Lecture Notes in Mechanical Engineering

João Onofre Pereira Pinto ·
Marcio Luiz Magri Kimpara ·
Renata Rezende Reis ·
Turuna Seecharan · Belle R. Upadhyaya ·
Joe Amadi-Echendu *Editors*

15th WCEAM Proceedings

Springer

Lecture Notes in Mechanical Engineering

Lecture Notes in Mechanical Engineering (LNME) publishes the latest developments in Mechanical Engineering—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNME. Volumes published in LNME embrace all aspects, subfields and new challenges of mechanical engineering. Topics in the series include:

- Engineering Design
- Machinery and Machine Elements
- Mechanical Structures and Stress Analysis
- Automotive Engineering
- Engine Technology
- Aerospace Technology and Astronautics
- Nanotechnology and Microengineering
- Control, Robotics, Mechatronics
- MEMS
- Theoretical and Applied Mechanics
- Dynamical Systems, Control
- Fluid Mechanics
- Engineering Thermodynamics, Heat and Mass Transfer
- Manufacturing
- Precision Engineering, Instrumentation, Measurement
- Materials Engineering
- Tribology and Surface Technology

To submit a proposal or request further information, please contact the Springer Editor of your location:

China: Ms. Ella Zhang at ella.zhang@springer.com
India: Priya Vyas at priya.vyas@springer.com
Rest of Asia, Australia, New Zealand: Swati Meherishi at swati.meherishi@springer.com
All other countries: Dr. Leontina Di Cecco at Leontina.dicecco@springer.com

To submit a proposal for a monograph, please check our Springer Tracts in Mechanical Engineering at https://link.springer.com/bookseries/11693 or contact Leontina.dicecco@springer.com

Indexed by SCOPUS. All books published in the series are submitted for consideration in Web of Science.

More information about this series at https://link.springer.com/bookseries/11236

João Onofre Pereira Pinto ·
Marcio Luiz Magri Kimpara ·
Renata Rezende Reis · Turuna Seecharan ·
Belle R. Upadhyaya · Joe Amadi-Echendu
Editors

15th WCEAM Proceedings

Springer

Editors
João Onofre Pereira Pinto
Oak Ridge National Laboratory
Oak Ridge, TN, USA

Renata Rezende Reis
Federal University of Mato Grosso do Sul
Campo Grande, Brazil

Belle R. Upadhyaya
Department of Nuclear Engineering
University of Tennessee at Knoxville
Knoxville, TN, USA

Marcio Luiz Magri Kimpara
Federal University of Mato Grosso do Sul
Campo Grande, Brazil

Turuna Seecharan
Department of Mechanical and Industrial Engineering
University of Minnesota, Duluth
Duluth, MN, USA

Joe Amadi-Echendu
Graduate School of Technology Management
University of Pretoria
Hatfield, Pretoria, South Africa

ISSN 2195-4356 ISSN 2195-4364 (electronic)
Lecture Notes in Mechanical Engineering
ISBN 978-3-030-96793-2 ISBN 978-3-030-96794-9 (eBook)
https://doi.org/10.1007/978-3-030-96794-9

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

As the title implies, 15th WCEAM Proceedings arises from events and activities marking the fifteenth edition of the International Society of Engineering Asset Management (ISEAM) flagship World Congress on Engineering Asset Management (WCEAM) series. The 15th WCEAM was organised by ISEAM and hosted by the Federal University of Mato Grosso do Sul (UFMS), Brazil, 15–18 August 2021. The quality of 15th WCEAM was assured through the work of a number of organising sub-committees comprising members of ISEAM Board of Directors and UFMS representatives as well as a technical review panel constituted by ISEAM Fellows and Members.

The initial theme for 15th WCEAM was conceived in 2019 so as to be deliberated upon in 2020, however, the incidence and spread of coronavirus necessitated postponement until 2021. Nevertheless, 15th WCEAM provided a forum for high quality pre- and post Covid-19 deliberations on technical and economic issues regarding sustainable management of engineered assets in the era of Society 5.0 concomitant with fourth industrial revolution technologies and globalisation 4.0 business models. The host city for 15th WCEAM was Bonito, Brazil but, due to the restrictions imposed as a consequence of the Covid-19 pandemic, the events of 15th WCEAM were held online with virtual participation by delegates physically located across the globe between UTC-7 and UTC+10 time zones. The events of 15th WCEAM included keynote addresses, panel discussions, workshops and paper presentations by distinguished practitioners and scholars and many of the ideas discussed are noted in this Proceedings e-book. Although the corresponding discourse is not included in the Proceedings, however, it is worth highlighting the topics and presenters of the keynotes, workshops and panel discussants viz:

- Challenges with infrastructure to support Covid-19 vaccination roll-out by Luiz Henrique Mandetta, former Health Minister, Brazil
- Managing Airport Infrastructure during Covid-1 by Dr Kofi Smith, CEO Keystone Management & former CEO Atlanta Airlines Terminal Company, USA

- Asset Control, Monitoring and Diagnostic in Industry 4.0 - Examples in the Oil&Gas Industry by Mario Campos, CEO SmartAutomation, Brazil
- Decision-Making Tools for PV Inverter Maintenance in Solar Power Plants by Prof Huai Wang, Aalborgh University, Denmark
- Motion Amplification and Machine Condition Monitoring by Dan Nower P.E. RDI Technologies, Inc., USA
- On Climate Change Sustainability and Infrastructure Asset Management by David Hood, Chair, Long Future Foundation, Australia
- Reliability Excellence, including performance metrics and case studies by Ron Moore, Managing Partner, The RM Group, Inc., USA
- Research Trends in Asset Management - UFMS Brazil panel by Ruben Godoy, Luigi Galotto and Moacyr Brito

November 2021

Joe Amadi-Echendu
Chair, ISEAM Board of Directors

Preface

This e-book contains qualifying papers presented during 15th WCEAM 2021. Subsequent to the approval of the book proposal by Springer, and a publishing agreement between ISEAM and Springer in 2020, the qualification process involved the following four processes implemented between 2019 and 2021 using Springer's Online Conference System (OCS):

- Initial call for papers in 2019, then a repeat call in 2020, and review of abstracts consistent with the 15th WCEAM theme and topics;
- Submission of categories A, B and C papers and peer review acceptance for inclusion in Book of Abstracts and for presentation during 15th WCEAM;
- Submission of category A manuscripts and peer review acceptance according to the well-established WCEAM double-blind review guidelines that determine the selection of qualifying manuscripts for inclusion in the 15th WCEAM Proceedings;
- Editorial review of qualifying category A manuscripts for inclusion in the 15th WCEAM Proceedings e-book as published by Springer.

The peer reviews were conducted by the WCEAM technical panel primarily constituted by the members and fellows of ISEAM, while the review processes were coordinated by the 15th WCEAM paper submission sub-committee. A brief synopsis of 15th WCEAM Proceedings follows.

Section 1 of this e-book contains four chapters that discuss matters regarding strategy and standards; the section particularly includes a critique on auditing and certification based on ISO 5500x standards. It is not surprising that many of the chapters and corresponding discourse mention the supervening events and disruptions associated with the COVID-19 pandemic; thus, Section 2 contains six chapters that discuss the concepts of sustainability and resilience as applied towards the management of engineered asset systems. Section 3 contains three chapters that discuss servitisation and Industry 4.0 business models for the management of assets, especially in the manufacturing sector. Section 4 contains six chapters that examine conventional asset management issues of risk, reliability and maintenance.

Section 5 contains four chapters with discourse on asset information systems, while Section 6 comprises six chapters containing discourse on asset management decision making. Finally, Section 7 comprehends fourteen chapters addressing condition monitoring and assessment. Interestingly, climate change effects and the COVID-19 pandemic are ongoing; hence, the impacts of the associated volatilities, uncertainties, complexities and ambiguities will continue to generate curiosity, debates and discourse regarding multidisciplinary scholarship and practice of managing engineered assets.

We humbly thank ISEAM for nominating and entrusting us with the responsibility to edit this 15th edition of WCEAM Proceedings. We are also grateful for the support provided to us through the long-established relationship between ISEAM and Springer.

November 2021

João Onofre Pereira Pinto
Marcio Luiz Magri Kimpara
Turuna Seecharan
Renata Reis
Belle R. Upadhyaya
Joe Amadi-Echendu

Acknowledgment

ISEAM takes this opportunity to thank the city of Bonito Brazil and its tourism partners for the wholehearted expression of willingness to host in-person live events of 15th WCEAM but could not do so due to the COVID-19 pandemic. The society looks forward to another opportunity to collaborate with the city in the future.

The ISEAM Board of Directors, steering and organising committees acknowledge the benevolent support of UFMS staff and students for coordinating and managing the operations centre, as well as the collaboration with Doity for providing the online event management and streaming platform, and StreamYard for providing the paper presentation platform. ISEAM and UFMS thank all the keynote speakers, authors, panelists, workshops and paper presenters for their primary contributions to the success of 15th WCEAM.

ISEAM specifically acknowledges the ongoing partnership with Springer for the use of the online conference system for the abstract plus paper submission and review, editorial review of qualifying submissions and preparation of the chapter manuscripts, as well as proofing and publication of the 15th WCEAM Proceedings e-book.

Contents

Strategy and Standards

A Maintenance Management Improvement Framework for Asset Management

Renan Favarão da Silva(✉) and Gilberto Francisco Martha de Souza

University of São Paulo, Av. Professor Mello de Moraes, São Paulo, SP 2231, Brazil
renanfavarao@usp.br

Abstract. Asset management has become an active field of research with the consolidation of the international ISO 55000 series. It coordinates activities to realize value from assets in organizations. Maintenance is one of the main stages to deliver business outcomes from physical assets over its life cycle. In the fourth industrial revolution scenario, it could not be different. Establishing activities to address unwanted incidents, nonconformities, and opportunities for improvement is an important and required element in the maintenance management of an asset management system. The standards, however, are not specific and only determine what needs to be done, not how to do it. Accordingly, this paper proposes a framework for maintenance management improvement (MMI) based on the international standard ISO 55000 series for asset management. To this end, a four-step methodology was applied. First, an ISO 55000 series review focused on a broad understanding of the concepts and requirements for the improvements in an asset management system is presented. Then, a framework was developed for MMI and demonstrated through a maintenance case study application in a Brazilian hydroelectric plant. Finally, the framework processes were correlated and discussed with the improvement requirements of ISO 55001. As the main result, an MMI framework for addressing improvements has been demonstrated to maintenance management for asset management. It encompasses activities for elaborating control and corrective actions, dealing with consequences, addressing opportunities for improvement, critically assessing events, and analyzing root causes. It was evidenced the framework is able to address nonconformities, incidents, and opportunities systematically. Therefore, these findings are expected to contribute to the researchers and practitioners in the field of asset management as the proposed framework based on the ISO 55000 series is an approach to the achievement of continuous improvement in maintenance management and, consequently, in asset management.

1 Introduction

Since physical assets are part of an organization's portfolio, some level of performance is expected of them [1]. Therefore, asset management is not an idea that came out of nowhere, but a result of decades of new thoughts that led to the current state, even going through moments of not being well defined [2]. As organizations have realized both the risks and opportunities presented by how they manage their physical assets, the debate about asset management practice has intensified in the last decade [3].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 3–13, 2022.
https://doi.org/10.1007/978-3-030-96794-9_1

According to ISO 55000 [4], asset management is the coordinated activity of an organization to realize value from assets. It typically involves balancing costs, opportunities, and risks against the desired performance of assets to achieve organizational objectives. In asset-intensive infrastructure industries (e.g., water, gas, electricity, oil, or transportation), effective management of physical assets plays an increasingly important role in the optimization of business performance as stated by Konstantakos, Chountalas, and Magoutas [5]. As a consequence, asset management has become an active field of interest, especially with the publication of the international ISO 55000 series.

A large number of physical assets or diversity in the characteristics of these assets in an organization further increases the importance of having a systematic approach to managing them [6]. Among the asset management benefits may include, but are limited to, improved financial performance, informed decisions about investments in assets, managed risks, improvements in services and products, demonstrated social responsibility, and improved efficiency [4].

During the physical asset lifecycle, maintenance is one of the main stages to deliver business outcomes to organizations. Nevertheless, the maintenance management of modern production equipment is an increasingly important and complex process, especially given the reduced production margin in the global market [7]. In addition, the scenario of the fourth industrial revolution introduces an increase in automatization and creates a greater reliance on modern engineering systems. This reiterates the relevance of discussing models and frameworks for maintenance management that are in line with the asset management discipline.

Among the elements of an asset management system under the ISO 55000 series guidelines is the improvement process. As stated in ISO 55000 [4], an organization's asset management system is likely to be complex and constantly evolving to meet its context, organizational objectives, and changes in its asset portfolio. Therefore, the establishment of activities to address unwanted incidents, nonconformities, and opportunities for improvement closes the cycle of maintenance management and contributes to its continual improvement.

Although the requirements for the improvement process specifies what should be done in ISO 55001 [8], the standard does not tell you how to do it. According to Wijnia and De Croon [9], these standards only prescribe what needs to be in place, not how these requirements should be fulfilled and many organizations struggle with it.

In this context, this paper proposes a framework for maintenance management improvement based on the international standard ISO 55000 series for asset management. The proposed framework intends to meet the improvement requirements set out in ISO 55001 for the asset management maintenance stage. Then, it is demonstrated through a maintenance case study application in a hydroelectric power plant.

This paper is organized as follows: Sect. 2 presents a brief description of asset management improvement in the ISO 55000 series context. Section 3 describes the interfaces between maintenance management improvement and asset management. Section 4 describes the methodology for achieving the objective of this research. Section 5 presents the proposed framework for maintenance management improvement for asset management. Section 6 applies the method to a maintenance case study in a hydroelectric power plant and discusses its findings. Finally, Sect. 7 presents the authors' conclusions about the proposed method and the case study.

2 Asset Management Improvement and ISO 55000 Series

The asset management discipline, by itself, is quite comprehensive and challenging to be implemented in organizations. Increasingly, stakeholders (such as customers, the public, regulators, and shareholders) are looking for assurance that an asset management system will provide safety, service continuity, and financial performance [6]. Thus, an asset management system is used by organizations to direct, coordinate and control asset management activities. It ensures that the objectives of asset management are achieved consistently and sustainably over time [4].

In 2014, a new milestone was reached for asset management discipline with the launch of the ISO 55000 series. It is the first official international standard in the asset management discipline that has reached global consensus. The international cooperation during the preparation of these series of standards has identified common practices that can be applied to the widest range of assets, in the broadest scale of organizations, through the widest diversity of cultures [4].

