. Therefore, dependability was dropped from further analysis and from the model.

4.2 Correlation Analysis

This section focuses on the correlations between the variables covered in the study i.e. operations strategy performance, trade-offs and sustainability. The correlations were done as part of achieving some of the objectives of the study which was to develop a model. In order to achieve this, it was important to test the correlation between the independent variable, dependent variable and mediating variable. Table 3 does not include dependability because it had a Cronbach's alpha value below a 0.6 cut-off and so was dropped.

Table 3: correlations

		Speed	Flexibility	Quality	Cost	Trade-Off	Maintenance	Capital Expansion
Speed	**Pearson Correlation**	1						
	Sig. (2-tailed)							
Flexibility	**Pearson Correlation**	.384**	1					
	Sig. (2-tailed)	.006						
Quality	**Pearson Correlation**	-.061	.288*	1				
	Sig. (2-tailed)	.675	.042					
Cost	**Pearson Correlation**	.333*	.553**	.189	1			
	Sig. (2-tailed)	.000	.000	.001				
Trade-Off	**Pearson Correlation**	.301*	.150	-.209	.102	1		
	Sig. (2-tailed)	.034	.299	.144	.483			

Maintenance	**Pearson Correlation**	.302*	.439**	.215	.575**	.002	1	
	Sig. (2-tailed)	.033	.001	.133	.000	.990		
Capital Expansion	**Pearson Correlation**	.387**	.634**	.223	.611**	-.017	.618**	1
	Sig. (2-tailed)	.006	.000	.088	.000	.907	.000	

** Correlation is significant at the 0.01 level / *. Correlation is significant at the 0.05 level, p = 0.1 the effect size is small, p = 0.3 the effect size is medium and p = 0.5 is large effect size.

Speed is positively statistically significant correlated (medium effect size) with flexibility, negatively statistically insignificant correlated (small effect) with quality, positively statistically significant correlated (medium effect size) with cost, positively statistically significant correlated (medium effect size) with trade-offs, positively statistically significant correlated (medium effect size) with maintenance and positively statistically significant correlated (medium effect size) with capital expansion.

Flexibility is positively statistically significant correlated (medium effect size) with quality, positively statistically significant correlated (large effect size) with cost, positively statistically insignificant correlated (small effect size) with trade-offs, positively statistically significant correlated (medium effect size) with maintenance and positively statistically significant correlated (large effect size) with capital expansion.

Quality is positively statistically significant correlated (small effect size) with cost, negatively statistically insignificant correlated (small effect size) with trade-offs, positively statistically insignificant correlated (small effect size) with maintenance and positively statistically insignificant correlated (small effect size) with capital expansion.

Cost is positively statistically insignificant correlated (small effect size) with trade-offs, positively statistically significant correlated (large effect size) with maintenance and positively statistically significant correlated (large effect size) with capital expansion.

Trade-offs is positively statistically insignificant correlated (small effect size) with maintenance and negatively statistically insignificant correlated (small effect size) with capital expansion.

Capital expansion is positively statistically significant correlated (large effect size) with maintenance.

4.3 The mediating effects of trade-offs in the relation between operations strategy performance and sustainability

Structural Equation Modelling was performed using AMOS 20 [38] for testing for mediating effects of trade-offs in the relationship between operations strategy performance objectives and water utility sustainability. The hypothesised model is a mediation model in which performance objectives of the operations strategy influence trade-offs, which in turn impact on water utility sustainability. The study has made a closer examination of the direct and indirect effects to evaluate their relative sizes. The study has hypothesised the model in figure 1. The structural equation modeling was conducted to test that model and the model in figure 2 below was produced. The model was found to be recursive which means that the causation flows is in one direction as opposed to non-recursive where causation can flow in more than one direction. This, therefore, suggest that operations strategy performance exert a causal effect on trade-offs but the inverse is not true. Furthermore, trade-offs exert a causal effect on sustainability while the inverse is not true. According to Hair et al [39], recursive models are better because it is less complex to identify as opposed to complex recursive models.

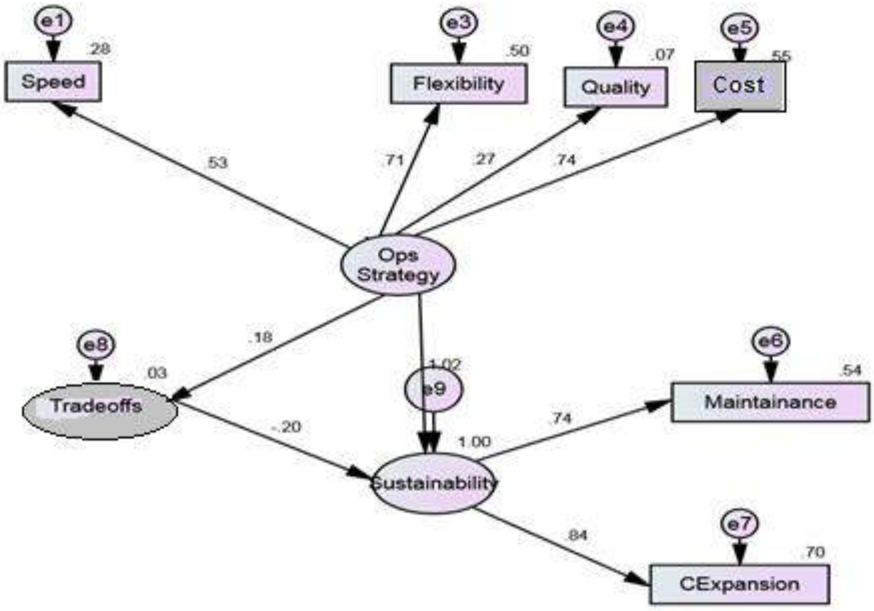


Figure 2: Model resulted from the SEM analysis

4.3.1 Predictive and comparative fit

A good fit for the proposed hypothetical model was obtained (see *Figure 2* χ^2 (18, *N* = 51) = 23.35, *p* = 0.18; $\chi2/df$ = 1.3 (recommended ≤ 3.00), the Tucker Lewis index (TLI) = .94 (recommended ≥ .90), the comparative fit index (CFI) = 0.96 (recommended ≥ .95), and the root mean square error of approximation (RMSEA) was 0.08 which is acceptable (recommended ≤ .08 and ≤ .05 is excellent). The recommended figures for SEM model were adopted from Yuan and Chan [40].

4.3.2 Direct, Indirect and Total Standardized Effects of Operations Strategy Performance objectives and Trade-offs

In order to test the mediation, the Beta-values for the standardized direct, indirect, and total were calculated. When analysing the mediator, there are two effects involved namely direct effect and indirect effect. The direct effect is the effect from the independent variable directly to a dependent variable, while the indirect effect is the effect from independent variable to a dependent variable that goes indirectly through the mediating variable. According to Zhao *et* al [41] in order to establish mediation, the indirect effect should be greater than the direct effect. However, for this study indirect effect (a relationship between operation strategy performance and sustainability via trade-offs) is -0.04 = (0.18*-0.20) while direct effect (a relationship between operation strategy performance and sustainability without trade-offs) is 1.02. This, therefore, suggests that direct effect is greater than the indirect effect which means there was no mediation established in this study.

4.3.3 Hypothesis Testing: required to determine the mediation effect

This study had three hypotheses in-line with figure 1 as they are highlighted in table 4 below. Multi-regression was conducted to test these hypotheses and the results are furnished in table 4.

Table 4: the hypothesis testing for a direct effect of operations strategy on sustainability

Hypothesis statement for path analysis	Beta Estimate	P-Value	Result on hypothesis
H1: operation strategy performance has a significant and direct effect on water utility sustainability (***P ≤ 0.001)	1.17	***	Supported
H2: operation strategy performance has a significant and direct effect on trade-offs	0.28	0.25	Not supported
H3: trade-offs have a significant and direct effect on water utility sustainability	- 0.15	0.07	Not supported

Inspection of table 4 suggests that p-values for both H2 and H3 are greater than 0.05 which means their direct paths are insignificant. Therefore, these hypotheses are not supported. This further suggests that the mediation did not take place.

5 DISCUSSION OF RESULTS

A hypothesised model of the relationship between operations strategy performance, trade-offs, and water utility sustainability was tested in a convenience sample (*n* = 50), and respondents were recruited from a number of the departments in some of the South African water utilities. The data did not support the model as it was hypothesised. However, when the mediating variable (trade-offs) was removed (table 5), the new model was supported by the data. The utility operations strategy performance influences their sustainability. Therefore, based on table 5 the model in figure 3 is the model that is proposed for the South African water utilities going forward.

Table 5: Standardized Direct Effects without trade-offs

Variables	Operations Strategy performance	Sustainability
Sustainability	1.02	0
Capital Expansion	0	0.84
Maintenance	0	0.74
Speed	0.53	0
Flexibility	0.71	0
Quality	0.27	0
Cost	0.74	0

Table 5 and table 4 suggest that operation strategy performance has a significant and direct effect on water utility sustainability (***P ≤ 0.001 and r = 1.02). This relationship resulted in a model presented in figure 3.

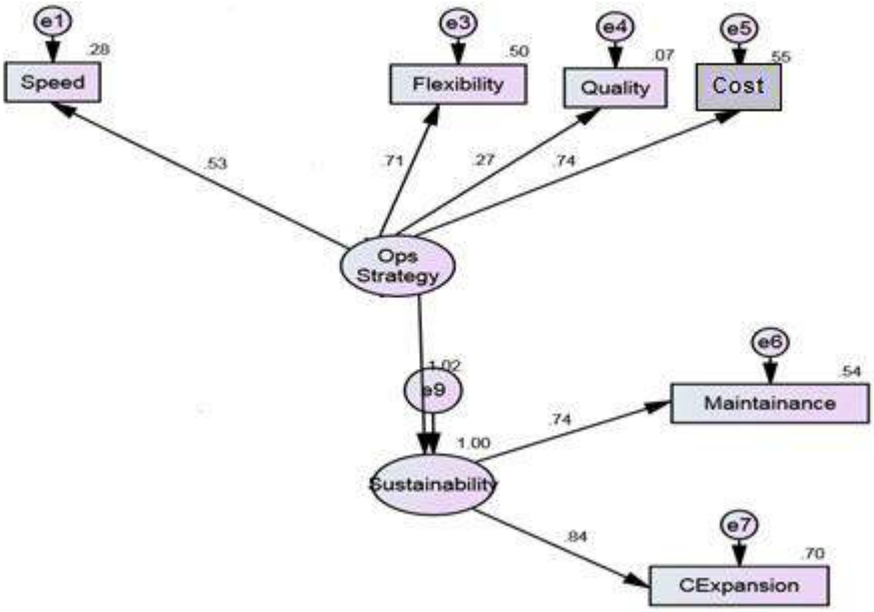


Figure 3: a final proposed model for SA water utilities

The model in figure 3 suggests that if the water utilities achieve robust operations strategy by good management of five operations strategy performance objectives while maintaining their plants well and managing their capital expansions, they will have sustainable utilities. The major limitation of this study was the sample size especially given that the new model was to be developed. Furthermore, given the complexity of the model in a sense that it contained the mediating variable, a big sample size would have been ideal. The study was conducted predominantly using some sections of a specific water utility using a convenience sample and other water utilities were difficult to gather the data from.

6 CONCLUSION

The primary drive for commissioning the study has been that water utility infrastructure is edging while the demand is increasing on the other side. However, there are many competing variables in efforts of sustaining the utilities such as cost, quality, flexibility among others. The study has examined the relationship between the water utility's operations strategy, and sustainability. It further examined the mediating role of trade-offs between operations strategy and sustainability. The study also aimed at developing a model for the relationship between sustainability and operations strategy of a water utility in South Africa. The model suggested that there is a statistically significant relationship between operations strategy and sustainability of a water utility in South Africa. However, due to smaller sample than ideally targeted the model did not support the mediating effect of trade-offs. This study has contributed to knowledge in that it adds the operations strategy model that can be used in South African water utility industry. There is no study in the past which has looked operations strategy in the context of performance objectives and which has defined sustainability of water utilities in terms of capital expansion and maintenance. Therefore, this study has closed that gap. Water utilities are encouraged to adopt this study in order to put in on a test. Further research on the same topic with bigger sample size needs to be conducted to evaluate the mediating role of trade-offs.

7 REFERENCES

[1] **Rhodes, C.** 2018. Infrastructure policies and investment: briefing paper, *House of Commons*, no. 6594.

[2] **Awwad, A.,S., Adel A. Al Khattab, A., A. & Anchor, J.R. 2013.** Competitive Priorities and Competitive Advantage in Jordanian Manufacturing. *Journal of Service Science and Management*, 6, pp 69-79.

[3] **Department for Business Innovation & Skills.** 2013. The smart city market: opportunities for the UK. London: BIS research paper no. 136.

[4] **Water Utilities Corporation**. 2012. Annual report for water co-operation building partnership.

[5] **Mlangeni, B**. 2013. Reflections on the impact of the Natives' Land Act, 1913, on local government in South Africa. Pretoria: Parliament of the Republic of South Africa.

[6] **Newbery, D. & Eberhard, A**. 2008. *South African infrastructure review*: electricity. Cape Town.

[7] **Mugabi, J., Kayaga, S. & Njiru, C**. 2007. Strategic planning for water utilities in developing countries. *Utilities Policy*, 15 (1), pp 1-8.

[8] **Slack, N. & Lewis, M**. 2002. *Operations Strategy*. Harlow: Pearson Education.

[9] **Russell, S., N. & Millar, H.** 2014. Competitive priorities of manufacturing firms in the Caribbean. *Journal of Business and Management*, 16(10), pp 72-82.

[10] **Kaviani, M., A. & Abbasi, M**. 2014. Analysing the operations strategies of manufacturing firms using a hybrid Grey DEA approach- a case of fars cement companies in Iran. *International journal of Supply and Operations Management*, 1(3), pp 371-391.

[11] **Skinner, W**. 1996. Manufacturing strategy on the 'S' curve. *Production and Operations Management*, 5 (1), pp 3–14.

[12] **Walters, D**. 2006. *Operations strategy*: Thompson learning: London

[13] **Jamaluddin, Z. *et* al.** 2016. Quality management practices and performance measurement in the manufacturing industry: an instrument validation. Sains Malaysiana, 45(6), pp 999-1006.

[14] **Sprague, L., G**. 2007. Evolution of the field of operations management. *Journal of Operations Management*, 25, pp 219–238.

[15] **Hill, T**. 1989. *Manufacturing Strategy*: text and cases, Irwin, Homewood.

[16] **Filippini, R., Forza, C., & Vinelli, A.** 1998. Trade-off and compatibility between performance: definitions and empirical evidence. *International Journal of Production Research*, 36 (12), pp 3379–3406.

[17] **Awwad, A., S**. 2007. The role of flexibility in linking operations strategy to marketing strategy. *18th annual conference*: Texas.

[18] **Amoako-Gyampah, K. & Acquaah, M.** 2007. Manufacturing strategy, competitive strategy and firm performance: An empirical study in a developing economy environment. *International Journal of Production Economics*, 111, pp 575–592.

[19] **Hallgren, M.** 2007. Manufacturing strategy, capability, and performance. Linkoping.

[20] **Indran, S.** 2007. Strategic roles and objectives of operations. accessed online, available from http://mindran.com/operation/027370849X_pp02.pdf

[21] **Black & Veach**. 2013. *Strategic directions in U.S. water industry*. Overland Park: Black and Veach Corporation.

[22] **Sarmiento, R.** 2011. A note on 'trade-off and compatibility between performance: definitions and empirical evidence'. *International Journal of Production Research*, 49(13), pp 4175-4183.

[23] **Jeena, U**. 2002. Selecting an appropriate process for the formulating of an operations strategy for Bridgestone/Firestone, Port Elizabeth plant, in a changed market environment.

[24] **Skinner, W.** 1992. *Missing the links in manufacturing strategy*. London: Chapman and Hall, pp 13–25.

[25] **Skinner, W.** 1974). The focused factory. *Harvard Business Review*, Vol. 52 No. 3, pp 113-22.

[26] **Clark, K.** 1996. Competing through manufacturing and the new manufacturing paradigm: is manufacturing strategy passe´?. *Production and Operations Management*, 5 (1), pp 42–58.

[27] **Msangi, J., R.** 2014. *Managing water scarcity to Southern Africa*: policy & strategies: combating water scarcity in Southern Africa. Springer briefs in environmental science, DOI:10.1007/978-94-007-7097-3_2. Pp 21.

[28] **Environmental Protection Agency**. 2013. Defining a Resilient Business Model for Water Utilities. US water research foundation.

[29] **Water and Sanitation Programme**. 2011. Water supply and sanitation in South Africa: Finance into Services for 2015 and Beyond. Nairobi: World Bank
[30] **Mannan, S**. 2012. *Lees' loss prevention in the process industries*: Hazard identification, assessment, and control. New York: Butterworth-Heinemann.
[31] **International Water Association**. 2011. Strategic management of water and wastewater assets in India: a primer for utility leaders.
[32] **Department of Water Affairs**. 2014. Annual report 2013/2014. Pretoria: Department of water affairs.
[33] **Andrew, D.,*et* al**. 2011. *Research methods and design in sport management*. London: Human kinetics
[34] **Steyn, H**. 1999. Practical significant relationship between two variables. *Journal of industrial psychology*, 28(3), pp 10-15.
[35] **Filed, A**. 2009. *Discovering statistics using spss*. London: Sage.
[36] **Ruperto, N. *et* al**. 2005. The pediatric rheumatology international trials organization criteria for the evaluation of response to therapy in juvenile systemic lupus erythematosus. *Arthritis & Rheumatism*, 52(9), pp 2854-2864.
[37] **Stacciarini, T.,S.,G., & Pace, A.,E**.2017. Confirmatory factor analysis of the appraisal of self-care agency scale. *Revista Latino-Americana de Enfermagem*.
[38] **Hair, J., F**. 2010). *Multivariate Data Analysis, Upper Saddle River*, NJ: Prentice Hall.
[39] **Arbuckle, J.L**. 2013. *Amos 20*. Chicago, IL: Smallwaters.
[40] **Yuan, K., & Chan, W**. 2016. Measurement invariance via multigroup sem: issues and solutions with chi-square-difference tests. *American Psychological Association*, 21 (3), pp 405-426
[41] **Zhao, X. , Lynch, J.G., & Chen, Q**. 2010. Reconsidering Baron and Kenny: myths and truths about mediation analysis. *Journal of consumer research*, 37, pp197-206.

8 ANNEXURE 1: SECTIONS OF THE QUESTIONNAIRE

- ***Section one*** - This section consisted of demographic information being respondent's age, gender, highest qualification, current position and department.

- ***Section two*** - *operations performance objectives:* This section used a five-point likert scale to test importance of each of the five operations strategy performance objectives (speed, cost, quality, flexibility and dependability).

- ***Section three*** – *trade-offs:* This section was used to identify the trade-off between the five operations strategy performance objectives. The respondents were asked to rank the performance objectives from 1 to 5 in the order they believe to be most important. This section had 5 items.

- ***Section four*** – *infrastructure status:* This section used a five-point likert scale to assess the status of the current infrastructure and capital expansion plan thereof. The section comprised of 10 items in total.

INDUSTRIALISATION 4.0 - THE IMPLICATIONS OF THE FOURTH INDUSTRIAL REVOLUTION FOR INDUSTRIALISATION IN AFRICA

B. Fynn-Hurwitz [1*] & B.P. Sunjka[2]

[1]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
fynn.belinda@gmail.com

[2]Department of Mechanical, Industrial and Aeronautical Engineering
University of Witwatersrand, South Africa
Bernadette.Sunjka@wits.ac.za

ABSTRACT

The traditional path to industrialisation by growing manufacturing output has leveraged two key sources of competitive advantage: (1) low-cost labour and (2) the economies of scale derived from centralised mass production. The maturing technologies of Industry 4.0 erode both of these foundational pillars. The implication for less developed countries is that the model for successful industrialisation going forward is likely to differ substantially from that of the past. This offers both challenges and opportunities, requiring a fundamentally different approach to driving industrialisation. This paper explores what this may mean for countries that are still seeking to become more industrialised.

[1] The author was enrolled for an MSC Eng degree in the Department of Mechanical, Industrial and Aeronautical Engineering, University of the Witwatersrand
*Corresponding author

1. INTRODUCTION

The question of how to best facilitate economic development is particularly pertinent in an African context, as one of the most pressing needs across the continent is to enable a large portion of the population to move from poverty, to a more developed or industrialised status [1]. Ten of the only thirteen countries that have managed to achieve uninterrupted GDP growth of higher than 7% for 25 years, did so through manufacturing-led growth [2]. At a high-level, the basis of this model was to attract investment into developing a country's manufacturing industry i.e. building large factories that create significant employment and become competitive exporters to the global market, by leveraging the twin competitive advantages of:

1. the availability of low-cost labour [3], and
2. economies of scale achieved through centralised manufacturing and mass production

In simplified format, some of the theoretical outcomes of this model are that the resulting increase in tax revenues from the export of finished good allows for increased government spending on infrastructure to support rapid urbanisation. This, together with the increase in spending power of an employed workforce allows service and support industries to grow and flourish. There are also significant learning [4] and ecosystem benefits obtained. At a point in time, labour costs rise to the extent that the competitive advantage of cheap labour no longer exists and industries move "off-shore" to other geographies in pursuit of the low-cost labour advantage, leaving behind them a significantly more "developed", "post-industrial" economy.

This is the approximate model that has been followed by China since 1978 [5]. As China transitions into a more high-wage economy with a greater focus on domestic consumption and services [6], there has been some speculation on which country will be the "next China". The "MITI-V group of countries (Malaysia, India, Thailand, Indonesia and Vietnam) have been positioning themselves to take over the mantle [7] as have various African countries, notably Ethiopia. The pertinent question however, is "Will this model of industrialisation succeed in the new context of the Fourth Industrial Revolution?"

As the technologies and ecosystems of the Fourth Industrial Revolution mature, they will have a significant impact on the two main sources of competitive advantage that underpin this model of manufacturing-led economic development [8]. This paper explores, through a critical review of the extant literature, the likely disruption of these two key levers of the traditional industrial development model, namely (1) the availability of cheap labour and (2) the competitive advantage of centralised manufacturing that is realised through economies of scale.

2. INDUSTRY 4.0 IN CONTEXT

To understand how the Fourth Industrial Revolution will disrupt traditional industrial development models, its broader context needs to be explored. In brief, the Fourth Industrial Revolution is the simultaneous maturing of a number of technologies with the potential to transform the manner in which industrial activity will take place going forward.

Seth Godin proposed that we are entering a "Fourth Industrial Revolution in his 2007 book "Meatball Sundae" [9], but the concept gained more rapid popularisation after the German Government endorsed the "Industrie 4.0" initiative, presented in 2011 by a group of prominent academics, as one of ten "future projects" forming Germany's High-Tech Strategy 2020 Action Plan [10]. In the words of German Chancellor Angela Merkel, Industry 4.0 is [11]:

'the comprehensive transformation of the whole sphere of industrial production through the merging of digital technology and the internet with conventional industry".

The World Economic Forum built on this, with the overarching theme for the 2016 Forum being "Mastering the Fourth Industrial Revolution", which focused on [12]:

"mastering the speed, scale and force at which the Fourth Industrial Revolution is reshaping the economic, social, ecological and cultural contexts in which we live".

Broadly speaking, Industry 4.0 encompasses the recent advances in the technologies of: artificial intelligence, robotics, the "Internet of Things", autonomous vehicles, 3-D printing, nanotechnology, biotechnology, materials science, energy storage, augmented reality and cognitive computing [13].

Technologies of the Fourth Industrial Revolution

- ❑ Big Data
- ❑ Cloud Computing
- ❑ Augmented Reality
- ❑ Cyber-Physical Systems (CPS)
- ❑ Next Generation Robotics
- ❑ Robotic Process Automation (RPA)
- ❑ Material Science
- ❑ Cognitive Computing
- ❑ Power Microgeneration
- ❑ Internet of Things (IoT)
- ❑ Digital Twin
- ❑ Smart Cities
- ❑ Additive Manufacturing / 3-D printing
- ❑ Blockchain
- ❑ Wearable Technologies
- ❑ Energy Storage
- ❑ Smart Grids
- ❑ Renewable Energy
- ❑ Genetic Engineering
- ❑ Biotechnology
- ❑ Smart Factory
- ❑ Drones
- ❑ Autonomous Vehicles
- ❑ Platform Business Models
- ❑ Digitisation
- ❑ Nanotechnology
- ❑ Data Analytics

Figure 1: Landscape of technologies of the Fourth Industrial Revolution

Although many of these technologies are not new, what is pertinent is that we are now reaching the inflection point[14] where the supporting ecosystems are maturing enough to allow them to be cost-effectively and widely adopted. Importantly, the way that they collectively interact is poised to revolutionise much of the world's current industrial landscape.

An example of this ecosystem effect relates to 3D printing. Although the first 3D printers came to market in the early 1980's, they needed specialised skills to operate and were expensive to buy. It took a series of advances in the overall supporting ecosystem for them to progress towards more widely spread, cost-effective adoption. Some of these advances included:

- Maturing Computer Aided Design (CAD) software packages that simplified the user interface (1990's) and made it possible for non-specialists to create designs.
- Expiring of "Fused Deposition Modelling (FDM)" printing process patents in 2009 leading to a wider range of 3-d printer models being released and printer costs falling [15].
- Improved general computing skills within the general populous and the rise of internet sharing and the open source community where designs and modifications could be easily shared and design requests could be cheaply crowd-sourced.
- The falling costs of input materials (in the form of plastic spools) and improving ease of sourcing through online shopping platforms.
- Refined and improved design resulting in improved speeds and quality of printing through different technology variations and also the ability to print using different materials (plastic, metallic, organic).
- The rise of crowd sourcing, internet commerce platforms and supporting small-scale distribution infrastructure (through couriers) that allows an individual to make a viable business by utilising their 3D printer to create custom ordered objects.

The theme of this is not that the technologies are new, but that the maturing of their supporting ecosystem means that an inflection point has been reached that will shortly be followed by rapid adoption as well as radical disruption to existing business models [16].

3. ERODING THE PILLARS OF "TRADITIONAL" INDUSTRIALISATION

To understand the impact that this will have on industrial development, we need to examine its effect on the underpinning pillars of "traditional" industrialisation mentioned above

3.1 The eroding of the low-cost labour advantage

The first is the eroding of the low-cost labour advantage. A number of technological advances in Industry 4.0 are making it easier to produce the same production output with fewer people. Examples of these advances include:

Smart Machines

These are factory (or other) machines that are equipped with sensors and software that allow them to self-diagnose faults and to exchange information with other machines thus, allowing automated handover along a production line and a reduced need for human operators

Digital Twin

This refers to factories for which a virtual copy or "digital twin" has been created. This involves visualising the real time operations of physical factory equipment as a "digital twin" on-screen display that updates in real-time to reflect actual production status and location of goods in progress. This potentially allows for centralised control from one location, instead of requiring a machine operator for each individual machine. It can also allow real-time monitoring of data to detect potential problems before they occur [17].

Quality monitoring via image processing

Advances in camera captured image processing means that visual quality checks can be carried out by a machine, removing the need for a person to perform visual inspection and allowing defective products to be automatically side-lined as reject if they do not meet the required tolerances.

Sensor-aided predictive maintenance

The cost of installing and monitoring predictive maintenance sensors is falling [18], allowing them to be more widely used. Examples include attaching heat monitoring and vibration sensors to equipment motors that will trigger an alarm if the values go outside of acceptable limits. This aids personnel in knowing where to focus their attention and reduces the man-hours needed to conduct maintenance inspections.

Augmented reality assistance

Superimposing computer-generated information over the user's view of the real world, for example using a head-up display, is being used in a number of applications to enable people to be more productive. An example is assisting a maintenance technician to complete a repair job faster by displaying the help manual and critical parts information in real time and leaving the technician with both hands available to conduct repairs to a piece of equipment. An extension of this technology allows a remote specialist technician to give real time coaching and advice, as they are able to see on their computer screen exactly what the maintenance technician can see, thanks to camera enabled streaming video linkage. All of this means that fewer skilled technicians can accomplish more work [19].

Automated flows

The ability to cost effectively embed tracking technologies into finished products (e.g. RFID tags) allows for automation of flows across the logistics chain, from warehousing to distribution. This reduces the number of operations staff required as warehouse picking of products for a specific delivery order can be largely automated, utilising forklift equivalents that can independently retrieve the required items.

Improved Human/Machine interfaces

As control software for machines continues to evolve, the human interface has also become more user friendly, requiring less specialised skills to utilise. For example, the programming of an industrial robot in the 1990's required specialist coding knowledge and time-consuming manual input of detailed code to make small changes to the basic programming [20]. The software interfaces of today have drastically simplified this, allowing for much faster changes to programming parameters, quick testing of new code in a virtual environment and a more intuitive interface that reduces the number of specialised technicians required to conduct changes.

Robotic process automation

The applications of Artificial Intelligence (AI) software has expanded to encompass many traditionally "white collar jobs" through the technologies collectively known as 'Robotic Processing Automation' (RPA). This includes the automation of traditional back-office support functions [21] to manufacturing such as: inventory ordering and optimisation, finance functions of accounts payable and receivable and even legal contracting. Through the use of Natural Language Processing, which enables chatbots, this has also had an impact on removing the need for a person manned help desk support for or sales queries.

A key point relating to all of these examples is that the price of the technologies required to accomplish this has been rapidly falling so that they have become more cost-effective to implement. This applies to both new factory installations and to retro-fitting existing manufacturing operations. Moreover, the applicability of automation is advancing into types of manufacturing that have not historically been automated, for example clothing production [22].

The net result is that when making an investment decision as to where to locate a new factory, the availability of cheap labour may no longer a key consideration. Increasingly, a large factory can be cost-effectively run by a minimal work force with medium to high-level skills required. As the salary and wages bill has historically been one of the biggest cost-drivers in manufacturing, this means any manufacturing operation that wishes to be able remain cost-effective is going to be forced into adopting these new technologies.

3.2 Improving the economics of the decentralised manufacturing model

Historically, global consumer goods manufacturing companies have tended to favour a centralised, rather than decentralised manufacturing model for production line type (continuous) manufacturing. A centralised model concentrates manufacturing into a few large factories with specialised production lines that are able to mass-produce specific items in large volumes, utilising extended production runs. This has traditionally resulted in a better Return On Capital Invested (ROCI) through high equipment utilisation rates and efficiencies, that were necessary to offset the high initial investment costs of factory production equipment.

A decentralised model on the other hand consists of creating a number of smaller manufacturing facilities or factories, located close to the markets that they intend to supply. These factories are typically much more flexible – producing multiple product types with much shorter lead times, but achieve less overall equipment utilisation due to the lost time needed to changeover between small production runs.

The disrupting effect of the Fourth Industrial Revolution is the way that new technologies are simultaneously eroding the benefits of centralised manufacturing and removing the disadvantages of decentralised or distributed manufacturing [23].