Among the elements that make up an asset management system, improvement addresses the nonconformities and corrective actions, preventive actions, and continual improvement. According to ISO 55001 [8], the organization shall react to the nonconformity or incident and evaluate the need for action to eliminate the causes of them when a nonconformity or incident occurs in its assets, asset management, or asset management system. Moreover, the organization also shall establish processes to proactively identify potential failures in asset performance and evaluate the need for preventive action and continually improve the suitability, adequacy, and effectiveness of its asset management and the asset management system [8].

Since nonconformities and unwanted incidents, including failures, can occur in its assets, asset management processes, and asset management system, the organization should be aware of the importance of the improvement element for asset management. It is recommended that the organization establish plans and processes to control nonconformities and their associated consequences, to minimize any adverse effects on the organization and on the stakeholder needs and expectations, as stated in ISO 55002 [10]. Accordingly, the elaboration of an improvement framework which is the main purpose of this work intends to fill this gap and propose how to make this process feasible for maintenance management.

3 Maintenance Management Improvement for Asset Management

Maintenance is defined by the Global Forum on Maintenance and Asset Management (GFMAN) as the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, restore it to, or replace it so that it can perform a required function [11]. Asset management, on the other hand, goes beyond acting on assets but using these to deliver value and achieve the organization's strategic objectives [12]. It allows the maximization of value and the achievement of strategic objectives by managing your assets throughout their life cycle [6].

According to IAM [12], the focus throughout the asset's life cycle is a differential of asset management in relation to other management systems. In Fig. 1, different stages of an asset's life cycle are illustrated based on IAM representations.

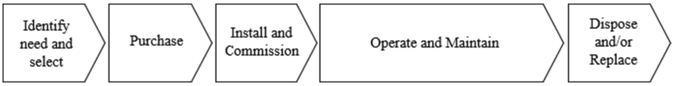


Fig. 1. Example of asset life cycle stages

It can be observed that asset management involves not only maintenance management but the management of other stages in the asset's life cycle. Nevertheless, the maintenance stage is one of the main levers to obtain business results, since about two-thirds of the total life cycle cost of the asset are consumed in the operation and maintenance phases [11]. In addition, evidence has shown that failures in the process can be a consequence of poor maintenance and, consequently, these failures can put people at risk, lose revenue and make operations unviable [13].

Thus, seeking improvements in the results of maintenance management sounds pertinent and appropriate to the achievement of the objectives of asset management and, consequently, of the strategic objectives of the organization. An improvement process can address unwanted incidents and nonconformities as well as opportunities for improvement in maintenance management. This corroborates the purpose of this research since it is intended to develop a framework to improve maintenance management based on ISO 55000 for asset management.

4 Methodology

As previously presented, this research aims to propose a framework for maintenance management improvement (MMI) based on the international standard ISO 55000 series for asset management. Accordingly, it is an applied research, whose objective is predominantly exploratory, as it seeks to provide greater familiarity with the research problem, in addition to proposing a framework as an applicable solution and demonstrating it. In order to achieve this objective, a four-step methodology was structured for the research, as shown in Fig. 2.

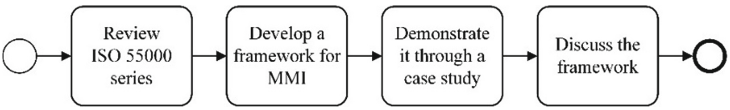


Fig. 2. The four-step research methodology

First, the ISO 55000 series for asset management was reviewed to gain a broad understanding of the concepts and requirements for implementing improvements in an asset management system. Subsequently, a framework was developed for maintenance management improvement following the ISO 55000 series guidelines for asset management. In other words, this second step proposed a framework that addresses improvements in maintenance management that is in line with the requirements set out in ISO 55001.

In the third step, the proposed framework was demonstrated through a maintenance case study of a Brazilian hydroelectric power plant. Lastly, the proposed framework was discussed in relation to the improvement requirements of ISO 55001 and the results of the case study in the fourth step.

5 Proposed Framework for MMI

In a management framework, several processes and activities are integrated to achieve their purposes. In this paper, the proposed maintenance management improvement (MMI) framework for asset management was developed through process modelling with Business Process Model and Notation (BPMN) as it provides a standardized graphical notation to represent business processes that are easily understood, as represented in Fig. 3. The proposed MMI framework is intended to treat different input events towards maintenance management improvement by deriving improvement actions from the analysis of these events that can modify the maintenance process and activities as well as the asset portfolio.

Three message start events are responsible for activating the framework's sequence of activities: unwanted incidents, nonconformities, and opportunities for improvement. A message start event means a message arrives from a participant (e.g., a business entity or a business role) and triggers the start of the sequence of activities in the process [14]. It is represented by a thin line circle with an envelope maker as those on the left side of the framework.

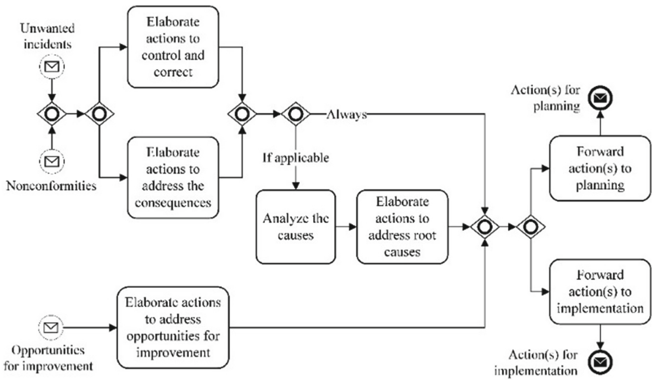


Fig. 3. The MMI framework

In this case, the messages to trigger these three start events come from other processes needed in an asset management system. These processes were not modelled and detailed as done for the MMI since it is not the purpose of this paper. Nevertheless, for better understanding, the sequence flows between them as well as the associated message flows were detailed in Fig. 4.

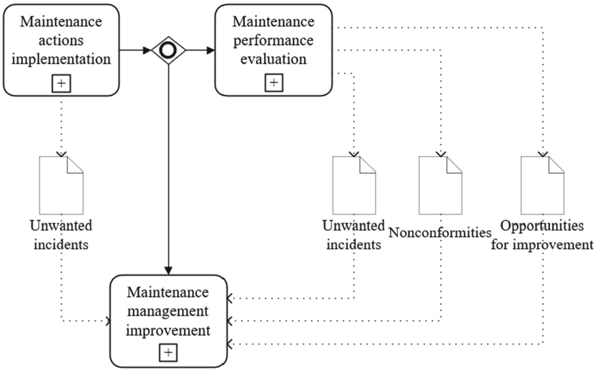


Fig. 4. Sequence and message flows interfacing the MMI process

As can be seen in Fig. 4, unwanted event messages could come from both the ‘Implementation of actions’ and ‘Performance evaluation’ processes. It means these two processes can notify the need to improve maintenance management when unwanted events occur. On the other hand, messages of nonconformities and opportunities for improvements are limited to the results of the performance evaluation process.

Once a start message triggered one of the start events, the activity flow detailed in the MMI framework begins (Fig. 3). Unwanted events and nonconformities have the same activity sequence flow. For them, it starts with the elaboration of actions to control and correct these demands, and/or actions are elaborated to address the consequences of these occurrences according to the pertinence to maintenance management. In a different sequence flow, the opportunities for improvement activate the elaboration of actions to address these demands for improvement.

It should be noted that the sequence flow for unwanted events and nonconformities has intermediate activities before the end events (thick line circles). If it is relevant for maintenance management, the responsible should investigate the causes for the incidents and nonconformities that occurred. Once the root causes are analysed, actions can be elaborated and integrated into the other improvement actions.

The proposed MMI framework presents two different end events. All actions elaborated to improve maintenance management can be forward to planning or implementation. Although it is recommended that actions be forwarded to planning as best practices, the framework allows a direct sequence flow for its implementation. This decision may depend on each elaborated action and the context of the organization.

6 Case Study

In this paper, the proposed framework was demonstrated in an around 200 MW installed hydroelectric power plant composed of 4 Kaplan turbine generating units. Hydroelectric plants are of great importance to Brazil, as around 67% of electricity is generated by this source [15]. Therefore, the choice of this application for the case study is quite relevant to the Brazilian energy context.

This plant is characterized by a centralized maintenance that serves it entirely. Its computerized maintenance management system (CMMS) is SAP Plant Maintenance (SAP PM) which is used to manages all maintenance activities in the plant. This particular hydroelectric power plant has undergone other studies to improve maintenance management under the guidelines of the ISO 55000 series for asset management [16, 17].

The MMI framework was applied to this hydroelectric power plant to demonstrate its activities sequence flow. An unwanted incident message from this plant was selected to demonstrate and trigger one of the start events of the improvement process. The communication of the unwanted incident consists of an automatic shutdown of the generating unit 1 (GU1) that happened in late 2018. In this case, it came from the 'Implementation of actions' (Fig. 4) process during an operational monitoring routine.

As soon as those responsible for the improvement process have been notified of this incident, the sequence of activities begins as provided for in the MMI framework (Fig. 3). First, the organization shall react to this incident taking actions to control and correct it and/or deal with the consequences. Both are the initial activities of the unwanted incident sequence flow in the MMI framework which can be carried out in parallel simultaneously or not.

In this occurrence, the maintenance team was called in to control and correct the incident. It was identified through the operational monitoring parameters that automatic shutdown was due to loss of power in the GU01. Thus, among the elaborated actions were: unblocking the water intake grills by cleaning (action 1) and inspecting the condition and the activation of the security systems associated with an automatic shutdown (action 2). In addition, an action to reset the conditions of the generating unit (GU1) for ramp-up (action 3) was prepared for the operation team due to the consequences of the incident.

Following the MMI activities, the flow is divided in two according to an inclusive gateway right after the elaboration of these actions (Fig. 3). Since each path in an inclusive gateway is considered to be independent, all combinations of the paths may be taken (from zero to all), creating alternative but also parallel paths [14]. Then, the three actions previously elaborated (actions 1–3) follow the upper flow (stated "Always") to the final gateways, where the actions are divided between forwarding for planning or implementation. In parallel, an evaluation of applicability had to be made in order to activate the second possible path (stated "If applicable") in this gateway.

In this case, all three actions were sent for implementation due to the emergency nature of the incident ending the sequence of activity through this process flow path. The other path was also activated for this hydroelectric plant due to its cause analysis policy. According to it, as this unwanted incident was an automatic shutdown and directly affected availability in power generation, it applies to the organization to proceed with an investigation of the incident to identify the contributing causes and prevent recurrences.

Thus, after the organization has reestablished power generation almost 8 h later, the maintenance and operation leaders started the cause analysis activity. As its standard tool, a cause-and-effect diagram (also known as Ishikawa diagram or Fishbone diagram) was used. Figure 5 reproduces the diagram highlighting the causes identified according to the 6 categories of the tool.

Through the cause analysis, it became evident that the clogging of the water intake grills led to a reduction in power in GU1. This occurs, for example, due to deforestation that produces organic waste such as grass, and is transported by the river to the hydroelectric plant reservoir. Clogging these grids is a gradual process and is compensated by the speed regulation system. However, when this is no longer possible and affects safe operating conditions, it is activated a protection system that shuts down the generating unit in order to avoid greater impacts to the GU.

According to the MMI process, once the causes have been identified, the next activity of the flow is the elaboration of actions to address the root causes. Other two actions were proposed: cleaning of the water intake grills when the power is less than 40 MW (action 4) and evaluating the installation of an automatic cleaning system for the water intake (action 5). In this flow, both actions were forwarded to planning since they are not immediate. All actions from this incident were organized in Table 1 for further discussions.

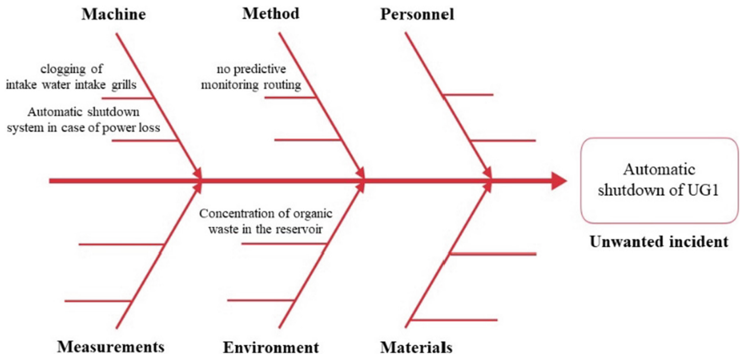


Fig. 5. Cause-and-effect diagram analysis

Table 1. Action derived do MMI process for the unwanted incident at GU1

#	Action description	Forward to	Type of action
1	Unblocking the water intake grills by cleaning	Implementation	Immediate
2	Inspect the condition and the activation of the security systems associated with the automatic GU shutdown	Implementation	Immediate
3	Reset the conditions of the GU for ramp-up	Implementation	Immediate
4	Clean of the water intake grills when the GU power is less than 40 MV	Planning	Preventive
5	Evaluate the installation of an automatic cleaning system for the water intake	Planning	Corrective

It can be seen from Table 1 that the actions were classified into three categories: immediate, preventive and corrective. According to ISO 55000, Corrective action is action to eliminate the cause of a nonconformity or unwanted incident and, consequently, prevent its recurrence while preventive action is to eliminate the cause of a potential nonconformity or unwanted incident. It means preventive actions are performed to prevent the occurrence of this event and when the system is functionally available [4]. Immediate actions are those that must be taken when observing nonconformities and unwanted incidents and should be known by all involved [6]. In the MMI framework, immediate actions are taken to control and correct and/or address the consequences of unwanted incidents or nonconformities.

Given that the flow of activities in the MMI framework has been demonstrated and closed with the forwarding of all elaborated actions, the activities of the proposed framework in relation to the requirements of ISO 55001 are discussed. This standard has a specific Section (10) for improvements which is divided into three subsections [10]. Subsection '10.1 Nonconformity and corrective action' requires what should be done when a nonconformity or incident occurs such as react to it, evaluate the need for action to eliminate the causes, implement any action needed and others.

In the second subsection, '10.2 Preventive action', ISO 55001 specifies that the organization shall establish tasks to proactively identify potential failures in asset performance and evaluate the need for preventive action. It can generally be addressed during the planning of maintenance tasks (e.g., maintenance plans) for the asset portfolio. In the MMI process, this section relates to the elaboration of new preventive tasks when applicable. Finally, Subsection 10.3 'Continual improvement' requires the organization to continually improve the suitability, adequacy, and effectiveness of its asset management, that in this case study it is equivalent to the maintenance management, and its asset management system [10]. In the framework, this last section is represented by the flow of opportunities for improvement.

For a better understanding, the activities of the MMI framework were correlated with the improvement subsections in Table 2.

Table 2. Correlating ISO 55001 requirements and MMI framework

Activity of the MMI framework	ISO 55001 subsection(s)
Elaborate actions to control and correct	10.1
Elaborate actions to address the consequences	10.1
Analyze the cause	10.1
Elaborate actions to address root causes	10.1, 10.2
Elaborate actions to address opportunities for improvement	10.2, 10.3
Forward action(s) to planning	10.1
Forward action(s) to implementation	10.1

Finally, in this case study, the other two start events did not have their sequence flows demonstrated due to the limitations of the paper extension. As shown in Fig. 3, the activities sequence flow for nonconformity is the same for the unwanted incident case presented while the opportunities for improvement have a simplified one. Thus, this single demonstration example is sufficient to the understanding of the entire framework.

7 Conclusions

Asset management enables analytical applications to manage an asset during different phases of its life cycle. Improvement is essential in asset management and an element of an asset management system. Through it, organizations can address unwanted incidents, nonconformities, and opportunities for improvement in their assets, asset management, and asset management system.

Since maintenance has major impacts on asset management, addressing maintenance management improvement is relevant for practitioners in achieving the organizational goals. In this context, this paper proposed an MMI framework for asset management based on the ISO 55000 series. Thus, the proposed framework aims to improve maintenance management by treating the causes of the maintenance unwanted incidents (e.g., unexpected failure) and maintenance nonconformities (e.g., underperformed indicator) and addressing maintenance opportunities to improvement (e.g., new technologies or proceedings) through improvement derived actions that modify the asset portfolio or their maintenance processes and activities.

For that, a four-step methodology was structured for the development of the research: review ISO 55000 series, develop a framework for MMI, demonstrate through a case study, and discuss the framework. The demonstration through a maintenance case study consisted of applying the framework to a hydroelectric power plant. As the main result, an MMI framework for addressing improvements systematically has been demonstrated to maintenance management for an asset management system. It encompasses activities for elaborating control and corrective actions, dealing with consequences, addressing opportunities for improvement, critically assessing events, and analysing root causes.

Although the case study was applied to the asset management maintenance stage, the MMI framework was elaborated in a generic way. Therefore, it is likely to be replicated for other stages of the life cycle, since the requirements of the reference ISO 55001 are the same regardless. As a suggestion for future work, the pertinence and validation in other life cycle stages can be verified.

Finally, it is expected to contribute to the researchers and practitioners in the field of asset management as the proposed framework is an approach to the achievement of continuous improvement in maintenance management and, consequently, in asset management based on the ISO 55000 series.

Acknowledgments. The authors gratefully acknowledge the financial support provided by FDTE (Fundação para o Desenvolvimento Tecnológico da Engenharia) for the development of the present research.

References

1. Calixto E (2016) Gas and oil reliability engineering: modeling and analysis. Gulf Professional Publishing, Houston
2. Bo DB (2015) Application of the ISO 55000 suite for a land-based manufacturing organization. University of Stavanger
3. Hodkiewicz MR (2015) The development of ISO 55000 series standards. In: Tse PW, Mathew J, Wong K, Lam R, Ko CN (eds) Engineering Asset Management - Systems, Professional Practices and Certification. LNME. Springer, Cham, pp 427–438. https://doi.org/10.1007/978-3-319-09507-3_37
4. International Organization for Standardization – ISO (2014) ISO 55000, Asset management – Overview, principles and terminology
5. Konstantakos PC, Chountalas PT, Magoutas AI (2019) The contemporary landscape of asset management systems. Quality 20(169):10–17
6. The Institute of Asset Management – IAM (2008) Asset management – Part 1: Specification for the optmized management of physical asset. British Standards Institution
7. Wisniewski Z, Blaszczyk A (2018) Changes in maintenance management practices - standards and human factor. In: Trzcielinski S (ed) Advances in Ergonomics of Manufacturing: Managing the Enterprise of the Future. Springer International Publishing, Cham, pp 348–354. https://doi.org/10.1007/978-3-319-60474-9_32
8. International Organization for Standardization – ISO (2014) ISO 55001, Asset management – Management systems – Requirements
9. Wijnia Y, de Croon J (2015) The asset management process reference model for infrastructures. In: Amadi-Echendu J, Hoohlo C, Mathew J (eds) 9th WCEAM Research Papers. Springer International Publishing, Cham, pp 447–457. https://doi.org/10.1007/978-3-319-15536-4_35
10. International Organization for Standardization – ISO (2014) ISO 55002, Asset management – Management systems – Guidelines for the application of ISO 55001
11. Global Forum on Maintenance and Asset Management (GFMAN) (2016) The maintenance Framework. GFMAN
12. Institute of Asset Management (IAM) (2015) Asset management: an anatomy
13. Ihemegbulem I, Baglee D (2017) ISO55000 standard as a driver for effective maintenance budgeting. In: Proceedings of 2nd international conference on maintenance engineering, Manchester, UK, 5–6 September 2017 (2017)
14. Object Management Group (OMG) (2011) Business Process Model and Notation (BPMN). OMG
15. Brazil (2019) Brazilian Energy Balance 2019 Year 2018. Empresa de Pesquisa Energética
16. da Silva RF et al (2019) Defining maintenance significant items based on ISO 55000 and AHP: a hydropower plant case study. In: Proceedings of 29th European safety and reliability conference, Hannover, Germany, 22–26 September 2019
17. da Silva RF et al (2020) Defining maintenance performance indicators for asset management based on ISO 55000 and Balanced Scorecard: a hydropower plant case study. In: Proceedings of 30th European safety and reliability conference, Venice, Italy, 21–26 June 2020

Case Study Critique of ISO 550xx Auditing and Certification

Joe Amadi-Echendu(✉), Kolomane Khoarai, and Mapule Lebata

Department of Engineering and Technology Management, University of Pretoria, Pretoria, South Africa
joe.amadi-echendu@up.ac.za

Abstract. The main proposition of the ISO 550xx series of standards is that any organization that intensively deploys engineered assets should implement a 'management system for asset management'. As human impact on earth's geology and ecosystems transcend the era of Society 5.0 powered by fourth industrial revolution technologies, curiously, the composite ISO 550xx series of standards are tantalizingly applied for auditing and certification in engineering asset management. What does it really mean to audit and certify on the basis of the ISO 550xx series of standards? The discourse in this paper uses empirical evidence from two case studies to examine the conundrum of auditing and certification according to the ISO 550xx series of asset management standards.

1 Introduction

A long-winded definition of certification states that it is a.

> *"formal procedure by which an accredited or authorized person or agency assesses and verifies (and attests in writing by is-suing a certificate) the attributes, characteristics, quality, qualification, or status of individuals or organizations, goods or services, procedures or processes, or events or situations, in accordance with established requirements or standards"* [1].

In simpler terms, certification is about a formal confirmation that a person or an entity functions and/or performs in conformance to requirements specified in pertinent norms and standards. The following lemmas emanate from the definition of certification. Firstly, the confirmation must involve a process of audit and accreditation, and the audit must be carried out by an authorized person or agent, that is, an apriori certified or accredited person or agent. Secondly, self-certification or self-accreditation is an anomaly, or a misnomer. Thus, in the legal parlance, the authority to conduct an audit is *ab initio* automatically vested in a pre-accredited/certified third person. Thirdly, the audit procedure encompasses a standardized process of due diligence, that is, the due diligence process must also be auditable. The fourth lemma arises from the fact that certification for professional practice tends to be regarded within a jurisdiction on the basis of an accredited/authorized body of knowledge (BoK). The reasons for this are beyond the scope of this discourse. Notwithstanding, a fifth but latent lemma is that auditing and

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 14–21, 2022.
https://doi.org/10.1007/978-3-030-96794-9_2

certification should be founded upon an authorized normative body of knowledge. Thus, a professionally accredited/certified legal person is expected to demonstrate 'maturity' that encompasses knowledge, skills, functioning, and performance at recognized normative levels. For an individual, professional accreditation/certification in a particular discipline implies that the individual has acquired knowledge and developed skills and competences founded upon a comprehensive, coherent and accredited BoK.

A fascinating feature of the era of Society 5.0 [2] as powered by the fourth industrial revolution (4IR) technologies is the rapidity of evolutionary and trans-formative change. The fusing of subject matter from many conventional disciplines more or less challenges the notion of a normative or standardized body of knowledge. This is more pronounced in the case of asset management, a field of endeavor which demands behaviours, competences, skills and performances that encompass cross-, multi-, and trans-disciplinary scope of knowledge. It is an intriguing challenge to harmonize such a wide scope of knowledge, skills and competences within the times scales of the evolutionary and transformative changes that are apparent in the era of Society 5.0 and 4IR. Accreditation and certification requirements, auditing procedures and due diligence processes can be re-specified in new standards or by updating existing standards at a pace that tracks the rapidity of evolutionary and transformative change. Therefore, auditing should be founded on BoK that is concurrently updated and authorized, otherwise, auditing, accreditation and certification may be compromised or vitiated.

ISO 550xx [3] intentionally promotes a management system approach especially to the management of engineered assets. Here, we denote that engineered assets encompass all man-made artefacts ranging from personalized gadgets, to material components and spare parts of equipment, machinery, to infrastructural facilities, industrial plant and cyber physical systems. The management of such a very wide range and variety of engineered assets across all sectors of human endeavor demands a very broad scope of knowledge. The BoK challenge is compounded by the fact that knowledge compounds, evolves, and transforms very rapidly in the era of Society 5.0 and 4IR.

An issue that prompts this discourse is the emergence of professional certifications in asset management by various associations. Most of the associations claim that their respective certifications are based on ISO 550xx standards, even though the standards do not specify an explicitly accredited and auditable body of knowledge for asset management. This raises the question as to whether such varied certifications in the emergent asset management profession deserve the same level of regard when compared to certifications that are based on well-defined, accredited, auditable and specified bodies of knowledge in other areas of professional practice involving disciplines such as in accounting, engineering, health, law, et cetera.

The content of this paper is structured as follows. Section 2 contains a brief but critical review of the developing ISO 550xx standards from the body of knowledge stance. Two case studies regarding applications of the standard are briefly presented in Sect. 3. The discussions and concluding remarks in Sect. 4 focus on the conundrum posed by auditing, accreditation and certification ac-cording to ISO 550xx standards.

2 A Management System is not a Body of Knowledge

ISO defines a management system as "the way in which an organization man-ages the interrelated parts of its business in order to achieve its objectives." So far, the management system for asset management comprises the following set of published standards [3]:

- ISO 55000: 2014 Asset management – Overview, principles and terminology;
- ISO 55001: 2014 Asset management – Management systems – Requirements;
- ISO 55002: 2018 Asset Management – Management systems – Guidelines for the application of ISO 55000;
- ISO 55010: 2019 Asset management—Guidance on the alignment of financial and non-financial functions in asset management.

The ISO 550xx standard "is intended to be used for managing physical assets in particular, but it can also be applied to other asset types." It is unclear whether physical assets are limited to engineered artefacts such as equipment and machinery, or whether they also include natural resources such as air, land, minerals and water. ISO 550xx is also intended to promote a management systems approach for managing assets guided by narrow delineations of value but, the standard is mute in terms of the broader ethos of value. The requirements stated in ISO 55001 are grouped under seven headings viz – (i) context of the organization; (ii) leadership; (iii) planning; (iv) support; (v) operation; (vi) performance evaluation; and (vii) improvement. ISO 55002 provides guidance towards satisfying the requirements stated in ISO 55001 and affirms that "an asset management system is used by the organization to direct, coordinate and control asset management activities to realize value from its assets for the organization and for its stakeholders." Thus, it is not uncommon that stakeholders tend to be defined only with regard to the entity (i.e., a natural or juristic person) that controls the use of engineered assets. The reality is that various stakeholders relate directly either to the asset, or to the asset controlling entity, or to both.

It is reasonably acknowledged that the requirements stated in ISO 55001 ap-pear less prescriptive than those specified in PAS 55, the predecessor to ISO 550xx. The release of ISO 550xx was subsequent to the publication of the ISO 31000 Risk Management standard. Hence, it was apparently unnecessary to in-corporate detailed risk management requirements into ISO 55001. Nevertheless, ISO 55002 amplifies the need to address risk in asset management. The indication is that inter alia, the BoK for managing assets must invariably incorporate knowledge of how to manage risk. The value ethos plus aforementioned principles of alignment, assurance, and leadership indicate that the subject matter of asset management encompasses knowledge areas from many conventional disciplines such as communications, economics, engineering, finance, humanities, law, philosophy, science, sociology, and technology. Consequently, the BoK that founds asset management is indeed, very broad and deep; as discussed in detail in reference [4]!

Subsequent to the release of ISO 550xx, various organizations (local and international) continue to promote variations of asset management BoK based on their respective preferences and persuasions. Notable among the BoK semblance is the '39 asset management subjects' described in [5]. The framework is "intended to provide a common understanding of the scope of asset management." Reference [5] categorically states that it is impossible to understand asset management without addressing the 39 subjects holistically as an integrated BoK. The implication is that auditing procedures and due diligence processes must be used to establish the extent to which a natural person or a juridical person understands the 39 subjects holistically. Curiously, only nineteen requirements specified in the ISO 550xx are identified as matching subjects in the associated competency specification for an ISO 55001 asset management system auditor/assessor [6]. Does this suggest that the remaining twenty subjects are less important for auditing, accreditation and certification of natural and juridical persons?

Another published version of asset management BoK [7] emphasizes the concept of 'maturity' – that is, the extent to "which leadership, culture, human performance and the asset management system are integrated into the whole organization…" This definition of maturity implies that certification must be graded ac-cording to a normative scale. Such a requirement for a graded normative scale of maturity is not specified in ISO 550xx. So, it is rather difficult to compare the maturity assigned by various agencies that currently conduct auditing, accreditation and certification in terms of ISO 550xx. The argument here is that a coding of maturity grades should be specified, so that the description of, say, code 'A' is consistent and standardized across the board. This will facilitate proper bench-marking of auditing, accreditation and certification in conformity with ISO 550xx.