A historical reason for centralised manufacturing and building a large factory, is that the cost of producing the equipment needed to produce even just one product item was so high that it required manufacturing large volumes in order to achieve a reasonable payback period on capital invested. New technologies are changing this by:

Reducing factory building costs

The design process of the factory can now be done virtually: Full computer aided design and simulation which reduces the final capital costs of building and equipment by ensuring that layouts are optimised and minimal changes to design are required once construction has started.

Reducing factory machinery costs

The man-hours required to both design and fabricate specialised manufacturing equipment is reducing thanks to the productivity enhances of Computer Aided Design (CAD) and improving Computer Aided Manufacturing (CAM) technologies (e.g. better interfaces to computer driven CNC (Computer Numerical Control) and metal working lathe machines). This means that Original Equipment Manufacturers (OEMs) are able to provide machines at lower costs.

Reducing prototype and tooling costs

3-D printing reduces costs to make prototypes and jigs [24]. This has historically been a large initial outlay cost, particularly in the automotive industry, where "tooling up" costs form the major portion of the Capital Expenditure (CAPEX) to produce each new model in an existing assembly facility.

Improving changeover efficiencies

Previously, long production runs were more efficient because of the amount of non-productive time and manual effort needed to changeover to produce a different product. With improvements in technology, this can now often be automated.

The consequence of this is that the capital investment cost needed to manufacture a small number of units has reduced. This means that there is less pressure to recoup sunk costs by producing large volumes – i.e. less benefit derived from maximising economies of scale. These advances also remove the disadvantages associated with small scale decentralised manufacturing, by reducing the capital costs and improving the efficiencies.

In this changing environment, decentralised small-scale manufacturing environment provides distinct advantages. Smaller production runs and greater changeover flexibility [25] means that the factory is able to be more responsive to swings in market demand. In a highly automated environment, it may also be easier to dial up production output incrementally, as utilising machinery longer does not necessarily require an additional labour shift. When logistics costs replace labour costs as a key driver, it is more efficient to transport finished goods short distances than long distances. Whilst this may require a more sophisticated input material supply chain, input materials are generally more efficient to transport in their raw, bulk format as there is less wasted space caused by packaging and less need to add padding and special measures to protect from transport damage. All of this adds up to less cost per tonne moved.

4. THE IMPACT ON THE FUTURE OF MANUFACTURING

The net result of this is that the future of some manufacturing industries that are currently dominated by the centralised manufacturing model are likely to shift future production to smaller, decentralised factories. This brings with it its own challenges of managing decentralised operations, quality control and intellectual property protection, but a tipping point is being reached where the benefits are starting to outweigh the challenges, meaning that manufacturers who get this right will reap the rewards of changing competitive advantage.

Another name that has been used to refer to this trend is “reshoring”. The advantages of smaller scale production being located closer to market is being recognised and acted upon [26]. The changes in manufacturing cost drivers is also helping to drive the “mass-customisation” trend away from mass produced goods, towards goods specifically customised to end-user requirements.

The relative advantages and disadvantages of small-scale decentralised manufacturing obviously vary per industrial sector. The diagram below shows illustrates a prototype model for which types of manufacturing could have early potential to move towards a decentralised, small-scale manufacturing model.

In this model, types of manufacturing operations have been loosely ranked (free-form comparative ranking) on two axis. The first axis compares centralised vs. decentralised advantages: i.e. where that type of manufacturing would benefit more from being closer to input materials or closer to end market in order to maximise logistics savings. This is based on compactness and ease of transport of finished goods vs. input materials. The second axis compares comparative cost advantages of producing at scale.

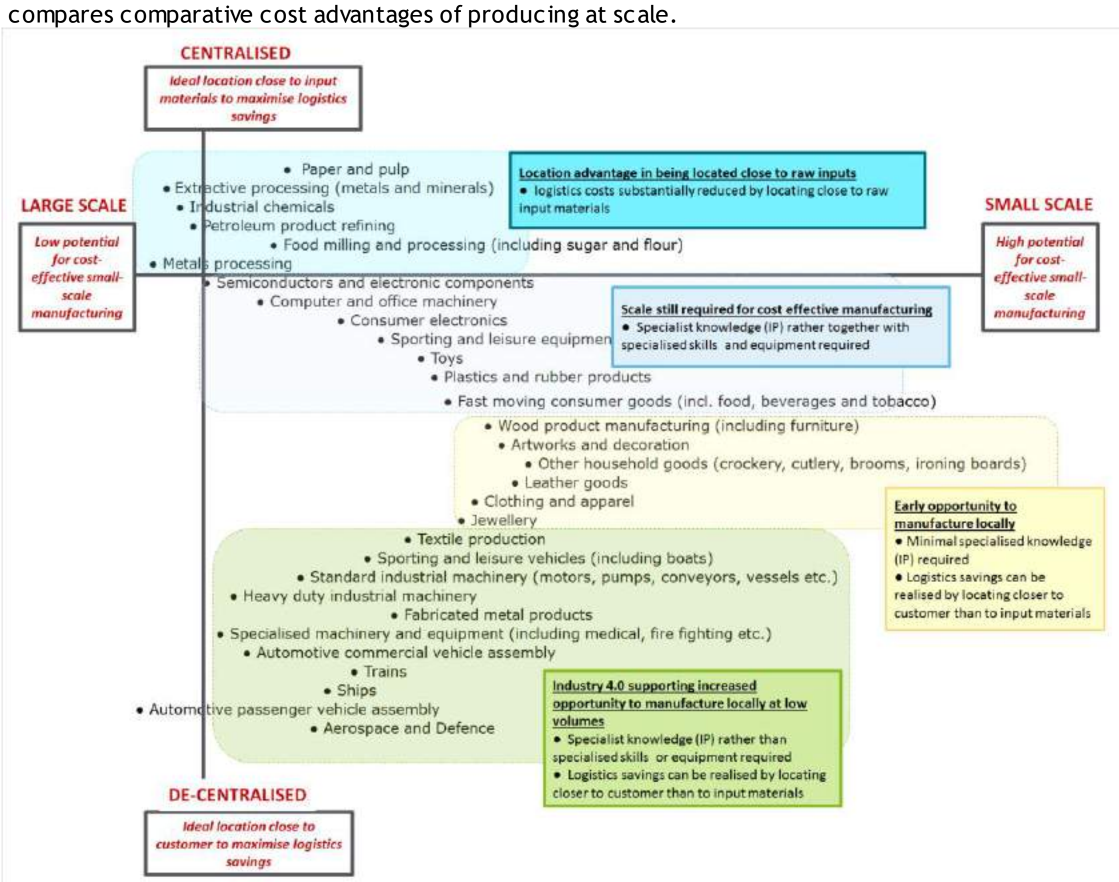


Figure 2: Illustrative view of manufacturing sectors that could have early potential to move towards a decentralised, small-scale manufacturing model

For example, if one looks at the paper and pulp industry, from a logistics efficiency perspective, it makes sense to locate the wood processing mills close to where the wood is felled, rather than close to the end markets. This is because processed paper for example is much more cost effective from a weight and space utilisation perspective to transport over long distances than raw, unprocessed timber. Paper and pulp processing/manufacturing therefore currently favours a centralised manufacturing model. On the other hand, in automotive assembly, where many different components come together from different sources to form the finished vehicle, there is no logistical advantage gained by locating an assembly plant nearby to the original raw material source of any individual component – i.e. no advantage in locating close to an iron ore mine or steel mill in order to minimise steel transport costs. It makes better sense to locate an automotive assembly plant where the route to market logistics costs for the final product are minimised – i.e. a decentralised manufacturing model. Looking at the horizontal axis, in the case of automotive assembly, current technologies still favour large scale production facilities as tooling up costs to create the first vehicle need to be spread out over a large production run to make economic sense. The interesting consideration is how the horizontal positionings of different on this diagram may shift as the impacts of Industry 4.0 technologies reduce the cost efficiencies to be gained from large scale manufacturing. For example, a large FMCG goods manufacturer currently will typically have one centralised factory making branded shampoo and supplying an entire region. The main component of shampoo by weight is water. It is not efficient from a logistics perspective to be transporting soapy water in bottles by road or rail from one central location to outlying regions. Emerging technologies may well mean that the model of the future is to have a small scale, highly flexible shampoo-making factory located in every major metropolitan area.

5. CONCLUSION

The key takeaway from the above, is that the route to furthering industrialisation in African (and other) countries the model for successful industrialisation going forward is likely to differ substantially from that of the past. In particular, there may be few, if any large scale industrial projects that will make a significant dent in unemployment. A new pathway needs to be forged.

While this is daunting, it is also exciting. The fact that these two historical sources of competitive advantages are being eroded means that developing economies (the majority of Africa's economies) are not necessarily starting at a disadvantage. They may even have certain advantage as there is less inertia and pushback to cling to the old, and more willingness to make use of any available opportunities to improve living conditions

6. REFERENCES

[1] The Third Industrial Development Decade for Africa (2016-2025): Joint Communique | UNIDO', accessed 26 June 2018, https://www.unido.org/news/third-industrial-development-decade-africa-2016-2025-joint-communique.

[2] Nimrod Zaik, 'What Is the Role of Manufacturing in Boosting Economic Growth and Employment in South Africa? | Econ3x3', Econ 3x3, February 2014, 4, http://www.econ3x3.org/article/what-role-manufacturing-boosting-economic-growth-and-employment-south-africa.

[3] Ronald W. Jones, 'Heckscher–Ohlin Trade Theory', in The New Palgrave Dictionary of Economics (Palgrave Macmillan, London, 2008), 1–13, https://doi.org/10.1057/978-1-349-95121-5_1116-2.

[4] Joseph E. Stiglitz, 'From Manufacturing Led Export Growth to a 21st Century Inclusive Growth Strategy for Africa' (JICA International Advisory Board Meeting, Tokyo, Japan, 27 March 2018), https://www8.gsb.columbia.edu/faculty/jstiglitz/sites/jstiglitz/files/Stiglitz%20March%2027%20Manufacturing%20Led%20Growthtokyomar2018.pdf.

[5] Yi Wen, 'The Making of an Economic Superpower —Unlocking China's Secret of Rapid Industrialization' (Federal Reserve Banks of St. Louis, 19 August 2015), 10, https://ssrn.com/abstract=2646145.

[6] McKinsey & Company, 'What's next for China?' (McKinsey & Company, January 2013), https://www.mckinsey.com/featured-insights/asia-pacific/whats-next-for-china .

[7] Craig A. Giffi et al., 'Deloitte Global Manufacturing Competitiveness Index 2016' (Deloitte Center for Industry Insights, 2016), 1, https://www2.deloitte.com/sg/en/pages/manufacturing/articles/2016-global-manufacturing-competitiveness-index-pr.html .

[8] Mary Hallward-Driemeier and Gaurav Nayyar, Trouble in the Making?: The Future of Manufacturing-Led Development (The World Bank, 2017), 1, https://doi.org/10.1596/978-1-4648-1174-6

[9] Seth Godin, Meatball Sundae: Is Your Marketing out of Sync? (New York: Portfolio, 2007), chap. The fourth industrial revolution.

[10] Industrie 4.0 – Smart Manufacturing for the future. Germany Trade and Invest. 2014. https://www.gtai.de/GTAI/Content/EN/Invest/_SharedDocs/Downloads/GTAI/Brochures/Industries/industrie4.0-smart-manufacturing-for-the-future-en.pdf, accessed August 2, 2017

[11] 'Speech by Federal Chancellor Angela Merkel to the OECD Conference' (Growth, Innovation and Competitiveness: Maximising the benefits of knowledge-based capital, Paris, 19 February 2014), https://www.bundesregierung.de/Content/EN/Reden/2014/2014-02-19-oecd-merkel-paris_en.html.

[12] 'World Economic Forum Annual Meeting 2016- Mastering the Fourth Industrial Revolution' (Switzerland: World Economic Forum, n.d.), 3, http://www3.weforum.org/docs/WEF_AM16_Report.pdf.

[13]Klaus Schwab, 'The Fourth Industrial Revolution: What It Means and How to Respond', World Economic Forum, 14 January 2016, https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond .

[14]Dr. Takis Damaskopoulos, The Fourth Industrial Revolution: Inflection Point, Entropy, or Interregnum? – Working draft study overview, European Institute of Interdisciplinary Research, 7 July 2017, 5, https://eiir.ca/wp-content/uploads/2017/08/IR4-STUDY_Overview-2_07-2017.pdf, accessed 9 September 2018.

[15]Hannah Bensoussan, 'The History of 3D Printing: From the 80s to Today', 3D Printing Blog: Tutorials, News, Trends and Resources | Sculpteo, 14 December 2016, https://www.sculpteo.com/blog/2016/12/14/the-history-of-3d-printing-3d-printing-technologies-from-the-80s-to-today/.

[16]K. Schwab, The Fourth Industrial Revolution (World Economic Forum, 2016), chap. 2.2 Tipping points, https://books.google.co.za/books?id=mQQwjwEACAAJ.

[17]Bernard Marr, 'What Is Digital Twin Technology - And Why Is It So Important?', Forbes, 6 March 2017, https://www.forbes.com/sites/bernardmarr/2017/03/06/what-is-digital-twin-technology-and-why-is-it-so-important/.

[18]Mario Honrubia, 'Industrial IoT Is Booming Thanks to a Drop in Sensor Prices', Ennomotive, 17 August 2017, https://www.ennomotive.com/industrial-iot-sensor-prices/.

[19]John Ahmet Erkoyuncu et al., 'Improving Efficiency of Industrial Maintenance with Context Aware Adaptive Authoring in Augmented Reality', CIRP Annals 66, no. 1 (2017): 465–68, https://doi.org/10.1016/j.cirp.2017.04.006.

[20]Zengxi Pan et al., 'Recent Progress on Programming Methods for Industrial Robots', Robotics and Computer-Integrated Manufacturing 28, no. 2 (April 2012): 87–94, https://doi.org/10.1016/j.rcim.2011.08.004.

[21]David Schatsky, Craig Muraskin, and Kaushik Iyengar, 'Robotic Process Automation - a Path to the Cognitive Enterprise', 14 September 2016, https://www2.deloitte.com/insights/us/en/focus/signals-for-strategists/cognitive-enterprise-robotic-process-automation.html.

[22]21 b Wolfgang Lehmacher, Francisco Betti, Paul Beecher, Christian Grotemeier, Maike Lorenzen, 'Impact of the Fourth Industrial Revolution on Supply Chains', World Economic Forum, October 2017, 10, http://www3.weforum.org/docs/WEF_Impact_of_the_Fourth_Industrial_Revolution_on_Supply_Chains_.pdf, Accessed 9 September 2018

[23]22bDaniel Küpper, Ailke Heidemann, Johannes Ströhole, Daniel Spindelndreier, Claudio Knizek, 'When Lean Meets Industry 4.0 Next Level Operational Excellence', Boston Consulting Group, 14 December 2017, https://www.bcg.com/en-gb/publications/2017/lean-meets-industry-4.0.aspx, accessed 9 September 2018.

[24]Jana Bukolovska, 'Filling Automation Gap in Garment Manufacturing', 23 June 2017, https://www.innovationintextiles.com/filling-automation-gap-in-garment-manufacturing/.

[25]Carrie Wyman, 'Opel Cuts Assembly Tool Production Costs Up to 90% with 3D Printing', Stratasys Blog, 18 November 2015, http://blog.stratasys.com/2015/11/18/opel-3d-printing/.

[26]'Coming Home - Reshoring Manufacturing', The Economist, 19 January 2013, https://www.economist.com/special-report/2013/01/19/coming-home.