It is worth reiterating that the adoption of ISO 550xx represents a phase in continuing attempts at formalizing the age-old human endeavor of managing assets. ISO 55002 reiterates the need to maintain documentary evidence so as to provide proof against each of the seven categories of requirements specified in ISO 55001. Understandably, such documentary evidence generally remains confidential. Other formalizations that precede ISO 550xx include guidelines designed to ensure compliance, especially in terms of local legislation (see, for example, [8]), as well as conformity to international norms (see also, [9]) which regulate how engineered assets should be managed. Obviously, such guidelines are generally more prescriptive than the ISO 550xx since there is a consequence for non-compliance to law or for non-conformity to international regulations. Unlike ISO 550xx, statutory and regulatory guidelines tend to specify 'what-to-do' and 'how-to-do' so. Thus, such guidelines articulate minimum scope of knowledge, skills, and competences. Furthermore, statutory and regulatory compliance/conformity means that redacted versions of the associated documentary evidence about audits conducted to justify accreditation and certification of a natural or juristic person may be publicly available.

Many agencies have developed proprietary methods of auditing and certification based on the ISO 550xx standards. Tantalizingly, there is a myriad of claims as to the benefits of auditing and certification according to ISO 550xx standard (e.g., as cited in [10]). Details regarding most claims remain confidential and somewhat unsubstantiated. Some claims that are publicly disclosed appear in promotional material advertised by consulting firms, software vendors and service providers, so it is challenging to ascertain whether or not the purported benefits can be exclusively attributed to auditing and certification according to ISO 550xx standard. For these reasons, we briefly describe two of the many first-hand applications of the ISO 550xx standard. The following discourse is limited to assessments of asset management at two case study organizations – a power generation utility, and a municipality. The assessments were carried out by the respective researchers and do not qualify as certifiable audits. Nonetheless, each case study involved considerable examination of the respective organization.

3 Case Studies

3.1 Asset Management Maturity Assessment – Power Utility

For this case study, the assessment was carried out using an audit method that combined the approaches outlined in references [5] and [6]. As with other proprietary methods supposedly aligned with ISO 550xx, the methods in references [5] and [6] also specify 'maturity' levels. Two of the sources of data and information were archival records and semi-structured interviews. The evidence included data and information about the portfolio of assets, maintenance history records, capital replacement records, risk records, and narrative responses from interviews. The interviewees were purposively limited to eight persons with job positions as follows – one technician, three shift supervisors, a senior engineer, a maintenance planner, a senior accountant, and a manager. The transcribed responses were thematically analysed to extract respondents' aggregate indication of maturity. A triangulation of the researcher's first-hand observations with the respondents' feedback and also with the documentary evidence resulted in the determination of a maturity level for the organization.

3.2 Asset Management Review – Municipality

For the second case study, the researchers reviewed legislation, guidelines and standards that were purportedly applied by a municipality to manage infrastructure assets. Given that the range of public infrastructure assets is quite wide (covering services such as education, energy, health, transportation, water, etc.), the re-view was grouped in terms of (i) governance, and (ii) non-governance types of as-set management activities. In this case, the researchers were guided by experiential understanding of the requirements of ISO 550xx, and did not apply any proprietary method. Primary data and information were collected using unstructured interviews of purposive respondents – hence a narrative enquiry process. Secondary data and information were adduced from archival records that were accessible such as policy documents, statutory and non-statutory reports. Similarly, for this case study, a triangulation of the researchers' first-hand observations with the respondents' feedback and also with the documentary evidence resulted in the determination of a maturity level for the organization.

3.3 About the Case Study Field Researchers

For brevity, the field researchers involved in the case studies have a minimum of bachelor degrees in engineering. Each field researcher had at least five years of work experience post-graduation. The field researchers were concurrently employed within the same organization where the respective case studies were conducted. The case studies were carried out as a capstone project in partial fulfilment of the requirements for a Master degree in Engineering and Technology Management. The duration for each capstone project was about twelve months and the data and information were typically collected within a month. The Master degree programmes include eight taught modules plus a mini-dissertation thesis on the capstone project. The scope of Master degree curriculum is extensive such that the researchers were exposed to a broad range of topics, subjects and knowledge areas within the scope of Engineering and Technology Management. Thus, the case studies summarised above were based on the individual knowledge, skill, competence and experience of the respective researchers.

4 Discussion and Concluding Remarks

During the course of the respective capstone projects each researcher became well acquainted with the contents of ISO 550xx standard, so, it can be assumed that the researchers understood how to apply the standard for their case studies. Incidentally, the epistemology of the assessments was overwhelmingly biased to-wards the respective subjectivities (i.e., attitudes, behaviours, and performances) of the researchers. Despite the researchers' appreciable academic background and work experiences, the researchers realised that application of the ISO 550xx standard demands broad and deep knowledge of many subjects encapsulated in several traditional disciplines. For example, to interview respondents in order to corroborate documentary evidence required communication, cross-examination, as well as non-verbal interpersonal skills.

The researchers also found that they required a reasonable depth of technical knowledge about the portfolio of assets in order to conduct assessments according to the ISO 550xx standard. In some instances, the knowledge associated with a particular asset had to be very narrow and deep in order to make sense of the evidence, e.g., data and information regarding condition monitoring, diagnosis and prognosis of machines, econometric and financial models supporting capital investment and replacement decisions, legal rules stipulated in asset purchase contract clauses about guarantees and warrantees, cultural norms influencing the relationships between individuals and functional organisational structures, etc.

Given the extensive scope of the BoK that they encountered, the researchers inferred that an audit based on ISO 550xx should be carried out by a team of professionals whose combined knowledge, knowhow, skills, competences embrace the scope of the BoK implied but, not specified in ISO 550xx. This inference raises concerns as to the designation of an individual as a certified ISO 550xx asset management assessor/auditor without a corresponding specification of an accredited/authorized BoK. It is remarkable that reference [6] states that, in general, a certified ISO 550xx asset management system auditor/assessor should "demonstrate a comprehension of asset management as a set of integrated technical and financial risk-based processes." Such narrowing of the competency requirements for asset management system auditor/assessor to technical and financial risk-based processes creates ambiguity. Also, it is perplexing that ISO 55001 does "not specify financial, accounting, or technical requirements for managing specific asset types" even though ISO 55002 supposedly provides in-formation on "relevant asset management subject areas." Interestingly, the issue of financial alignment has been seemingly addressed in the ISO 55010 standard.

ISO 55002 affirms that “an asset management system is used by the organization to direct, coordinate and control asset management activities to realize value from its assets for the organization and for its stakeholders.” In simple terms, the reason for the assets is so that the organization and its stakeholders can realise value, irrespective of how value is defined. The paramountcy of the value ethos is that the asset management system inextricably interweaves all activities and functions of the organisation [11]. Such supremacy of the asset management system means that all activities and functions serve to ensure that the assets are deployed to provide value to the organisation and its stakeholders, that is, all other systems, “for example, quality, finance, accounting, safety, risk and human re-sources” form part and parcel of the asset management system [12]. An inference here is that the competency requirements for auditing and certification to the ISO 550xx standard are really onerous! Therefore, any attempt at narrowing the ISO 55001 planning principle requirements to just planning for maintenance intervention only vitiates the scope of the principle and raises concerns about the ‘maturity’ scale for asset management planning. Here maturity represents an aggregated measure encompassing attitude, behaviour, knowledge, skill, competence, and performance.

The ISO 550xx standard was released in 2014 but, within a relatively short time scale, there seems to be a myriad of claims as to the benefits of auditing and certification according to ISO 550xx standard. The proposal for more ISO 550xx standards is proof of the extensive BoK required for asset management. The requirements in each ISO 550xx standard inadvertently interweave many disciplines and varied persuasions, thus making it difficult to substantiate many of the current claims about benefits obtained from auditing and certification based on ISO 55000, 55001 and 55002 standards. Furthermore, the concept of ‘fusing’ that characterises the era of Society 5.0 and 4IR reinforces the notion of inter-connected asset systems (re: cyber physical systems). The value ethos coupled to this notion of asset systems challenges the traditional view as to what constitutes an engineered asset. The annexures in ISO 55002 appear to address this conundrum. Interconnectivity, inter-communicability and interdependency are intrinsic features of asset systems, thus capabilities to manage risk, resilience and vulnerability should be included in any accredited/authorized BoK for asset management as well the ISO 550xx competency specifications. In conclusion, longitudinal studies based on sound research methodologies are required so as to eliminate un-substantiate claims as to the benefits of auditing and certification according to ISO 55000, 55001 and 55002 standards, including the latest addition to the series, i.e., the 55010 standards.

References

1. anon. http://www.businessdictionary.com/definition/certification.html
2. Salgues B (2018) Industry of the Future, Technologies, Methods and Tools. Society 5.0, vol 1. Wiley, Hoboken
3. ISO 2014–2020. ISO 550xx Asset Management System Standards including 55000, 55001, 55002, 55010
4. Gavrikova E, Volkova I, Burda Y (2020) Strategic aspects of asset management: an overview of current research. Sustainability 12(15):5955. https://doi.org/10.3390/su12155955
5. GFMAM (2014) The Asset Management Landscape. ISBN 978-0-9871799-2-0

6. GFMAM (2014) Competency Specification for an ISO 55001 Asset Management System Auditor/Assessor, 1st Edition. ISBN 978-0-9781799-3-7
7. Asset Management Council (2014) Framework for Asset Management, 2nd Edition. ISBN 978-0-9870602-6-6
8. National Treasury, South Africa (2015) Local Government Capital Asset Management Guideline. ISBN 978-0-621-38093-4
9. IPWEA (2015b) International Infrastructure Management Manual, 5th Edition
10. Alsyouf I, Alsuwaidi M, Hamdan S, Shamsuzzaman M (2021) Impact of ISO 55000 on organisational performance: evidence from certified UAE firms. Total Qual Manag Bus Excell 32(1–2):134–152. https://doi.org/10.1080/14783363.2018.1537750
11. Asset Leadership Network (2018) Asset Management in High Performance Organizations. www.assetleadership.net
12. Amadi-Echendu JE (2017) Developments in the management of engineered assets: before and beyond ISO 550xx. Keynote 4, International Conference on Industrial Engineering and Operations Management, Rabat, Morocco, 11–13 April 2017

Continuous Quality Improvement and Business Performance: The Mediating Role of Physical Asset Management

Damjan Maletič[1](✉), Matjaž Maletič[1], Basim Al-Najjar[2], and Boštjan Gomišček[3]

[1] Faculty of Organizational Sciences, University of Maribor, Kidričeva c. 55a, 4000 Kranj, Slovenia
{damjan.maletic,matjaz.maletic}@um.si
[2] Department of Mechanical Engineering Faculty of Technology, Linnaeus University, Building M, 351-95 Växjö, Sweden
basim.al-najjar@lnu.se
[3] School of Business & Quality Management, Hamdan Bin Mohammed Smart University, Dubai Academic City, 71400, Dubai, United Arab Emirates
b.gomiscek@hbmsu.ac.ae

Abstract. This study examines the mediating effects of physical/engineering asset management on the relationship between continuous quality improvement and business performance. Using empirical data based on survey data from six European countries (i.e. Greece, Poland, Slovakia, Slovenia, Sweden and Turkey), this study used mediation analysis to address the research problem. A macro for SPSS was used to estimate the size of an indirect effect of continuous quality improvement on business performance by a proposed mediator. The results of this study show that physical asset management mediates the effect of continuous quality improvement on business performance. This study provides valuable insights into mechanisms that have the potential to improve business performance. The results contribute to a better understanding of how companies could achieve higher performance outcomes through the introduction of continuous quality improvement and through physical asset management practices.

Keywords: Physical/engineering asset management · Business performance · Continuous quality improvement · Mediation

1 Introduction

Physical/engineering asset management (PAM) is a growing area of interest in business and research today because of its critical impact on organizational performance and success [1, 2]. Physical assets, also known as engineering assets, are important for creating tangible value for an organization in a variety of industrial settings such as manufacturing, power supply, water supply, construction, mining, transportation services, and various other sectors [3]. Most studies on PAM have focused on understanding how organizational factors, including leadership, policy and strategy, risk management, life

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 22–32, 2022.
https://doi.org/10.1007/978-3-030-96794-9_3

cycle management, and performance evaluation, influence performance outcomes [4, 5]. Less is known about the synergies of PAM and other organizational disciplines and how PAM might facilitate performance improvement. Although there is a growing literature on various aspects of PAM (e.g., [1, 6, 7]), the discipline of PAM is relatively new from a scientific perspective. Therefore, further studies are needed to investigate different aspects such as interactions and mechanisms that lead to improved performance.

PAM itself is multidimensional as a concept and encompasses various theoretical and practical lenses. According to the Hastings [8], PAM could be defined as: "a set of activities associated with identifying needed assets, funding requirements, ways of acquiring assets, provision of maintenance and logistic support systems for assets, and to dispose or renew them in order to meet desired goals effectively and efficiently." Indeed, PAM denotes the systematic and coordinated set of practices through which an organization optimally and sustainably manages its assets, systems, associated outputs, risks, and expenditures over their life cycles to create value and realize organizational strategic plan [9]. PAM is thus an approach that provides support in creating value throughout the entire life cycle of assets. Therefore, organizations seek to maximize the value extracted from their assets by investing efficiently in asset management to generate better returns for their business [10]. Realizing the value from assets is a holistic approach that considers the complex expectations of stakeholders and provides the organization with a competitive advantage [11]. It encompasses a wide range of disciplines and processes that cover the lifecycle phases of creating, establishing, using, and disposing of a physical asset in a balanced manner to meet the continuum of constraints imposed by business strategy, economics, ergonomics, technical and operational integrity, and regulatory compliance [12].

Researchers [1, 13] have argued that the ultimate goal of PAM is to realise the value of an organisation's assets. Furthermore, Alsyouf et al. [13] contend that the use of an asset management system (AMS) helps organisations improve financial performance, decision-making processes, risk management, services and outcomes, social responsibility, and organisational reputation and sustainability. In essence, PAM requires an interdisciplinary approach, where synergies should exist between traditional disciplines such as accounting, engineering, finance, humanities, logistics and information systems technologies [14]. It is argued that organizations have adopted an integrated and holistic approach to PAM to increase the performance of physical assets, reduce business risk and reduce the whole-life costs of assets [15].

The early stages of empirical research in PAM were almost exclusively limited to attempts to delineate the relevant constructs and frameworks. For example, several initiatives were developed to contribute to the knowledge base in the emerging discipline of asset management [16–18]. More recently, studies have empirically examined the relationship between PAM and performance outcomes [1]. Existing work by researchers thus helps us understand how organisational performance can be improved through PAM and quality improvement efforts.

To date, however, few studies have been conducted to investigate the mediating effect of PAM on the impact of continuous quality improvement on organizational performance. It is vital to identify the factors by which organizational performance can be improved and to determine whether PAM can effectively contribute to business performance. Therefore, the primary objective of this study is to investigate the mediating role of PAM on the relationship between continuous quality improvement and business performance. Accordingly, this study contributes to the theoretical development of the relationships between quality improvement, PAM and business performance, which have rarely been examined in a single study. Given that nowadays many organizations operate in asset-intensive industries, the role of PAM in this relationship is even more important.

2 Methods

The main objective of this section is to describe important aspects of the research methodology of the study. Thus, this section outlines the research design as a framework for sampling, data collection, measurement and data analysis. In line with the research objective of this study, empirical research is considered as an appropriate approach enabling the researcher to focus on the phenomena under study and test the validity of the proposed model [19].

2.1 Sample and Data Collection

This cross-sectional study is based on mailed survey of a sample of European organizations which cover various manufacturing and service sectors. In order to achieve the objectives of the present study, a research study was carried within a research project conducted by a team of international researchers in the field of maintenance and asset management. Data were collected through an online survey using the 1ka web survey platform (https://www.1ka.si/d/en). An online questionnaire was distributed to organisations in six countries. The recommendations of the authors of Dillman et al. [19] were followed in the design and implementation of the survey. In each country, the questionnaire was addressed to a manager familiar with the organisation's practises related to quality and PAM and the evaluation of performance outcomes. Organisations were selected through national business registries and email addresses from participating universities' contact databases. Each survey coordinator had to ensure that a list of population members to be surveyed was both up-to-date and from a reliable source (business register). The empirical investigation served to deepen the understanding of mechanisms between constructs. Considering the sampling strategy, manufacturing and other asset related industries were considered most suitable for this study to represent the population frame. A total of 138 completed questionnaires were received, yielding a response rate of 13%. The questionnaire was responded by organizations that are located in Poland, Slovenia, Greece, Sweden, Turkey and Slovakia, in portion of 34.1%, 31.9%, 16.7%, 6.5%, 5.8% and 5.1%, respectively. Approximately 48.7% of the organizations are classified as small and medium-sized enterprises, while 17.4 employ less than 50 employees, 21.7% organizations employ 251–500 employees, 12.2 employ more than 500 employees. The list of industry sectors captured in the sample is presented in Table 1.

Table 1. Sample distribution by industry type

Industry type (in %)	Manufacturing	39.3
	Wholesale and Retail Trade, Repair of Motor Vehicles and Motorcycles	16.2
	Construction	6.8
	Mining and Quarrying	6
	Electricity, Gas, Steam, and Air Conditioning Supply	2.6
	Other	29.1
	Total	100 (N = 138)

2.2 Measures

The design of the structured questionnaire was based on a comprehensive literature review. Previous studies have provided theoretical and empirical grounded measures that were used as the basis for drawing the questionnaire items. A five-point Likert scale was used to measure subjectively all the variables – questionnaire items. The respondents were requested to indicate their level of agreement or disagreement with the items for continuous quality improvement and physical asset management and the level that accurately reflects the extent of their organization's business performance over the last three years. The measures used in this study for physical asset management were all based on those used in previous studies on similar topics in order to ensure their content validity [2, 16, 20]. In particular, the following PAM dimensions were considered: Policy & Strategy, Risk Management, Life Cycle Management, and Performance Assessment. The continuous quality improvement scale comprises items that are derived and customized based on previous studies [21–23]. Business performance is operationalized in this study to reflect the key competitive priorities (i.e. sales, profit, and market share) that are broadly recognized in the literature (e.g., [24]). The list of all items can be seen in the Appendix A.

2.3 Mediation Analysis

To test the mediation effects of physical asset management on the relationship between sustainability and operational performance, we used the SPSS procedure (SPSS macro) proposed by Preacher and Hayes [25] to estimate indirect effects in simple mediation models. The macros provide unstandardized coefficients as required for the mediation test [25, 26]. Path a represents the effect of X on the proposed mediator, while path b is the effect of M on Y, which partializes out the effect of X (Fig. 1B). All of these paths would typically be quantified using unstandardized regression coefficients. The indirect effect of X on Y through M can then be quantified as the product of a and b (i.e., ab). The total effect of X on Y is quantified by the unstandardized regression weight c (Fig. 1A). The total effect of X on Y can be expressed as the sum of the direct and indirect effects: $c = c' + ab$.

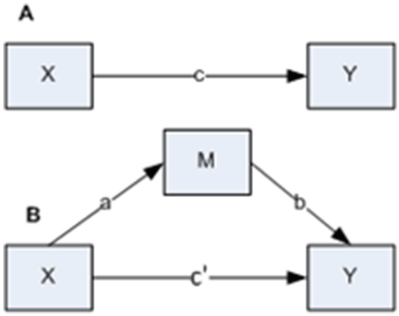


Fig. 1. (A) Illustration of a diect effect. X affects Y. (B) Illustration of a mediation design. X is hypothesized to exert an indirect effect on Y through M [25, 26].

2.4 Measurement Reliability and Validity

The summary of the reliability and validity assessment is shown in Table 2. First, exploratory factor analysis (EFA) was applied to refine the latent factors of studied constructs and to assess their preliminary construct validity. The result of the EFA supports the validity of the constructs as indicated by the amount of variance explained, which exceeded 50%, and the loading factors of all items within each scale exceeded 0.5 [27]. The scale reliability was tested by calculating its Cronbach's alpha. Indeed, the most common method for measuring reliability of survey questionnaires involves estimating internal consistency and Cronbach's α coefficient [28]. The Cronbach's a coefficients range from 0.807 to 0.936 indicating the internal consistency reliability of the measures as they exceed the 0.70 cut-off limit [27].

Table 2. Summary of the reliability and validity results

Latent variable (number of items; % of variance)	Factor loadings (interval)	Cronbach's alpha
Continuous quality improvement (6; 76.28)	0.821 – 0.928	0.936
Business performance (5; 68.55)	0.675 – 0.899	0.866
Physical asset management (4 factors (F); 66.90)	0.468 – 0.947	F1 = 0.926; F2 = 0.933; F3 = 0.877; F4 = 0.807

Notes. F1 - Risk Management; F2 - Performance Assessment; F3 - Life Cycle Management; F4 - Policy & Strategy

In summary, the results of the validity tests provide sufficient evidence for the validity of the measurement model and thus support the empirical rationale for combining the items of the individual constructs into a composite variable. Accordingly, mean scores were calculated from the items in the scale to generate the composite scores for the physical asset management construct as well as for other variables examined.

3 Analysis and Results

We have followed the suggestion of Baron and Kenny [29], who recommend that a mediator rather than a moderator function is better when there is a strong relationship between the predictor and the criterion variable. Therefore, we assume that the predictor variable "physical asset management" is related to the criterion variable "business performance", and we assume that "physical asset management" has a mediator function for the relationship between continuous quality improvement and business performance. As described above, the mediation analysis was tested in the following model (Table 3).

Table 3. Description of the analyzed mediation model

Model	Independent variable	Proposed mediator	Dependent variable
Model	Continuous quality improvement	Physical asset management	Business performance

The results of the mediation analysis are summarized in Table 4. Regarding the mediation model, the results show that the direct effect is not statistically different from zero, indicating that after controlling for the mediator, there is no relationship between continuous improvement practices and business performance (c' = 0.0488, $p > .05$). This suggests that asset management fully mediates the effect of continuous improvement practices on business performance. The latter is consistent with Baron and Kenny [29] who suggest that perfect mediation occurs when c' becomes insignificant after controlling for M, which is the case in our paper.

Table 4. Mediation of the effects of the continuous quality improvement on business performance through physical asset management

Coefficients			
(a paths)	(b paths)	Total effect (c path)	Direct effect (c-prime path)
Model ($F = 10.4477$, $p < 0.01$)			
0.4998, $p = 0.0000$	0.4853, $p = 0.0022$	0.2913, $p = 0.0017$	0.0488, $p = 0.6772$

As can be seen from the results (Table 5), the true indirect effect of physical asset management is estimated to be between 0.1088 and 0.3977.

Overall, the directions of the a and b paths are consistent with the interpretation that a greater commitment to continuous quality improvement leads to greater asset management performance, which in turn leads to greater business performance.

4 Discussion and Conclusion

Empirical evidence of the role of PAM in enhancing business performance is scarce and incomplete [4, 30]. The present survey highlights the relationship between continuous

Table 5. Bootstrap estimates of the mediated effect and its standard error

	Product of coefficients		Bootstrapping/BCa 95% CI	
Point estimate	SE	Z	Lower	Upper
Model				
0.2426	0.0810	2.9945	0.1088	0.3977

Notes. Bca CI -Bias Corrected and Accelerated Confidence Intervals, 1000 bootstrap samples

quality improvement and business performance by studying the PAM as the mediating role. Moreover, from a theoretical perspective, our results show that continuous quality improvement contributes significantly and positively to business performance. This is consistent with current literature suggesting that firms can benefit from implementing quality management practices in their processes [31]. To better understand this phenomenon, we also examined the mediating effects of PAM on the relationship between continuous quality improvement and business performance. We found that PAM serves as a full mediator in the link between continuous quality improvement and business performance. Our results show that organizations can benefit by focusing on quality management, thereby creating an environment in which organizations can successfully build and deploy continuous improvement activities. The results also clearly show that quality management activities can facilitate the organization's pursuit of affective asset management, which in turn improves organizational performance. Accordingly, the study highlights the need to recognize the importance of quality management activities, particularly their role in influencing PAM outcomes.

The results represent a major contribution to the asset management literature by providing insights into the mechanisms of mediation effects given by mediation analysis. The results of this research therefore contribute to the current literature on asset management. [1, 4, 5, 32]. It can be argued that incorporating, integrating, and managing PAM in organizations is still under-researched. This paper has aimed to provide insights into this topic. Moreover, our findings support the idea of conceptualizing and operationalizing PAM in a holistic manner, thus supporting previous studies advocating holistic and systematic approaches focused on continuous improvement and PAM [32, 33]. The additional theoretical implication, regarding the performance improvement, could be seen in terms of PAM being the precondition for achieving superior performance results. This result is in-line with previous research that highlighted the importance of performance-driven PAM [34].

From a practical perspective, our findings broadly support the implementation of continuous quality improvement and PAM-focused practices as part of a strategy to create a performance-driven culture in the organization. To the extent that they can be generalized, we believe that the present findings are instructive for organizations implementing the PAM practices covered in this study, namely Policy & Strategy, Risk Management, Life Cycle Management, and Performance Assessment. Our conceptualization of PAM practices shows strong coherence with globally accepted asset management frameworks.

In conclusion, some limitations of the study must be acknowledged. Clearly, causal inferences about the relationship between continuous quality improvement practices, PAM and business performance must be made with caution, particularly because the data are cross-sectional. Furthermore, it is possible that the relationships observed in the present study are inflated by the common method variance associated with perceptual measures. Future research should use longitudinal studies to reduce potential common method variance by separating the measurement of the independent and dependent variables [35]. In addition, a conceptual framework could be developed that would propose PAM factors that lead to improved organizational performance, with potential moderators such as asset management culture. Moreover, our research model can be an impetus for future research to further develop similar research models that explain how PAM contributes to firm performance in a single business context, such as a particular service or industry. Future studies could frame other contemporary issues within PAM, such as maintenance management [36] and digitalised maintenance [37]. Apart from business performance, which was used in this research, future studies could investigate the relationship between PAM and other aspects of performance, such as innovation performance. Additionally, organizations are expected to be sustainable in environmental and social sense. In this regard, future studies could focus on the integration of sustainability aspects into asset management activities.

APPENDIX A: Measurement Scales

Physical asset management *(Risk Management, Performance Assessment, Life cycle Management, Policy & Strategy).*

Risk Management: RM1: We embed risk into all activities which could affect assets performance; RM2: We analyse IT-system, business system, human resources, competence, etc. and address risk; RM3: We analyse operation, production, quality and logistic process and address risk; RM4: We perform risk assessment in order to minimize business losses; RM5: Risk management is an integrated part of asset management strategy; RM6: We analyse equipment failure causes and effects to address risk.

Performance Assessment: PA1: We exploit asset history to enhance asset knowledge; PA2: We regularly review overall effectiveness of asset management activities; PA3: We undertake benchmarking to support asset management activities; PA4: We monitor key performance indicators (KPIs) to verify the achievement of organization's asset management goals; PA5: We proactively pursue continuous improvement of asset management activities; PA6: Company collects and analyses data related to asset management activities; PA7: We regularly review overall efficiency of asset management activities; PA8: We exploit information systems to support asset management activities (ERP, CMMS, AMS, or similar ones); PA9: We monitor condition of critical assets.

Life Cycle Management: LM1: We continuously modernise our assets in accordance with our renewing/revision plans; LM2: We continuously rationalise our assets to reduce production cost; LM3: We assure quality of our assets during the whole life cycle phases; LM4: We assure execution of maintenance processes within all assets' life cycle phases; LM5: We execute disposal of assets in accordance with the asset management plan.

Policy and Strategy: PS1: We execute asset management strategy; PS2: We undertake analyses of asset management policy to determine future production capacity; PS3: We apply asset management policy; PS4: We develop asset management objectives.

Continuous quality improvement - CQI1: We continuously invest effort in searching for losses in all internal processes; CQI2: We use specific organizational structures (quality committee, improvement teams, etc.) to support quality improvement; CQI3: We systematically identify areas for improvement; CQI4: We systematically monitor proposals and suggestions for improvements; CQI5: We deploy quality management approaches (standards, methods, tools) to improve our processes.

Business Performance - BP1: Sales growth has increased above industry average during the last 3 years; BP2: Market share has increased during the last 3 years; BP3: Profit growth rate has increased above industry average during the last 3 years; BP4: Return on assets (ROA) has increased above industry average during the last 3 years; BP5: Return on investment (ROI) has increased above industry average during the last 3 years.