CHAINING THE BUILDING BLOCKS FOR BLOCKCHAIN IMPLEMENTATIONS IN SOUTH AFRICA'S PUBLIC SECTOR

Q. van Heerden[1*], A. Steenkamp[2], M. van Heerden[3]

[12]Council for Science and Industrial Research
[1]qvheerden@csir.co.za, [2]asteenkamp@csir.co.za

[3]Department of Public Administration and Management
UNISA, South Africa
vheerm@unisa.ac.za

ABSTRACT

The emergence of the 4th industrial revolution has brought about several technological advancements in which blockchain fulfils a prominent role. To most people, the notion of blockchain is merely synonymous with the popular cryptocurrency Bitcoin. However, blockchain is not a cryptocurrency *per se*, but rather provides the backbone for the existence of cryptocurrencies. Blockchain has multiple inherent and potential benefits, especially so for governmental use-cases with the governments of Dubai, Estonia and Gibraltar already investing in such implementations. Blockchain applications include healthcare, identity management, voting, and banking services, among others. Since records on the blockchain cannot be altered, it is deemed to be a secure option for many of these implementations. The contribution of this paper is twofold. Firstly, it provides a description of the use of smart contracts, a form of self-executing contract, based on predefined rules between buyers and sellers. It describes how these contracts can potentially be used in South Africa to eradicate corruption in governmental processes by improving accountability and transparency. Secondly, it provides a blueprint of the use of Industrial Engineering tools and techniques to adequately design and implement blockchain use-cases, especially in the public sector.

*Corresponding author

1. INTRODUCTION

It is almost daily that we hear of, or experience, one or more public service delivery protests. The delivery of essential public services is either lacking in quality, does not have a transparent process, or has an obvious corrupt element involved. To better respond to basic service demands, with more transparency and trust, government needs assistance. The advances in technology, more specifically related to the fourth industrial revolution, could potentially assist in this regard.

In the Industry 4.0 report, Deloitte states that South Africa, like other developing countries, still faces challenges in terms of connectivity and accessibility, which slow down advanced technological adoption [24]. While these phenomena are evident, applications could be developed in such a way as to ease the learning curve, increase accessibility, and make it user-friendly to interact with. The report, furthermore, states that industry leaders and policy makers are increasingly acknowledging that increased effort is required to adopt these technologies. Affordable internet and smartphone penetration are increasing and advanced technological solutions should be developed well in advance to ensure that when internet and mobile usage become common-place, the solutions are in place. Smartphones are key to the adoption since payment gateways, applications, and access to information are increasingly moving to mobile phones for increased accessibility.

The fourth industrial revolution has seen various technological advancements in multiple domains, such as: machine learning, the Internet of Things (IoT), and blockchain. However, not all these advancements have been favourably received, especially not by governments.

1.1 Governments and the use of blockchain

Even though some governments have banned certain blockchain applications, there are other governments that have become increasingly interested in the technology as a means to better regulate, maintain, control, and report on certain public services. For example, the government of Gibraltar is the first government to approve a blockchain exchange to trade cryptocurrencies on. They are also the first to issue licenses for fintech firms to operate using the blockchain and cryptocurrencies as a recognised means of transaction records and an accepted mechanism of payment [1].

Estonia has already implemented multiple blockchain solutions in an integrated electronic solutions environment, with applications ranging from legislation, to managing court procedures, to more effective policing systems [2]. Estonia is recognised as a leader in eGovernment, with their most prominent contribution being the development of a digital signature system, *Keyless Signatures Infrastructure®* (*KSI*), a real-time authentication system for digital assets. This system is used to verify government processes and improves trust and transparency in services to the public.

Dubai is set to become the first blockchain-powered government by 2020 through its blockchain strategy, which was developed in 2016. Dubai plans to harness blockchain technology for everything through a Smart City Platform that integrates everything from license renewals, to visa applications and the payment of bills and eventually to be completely paperless [4].

1.2 Other smart city blockchain implementations

Digital Town aims to provide turn-key solutions that relate to smart governments, civic engagement, digital inclusion, and smart tourism. The platform allows the city and its citizens to connect digitally and to promote local production and consumption. Citizens can be rewarded in fiat currency or cryptocurrency, which they can again spend at local suppliers; this promotes the development of the local economy. The platform, furthermore, connects municipalities, local businesses, e-commerce merchants, residents, developers, and start-ups to create a large ecosystem, all powered by the blockchain [5].

1.3 Blockchain adoption in South Africa

The South African Reserve Bank has recently started to test the blockchain as a means to transfer funds and settle transactions in a quicker and cheaper manner than what is currently available [6]. This is, however, aimed at developing an understanding of the pros and cons of using blockchain for financial transactions so as to develop relevant policies and regulations on the use of cryptocurrencies.

The South African government has also implemented some electronic systems to improve certain processes, with the most prominent being the smart identification card system, which replaced the previous identity books. The major hurdle, however, still remains in the processes that govern some of the systems and the many processes that aren't transparent. It is still too easy to circumvent policies and procedures, since the systems that govern the processes can be tampered with or ignored completely. Various opportunities are still available to use blockchain technology in South Africa to eradicate corruption and fraudulent transactions in the public sector.

The contribution of this paper is twofold. Firstly, it aims to give an overview of governmental processes that are prone to corruption and that face the most protest action. Secondly, it aims to describe the working of blockchain technology in terms of transactions, transfer of ownership, and smart contracts, to set the scene for how these can be used to improve governmental processes in terms of transparency, public participation, and improved efficiency. It indicates that Industrial Engineers are well-suited to play various roles in implementing blockchain solutions in the public sector in South Africa.

2. GOVERNMENT PROCESSES

Section 195(1) of the Constitution of the Republic of South Africa, 1996, governs the activities and services of public institutions and places an ethical duty on all spheres of government to achieve and uphold a fair, transparent and honest administration, which serves the general interests of the public. In terms of section 85 of the 1996 Constitution, all government institutions must provide public services impartially, fairly, equitably and without bias. People's needs must be responded to in an accountable manner because government creates expectations among the public that public services will be provided. Government may not fail to deliver public services because this will disrespect the principles of the Bill of Rights in chapter 2 of the Constitution, 1996 [7].

Unfortunately, today and for a number of years, South Africa experiences and has experienced public service delivery as insufficient and with a high level of corruption. This situation usually leads to protests against poor public service delivery. Government institutions are unable or unwilling to prioritise the welfare of the public. It also appears that all levels of government are relatively unwilling to make use of, or are uninformed about, the availability of alternative public service delivery techniques that can improve the public service delivery process. Furthermore, the government does not appear to have any intention of addressing the reasons and underlying causes for the protests. This is clearly a lack of responsiveness and accountability towards the public and also a matter of moral concern because the ministerial head of departments must ensure that government functions are properly performed. Should this process of service delivery fail, public service delivery together with government's legitimacy suffers [8].

To ensure the proper use of public money, Government established the office of the Auditor-General in terms of the 1996 Constitution, Section 181(1)(e). In an interview with City Press (a South African Sunday newspaper) on 25 May 2018, the Auditor-General, Mr Kimi Makwetu, presented facts exposing irregular expenditure at municipalities. He stated that poor governance and leadership were the main factors that caused municipalities to malfunction. Although the Auditor General has given advice to these institutions, such advice was not being implemented or is ignored as the same bad practices persist. He mentioned that many financial officials were placed in their positions without having the required skills, thereby enabling people inside and outside the municipality to abuse its coffers and create a "free for all" situation. South Africa has 257 municipalities and the Auditor-General exposed an increase of 75% in irregular expenditure (amounting to R28-billion) for the financial year of 2016-2017. He pointed out that the South African fiscus simply cannot afford to lose such finances due to the fact that total government debt is currently R2.3 trillion, which accounts for 50% of the Gross Domestic Product (GDP) [9].

Public institutions are obliged to ensure that all public activities and expenditures are exercised with adherence to supply chain management regulations. There is also a distinction between the way in which authority is structured and the way in which it is applied. The first is dependent on formal organisational structuring whereas the latter is a personal orientation [8]. This indicates that a balance is expected to be maintained between the authority that is granted to the executive institution and the manner in which such authority is exercised.

The Auditor-General mentioned that the worsening financial state of the country's municipalities has become a concern [10]. Of the R350-billion total expenditure budget for 257 municipalities, 3.5% (or R12.2-billion) of the budget was irregular expenditure mainly due to non-compliance with supply chain management regulations. This meant that support programmes to improve financial management in municipalities had to be introduced. In a situation such as this, irregular expenditure and non-compliance with regulations can also be linked to irresponsible, unaccountable and corrupt activities that need to be curbed and prevented.

Besides the South African municipalities, the South African education system is also in a crisis situation. The number of state owned schools declined by 12% between the years 2000 and 2016. In the same period, the number of private schools increased by 91%, from a number of 971 to 1 855 schools [11]. The problem is that private schools have to generate their own income by means of school fees and these are usually higher than that of state owned schools. There are, however, some private schools that charge fees that are affordable to parents. The additional problem is that parents need to have a choice of the school for their children. The Institute of Race Relations proposes the introduction of a voucher system, whereby parents are given a voucher for each of their children. This is to be taken out of the Government Education Budget and a voucher will be to the value of about R12 000. Parents will then have the choice about which school to send their children to and will be able to use the voucher at private or state schools [12]. This financial procedure is another example that needs careful, responsible and accountable management as well as transparency or it could easily become another mismanaged and corruptive practice.

All public procedures, especially those involving public finances, must be transparent and served by trained, informed and honest public officials. Faced with these challenges, governance in South Africa has to comply with the demands. The time has come for the South African public bureaucracy to transform into a dynamic and flexible organisational entity capable of responding quickly to a changing environment [8]. Such a transformation should include modern and current systems such as blockchain to create surety in processes and to eradicate corruption in governmental processes by improving accountability and transparency, which, in turn, will improve trust from citizens.

3. BLOCKCHAIN AND THE NOTION OF SMART CONTRACTS

This section provides an overview of blockchain technology with specific emphasis on how it enables smart contracts. A more technical and detailed discussion is also provided on smart contracts with an emphasis on key features that can facilitate governmental processes.

3.1 Overview of the workings of the blockchain

The notion of *blockchain* can be divided into two parts: *blocks*, which refer to a specified number of validated transactions; and *chain*, which refers to all the historical blocks of transactions chained together. There is no central authority (such as a bank) that controls blockchain transactions. The blockchain is, rather, an example of distributed ledger technology (DLT) in which storage and computation are shared between the member users (referred to as the nodes) who are connected on a peer-to peer network; the members contribute to and control the network [14]. It can therefore be thought of as a distributed database that records every single transaction that has ever occurred on the blockchain network. Once a block has been created and appended to the blockchain, it is almost impossible to change or reverse these transactions. An illustrative example of how the blockchain works is provided in Figure 1 and its content is discussed in the remainder of this subsection.

A ledger is a term used in accounting that describes a document that reflects all of the completed transactions of a business. As mentioned, blockchain is an example of a distributed ledger and thus implies that all of the transactions on the blockchain are similarly recorded as would be the case in an accounting system. The ledger in the case of blockchain is, however, not maintained by any central authority. Any updates to the ledger are constructed independently and also recorded by each node. Currently, the majority of cryptocurrencies use the concept of *Proof of Work* (PoW) to validate a block of transactions. This requires a complex mathematical problem to be solved, which is known as a *hash function*. A hash function is any function that can be used in order to map data of an arbitrary size to data of a fixed size. A hash function is said to be secure if the output is indistinguishable from random.

Transaction validators (other participants on the network), also referred to as nodes, all then try to solve this complex mathematical problem, since a reward is offered to the node that first validates the transaction. This problem requires a lot of computational resources in the form of electricity and CPU processing time. The difficulty of the calculation that has to be solved is adjusted according to the demand and number of nodes. Once a node has obtained a solution to this mathematical problem, the solution is broadcasted to the network in order for all of the other nodes to validate that the solution is correct. The nodes then vote in order to determine if the transactions are indeed authentic. This process of voting and agreement between the nodes on a version of a ledger is referred to as *consensus*. An alternative to PoW, is *Proof of Stake* (PoS), which does not require any additional work (as is the case with PoW), but instead investors are rewarded simply based on the number of coins that they hold. The interested reader is referred to Zheng et al. [15] for more information on these concepts.

Blockchain has attracted a lot of attention due to its potential for facilitating transactions among unknown parties, without the need for a trusted third party [16]. This is due to its key characteristics of decentralisation, persistency, anonymity and auditability [15]. This removal of the need for a third party significantly reduces the cost and time associated with completing a transaction and also significantly expands the potential market, as the risk is reduced or eliminated and transparency is increased.

A blockchain can be classified as a private or public blockchain. In the case of a public blockchain, any anonymous user is allowed to join the network, read the blockchain's content, participate in a transaction, or be allowed to verify the correctness of each new block. Bitcoin and Ethereum are examples of public blockchains. Private blockchains can only be accessed by users with permission. A company, or a group of companies, usually gives this permission. An example of such a private blockchain is that of *Ripple*.

It is frequently stated that the blockchain is completely immutable and unchangeable, but this is slightly misleading. A more accurate description would be that it is "Mutable-By-Hashing-Power", which means that if (and only if) one entity owns more than 50% of all computing power that contributes to the blockchain, then only can the blockchain be changed by someone. This is, however, a highly unlikely scenario which has not been encountered by any current blockchain network [14].

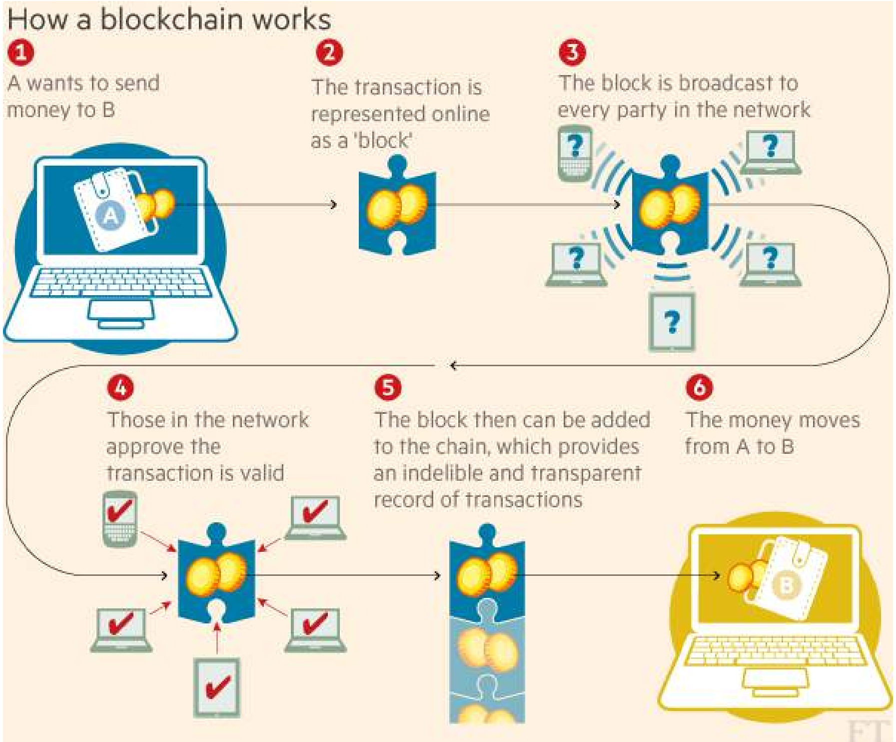


Figure 1: High level workings of the blockchain [23]

3.2 Smart Contracts

Smart contracts are an application of blockchain technology that is believed to have the potential to revolutionise the way in which business and administration is done across the globe. A smart contract is defined as an executable piece of code that is run on the blockchain and used to execute and enforce the terms of a specific agreement [16]. Smart contracts allow for the automatic execution of the terms of an agreement as soon as specified conditions are met.

This technology, therefore, eliminates the need for expensive and time-consuming third party legal systems in order to monitor the progress of a contract and to enforce its terms. It also promises to significantly reduce the cost and risk of conducting business across the globe. This could be specifically advantageous to Small, Medium and Micro-sized Enterprises (SMME's) that do not have the funds to obtain substantial legal support and are therefore often prone to exploitation. Smart contracts can be used for a myriad of applications including the transfer of any physical asset such as a house and even to act as a voting system [17]. It is thus a prime candidate to facilitate a number of public processes in South Africa.

The concept of a *smart contract* has been around since 1994 and was originally proposed by Szabo [18]. As with current smart contracts, he proposed that cryptographic and other security mechanisms could be used to enforce electronic contracts. The concept was, however, not explored to any great extent until the emergence of blockchain technology, which provided the necessary functionality [16].

A smart contract consists of an account balance, private storage of funds as well as executable code. A contract's state consists of the storage and the balance of the specific contract. It is this state that is stored on the blockchain and is updated every time that the contract is called or invoked.

Once the contract has been deployed onto the blockchain, the content or code of the contract cannot be changed or tampered with. Every contract is assigned to a unique 20 bytes address that acts as the account for the contract. The code is executed when a transaction is sent to the contract's unique address. Each node on the blockchain will then run the code in order to reach a consensus on the output. Once consensus is reached, the contract will be in a new state, which will be updated accordingly on the new block created.

Smart contracts can be classified either as deterministic or non-deterministic. A deterministic contract does not require any additional information external to the blockchain in order to execute any terms. A non-deterministic contract requires information from a party external to the blockchain. Necessary mechanisms should, thus, be put in place in order to execute contracts that require external input.

Non-deterministic contracts currently pose a challenge, since the verification of the accuracy of external data can be troublesome [17]. It is, however, believed that with the rise of the 4th industrial revolution and the Internet-of-Things (IoT) coupled with advances in machine learning, that this would become less of a stumbling block in the near future.

3.3 Cryptocurrency

One of the first applications of blockchain technology was the development of cryptocurrency as a peer-to-peer (P2P) payment solution. The first cryptocurrency to be developed after the creation of blockchain technology was Bitcoin and was created by the mysterious Satoshi Nakamoto in 2008. Since then many new cryptocurrencies have been developed, which are known as *altcoins* (alternative coins), which include Ethereum and Ripple. These altcoins still use similar cryptographic technology but have employed different algorithmic designs in order to address some of the perceived concerns and limitations of Bitcoin. At its peak, cryptocurrencies had a market valuation in excess of $800 billion [19].

Cryptocurrency is a subset of digital currency, which is not issued by a central authority and thus is not confined to a specific geographical area and is not tied to any specific fiat currency. An example of a centralised electronic currency is a loyalty token or some sort of reward points that are earned. The use of a decentralised P2P payment system promises to resolve many of the shortcomings of centralised fiat currency systems such as increased capacity, better security, faster settlement and lower transaction costs [19].

Since cryptocurrencies run on P2P networks, transactions take place directly between users without an intermediary party. The transactions are recorded on the ledger after verification as discussed earlier, which prevents double spending and also allows the prevention of any alterations to a transaction. These cryptocurrencies therefore allow for transparent, flexible, decentralised transactions with low transaction fees. This is an ideal payment system for use in the public sector.

3.4 Transfer of ownership and voting

Cryptocurrencies can be used to transfer the ownership of assets including financial assets or even to cast a vote. As indicated earlier, blockchain technologies have several advantages over a centralised system. The advantages over a traditional system will be explained with the use of a few simple examples.

The first example is that of the differences between the traditional trading securities market and the securities market as created by smart contracts. The former is very complex and requires a plethora of intermediary processes, brokers, and institutions. The typical process is illustrated in Figure 2.

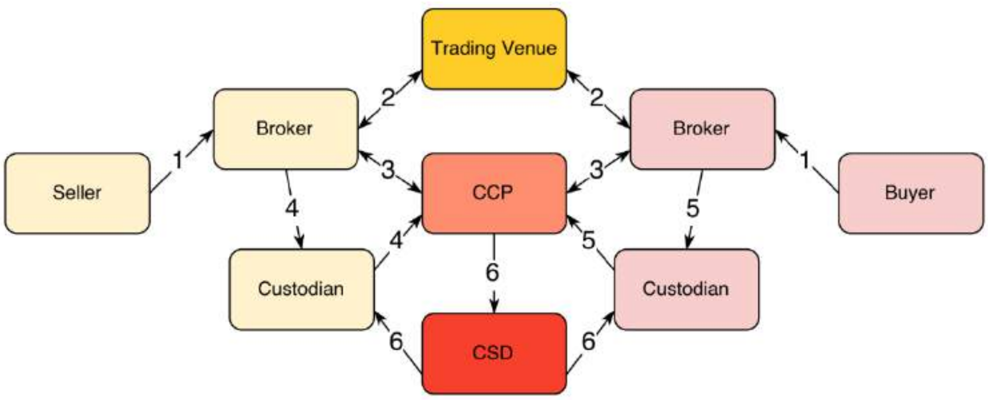


Figure 2: Illustration of the financial entities involved in a typical securities trade [20]

As illustrated in Figure 2, the buyer and seller first have to contact their own stockbrokers. The two stockbrokers will charge a commission and then subsequently introduce the buyer and seller to the second middle man, which is called the Central Counterparty Clearing House (CCP). The CCP's job is to ensure that none of the contracting parties default on their obligation. The CCP will charge its own commission in order to proceed to taking the asset from the seller's broker through the seller's custodian and also then receive money from the buyer's broker through the buyer's custodian. The CCP will then instruct the Central Security Depository (CSD) to credit the buyer's custodian with the transferred assets and the money in the case of the seller's custodian in order to conclude the transaction [17]. This is a very complex, time consuming and expensive process, which is needed in order to minimise risk between the two parties.

In the case of a smart contract, no intermediaries are required at all. Instead, the smart contract code is executed on the blockchain and the transactions and transfer of ownership are recorded on the ledger. It is, therefore, evident that this would lead to a substantial reduction in processing time as well as much lower transaction fees, which increases market size and access.

Smart contracts are also ideally suited to develop voting systems on the blockchain as a transaction can simply be sent to the smart contract to cast a vote. Various smart contract code options are available but the generic process will subsequently be discussed.

The election administrator owns the smart contract that will facilitate the voting process. The election administrator is responsible for enlisting the candidates or possible resolutions, authenticating all the voters by using their user-controlled account and subsequently updating the list of eligible voters. Certain timers are specified in the smart contract to ensure that the process occurs in a timely manner and has a definite stop time after which votes should be tallied. Only eligible voters are allowed to register and subsequently cast their vote. The administrator will then change the state from registration to voting. At this point the registered voters will be allowed to cast their single vote. This vote can either occur with the transfer or a specified amount of cryptocurrency or by publishing a unique key assigned to each voter during registration [21].

Due to the fact that each registered voter's identity is stored on the blockchain via the smart contract, no duplication of votes can occur. The system can be made even safer by having an end-to-end system where the results are sent to the voter, who can then validate that their vote has been correctly recorded.

4. BLUEPRINTS OF USING BLOCKCHAIN IN THE PUBLIC SECTOR

Smart contracts have the potential to substantially reduce corruption and irregular spending in South Africa's public sector whilst simultaneously improving service delivery. The auditability of the blockchain also implies that the public sector can easily be audited, which would bring about significantly more accountability in South Africa. Blockchain technology has already been successfully used by the South African Reserve Bank during June 2018 in order to deal with real-time gross settlements. The project was a huge success and the typical transactional volume experienced in South Africa was processed in less than two hours [6]. This will set the scene for blockchain regulation and the use of blockchain for other applications in the public sector.

The rest of this section provides a few ideas as to how blockchain can be used in the public sector. It is, by no means, considered the golden bullet to completely eradicate corruption nor solve all problems, but it will provide a means to improve the control of public sector processes, enhance transparency, and reduce the opportunity for corruption. At the very least, if it does not reduce corruption, it will enhance accountability through the fact that everything that gets captured on the blockchain, is forever stored and malicious processes will be flagged and denied.

4.1 National elections and voting in public hearings

Blockchain technology also provides superior performance in the case of voting. In traditional voting systems, there is a need for an independent third party that will facilitate the entire process, for example the Independent Electoral Commission (IEC). This third party is responsible to verify the identity of voters, tally the votes, and audit the results, which can be a tremendously tedious exercise. This third party must then also create all of the material necessary to record votes, ensure no double voting occurs, no unrecorded votes enter the system and no votes are removed or tampered with. They are also responsible for counting the votes and making the results known. It is evident that this system is very complex and introduces numerous possible points of failure and is very susceptible to corruption and tampering.

In the case of a blockchain-based voting system, various options are available to facilitate a more efficient process. Firstly, the Know Your Customer (KYC) capabilities can be used to verify identities of individuals by utilising a service such as *Civic*, a blockchain identity verification technology, conceived and co-founded by South Africa's Vinny Lingham [22]. A cryptocurrency transaction between the voter and the voting system can be used to cast a vote. Due to the various key properties of the blockchain, this vote would be free of any possibility of duplication, tampering or elimination as the transaction is recorded on the distributed ledger. The transaction is also easily auditable and the network has built-in capabilities to automatically tally the votes as they are cast.

In the case of public hearings, it is often the case that few to no people at all pitch up, since the hearings aren't advertised properly. The public is often disillusioned due to the fact that they did speak up and voice their opinions, but that these weren't adequately captured and reported and that the public sector just implemented projects according to their own political agendas. By using the blockchain in public hearings, would again improve the process tremendously as follows. Firstly, the event itself can be put on the blockchain to ensure that there is a record that it took place. KYC can again be used to verify the identity of those present. The smart contract can be set up in such a way as to require a minimum number of public participants at the event or the event has to be rescheduled, which would prompt the public sector to make a bigger effort to involve the public.

Once the proposed projects have been presented, the participants can give inputs that are recorded on the blockchain, whereafter participants can cast their votes. The votes are automatically tallied and the outcome determined by the smart contract. It would be possible to still allow a central authority, such as National Treasury, to set up the smart contract, but even in this case it would be beneficial to involve the public to set the rules of the hearings and in doing so enhance the trust and participation of the public.

The potential process for such voting is presented using the Business Process Modelling Notation in Figure 3. The two swimlanes with grey blocks and a star (*) next to the role player are functions that could be executed purely on the blockchain with minimal external input.

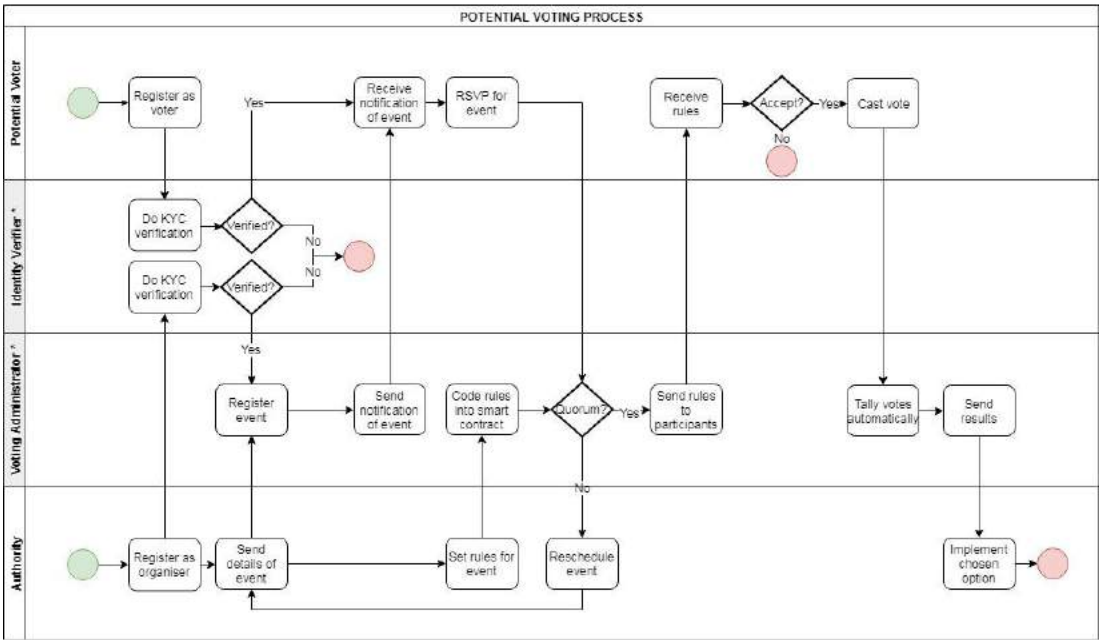


Figure 3: Potential voting process with blockchain as the backbone

4.2 Subsidies, housing provision, bonds

The backlog of housing provision is causing a lot of uprise. Furthermore, government is struggling to keep track of who has applied for subsidised housing, how long they have been waiting, and when transfer of ownership took place. This also makes it extremely difficult to report on progress made in line with what was planned in the major infrastructure projects, such as the National Development Plan [25] and the 9 Point Plan [26].

By introducing a blockchain application, each applicant can be registered on the system using KYC technology again. If this application is integrated with the national identification system and South African Revenue Service, up to date information on household income levels can also be automatically obtained to determine whether the household qualifies for subsidised housing; the smart contract will take care of all of these checks. Finally, when a house or a subsidy is awarded to a citizen, this is recorded on the blockchain and this will eliminate duplicate awards or corrupt transfer of funds.

Blockchain could also play a valuable role in the "willing buyer, willing seller" housing market. During the 4th quarter of 2017, the average value of a residential home bond in South Africa was valued at R1.11 million for 36 943 bonds [28]. Assuming these values remained constant during 2017, approximately R165 billion worth of bonds were approved during 2017. The typical costs associated with obtaining a bond of R1.11 million at one of South Africa's largest banks is provided in Figure 4, using an online calculator [29]. The table contains a comparison between the costs associated with a theoretical bond with and without using blockchain.

Using blockchain technology, all (or most) of the administrative functions can be automated and therefore the administrative costs reduced. By saving the initial admin fees and saving service fees, it is expected that the total cost can be reduced by approximately 5.4% if all other factors remain constant. This calculation does not even consider the costs associated with respect to time and inconvenience for the applicant.

Extrapolating these savings to the approximate R165 billion worth of bonds approved in 2017, the savings for residential bonds could amount to R21.2 billion rand over a period of 20 years or R1.06 billion per annum. These savings would result in higher disposable income for households and could stimulate the economy. This just highlights the possible impact of blockchain technology on one small case study, not to mention the potential savings from other forms of asset sales including automobiles and commercial property. Savings of this magnitude would have a substantial impact on the South African economy.

Bond Cost Comparison with and Without Blockchain		
	Bond with all costs	**Bond without admin costs**
Value of the bond required	R 1 110 000	R 1 110 000
Deposit	N/A	R 53 697
Value of bond taken	R 1 110 000	R 1 056 303
Interest rate	10%	
Bond period	240 months	
Once off fees		
Transfer costs	R 23 830	N/A
Bond costs	R 23 830	N/A
Initiation fee	R 6 037	N/A
Transfer duty	R 6 000	R 4 390
Total including transfer duty	R 59 697	R 4 390
Total excluding tranfer duty	R 53 697	R 4 390
Monthly payment		
Repayments	R 10 615.24	R 10 097.05
Service fees	R 69	N/A
Total	R 10 684.24	R 10 097.05
Bond comparison		
Total payments	R 2 623 914.60	R 2 481 379.00
Total savings		R 142 535.60
% savings		5.4%

Figure 4: Bond cost comparison with and without blockchain

4.3 Local economies and loyalty programmes

As mentioned in Section 1, the Institute of Race Relations proposes the introduction of a voucher system for education purposes, whereby parents are given a voucher for each of their children. By using smart contracts for these purposes, predefined rules can be put in place that will ensure that parents do not receive more than they should, cannot override the system or use the money for anything other than education purposes.