References

1. Maletič D, Maletič M, Al-Najjar B, Gomišček B (2020) An analysis of physical asset management core practices and their influence on operational performance. Sustainability 12(21):Art no 21. https://doi.org/10.3390/su12219097
2. Emmanouilidis C, Komonen K (2013) Physical asset management practices in industry: comparisons between greece and other eu countries. In: Prabhu V, Taisch M, Kiritsis D (eds) APMS 2013, vol 415. IAICT. Springer, Heidelberg, pp 509–516. https://doi.org/10.1007/978-3-642-41263-9_63
3. de Almeida NM, Vieira J, Silva JG, e Castro C (2021) The impact of asset management development programs in infrastructure organizations. In: Rodrigues H, Gaspar F, Fernandes P, Mateus A (eds) Sustainability and Automation in Smart Constructions. ASTI. Springer, Cham, pp 247–258. https://doi.org/10.1007/978-3-030-35533-3_29
4. Maletič D, Maletič M, Al-Najjar B, Gomišček B (2018) Development of a model linking physical asset management to sustainability performance: an empirical research. Sustainability 10(12):Art no 12. https://doi.org/10.3390/su10124759
5. El-Akruti K, Dwight R (2013) A framework for the engineering asset management system. J Qual Maint Eng 19(4):398–412. https://doi.org/10.1108/JQME-01-2012-0002
6. Trindade M, Almeida N, Finger M, Ferreira D (2019) Design and development of a value-based decision making process for asset intensive organizations. In: Mathew J, Lim CW, Ma L, Sands D, Cholette ME, Borghesani P (eds) Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies. LNME. Springer, Cham, pp 605–623. https://doi.org/10.1007/978-3-319-95711-1_60
7. Amadi-Echendu J et al (2010) What is engineering asset management? In: Amadi-Echendu JE, Brown K, Willett R, Mathew J (eds) Definitions, Concepts and Scope of Engineering Asset Management. Springer London, London, pp 3–16. https://doi.org/10.1007/978-1-84996-178-3_1
8. NJ Hastings (2015) Physical Asset Management. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-14777-2
9. Burnett S, Vlok PJ (2014) A simplified numerical decision-making methodology for physical asset management decisions. S Afr J Ind Eng 25(1):162–175

10. Lima ES, McMahon P, Costa APCS (2021) Establishing the relationship between asset management and business performance. Int J Prod Econ 232:107937. https://doi.org/10.1016/j.ijpe.2020.107937
11. Chattopadhyay G (2021) Asset management journey for realising value from assets. In: Misra KB (ed) Handbook of Advanced Performability Engineering. Springer, Cham, pp 429–450. https://doi.org/10.1007/978-3-030-55732-4_19
12. Amadi-Echendu JE (2004) Managing physical assets is a paradigm shift from maintenance. In: 2004 IEEE International Engineering Management Conference (IEEE Cat. No.04CH37574), October 2004, vol 3, pp 1156–1160. https://doi.org/10.1109/IEMC.2004.1408874
13. Alsyouf I, Alsuwaidi M, Hamdan S, Shamsuzzaman M (2021) Impact of ISO 55000 on organisational performance: evidence from certified UAE firms. Total Qual. Manag. Bus. Excel 32(1–2):Art no 1–2. https://doi.org/10.1080/14783363.2018.1537750
14. Frolov V, Ma L, Sun Y, Bandara W (2010) Identifying core functions of asset management. In: Amadi-Echendu JE, Brown K, Willett R, Mathew J (eds) Definitions, Concepts and Scope of Engineering Asset Management. Springer London, London, pp 19–30. https://doi.org/10.1007/978-1-84996-178-3_2
15. Mardiasmo D, Sampford C, Barnes P (2012) The exemplification of governance principles within state asset management laws and policies: the case of Indonesia. In: Mathew J, Ma L, Tan A, Weijnen M, Lee J (eds) Engineering Asset Management and Infrastructure Sustainability. Springer, London, pp 613–631
16. EFNMS-EAMC (2012) How organizations manage their physical assets in practice, EFNMS Asset Management Survey (EFNMS). EFNMS
17. GFMAM (2014) The Asset Management Landscape, Second Edition. The Global Forum on Maintenance & Asset Management (GFMAM). https://gfmam.org/sites/default/files/2019-05/GFMAMLandscape_SecondEdition_English.pdf
18. TheIAM (2015) Asset Management – An Anatomy. The Institute of Asset Management. https://theiam.org/media/1781/iam_anatomy_ver3_web.pdf
19. Dillman DA, Smyth JD, Melani Christian L (2014) Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Ed. Wiley, Hoboken
20. Maletič D (2015) Interaction between Quality Management, Production and Maintenance Performance. University of Maribor, Faculty of Organizational Sciences, Kranj
21. Kaynak H (2003) The relationship between total quality management practices and their effects on firm performance. J Oper Manag 21(4):405–435. https://doi.org/10.1016/S0272-6963(03)00004-4
22. Terziovski M (2006) Quality management practices and their relationship with customer satisfaction and productivity improvement. Manag Res News 29(7):414–424. https://doi.org/10.1108/01409170610690871
23. Jaca C, Viles E, Mateo R, Santos J (2012) Components of sustainable improvement systems: theory and practice. TQM J 24(2):142–154. https://doi.org/10.1108/17542731211215080
24. Prajogo DI (2016) The strategic fit between innovation strategies and business environment in delivering business performance. Int J Prod Econ 171:241–249. https://doi.org/10.1016/j.ijpe.2015.07.037
25. Preacher KJ, Hayes AF (2004) SPSS and SAS procedures for estimating indirect effects in simple mediation models. Behav Res Methods Instrum Comput 36(4):717–731. https://doi.org/10.3758/BF03206553
26. Preacher KJ, Hayes AF (2008) Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 40(3):879–891. https://doi.org/10.3758/BRM.40.3.879
27. Hair JF, Black B, Babin BJ, Anderson RE (2010) Multivariate Data Analysis, 7th edn. Pearson Prentice Hall, Upper Sadde River

28. Kafetzopoulos D, Psomas E, Skalkos D (2019) Innovation dimensions and business performance under environmental uncertainty. Eur J Innov Manag 23(5):856–876. https://doi.org/10.1108/EJIM-07-2019-0197
29. Baron RM, Kenny DA (1986) The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 51(6):1173–1182. https://doi.org/10.1037//0022-3514.51.6.1173
30. Schuman CA, Brent AC (2005) Asset life cycle management: towards improving physical asset performance in the process industry. Int J Oper Prod Manag 25(6):566–579. https://doi.org/10.1108/01443570510599728
31. Cua KO, McKone KE, Schroeder RG (2001) Relationships between implementation of TQM, JIT, and TPM and manufacturing performance. J Oper Manag 19(6):675–694. https://doi.org/10.1016/S0272-6963(01)00066-3
32. Gavrikova E, Volkova I, Burda Y (2020) Strategic aspects of asset management: an overview of current research. Sustainability 12(15):Art no 15. https://doi.org/10.3390/su12155955
33. Konstantakos P, Chountalas P, Magoutas A (2021) The contemporary landscape of asset management systems. Social Science Research Network, Rochester, NY. SSRN Scholarly Paper ID 3380765. Accessed 20 Apr 2021. https://papers.ssrn.com/abstract=3380765
34. Roda I, Garetti M (2015) Application of a performance-driven total cost of ownership (TCO) evaluation model for physical asset management. In: Amadi-Echendu J, Hoohlo C, Mathew J (eds) 9th WCEAM Research Papers. LNME. Springer, Cham, pp 11–23. https://doi.org/10.1007/978-3-319-15536-4_2
35. Podsakoff PM, MacKenzie SB, Lee J-Y, Podsakoff NP (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 88(5):879–903. https://doi.org/10.1037/0021-9010.88.5.879
36. Barberá L, Crespo A, Viveros P, Stegmaier R (2012) Advanced model for maintenance management in a continuous improvement cycle: integration into the business strategy. Int J Syst Assur Eng Manag 3(1):47–63. https://doi.org/10.1007/s13198-012-0092-y
37. Algabroun H, Bokrantz J, Al-Najjar B, Skoogh A (2020) Development of digitalised maintenance – a concept. J Qual Maint Eng ahead-of-print(ahead-of-print). https://doi.org/10.1108/JQME-04-2019-0039

A Strategic Asset Management Framework for Improving Transport Infrastructure: Analysis for Belgian Land Transport Modes

Laura Molinari[1,2](✉), Elvira Haezendonck[1], and Manuel Hensmans[2]

[1] Vrije Universiteit Brussel, Pleinlaan 5, 1050 Brussels, Belgium
laura.molinari@vub.be

[2] Université Libre de Bruxelles, Avenue Franklin Roosevelt 50, 1050 Brussels, Belgium

Abstract. In today's society, infrastructure asset management is a priority for multiple policymakers as it is key to guarantee high-quality transport infrastructure. While the relative quality of transport infrastructure in a number of Western European countries is deteriorating, the volumes of freight and passengers, as well as the expected service levels of all modes of transport for citizens and businesses, are increasing sharply. In response, infrastructure asset managers have developed and integrated technical and management system innovations. While short-term cost and damage control is taken better care of, a long-term asset vision and strategic principles supporting a strong future transport infrastructure network are still largely missing in many EU countries. In this paper, we analyze the strategic infrastructure asset management (SIAM) for Belgian road transport, rail and inland waterways through a cross-case analysis. Our literature study identifies strategic asset management principles, potential barriers and solutions for transport infrastructure assets in general, as well as for the different transport modes in particular. Through in-depth interviews with Belgian top administrators, the principles and SIAM frameworks for different types of mainland infrastructure are analyzed. We find, based on the studied Belgian cases, that 'one SIAM-model does not fit all', and that a variety of models, adapted to transport modes and the regional context, could better suit the strategic goals of different policies.

1 Introduction

Infrastructure, in Europe in particular, is ageing and is reaching the end of its lifespan [1–3]. Many of these assets date from the 1950s until the 1970s as a result of the post-World War II Marshall plan and are desperately in need of new investments [2, 4]. Contrary to the aforementioned need, expenditure in the EU has remained stable in 2016, the year that it reached its lowest point in 20 years. This while expectations from users and stakeholders regarding quality, reliability, and service are continuously increasing [5–7]. Expectations of infrastructure remain unanswered as long as investments are not forthcoming [2]. This is also the case beyond the EU, on a global scale. A report from McKinsey Global Institute indicates a yearly investment need in transport, utility and telecom infrastructure of $3.7 trillion until 2035 to be able to support the global economy [8]. The Global Infrastructure Outlook's report [9] extends the investment time range

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 33–43, 2022.
https://doi.org/10.1007/978-3-030-96794-9_4

until 2040 instead of 2035 with a necessary investment of $4.6 trillion by 2040. Global spending on infrastructure by 2040 is estimated at $3.8 trillion, which leads to a gap of $800 billion [9]. This gap is estimated at $93.4 billion by 2040 for European transport infrastructure assets, including ports, airports, roads and railway infrastructure [9]. Even though many EU-countries have recently taken steps to increase the infrastructure investments, cf. recent Eurostat figures, the huge gap between the current assets state and the service level demands in a competitive Europe will continue to exist or even sharpen again, if in the long run the assets are not managed strategically. Porter [10] defines infrastructure, in his "Diamond of national advantage" – a seminal framework to understand competitiveness of business clusters and regions, as one of the most salient factor conditions when determining and supporting regional competitiveness. Making sure infrastructure does not become a competitive disadvantage in times of crises or serious budgetary restrictions, requires tight control as well as a long-term plan and vision. Both are at the core of asset management objectives. Sound asset management matches user's demands with infrastructure's supplies by setting a strategic direction and supports decision-makers with taking the right decisions at the right time when executing this strategy. Principles and possible frameworks for good asset management are discussed in the next section, and when looking at the evolution of insights on this matter, the focus of management of long-life assets evolves from a technical approach to a strategic matter [11–13]. Yet, while (case-based) research contributions on the operational approach continue to grow, insights on what a strategic approach to transport infrastructure asset management should entail, are largely absent [14, 15]. This paper defines eight key factors for a sound SIAM for managing land transportation assets by public organizations and analyzes for three modes and three regions in Belgium how their SIAM can potentially be improved to ensure a better asset quality and service level in the long run.

2 Infrastructure Assets of a Region and Their Management

Infrastructure assets refer to assets with a physical rather than a financial nature. Uddin, Hudson and Haas [16] add the 'public' aspect and define them as *"…all these combined facilities that provide essential public services of energy, transportation, roads, airports, … Infrastructure also provides the physical systems used to provide other services to the public through economic and social actions. These infrastructure facilities and services are provided by both public agencies and private enterprises."*. This paper researches the aspects related to public infrastructure for transport or with 'transportation' as main function, based on the research of Baldwin and Dixon [17]. Roads, inland waterways, railways, bridges and tunnels are particularly considered as these assets seem to have suffered heavily from underinvestment during the last decades and are still the main transport vectors in society today. According to the Global Infrastructure Hub [9], the greatest investment gap is supposed to be in road infrastructure, where the gap between spending following the current trends and investment needs will be around 31%. Policymakers are therefore under pressure and consequently, it is important to advise them on their strategies in maintaining and if possible, improving these particular assets, beyond countering basic infrastructure concerns such as safety and availability.

Infrastructure Asset Management (IAM) may range from managing maintenance activities only in a narrow focus [18], to managing all activities related to the lifecycle [19]. The asset's life cycle consists of the following activities: (1) needs assessment & goals identification, (2) infrastructure planning, (3) infrastructure design, (4) infrastructure construction, (5) infrastructure operations, (6) infrastructure monitoring and inspection, (7) infrastructure preservation and (8) end of life [20]. Within preservation, the following subsections can be distinguished: regular maintenance, rehabilitation, replacement and upgrade. This can range from a small-scale maintenance intervention to the replacement of the asset by a more sustainable option [16, 21]. The International Organization for Standardization [22] states that IAM can be seen as all the activities that create value from an asset. But, since the demand for infrastructure assets can be viewed as a derived demand for transport or movements, assets do not express value on themselves. They can contribute to the value-creation for its users by enabling them to travel more rapidly or easily [23]. Better road infrastructure for example can increase this contribution [23].

Many variations of IAM definitions and objectives exist, resulting in the fragmentation of the concept [24]. Definitions of IAM of roads are focusing mainly on cost-effectiveness, while the objectives for rail and inland navigation infrastructure are respectively safety and reliability and service and availability [20, 21, 25–28]. As a consequence of this variety, the management of infrastructure can be compared with the iron triangle in project management, consisting of time, performance and cost. An asset manager should always consider the trade-off between cost-effectiveness, safety and reliability and service and availability. Only one objective can be constrained, a second one needs to be optimized and a certain level of the third needs to be accepted. Based on the literature it can be stated that road IAM constrains cost-effectiveness and that inland navigation and rail IAM are constraining respectively service and availability and safety and reliability. The other strategies will be defined accordingly to the organization's objectives. The different objectives may be rooted in current stakeholder expectations regarding a transport mode, and reflected through political priorities, but this is not the focus of our research and therefore this is taken as a basic assumption. The variance in objectives leads to the question if IAM principles and processes should be equal for each public body and each transport mode.