In the greater *smart city* realm, families or households can be rewarded with cryptocurrency for services that they render for the greater public good, such as recycling. This currency can then also be used to reduce school fees, for public transport or even for entrance to certain sites. It essentially acts as a subsidy, but with complete control as to where it can be spent and transparency on how and when it was spent. Reporting will also be greatly enhanced as all transactions are captured on the blockchain network and standardised reports can be developed on top of the network. This can facilitate improved monitoring and evaluation by municipalities on how funds are used and this information can better be used to report to National Treasury. An example of such planning and reporting by municipalities (with the role of Industrial Engineers) is that of the Built Environment Performance Plan (BEPP) and the subsequent implementation, monitoring and evaluation of infrastructure projects, as described in Van Heerden and Van Heerden [27].

All of this can be included in a drive to support local producers and shop-owners, essentially stimulating local economies. This could be highly beneficial for township economies, but also for cities that try to, in general, enhance the local economy and economic competitiveness. The Digital Town, as described earlier in this paper, could facilitate this drive to include public-private-partnerships, public sector initiatives, and the public themselves.

5. CONCLUSION

South Africa's public sector is often criticised for its inefficiencies and the number of public service delivery protests echo these issues. This paper touched on some of the most prominent contributors to these issues with special emphasis on processes that are prone to corruption and easy to circumvent.

Technological advances including the Internet of Things, machine learning, and blockchain enable many smart applications to be developed to potentially assist government to better deliver services. The problem, however, is that many of these applications are too complex to implement and require deep understanding of both the public sector processes and the technological solutions to be implemented so as to effectively design these solutions. This paper introduced blockchain as a possible solution to many of the problems faced by the public sector.

It is easy to get lost in the complexities of technology and we need to collectively steer the fourth industrial revolution to improve trust in government as well as assist government to improve service delivery. Industrial Engineers are well-suited to play various roles in implementing blockchain solutions in the public sector in South Africa. Firstly, through proper business process mapping and re-engineering. An example of a possible process was provided in this paper. Furthermore, since a lot of data and information will need to be captured, analysed, and maintained, requires proper user requirements analysis, database design, and analytics skills. Industrial Engineers can also play an interfacing role between various entities in the whole process, ranging from public sector officials, to technical implementation teams, to real estate developers, to project and programme management, to name a few.

REFERENCES

[1] **Jones, H.** 2018. Gibraltar launches financial services license for blockchain. Reuters. [Online]. Available at: https://www.reuters.com/article/us-gibraltar-regulator-blockchain/gibraltar-launches-financial-services-license-for-blockchain-idUSKBN1E81JO.

[2] **E-Estonia.** 2018. [Online]. Available at: https://e-estonia.com/ Accessed June 2018.

[3] **Guardtime.** 2018. eGovernment. [Online]. Available at: https://guardtime.com/solutions/egovernment Accessed June 2018.

[4] **Smart Dubai.** 2018. Dubai Blockchain Strategy. [Online]. Available at: https://smartdubai.ae/dubai_blockchain.php Accessed June 2018.

[5] **DigitalTown Inc.** 2018. People-centric cities. [Online]. Available at: https://digitaltown.com/ Accessed June 2018.

[6] **Vermeulen, J.** 2018. South African Reserve Bank blockhain project a huge success. [Online]. Available at: https://mybroadband.co.za/news/banking/263329-south-african-reserve-bank-blockchain-project-a-huge-success.html Accessed June 2018.

[7] **Republic of South Africa.** (1996), Constitution of the Republic of South Africa, 1996. Pretoria: Government Printer. (Government Gazette 17678 of 18/12/1996).

[8] **Minnaar F. & Bekker, J.C.O.** 2005. Public Management in the Information Age. Pretoria: Van Schaik.

[9] **Sathekge, B.** 2018. Municipalities drag SA down. Afro Voice. 24 May 2018.

[10] **Nkosi, N.** 2018. AG finds municipalities wanting. Afro Voice. 24 May 2018.

[11] **Roodt, M.** 2018. Black pupils hindered by poor education system. Cape Argus. 24 May 2018.

[12] **Vegter, I.** 2018. Daily Maverick. [Online]. Available at: https://www.dailymaverick.co.za/opinionista/2018-05-22-low-fee-private-schools-and-vouchers-can-save-sas-woeful-education-system/#.WyZ4z00UkiR Accessed June 2018.

[13] **SAPA.** 2013. Give feedback on e-toll public hearings: Outa. Times Live. 27 May 2013. [Online]. Available at: https://www.timeslive.co.za/news/south-africa/2013-05-27-give-feedback-on-e-toll-public-hearings-outa/ Accessed June 2018.

[14] **de Leon, D.C, Stalick, A.Q., Jillepalli, A.A., Haney, M.A., & Sheldon, F.T.** 2017. Blockchain: properties and misconceptions. Asia Pascific Journal of Innovation and Entrepeneurhsip, 11(3), pp. 282-285.

[15] **Zheng, Z., Xie, S., Dai, H., Chen, X., & Wang, H.** 2017. An Overview of Blockhain Technology: Architecture, Conensus, and Future Trends. 6th IEEE International Congress on Big Data.

[16] **Alharby, M. & van Moorsel, A.,** 2017. Blockchain-Based Smart Contracts: A Systematic Mapping Study. CoRR, Volume abs/1710.06372, p. e print: 1710.06372.

[17] **Eze, P.U., Eziokwu, T. & Okpara, C.R.** 2017. A Tripiclate Smart Contract Model using Blockhain Technology. Circulation in Computer Sceince, pp. 1-10.

[18] **Szabo, N.** 1997. Formalizing and securing relationships on public networks. First Monday, 2(9).

[19] **Kuo Chuen, D. L., Guo, L. & Wang, Y.** 2018. Cryptocurrency: A New Investment Opportunity?. The Journal of Altarenative Investments , 20(3), pp. 16-40.

[20] **Wall, E. & Malm, G.** 2016. Using Blockchain Technology and Smart Contracts to Create a Distributed Securities Depository, an MSc dissertation. Department of Electrical and Information Technology, Lund University.

[21] **McCorry, P., Shahandashti, S.F., & Hao, F.** 2017. A Smart Contract for Boardroom Voting with Maximum Voter Privacy. Conference of Financial Cryptography and Data Security.

[22] **Civic.** 2018. Blockchain-based identity management. [Online]. Available at: https://www.civic.com/solutions/kyc-services/ Accessed June 2018.

[23] **Financial Times.** 2018. How will blockhain technology transform financial services? [Online]. Available at: https://www.weforum.org/agenda/2015/11/how-will-blockchain-technology-transform-financial-services/ Accessed June 2018.

[24] **Deloitte.** 2018. Industry 4.0. Is Africa ready for digital transformation? [Online]. Available at: https://www2.deloitte.com/content/dam/Deloitte/za/Documents/manufacturing/za-Africa-industry-4.0-report-April14.pdf Accessed July 2018.

[25] **South African Government.** 2017c. The National Development Plan. Available at: http://www.gov.za/sites/www.gov.za/files/devplan_2.pdf. Accessed on 1 June 2017.

[26] **South African Government.** 2017e. The Nine-Point Plan. Available at: http://www.gov.za/issues/nine-point-plan. Accessed on 1 June 2017.

[27] **Van Heerden, Q. and Van Heerden, M.** 2017. The role of industrial engineering in public service delivery. In Proceedings of the 28th SAIIE Conference.

[28] **Private Property.** 2018. South African property trends Q4 2017. [Online] Available at: https://www.privateproperty.co.za/advice/news/articles/south-african-property-trends-q4-2017/6155 Accessed August 2018.

[29] **First National Bank.** (n.d.). Home Loan Bond Calculators. [Online] Available at: https://www.fnb.co.za/calculators/homeloan/BondCalculator.html Accessed August 2018.

A CASE STUDY ON RISK PREVENTION AND CONTROL MECHANISMS IN A LOGISTICS PROJECT

Dafeng Xu[1] , Leon Pretorius[2] & Dongdong Jiang[3]*

[1] Economics and Management Shandong Jiaotong University, China,
Department of Engineering and Technology Management
University of Pretoria, South Africa
u16403992@tuks.co.za

[2] Department of Engineering and Technology Management
University of Pretoria, South Africa
Leon.Pretorius@up.ac.za

[3] Department of Engineering and Technology Management
University of Pretoria, South Africa
dongdong.jiang@up.ac.za

ABSTRACT

This paper analyzes the definition and division of modern logistics in the logistics industry, and expounds the project management model faced by logistics issues in the transitional period. It further analyzes and interprets project management in the logistics industry, regulates and annotates the meaning of the project, and defines its specific operation contents as "5 + 1", which extracts the theoretical basis and practical needs of the project risk management mechanism. The risk categories are divided into two major categories, four sub-categories, eight classifications and 24 sub-items. Furthermore the probability of various risk occurrences is initially established and then the evaluation methods as well as responding solutions against various degrees of losses are designed so as to establish a set of independent risk prevention and control mechanisms for the logistics project - Five Steps Risk Management and "network" risk control mechanism. This study is carried out based on the case of Korea Hanjin Shipping Bankruptcy, which proves the necessity of the project management mode and risk prevention and control mechanism. Through the practical trial operation in the logistics enterprises, the study validates the feasibility of the theory contained in this paper in a real operation.

*Corresponding author

1. INTRODUCTION

With the acceleration of economic integration and the rising of e-commerce which lead to the emergence and explosive development of the logistics industry, traditional logistics has been gradually replaced by modern logistics. Opportunities and risks brought about by industry evolution are coexisting. The bankruptcy of the Korean logistics giant, Hanjin Shipping, illustrates the risk crisis in the current logistics project. Therefore, it urgently needs more reasonable risk prevention and control mechanisms to deal with the various risks facing the logistics industry.

With these problems, scholars around the world have carried out some of the following research work and made breakthroughs to achieve results of theoretical research, especially in the last decades: defined life cycle of logistics projects; proposed the advantages of introducing project management into the logistics project; analyzed the complexity of system construction in logistics project management; preliminarily classified the risk categories in logistics projects and finalized the direction and mode of risk control mechanism in logistics projects. However, it is a short space of time since the project management theory was introduced into the logistics industry; researches, especially domestic research in China is still in the exploratory stage which failed to establish a mature management system and mode of operation. It is necessary to combine the research of modern logistics, project management and risk prevention and control to establish a theoretical system of logistics project risk prevention and control, which is suitable for modern enterprises and has pertinence and applicability.

The research approach for this study is based on the theory of Project Management. The theories and methods of Comparative Management, Systems Engineering, Risk Management, Mathematics and Logic are also adequately and comprehensively applied in the study. Taking the logistics project under project management mode as the research target, and the risk prevention and control mechanism as the main topic, the study establishes the necessity and importance of the logistics project management mode by analyzing the current situation and development trend of the logistics industry. Based on the above researches, the study defines and discerns the logistics project risk, analyzes the causes and consequences of the logistics project risk, establishes a set of independent risk prevention and control mechanisms in the logistics project environment according to the attributes and characteristics of the logistics project risk. The study verifies the appropriateness of the theoretical research direction by analyzing the logistics control system of Korea Hanjin Marine Logistics Co., Ltd. as a case study; Furthermore, this study takes the New Jiahong Logistics Company as a partner case study and experimental target and conducts an experimental operation to confirm the feasibility of the logistics risk control mechanism. The abovementioned topics are now logically and sequentially discussed in the ensuing sections of this paper.

2. LOGISTICS REFORM AND PROJECT MANAGEMENT

2.1 Industry Division in Modern Logistics

The concept of logistics first appeared in the 1930s [1], about half a century later, it was introduced to China. According to the "National Standard of the People's Republic of China - Logistics Terminology" [2], logistics refers to the flow of goods from the supplier to the receiving entity, including the organic integration of transportation, storage, loading and unloading, handling, packaging, circulation processing, distribution and information processing based on actual needs. The traditional logistics means "physical distribution" or "goods delivery." Since the 20th century, the logistics industry was under the transformation from traditional logistics to the modern logistics [3] - the concept of logistics systematization was introduced into this industry: namely, social logistics and corporate logistics are combined - leading to the integration and extension of both ends of the logistics, and ultimately achieving the transformation of logistics from industry activities to a form of integrated organization management. As a result, project management is upgraded into a basic component of this industry rather than being only one of the management methods; the importance and significance thereof are increased as well.

At present, according to the nature of enterprises and market share, China's logistics enterprises can be broadly divided into four categories, namely: state-owned enterprises giants (such as Guoyuan Logistics, COSCO Logistics, China Railway Express, etc.); private giants (such as SF, YTO, ZTO, STO, Yunda, with express service as their major service in the market); Special enterprise (Only serving certain type of supply chains , such as fast food or medicine); small and medium enterprises, retailers (mostly for start-up companies or family business). From the maturity of management model's perspective [4], i.e., from the transformation extent of traditional logistics to pattern logistics industry, the Chinese logistics enterprises can be broadly divided into three categories: pre-transition, in-transition and leading transition enterprises.

It is worth mentioning that the first author, after interviewing nearly 30 logistics enterprises, concluded that most of the leaders want to develop their companies into large enterprises or giant of specialized enterprises. However 60-70% of the leaders (mainly in small and medium enterprises) have a lack of understanding and attention on the project management model.

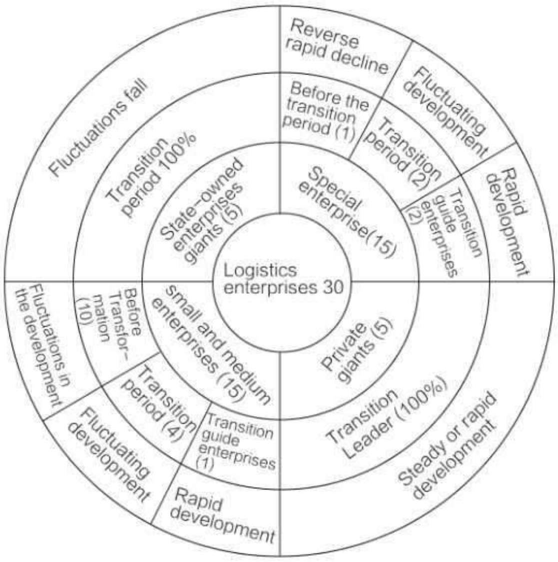


Figure 1: Results of the interviews

From the Figure 1 shown above we can see: the development level of project management model and enterprises development speed are closely related.

2.2 Logistics Project Management Model

The interpretation of the logistics project management pattern is: the logistics industry makes use of specialized knowledge, technology, methods and tools to: address the project activities in a given time, human, material and other resources to achieve or exceed the expected target in the course of business operations [5]. This is an overall design, monitoring and control of project activities. Project Management [6] (published in Science Press 2007) pointed out that project management includes five major tasks: leading, organizing, staffing, planning, controlling and other.