2.1 Developing a Strategic Infrastructure Asset Management (SIAM) Framework for Transport Infrastructure

Several frameworks have been developed to guide decision-makers with the process and implementation of IAM, mainly focusing on the technical side and optimization of IAM systems. Chen and Bai [29] analyzed 337 academic articles on optimization techniques for asset management and found that the number of articles on this subject only increased during the last years. Some do point at the lack of the strategic aspect and the connection with organizational objectives [14, 15], but few have addressed it from this lens. Some that have [12, 15, 22, 30–32], defined SIAM as *"A strategic and systematic process of optimizing decision-making in resources allocation with the goal of achieving planned alignment of infrastructure asset with service demand throughout its lifecycle"* by Too, Betts and Kumar [12]. After analyzing five main frameworks for

SIAM [12, 15, 30–32], eight key factors for a sound SIAM were defined: (1) the accountability of context factors' influences on the government policy, (2) the translation of the policy into organizational objectives, (3) the possibility of non-asset solutions, (4) the development of transport mode specific goals, (5) the alignment between the government strategy and asset strategy, (6) the optimization of options, (7) the introduction of feedback loops and (8) organizational and knowledge management. First, to define a governmental policy responding to the needs of a country's economy and society, context factors (for example user's needs and environmental factors) need to be considered (1). Next, the defined policy should be translated into specific organizational strategic management objectives and a corresponding strategy needs to be developed (2). After that, in a stage of strategic planning, the gap between the objectives and the current supply needs to be analyzed, which can consequently be solved with asset or non-asset solutions (3). Including non-asset solutions in the SIAM framework is essential as it offers the possibility to solve existing problems without large asset interventions. Furthermore, different solutions must be translated into goals for each transport mode (4), and asset solutions consequently into asset management goals and plans on acquisition, operation, maintenance and disposal (5). The particular transport mode goals are included as silo-mentality forms one of the greatest issues in managing infrastructure at this moment [2]. The development of strategic goals for each mode can facilitate the collaboration between departments. Thereafter, the possible options need to be optimized based on the defined objectives as cost, time, risk and quality (6). Finally, feedback loops and a constant organizational and information management is required to ensure a continuous optimal service (7)(8).

Applying these factors as management principles when managing infrastructure can contribute to a better IAM and thus, to a better service and a more cost-effective policy. These factors should be included in each SIAM framework, regardless of the mode and the objectives handled by the organization. Considering the different objectives, the focus lies on the translation of organizational objectives into asset objectives and optimization of options. In these two principles, the objective chosen to be constrained should be well incorporated as it is the key aspect that needs to be considered when deciding on an investment or a range of investments.

3 Methodology, Case Selection and Data Collection

3.1 Cross-Case Study of Land Transport Modes Managed by Belgian Regional Administrations

To identify the current and desirable future practices of transport IAM in Belgium, a case study method is used. Schramm [33] argues that *"the essence of a case study, the central tendency among all types of case study, is that it tries to illuminate a decision or a set of decisions: why they were taken, how they were implemented, and with what result"*. In that way the reason and the history behind principles used in the organizations can be discovered. More in detail, the study opted for a multiple case method and investigates the IAM principles of land transport modes, rail, road and inland navigation, managed by Belgian federal and regional administrations. The goal is to give an as extensive as possible overview of the current and future desirable situation, without comparison to other modes or countries. The literature review already suggested that different modes can have different objectives, and therefore also need different frameworks and principles. The characteristics of modes, regions and countries can be various which would make a comparison inappropriate. As different organizations and transport modes are included in this research, the exact research method can be called a holistic multiple case method [34].

Belgium was chosen as country to perform the case study. First, the researchers and research chair are based in Brussels, Belgium, which leads to an extensive network of public and private contacts. Next to that, a study by Meersman and Nazamzadeh [35] states that the network of roads, rail and ports are main indicators to drive the Belgian economy. But despite its importance, it is clear that Belgium needs help with their infrastructure management, especially in the case of transport infrastructure. Comparing 2008 with 2019, the position of Belgium in the Global Competitiveness Report for the road quality index decreased, this while The Netherlands improved their position and the world's median increased [36, 37]. Furthermore, expenses on maintenance and new investments as % of GDP have been the lowest in Belgium compared to peer-countries[1] from 2007 until 2017. Belgium spent on average 0.6% of their GDP on transport infrastructure investment and maintenance between 2010 and 2017 [38–40]. This is lower than the advised 1% of GDP by the European Conference of Ministers of Transport [41]. Between 2010 and 2017, 75% of the expenditure can be related to new investments and 25% to maintenance [38, 39]. Finally, according to Mr. Debrun, senior advisor at the research department of the National Bank of Belgium, the need for public investment in infrastructure in Belgium must be recognized [42]. Although the need is clear, political obstacles are holding back investments and improvements of the transport network. Belgium forms the perfect example to indicate these potential political obstacles, given the complexity of decision making and the short duration of government terms. The country is managed by six official organizations and four different policy leaders and the duration of a government term is five years. The democratic mandate, and the changing leaders and public opinion, creates a tension field between the political leaders and the top administrators. As a consequence of the regionalization of the ministry of public works in 1989, mobility and infrastructure became regional authorities, meaning that the responsibility of transport infrastructure lies with the regions, except for rail which stayed a federal authority. Belgium is divided into three regions, Brussels, Flanders and Wallonia. The federal government includes one minister of mobility to manage the railways and to keep an overview, and each regional government includes one minister of mobility and public works responsible for roads and inland waterways. Under the responsibility of a minister, there are administrative organizations led by a chief administrator. Six different organizations are mandated with the management of transport infrastructure over the country's three regions, Brussels, Flanders and Wallonia (see Table 1). Organizations in Flanders are managed by another overarching organization called Mobiliteit en Openbare Werken (MOW). The federal and regional governments form their own budget, based on their own incomes and expenses. Next to that, the federal government provides the regional governments with additional financial resources.

[1] France, Italy, Luxembourg, The Netherlands, Finland and Austria. There was no complete data available for other countries.

Besides other motives, this complexity, and by consequence possible silo mentality, shows the research interest for a Belgian case study.

Table 1. Organizations mandated with the management of transport infrastructure in Belgium

Region	Road	Inland waterways	Rail
Brussels	Bruxelles Mobilité/Brussel Mobiliteit	Port of Brussels	Infrabel
Flanders	Agentschap Wegen & Verkeer	De Vlaamse Waterweg	
Wallonia	Service Public de Wallonie Mobilité et Infrastructures (& SOFICO)		

3.2 Data Collection Methods

Scientific literature, desktop secondary data in the form of reports from organizations in the same field and policy documents, in-depth face-to-face interviews with top administrators and a focus group with top administrators and experts were used as data collection methods. The interviews covered nearly all regions and all transport modes in Belgium (except for the Port of Brussels). In total seven interviews took place. Based on scientific literature and our developed framework, a semi-structured qualitative survey, including some open-ended questions, was drawn up. After executing seven intensive two-hour interviews in the period October 2019 - November 2020, remaining questions, confirmation of the preliminary findings and potential solutions or ways ahead were discussed in the format of a focus group with all six top administrators responsible for the IAM at their regional level. After carrying out the interviews and gathering the data, the method for cross-case analysis of qualitative data described by Miles and Huberman [43] was followed. The information was first reduced and synthesized, then it was displayed using visuals in the form of matrices and finally, based on these matrices, relevant conclusions were drawn and validated [43].

4 Results

In this section, results of the cross-case study will be discussed per transport mode. For each Belgian organization mandated with the management of road, rail or inland navigation infrastructure, their management is compared against the identified key success factors of the framework for SIAM.

4.1 Road Infrastructure

At this moment, Brussels and Wallonia are considering user's needs before drawing up a long-term vision plan, using workshops or user surveys. Brussels followed this method for the first time during this government term, while Wallonia repeats it regularly. Flanders, on the other side, does not yet include methods to define the demand for

Table 2. Application of the key-success factors of IAM for road infrastructure in Belgium

	Brussels	Flanders	Wallonia
(1) the accountability of context factors' influences on the government policy	●		●
(2) the translation of the policy into organizational objectives	●	●	●
(3) the possibility of non-asset solutions			
(4) the development of transport mode specific goals	●	●	●
(5) the alignment between the government strategy and asset strategy	●	●	●
(6) the optimization of options		●	●
(7) the introduction of feedback loops			
(8) organizational and knowledge management		●	●

infrastructure. Each organization uses their long-term vision plan to translate policy objectives into organizational objectives. However, while focusing on infrastructure to solve the needs identified from the gap between supply and demand, non-asset solutions are not yet considered in the regions. In Wallonia, policy objectives are introduced in the Gestion des Projets routierS (GPS) system, which automatically results in candidate asset projects. Brussels and Flanders translate policy objectives into asset objectives. All regions set transport mode specific goals, in Flanders this is done under the supervision of MOW. SPW in Wallonia also manages inland waterways next to roads, in Brussels only roads are managed by Brussels Mobility. Depending on the scope of activities of each organization in Brussels, general policy goals are being implemented as transport mode specific goals. Railroads are never part of it as this is the responsibility of another organization named Infrabel. Whereas Flanders and Wallonia are already implementing organizational and information processes dedicated to IAM, Brussels is still setting up an asset management direction and collecting information and creating a database. Based on the available information, only Flanders (Pavement Management System - PMS) and Wallonia (GPS) are trying to optimize their interventions. Since road agencies have budget as their main constraint, a strongly embedded budget optimization would be expected in their IAM practices. In contradiction, budget optimization is only done for highways in Flanders. It can be noticed that none of the regions uses feedback loops (Table 2).

4.2 Inland Navigation Infrastructure

Focusing on inland navigation infrastructure, both Flanders and Wallonia are considering context factors and user's needs by having regular conversations with the users of the waterways. Using their long-term vision plans, they attempt to translate policy objectives resulting from the input on these context factors into organizational objectives, and eventually into asset objectives. For the same reason as for road infrastructure, common

Table 3. Application of the key-success factors of IAM for inland navigation infrastructure in Belgium

	Brussels	Flanders	Wallonia
(1) the accountability of context factors' influences on the government policy	No data	●	●
(2) the translation of the policy into organizational objectives		●	●
(3) the possibility of non-asset solutions			
(4) the development of transport mode specific goals		●	●
(5) the alignment between the government strategy and asset strategy		●	●
(6) the optimization of options			
(7) the introduction of feedback loops			
(8) organizational and knowledge management		●	(●)

transport mode goals are defined by the overarching organizations. Both regions already have a system in place to collect inspection data from bridges and to link quality indicators to it. While Flanders already has a database of their other assets, Wallonia is still creating one. The organizations are focusing on the development of an extensive database of all their assets to define the assets that are the most critical in delivering an optimal service and that need an intervention. The optimization of options is not yet done in Flanders, nor in Wallonia and an integration of budget is still work in progress. Equal to road infrastructure, the possibility of non-asset solutions and the introduction of feedback loops are also missing in the SIAM frameworks of inland navigation infrastructure (Table 3).

4.3 Rail Infrastructure

In collaboration with the Belgian railways, responsible for delivering rail service in Belgium, Infrabel defines the user's needs, to draw up the document 'Strategy GO' and to define its objectives. Afterwards, these specific objectives are being translated into asset objectives. Years ago, Infrabel implemented an information system. But, while a large quantity of data is available, the quality of the data is still lacking. Together with the optimization of options, the possibility of non-asset solutions and feedback loops are at this moment non-existing (Table 4).

5 Conclusion

Based on the results and the findings of this study it can be concluded that each transport mode and each region have its own objectives, good practices and challenges. A variety of asset management principles and processes exists, is possible and is potentially

Table 4. Application of the key-success factors of IAM for rail infrastructure in Belgium

	Federal
(1) the accountability of context factors' influences on the government policy	●
(2) the translation of the policy into organizational objectives	●
(3) the possibility of non-asset solutions	
(4) the development of transport mode specific goals	●
(5) the alignment between the government strategy and asset strategy	●
(6) the optimization of options	
(7) the introduction of feedback loops	
(8) organizational and knowledge management	●

even required based on the differences. A 'one fits all strategy' implementation does not exist, nor in theory, nor in practice. The framework provides general guidelines, retrieved from best practices abroad and scientific literature, that can be applied on each mode and each region and that can help policy makers and top administrators with the introduction and development of their custom asset management process. In this way, tension fields between policy makers and top administrators resulting from the short government terms can be reduced or even solved as clear principles are available. It is however necessary to include all these principles to achieve a sound SIAM. None of the studied organizations are considering non-asset options as possible solutions and are introducing feedback loops. Moreover, while option optimization and supporting databases should be well developed to focus on the chosen objective, this is only the case in respectively two and three out of the six organizations included in the research. Given the raising complexity in launching, budgeting and executing infrastructure assets, a strategic, and following operational asset management (including data management), would provide more resilient asset management and thus, a more stable and stronger infrastructure factor for competitiveness in a centrally located country in the European Union.

6 Research Financing

This research was conducted in the context of the research chair in Infrastructure Asset Management (Vrije Universiteit Brussel and Université Libre de Bruxelles) with the financial support of Antea Group Belgium, Arcadis Belgium, Sweco Belgium and Tractebel.