Qin Ligong argues that the life cycle of a logistics project refers to the various stages from the beginning to the end of the logistics project; it generally includes project initiation, project planning, project implementation, project control, project accomplishment and so on [7]. The Project accomplishment stage includes summarization and feedback. The experience and lessons learned from the project will be applied to the next project to form a recycling mechanism for the management work in the life cycle of the logistics project. See also Figure 2.

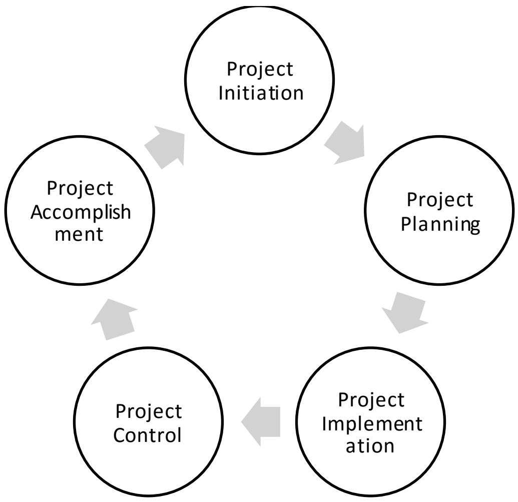


Figure 2: Lifecycle diagram of the logistics project [8]

Scholars around the world began to research the project management model of the logistics industry all at a relatively late stage. In China, Heping and Xiandong [9] suggested to apply the project management theory to logistics industry only since 2003. Up to now, worldwide research in this area is still in the preliminary stage; it does not always follow a quantitative research approach in the logistics industry. The research mainly focuses on the design and implementation of project management, which typically ignores the evaluation of the follow-up process, affecting the overall control of the project and risk prevention and control.

As mentioned earlier, modern logistics is a form of integrated organization and management; project management has been upgraded to be one of the basic components of the industry. Specifically, the application of the project management model in the logistics industry has two types: first is the main business management, this is to take logistics enterprises as a whole and the company's overall operations as a project; the focus is to make use of project management in daily operation. The second is single event management: to split the enterprise’s structure and operation into single activities to manage.

According to the theoretical research and industry situation analysis [10], project management in the logistics project actual operation, includes "5+1" activities: Project scope management, project cost management (human resource cost, material cost, time cost, etc.), project schedule management, project quality management, project human resource management. In addition, all the management mentioned above cannot ignore the risk prevention and control mechanism; therefore, project risk management runs through all the operational processes. The next section will now focus more on risk management issues also related to logistics.

3. RISK MANAGEMENT AND CONTROL MECHANISMS

3.1 Interpretation of Risk Control and Division

The risk of the logistics project is a kind of general risk, which is shown as the uncertainty of project cost and benefit. It is the difference between the target and the result in a given time and specific environment. It should be viewed as a kind of management risk. The reason for the risk occurrence is because of the number of encounters of the uncertainty in the business activities, thus affecting the realization of project objectives.

The occurrence probability of the unforeseen event and its influence degree is typically very uncertain; so it is also called event uncertainty in logistics. The occurrence of such an incident is an objective reality; it is not to be changed by individual will. In the meantime, such events under circumstances can be transformed; its state of existence and the consequences are variable. In summary, risks have objectivity, uncertainty and variability. The existence cycle of risk [11] is generally divided into three stages: latent stage, attacking stage and consequential stage. In this cycle, risk management presents continuity, coexistence and changeability.

According to probability of occurrence, risk is divided into five levels as indicated in Table 1:

Table 1: Project risk probability classification

Level	P1	P2	P3	P4	P5
Probability	Almost no	rarely	Common	Frequently	Almost inevitable

According to different division bases, management risks can be divided as indicated in Table 2.

Table 2: Management Risk Classification

Division Basis	Classification	Description
Source of risk	Natural risk	Natural changes or geological disasters, such as rain and snow weather, geological disasters.
	Human risk	Social or personal factors, such as policy changes, operational errors and so on.
Risk pattern	Static risk	Due to irregular changes in natural forces or human behavior errors
	Dynamic risk	Due to changes in the market environment or social environment risks
Risk opportunity	Pure risk	Cannot bring any profit, the consequences include two aspects: the loss or did not cause any damages.
	Speculative risk	Both potentially harmful consequences and the existence of possible risks of interest, there are three consequences: lead to profit, resulting in loss and no loss.
influence level	Acceptable risk	The consequences and the impact of the risk within the acceptable range, such as individual employee resignation, damage to less valued goods.

	Non - Acceptable risk	Serious consequences beyond the enterprises' ability to bear. Such as the disbandment of the core group, the transport system out of control.

According to Loss Degree, risk is typically divided into five levels as shown in Table 3.

Table 3: Project Risk Loss Degree

Level	L1	L2	L3	L4	L5
Loss degree	Almost none	slight	medium	Serious	disaster

Losses or consequences of project risks are generally divided into direct and indirect losses (also internal and external e.g. [12]), which include: financial and property losses (including loss of income and loss of assets), loss of personnel (including injury, death and loss of personnel), market losses (Loss of customers) and social reputation damage (including loss of credibility and legal consequences). Among them, there are possibility of mutual excitation and mutual conversion between various types of losses. For example personnel injuries and deaths may lead to loss of personnel; loss of credibility will inevitably lead to the loss of customers and so on.

As to the logistics project, the risk is typically classified as categorised in Table 4:

Table 4: Logistics Project Risk Classification

Ⅰ.classification	*Ⅱ.classification*	*Ⅲ.classification*	***Details***
Internal risk	Management risk	Personnel risk	mechanism personnel ability changes in management personnel
		Decision risk	Reasonable degree feasibility execution
	Operational risks	Facility risk	Advance degree operation operation status
		Human risk	Operator ability Operator change Operator safety
extraneous risks	surroundings risk	Society risk	Policies and regulations Economic situation Social Personnel behavior changes
		Market risk	temperature bad weather disaster
	market risk	Competition risk	Competitor Supply and demand Enterprises' adaptability
		Customer risk	Degree of morality Satisfaction degree Affordability

3.2 Five-step Risk Control and "network" Mechanism

The work system of project risk management is typically as follows[13]: applying project management idea, aiming at the objectivity of project risk, adopting the means of identification and analysis to identify and evaluate the type and coefficient of risk; based on the uncertainty of risk , adopt or formulate the relevant technology or measure, make out the responding plan to deal with the occurrence of risk; through the implementation of the solution above to minimize the risk degree.

As mentioned earlier, the life cycle of a logistics project is divided into five stages. The work procedure of prevention and control of project risk can also accordingly be divided into five stages: Risk factor prediction, risk control program production, risk prevention and control operations, risk consequence handling and summary of risk control work. This "five-steps risk control method" constitutes the entire life cycle of risk management; it should be coherent to form an open cycle of the entire mechanism. Among them, the risk assessment and risk control scheme production involves many disciplines and different elements, which determine the direction and results of risk control work. It is also the theoretical basis and key difficulty of the overall work. See also Figure 3.

Through the analysis of the content of the risk management of the logistics project, the risk management technology can be optimized; the rule of management can be mastered and the handling ability of risk response can be improved. Then, an operational mode and system of risk prevention with universality, pertinence and guidance will be formed, which is the risk prevention and control mechanism of the logistics project.

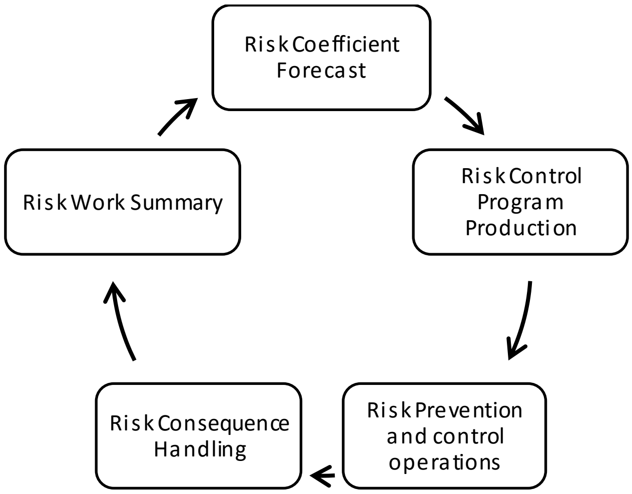


Figure 3: Full Life Cycle of Risk Management

3.2.1 Risk Factor Prediction

Risk factor prediction includes risk identification and risk assessment [6]. Risk identification is based on industry conditions and project environment; the possible occurrence of risks within the project cycle can be identified and listed; risk assessment is based on risk identification, taking full account of the external environment and internal factors, the integrated use of probability and statistics theory and method to calculate and count the occurrence probability, occurrence time and loss degree of project risk. Risk assessment methods are divided into quantitative assessment and qualitative assessment: Quantitative assessment is to quantify the probability of occurrence and loss of risk; qualitative assessment is to draw the risk rating and risk ranking. In general this is summarised in Figure 4.

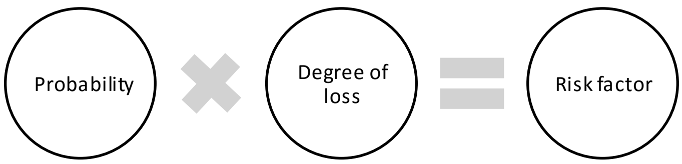


Figure 4: Risk Factor Formulation

The higher the risk factor, the more attention should be paid. Under the comprehensive utilization of these two, the risks can be listed, sorted out and rated comprehensively, thus laying a solid foundation for the design of a targeted, definite and feasible risk prevention and control scheme.

3.2.2 Risk Control Program Production

The key work of the ***Risk Control Program Production*** is to develop the corresponding risk response measures according to the data results of the risk assessment, that is, to design the responsive scheme to deal with the project risk. The general responsive scheme to the risk in logistics projects include: risk avoidance, risk prevention, risk reduction, risk transfer, risk acceptance and contingency planning [6].

Table 5: Risk Responsive Measures

Measures	*Details*
Risk avoidance	Cancelling an undeveloped project, terminates an ongoing project
Risk prevention	Establish regulations and mechanisms to strengthen management and control Carry out thought and experience to improve quality and skill Interpreting policy and normative research forms and contexts Improve the technology and methods to carry out public relations and publicity
Reduced risk	Ask for help to reduce risks
Risk transfer	Project outsourced or sold for security or insurance
Risk acceptance	Self-incurring losses and the responsibility of internal digestion and apportionment
emergency plan	Start the Risk Control Release Process to restart the risk management process

3.2.3 Risk prevention and control operations

In the stage of risk prevention and control operation, all participants in the logistics project are the actual operational personnel of risk prevention and control; their professional ability and professional attitude are the key factors in this work stage.

A qualified risk control operator should have the following qualities:

- High corporate loyalty.
- High passion towards the career
- Rich professional knowledge.
- Strong professional ability.
- Pay attention to the role of risk management.
- Familiar with risk management.
- Having the quality of obeying and adaptability.
- Good at thinking and communication.
- To be able to identify problems and report soonest, if necessary, to make their own favorable decisions and take reactions.

In addition, there were results shown that risk has the highest probability of occurrence in the early stages of logistics projects, resulting in a relatively high degree of damage[4]. Therefore, the earlier of implementation risk control operation, the more conducive for enterprises to avoid risks and reduce losses.

3.2.4 Risk consequence handling

The ***Risk consequence handling*** is the responsive action taken by the logistics enterprise in the stage of risk consequence. That is to say, the risk consequence can be evaluated and dealt with under the condition that the risk has already happened and the consequence has occurred so as to reduce the influence of risk loss as far as possible [14].

In addition, the main parties involved in the activities in logistics project in addition to logistics enterprises are the manufacturers, sellers and end customers, the logistics enterprises in the logistics process will have direct contact and interactions with the other three entities [15]. The occurrence of risk will affect the other three entities. Thus the risk handing will take all these three parties' benefits into consideration as indicated in Figure 5.

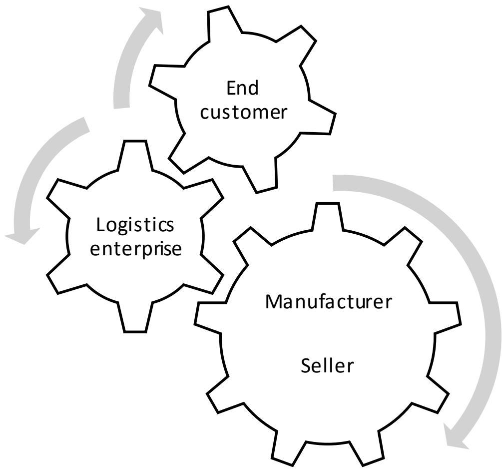


Figure 5: Logistics Market Tripartite Relationship

3.2.5 Summary of Risk Control Work

The completion of a logistics project cycle means the end of a round of risk management [16]. In the final stage, the overall work needs to be reviewed, inspected, researched and summarized in order to learn the lessons by analyzing the shortfalls, to gain the experience by analyzing the achievement, to make use of the advantages and bypass the disadvantages. It can guide us to do the next cycle, and to improve the whole working standard of the logistic enterprises.

In view of characteristics of logistics enterprises, each round of its work cycle has a high degree of similarity or comparability. Therefore, the Summary of Risk Control Work of the logistics enterprise is particularly important [17].

3.2.6 "Network" risk control mechanism

The logistics project risk prevention and control mechanisms should have the following elements:

- Follow the project principles of management principles.
- In line with the original intention of risk management.
- Suitable for the modern logistics industry model.
- Applicable to the main entities which has established the mechanism of logistics enterprises.
- Make full use of management, comparative study and other disciplines theory.
- To use precisely the skills and methods of mathematics and logic disciplines.
- Taking full account of the external environment, internal mechanisms and other factors.
- Effectively matching the entire lifecycle of the logistics project.
- To form a cycle of risk management.
- To cover all aspects of logistics projects and risk management.
- To have the characteristics of preciseness, feasibility, flexibility and so on.
- Trying to control the operational costs of the mechanism which required for human and material .

Based on the above principles, combined with the modern logistics status and project management pattern, and with the cyclical pattern of logistics projects and risk management, a complete set of logistics project risk prevention and control mechanisms can be known or determined, which should be a set of external presentation of a complete cycle, with internal intricately related mesh patterns. [18]. These related theories including project management and risk for logistics is now applied to a case study in the next section.

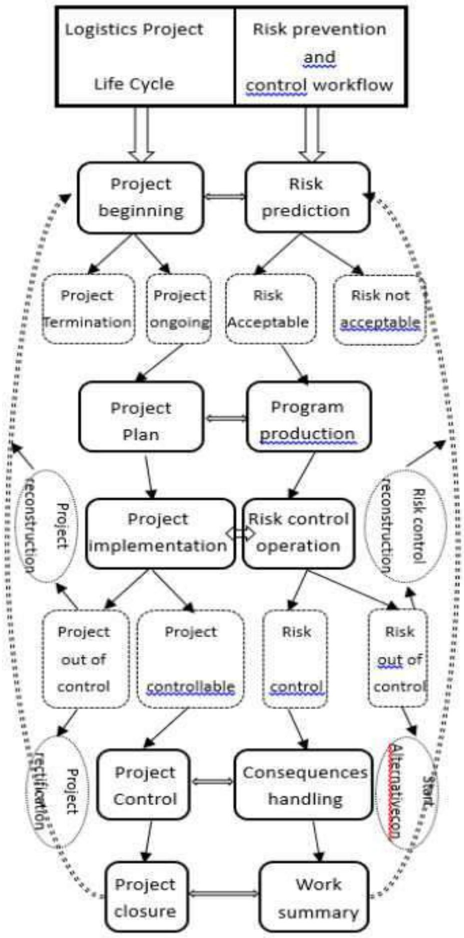


Figure 6: Mesh mode

4. CASE STUDY OF HANJIN SHIPPING

4.1 Hanjin Shipping Bankruptcy

Hanjin Shipping, established in 1977, was Korea's largest shipping company, which was one of the world's top ten shipping companies (ranked seventh). Its mother holding company, Hanjin Group, was one of the global logistics and transportation giants. Hanjin Shipping had more than 200 vessels ; the total shipping capacity reached the world's No. 7, operating more than 60 routes in the global range with an annual transport volume of more than a few hundred million tons of cargo around the whole world. Hanjin Shipping had established a global network of four regional headquarters, more than 200 branches in more than 30 companies over the global operational network, with 13 professional wharves and 6 inland logistics bases, of which two were based in Shanghai, Qingdao of China. Hanjin Shipping's main business was divided into six categories; in addition to ship repair and ship building business, all others were around the logistics projects launched, including: electronic platform services, freezers business, containers business, bulk shipping business, terminal operations. Its branches responsible for three-party logistics business had subsidiaries in 25 countries, a total of more than 150 global customers with aviation and maritime transport, inventory management and other transportation services. Its operation system covered the order management, warehouse management, transportation management and the service management systems, which were the more typical new management systems.

For the past years, Hanjin Shipping business was in trouble due to heavy debt. According to statistics, by the end of 2015, the company had debt of more than 40 billion yuan with debt ratio close to 850%; in the first half of 2016, the company lost about 2.8 billion yuan. On the afternoon of August 31, 2016, Hanjin Maritime filed an application to the court for bankruptcy protection, which was recorded and written by the media as "the largest bankruptcy in the history of global shipping." After the news, a series of reactions were triggered: a number of cooperative companies of the enterprise, such as China Ocean Shipping, Taiwan Evergreen Marine, Japan Kawasaki Steamboat, swiftly announced a comprehensive suspension of cooperation with the Hanjin Shipping; Concerted actions were taken in global port around the world not to accept Hanjin shipping's ship anchored and containers into the port, or even withholding Hanjin's ship; The enterprise came to a complete standstill and the total loss valued about 14 billion US dollars of consignment went into custody. Shanghai Shipping Exchange researcher Zhou Shu said that Hanjin shipping bankruptcy had a huge impact on the global shipping industry in the short-term, involving the owners, ports, freight forwarders, trailer companies and other industrial chains of various suppliers. For the goods owners, the current containers in transit would face the risk of detention, or being unable to be delivered on time. Another expert believed that the actual impact of Hanjin shipping bankruptcy was not only limited to the logistics industry, it would affect the Korean economic development. On September 13, 2016, South Korean President Park Geun Hye said publicly that Hanjin shipping bankruptcy led the logistics industry into chaos, Hanjin Group should bear the consequences for this.

In summary, Hanjin Shipping had a long history, with a strong mode of advanced logistics enterprises. Its bankruptcy took the entire logistics industry in South Korea into market shock and lack of confidence, which also spread to the world. Therefore, it is necessary to study this case by using the theory of project management. Analyzing the reasons behind Hanjin shipping bankruptcy and management loopholes should provide meaningful experience for the contemporary logistics enterprises.

4.2 Weak Risk Control Made Hanjin collapsed

The experts in this industry believe that the main reasons of Hanjin shipping's bankruptcy are as follows:

- Frequent changes of President: Hanjin shipping was founded by Cho Jung-Hoon; in 2003, Cho Jung-Hoon's Son Cho Shu-ho took over the president; in 2007, Cho Shu-ho passed away and his wife Cuiyin Ying inherited the duties; in 2013, Cho Yang-ho took over Hanjin. The actual frequent changes of CEO made the management very difficult.

- The financial crisis: In 2008, the financial crisis swept the world; global trade was heavily affected. According to the data at that time, Hanjin Shipping's revenue declined in 2008; by the first quarter of 2009, it was showing a huge loss.

- Industrial cost adjustment: Before 2008, maritime logistics was in the golden age, which led to the rapid expansion of many shipping companies. After 2008, the maritime transport logistics industry was showing signs of overcapacity; shipping enterprises began to enhance the price war to improve the competitiveness of the market. At the same time, maritime technological progress also reduced the cost of shipping logistics which resulted in freight down step by step. In contrast, the charter costs still remained at a high price. In 2015, for example, when Hanjin seaborne turnover was 7.7 trillion won, the charter costs were up to 1.1 trillion won.

- Failure of a self-rescue plan: After Hanjin had the crisis, the leadership adopted some post-remedial strategies, such as selling assets, selling shares and applying for loans, but failed to solve the problem. In particular, the 500 billion won self-rescue plan was rejected by the creditor, which led to the final crash of Hanjin Shipping.

In accordance with the risk division (see Figure 3), after 2003, Hanjin shipping certainly suffered from management risk under the risk of internal personnel changes, a reasonable degree of risk for decision-making, external risk fluctuations in the economic situation, the strength of competitor's risk, risk of supply and demand change, and customer satisfaction. There is a certain probability that Hanjing also suffered the risk of the quality of management personnel in the internal risk, the risk of decision-making at a feasible degree, the risk of adjustment of policies and regulations in external risks and the enterprise adaptability [19]. Hanjing Shipping was subject to the occurrence of all these risks itself, that runs through the entire logistics project of the enterprise's total life cycle (especially after 2008).

From the perspective of probability of occurrence and loss suffered (refer to Table 1 and Table 2), since 2008, due to frequent changes in the high level management personnel, Hanjin logistics' probability of management risk is on P5 level (close to inevitable); the financial crisis Hanjin logistics suffered from is environmental risk which is P4 (common); due to increased competition in the industry, Hanjin logistics probability of market risk is P4 level (common). These three risks together caused loss of market value, loss of customers, social loss (loss of reputation and legal consequences), the ultimate loss level is L5 (disaster level (loss of assets), which is beyond the enterprise's ability to handle.

Hanjin Marine confirmed that the risk mitigation measures included: risk reduction (seeking loan support), risk transfer (selling equity) and risk retention (selling assets to make up for loopholes). These measures are mostly remedial measures. It means that Hanjin shipping tended to start the risk management in the early stage when risk occurred, which is against the principle of risk forecast, prevention and control operations. From the effect perspective, selling the equity and the assets was not enough to make up for losses, and the search for loan support ended in failure, so these measures did not play a substantive role and failed to help the enterprises to complete the project objectives.

In summary, Hanjin shipping's bankruptcy had both internal and external factors. According to the project management theory, the main mistakes of the enterprise in the project risk prevention and control were: failing to complete the risk factor prediction in time; failing to take the appropriate risk responsive measures; failing to properly deal with the risk consequence; failure to effectively accomplish and utilize Summary of Risk

Prevention and Control; failure to properly match and integrate the project life cycle with the risk management process. In short, Hanjin logistics failed to establish a sound project risk control mechanism; it could not effectively avoid or respond to the occurrence of risk, resulting in catastrophic consequences.

5. A TRIAL OPERATION WITH SOME INITIAL RESULTS

Shandong New Jiahong Company as a further case study to interpret some of the theory presented in previous sections is a logistics enterprise, its predecessor, Jiahong Biology, is a sales company specialised in the sale of the biotechnology products. These cases are evaluated in this section .The case company Jiahong Biology began the transformation into a logistics company at the end of 2013. The company is headquartered in Heze in China; the office area is about 20,000 square metres, and it now has approximately 35 employees, 8 different types of vehicles used for transport, a conventional warehouse area that is approximately 5,000 square metres and cold storage that is approximately 600 square metres. In 2015, the turnover was 170 million yuan, which is a typical small and medium-sized new logistics enterprise.

According to this study and theory considered, the new Jiahong as a case study presented in this section is under the transformation from the traditional to the modern logistics. The company's core management team attaches great importance to the concept of project management, but they lack management expertise and system concept appreciation. Its principal leadership endorsed the theory and vision set out in this paper and agreed to try the mechanism described in this paper, and formally commenced trial operation at the company in October 2016. The main resulting work packages are as follows:

- Establishment of risk control management group; the company's core leadership is directly responsible.
- Appointed a head of the risk control management group to monitor the logistics project risk prevention and control system.
- Conduct public and educational activities to educate all employees of the company on project management and risk prevention and control.
- Introduce the thoughts on project management and the theory of risk prevention and control as described in this paper, and design a set of feasible "network" risk prevention and control mechanism according to the specific situation of the company as an interim regulation.

According to the feedback from New Jiahong Company, by the end of 2016, the company had already adhered to the trial operation on the new risk prevention and control mechanism for three months. During this period, the company turnover increased by 17%, net profit rose by 23% and customer retention and renewal rate increased by approximately 10%. Through the work of the risk control management group, the company avoided one major degree of loss, two moderate losses, 10 slight degree of losses, the loss rate decreased by nearly 50%. Taking one of the classic operation case logistics activities and analyzing it resulted in the following:

In October 2016, the company leadership required the company to participate in a project bidding activity. The risk control group found that in the course of this project, a moderate probability of encountering policy changes, social risk could result in L4 level degree of loss, and risk could not be transferred. The company leadership decided to give up on the bid. Eventually, the project was won by a larger rival company L. In November, during the implementation of the project, the relevant departments made policy adjustments related to the project, which led to risk loss on L company.

In December 2016, the company received a single bulk cargo of the consignment services; the customer required to ship the goods to Yantai within a limited time. The risk control group carried out risk analysis and estimated that the project had a higher probability of natural risk and customer risk, resulting in a slight loss of L2 level. They developed an emergency plan to this potential risk. The day when goods had to be dispatched, Shandong Province experienced a large-scale haze weather with some portion of high way closed. The drivers adjusted the departure time and driving routes according to the emergency plan; marketing department arranged professional staff to communicate with customers, and ultimately the delay time was controlled within the customer acceptable range. The final result is: the customer gave up on claiming any damages due to the delay and they signed a long-term cooperation agreement with New Jiahong Company. In this project, New Jiahong Company made use of project management ideas and risk control mechanism, which transformed the pure risk into speculative risk and they benefited from it.

The leaders of New Jiahong Company said that in February 2017, the project risk prevention and control mechanism would be adopted from trial to formal implementation, and written into the articles of association.

6. CONCLUSION

The following conclusions can be summarised from the research presented in this paper:

(1) Logistics transformed from the traditional logistics to modern logistics and from industrial activities to a set of integrated organizational form. Based on the scales, China's logistics is divided into state-owned enterprises giants, private enterprise giants, specialized enterprises, small and medium enterprises and retailers. According to the degree of transition, it is divided into pre-transition enterprises, transition enterprises and transition leading enterprise. Based on the interviews and statistics, it can be seen that the level of development of project management and enterprise development speed are closely related. Project management has become one of the fundamental elements of the modern logistics industry. Its practical operation content should be in the "5 + 1" mode, namely: (project scope management + project cost management + project schedule management + project quality management + project human resources management)+ project risk management = overall project management.

(2) Logistics project risk is a kind of management risk, with objectivity, uncertainty and variability. Risk level and probability of occurrence can be divided into five levels with the common basis of division including the source, shape, consequences and effects. Specifically, to the logistics project activities, risk categories are divided into 2 categories, 4 sub-categories, 8 classifications and a total of 24 sub-items. The work procedure of prevention and control of project risk can also be divided into five stages accordingly: Risk factor prediction, risk control program production, risk prevention and control operations, risk consequence handling and summary of risk control work. It is therefore named: Five-Step Risk Control. Through the analysis and introduction of risk management, the "network" risk prevention and control mechanism of the logistics project can be extracted.

(3) Through the analysis of Hanjin shipping bankruptcy as a case study, it can be seen that the reasons of Hanjin shipping's bankruptcy had both internal and external factors. According to the project management theory, the main mistakes of the enterprise in the project risk prevention and control were: failing to complete the risk factor prediction in time; failing to take the appropriate risk responsive measures; failing to properly deal with the risk consequence; failure to effectively accomplish and utilize Summary of Risk Prevention and Control; failure to properly match and integrate the project life cycle with the risk management process. In short, Hanjin logistics failed to establish a sound project risk control mechanism; it could not effectively avoid or respond to the occurrence of risk, resulting in catastrophic consequences.

(4) Shandong New Jiahong Co., Ltd. tried the project management concept and risk prevention and control mechanism shown and described in this paper. During the 3-month trial period, it effectively improved the response ability and processing ability to the project risk, thus improving the company's performance. Trial results showed that the "network" risk prevention and control mechanisms might help the logistics enterprises to improve risk aversion and the ability to respond to risks, so that enterprises may get more benefits.

7. REFERENCES

[1] Ballou, R.H., *The evolution and future of logistics and supply chain management*. Production, 2006. *16*: p. 375-386.
[2] China National Standardization Administration, 2016. National Standard of the People's Republic of China - Logistics Terminology(GB/T 18354-2006),Revised edition [M]Logistics Terms PP6-7
[3] Dong, Z., *The Technical Conditions of Modern Logistics*. Open Journal of Social Sciences, 2013. *01(*05): p. 19-22.
[4] Radosavljevic, M., et al., *Supply chain management maturity assessment: challenges of the enterprises in Serbia*. Journal of Business Economics and Management, 2016. 17(6): p. 848-864.
[5] Liu Xiaohong, Xu Jiuping.2008. Project risk management. *Beijing: Economic Management Press.*
[6] Qi An-bang .2007.*Project* Management[M].Science Press.*pp12-13*
[7] Qin Ligong.2004. Risk Management of Logistics Project Lifecycle (TLC).*Logistics Management,* 10, pp C39-C63.
[8] Jin Chenhe,Gao Juhong.2011. Research on the Life Cycle of China's Logistics Industry[J].Value Engineering,2011 , 30 (25) ,pp 26-27
[9] **Z***hang Heping,Zhang Xiandong.2003.Project Management Technology for Logistics Organization Design.Logistics Technology(5) .pp 29-31*
[10] Disterer G.2002. Management of project knowledge and experiences. *Journal of Knowledge Management,* 6(5), pp 512-520.
[11] Jayant, A., P. Gupta, and S.K. Garg, *Perspectives in Reverse Supply Chain Management(R-SCM): A State of the Art Literature Review*. Jordan Journal of Mechanical & Industrial Engineering, 2012. 6(1): p. 87-102.
[12] Wu, T., J. Blackhurst, and V. Chidambaram, *A model for inbound supply risk analysis*. Computers in Industry, 2006. 57(4): p. 350-365.
[13] LI Yan-mei.2011. Logistics Project Management Research. Logistics Technology, (6), pp 32-33.

[14] Chen W, Feng Q, Xu Q.2010. Emergency Logistics Risk Assessment Based on AHM. Information Science and Management Engineering (ISME), 2010 International Conference of. IEEE, 2,pp 59-61.
[15] Wang Y.2013. Research on Risk Control of Logistics Park Construction Project Based on Reliability Theory. *Procedia-Social and Behavioral Sciences,* 96, pp 2194-2200.
[16] Geraldi J, Lechler T.2012. Gantt charts revisited: A critical analysis of its roots and implications to the management of projects today. *International Journal of Managing Projects in Business,* 5(4), pp 578-594.
[17] Mobey A, Parker D.2002. Risk evaluation and its importance to project implementation. *Work Study,* 51(4), pp 202-208.
[18] WENG Y, SONG S, WANG J.2010. Research on the Models of Logsitics Safety Support System. *Logistics Technology,* 5, p 008.
[19] Jianwei Z, Minjie Z, Liwei Z.2011. Risk evaluation of the logistics ecological environment system based on FAHP. *Procedia Engineering,* 15, pp 381-385.

[Intentionally left blank]

SOUTHERN AFRICAN INSTITUTE FOR
INDUSTRIAL
ENGINEERING