References

1. OECD (2012) Strategic Transport Infrastructure Needs to 2030
2. Parlikad AK, Jafari M (2016) Challenges in infrastructure asset management. IFAC-PapersOnLine 49(28):185–190

3. van Breugel K (2017) Ageing infrastructure and circular economy: challenges and risks. In: Proceedings of the 2nd world congress on civil, structural, and environmental engineering, Barcelona
4. World Economic Forum (2014) Strategic Infrastructure Steps to Operate and Maintain Infrastructure Efficiently and Effectively
5. Directorate-General for Mobility and Transport of the European Commission (2019) Transport in the European Union Current Trends and Issues
6. Shah R, McMann O, Borthwick F (2017) Challenges and prospects of applying asset management principles to highway maintenance: a case study of the UK. Transp Res Part A Policy Pract 97:231–243
7. Wijnia Y, Herder P (2010) The state of asset management in the Netherlands. In: Engineering asset lifecycle management, pp 164–172
8. Woetzel J, Garemo N, Mischke J, Kamra P, Palter R (2017) Bridging Infrastructure Gaps: Has the world made progress?
9. Global Infrastructure Hub (2017) Global Infrastructure Outlook
10. Porter ME (1990) The Competitive Advantage of Nations. Free Press, New York
11. Arif F, Bayraktar ME (2018) Current practices of transportation infrastructure maintenance investment decision making in the United States. J Transp Eng Part A Syst 144(6):04018021
12. Too EG, Betts M, Kumar A (2006) A strategic approach to infrastructure asset management. In: Proceedings BEE postgraduate infrastructure theme conference, Queensland University of Technology, Faculty of Built Environment and Engineering
13. Too E, Too L (2010) Strategic infrastructure asset management: a conceptual framework to identify capabilities. J Corp Real Estate 12(3):196–208
14. El-Akruti K, Dwight R, Zhang T (2013) The strategic role of engineering asset management. Int J Prod Econ 146:227–239
15. Laue M, Brown K, Scherrer P, Keast R (2012) Integrated strategic asset management: frameworks and dimensions. In: Third international engineering systems symposium, Delft
16. Uddin W, Hudson WR, Haas R (2013) Public infrastructure asset management. McGraw-Hill Education
17. Baldwin J, Dixon J (2008) Infrastructure capital: what is it, where is it, how much of it is there? Can Prod Rev
18. Shewen E, Kovacs E (1995) Enterprise-wide integrated infrastructure asset management. Public Works 126(10):66–68
19. Weninger-Vycudil A, Hanley C, Deix S, O'Connor A, Pakrashi V (2015) Cross-asset management for road infrastructure networks. Proc Inst Civ Eng Transp 168(5):442–456
20. Sinha KC, Labi S, Agbelie BR (2017) Transportation infrastructure asset management in the new millennium: continuing issues, and emerging challenges and opportunities. Transportmetrica A Transp Sci 13(7):591–606
21. Zoeteman A (2001) Life cycle cost analysis for managing rail infrastructure Concept of a decision support system for railway design and maintenance. Eur J Transp Infrastruct Res 1(4):391–413
22. International Organization for Standardization (2014) Asset management - Overview, principles and terminology ISO 55000
23. Hartmann A, Ling FYY (2016) Value creation of road infrastructure networks: a structural equation approach. J Traffic Transp Eng (Engl Ed) 3(1):28–36
24. Schraven D, Hartmann A, Dewulf G (2011) Effectiveness of infrastructure asset management: challenges for public agencies. Built Environ Proj Asset Manag 1(1):61–74
25. Bittner J, Rosen H (2004) Transportation asset management overview. Public Works Manag Policy 8(3):151–155
26. Yin Y, Madanat SM, Lu XY (2009) Robust improvement schemes for road networks under demand uncertainty. Eur J Oper Res 198:470–479

27. Sen D, et al (2010) Developing waterways asset management program that incorporates differences between highly urban and semi-urban or rural watersheds in risks and management strategies. In: Water environment federation's technical exhibition and conference
28. Rama D, Andrews JD (2016) Railway infrastructure asset management: the whole-system life cost analysis. IET Intel Transport Syst 10(1):58–64
29. Chen L, Bai Q (2019) Optimization in decision making in infrastructure asset management: a review. Appl Sci (Switzerland) 9(7)
30. Tranfield D, Denyer D, Burr M (2004) A framework for the strategic management of long-term assets (SMoLTA). Manag Decis 42(2):277–291
31. Too EG (2010) A framework for strategic infrastructure asset management. In: Amadi-Echendu JE, Brown K, Willett R, Mathew J (eds) Definitions, concepts and scope of engineering asset management. Springer London, London, pp 31–62. https://doi.org/10.1007/978-1-84996-178-3_3
32. Brown K, Laue M, Tafur J, Mahmood MN, Scherrer P, Keast R (2014) An integrated approach to strategic asset management. In: Gheorghe AV, Masera M, Katina PF (eds) Infranomics, vol 24. TSRRQ. Springer, Cham, pp 57–74. https://doi.org/10.1007/978-3-319-02493-6_5
33. Schramm W (1971) Notes on Case Studies of Instructional Media Projects
34. Yin RK (2003) Case study research: Design and methods. Sage Publications, Thousand Oaks
35. Meersman H, Nazemzadeh M (2017) The contribution of transport infrastructure to economic activity: the case of Belgium. Case Stud Transp Policy 5:316–324
36. World Economic Forum (2008) The Global Competitiveness Report 2008–2009
37. World Economic Forum (2019) The Global Competitiveness Report 2019
38. OECD data Infrastructure investment (indicator) (2021) OECD. https://data.oecd.org/transport/infrastructure-investment.htm. Accessed Apr 2021
39. OECD data Infrastructure maintenance (indicator) (2021) OECD. https://data.oecd.org/transport/infrastructure-maintenance.htm. Accessed Apr 2021
40. World Bank Open Data GDP (current US$) (2021) World Bank. https://data.worldbank.org/indicator/NY.GDP.MKTP.CD. Accessed Apr 2021
41. European Conference of Ministers of Transport (2003) Principal Acts of the ECMT
42. Trandafir A (2020) A prosperous new year thanks to public investment? American Chamber of Commerce in Belgium. https://amcham.be/blog/prosperous-new-year-thanks-public-investment. Accessed Apr 2021
43. Miles M, Huberman M, Saldana J (2014) Qualitative Data Analysis: A Methods Sourcebook. Sage publications, London

Sustainability and Resilience

A Sensor-Less Daylight Harvesting Approach Using Calibration to Reduce Energy Consumption in Buildings

Brenden Harris[1(✉)] and Juan Montes[2]

[1] Queensland University of Technology, Brisbane, Australia
bharris@fredon.com.au
[2] Fredon QLD Pty Ltd., Underwood, Brisbane, Australia
fredoncad@fredon.com.au

Abstract. Using daylight to offset renewable power generation and energy consumption in buildings has attracted significant attention in recent years. More specifically, daylight harvesting (DH) systems have attributes that reduce energy consumption, but there has been limited uptake because compliant DH designs are rarely explored during design phases. A contributing factor is that the cost of equipment, and the installation of such, rarely achieves satisfactory investment returns. To demonstrate an economic benefit, this paper presents a predictive energy saving simulation that can be applied to artificial lighting designs. The results show that when using a sensor-less daylight harvesting (SDH) algorithm to predict energy savings, satisfactory financial returns can be achieved by modifying existing lighting control systems (LCS). To support this finding, a simulation using a prototype LCS, and real-world case study on a high-rise building was carried-out to demonstrate the financial assessment methodology that is used to predict energy savings.

1 Introduction

Designers of commercial buildings reportedly consider the impact that daylight has on a building work space [1]. Harnessing natural light is possible using lighting control systems to reduce energy consumption. This is important because power demand estimates for artificial lighting is approximately 20% to 30% of the consumed power in a commercial building [2–4]. Herein lies the opportunity for daylight harvesting (DH). The lamp power consumption is "roughly proportional" to the output illuminance [5]. Thereby providing a significant energy cost and whole of life saving if daylight can be efficiently harvested. To this point, research by Li et al. [3] used daylight factor (DF) as a medium to dim lights. This was supported by other studies that decreased energy cost by dimming lights [2, 6]. However, has become an active field of interest, especially with the publication of the international ISO 55000 series.

A large number of physical assets or diversity in the characteristics of these assets in an organization further increases the importance of having a systematic existing DH theory require lux sensors to be installed at intervals adjacent to the external windows,

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 47–56, 2022.
https://doi.org/10.1007/978-3-030-96794-9_5

or include built-in meshed sensors to network each light fitting [7]. The problem with these approaches is that they rarely provide a satisfactory return on investment (ROI) due to the high cost of installing sensors and extensive calibration, thereby limiting the use of energy-saving DH technologies.

Observations revealed that DH is not regularly adopted in industry practice. Researchers suggest that "lighting control is not popular" which is in part due to limited artificial lighting cost and performance data analysis [3]. This paper advances the application of SDH [8] by presenting DH analysis and calibration techniques to remove internal DH sensors. Furthermore, to establish the viability of installing SDH, a prototype is presented that integrates digital addressable lighting control (DALI) technology. The new software model (in MATLAB) and communication protocol produced a functioning SDH system to practically realise energy savings.

1.1 History of Daylight Harvesting

Daylight factor (DF) is identified as the "minimum legal daylight standard" [9] and is widely used to assess daylight penetration into an occupied building space. The 15 different sky types included in lighting software programs use DF to estimate available daylight in simulations [10]. However, the shortcomings of DF is well known and as such the Illuminating Engineering Society of North America (IESNA) Daylighting Subcommittee introduced algorithms that measure the impact of direct sun light [9]. Although, detailed daylight analysis appears to be used more extensively in architectural rendering practices [11].

Using lighting software to simulate the financial benefits from harness daylight is not required in design practice. This is because of questionable investment returns and the elements such as sensors and practical applications associated with DH technologies that are not adequately understood by practitioners [12]. However, the treatment of artificial light using sensors has received significant attention in literature [12–16]. The isolated research into sensor-less DH technology is by Yoo et al. [17]. They use a small control sample which limits the use of photovoltaic (PV) as an external sensor to dim 3 lights. Unlike SDH, their application to DH doesn't address an open plan office layout that requires an array of light fittings to be dimmed. To control an array of lights, Yoo et al. [17] identified lux sensors to be positioned approximately 4 m apart and 2–3 m from the window. Therefore, PV cells would require similar spacing. Furthermore, as the area of the sampled space increases, the decay of light is poorly represented when using an average daylight factor to determine light at the furthest artificial illumination position from the window. This deficiency cannot be ignored, and as such the SDH algorithm considers the factor of light at any point in the room through calibration.

2 Internal Daylight Simulation and Prediction

The deficiencies in current lighting design practices is that compliant designs provide little guidance on potential energy savings. To accurately forecast energy consumption savings predicting internal lux levels, the SDH method was used. Furthermore, to test the accuracy of SDH modelling, the AGi32 lighting software program using a transition glass setting was compared to the SDH calculation. A comparison between simulated and real world lux levels was not completed because calibration of internal light measurements was verified by Kacprzak and Tuleasca [2].

To demonstrate the relationship of daylight being external and internal to the building, three plot (x) locations (Fig. 1) were simulated in AGI32. This test verified existing daylight harvesting literature, and revealed a scalar factor (SF) that is used as a coefficient when calculating SDH [8]. Furthermore, the simulation (Fig. 2) ascertained an exponential decay of daylight that is consistent with existing literature [18].

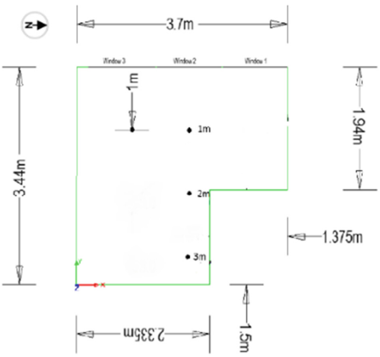


Fig. 1. Room layout to test AGI32 simulation

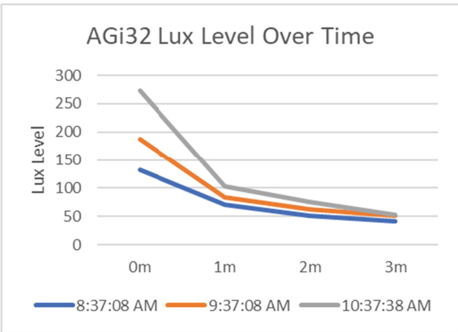


Fig. 2. AGI32 lux level in room exponential shape test

2.1 Daylight Modelling

The DF equation is not predictive because the influence from reflective surfaces on daylight distribution in a building remain unaccounted for. In addition, the room layout and furniture location will impact the distribution of daylight. These variables will either increase or decrease the amount of available daylight. To adequately predict daylight at a point in a room, these variable need to be factored. The other variable that needs to be considered is atmospheric being the position of the sun and the sky condition. To simulate these conditions, AGI32 and the like have standardised these variables using overcast sky conditions [19].

$$DF = \frac{\text{internal illuminance}}{\text{external illuminance}} \times 100\% \tag{1}$$

The AGI32 software was used to provide a conservative estimate of natural light using CIE overcast sky conditions. The hypothesis is that irrespective of the weather condition, if the variables such as reflection percentage of ceilings, wall, desks, and glass translucency remain constant, a linear relationship exists between the amounts of external illuminance in proportion to the internal illuminance at a particular point. This scalar value (SF) relationship is dependent on the exponential function of the decaying luminance intensity, and is different (exponentially) at any given distance from the window [20].

$$SF(x) = \frac{EL(t)}{IL(t, x)} \tag{2}$$

where x denotes the distance away from the window and t denotes time. The key hypothesis is that SF depends only on x and does not depend on time, i.e. ratio of external lux (measure at the weather station) to internal lux is constant for a particular distance from the windows.

To derive a financial prediction using SDH, the "base-intercept" or in the "relative-rate-intercept" form (IL = y) is used to determine the decay of daylight at any point in a room to establish a basis for dimming a light fitting.

$$IL(t, x) = IL(t, 0)a^x \text{ or } IL(t, x) = IL(t, 0)e^{rx} \tag{3}$$

2.2 Dimming Control with Multiple Natural Light Sources

The variable to calculate the dimming value (DimX) at each sample location changes when a room records light entering from multiple directions. To reflect this internal to daylight relationship [2], AGI32 software can simulate the lux at each lighting point to establish SF. In practice, to identify the dimming value (DimX), the internal lux level at each lighting location needs to be known. The combining of natural lux and artificial lux is used to maintain the design value (DV) that represents the specified working lux level. The DimX percentage is prioritised to the natural lux readings in that it will only identify a dimming percentage once the internal light falls below the specified DV lux level.

$$DimX = 1 - \left\{\frac{1}{DV} \times \left(\frac{Elx}{SF}\right)\right\} \text{ or } 1 - \left\{\frac{1}{DV} \times \left(\frac{Elx}{\frac{Elx}{Ilx}}\right)\right\} \tag{4}$$

2.3 Model Calibration in the Presence of Multiple Light Entries

When calibrating internal light levels, there needs to be a minimum of three sample plots which ideally will correspond to the dimming locations (Fig. 3). When the daylight enters from multiple direction, the sample lux plot is deemed sufficient because the decay of light at that point is captured by the sensor. Albeit when there is no light obstruction, the sample needs to be taken from the center point with sensors position at intervals to the window. By sampling the daylight values, window obstructions and the reflectance values are captured. This allows SF to be determined and hence a dimming compensation value can be derived to predict energy comparison from SDH to on/off state controls.

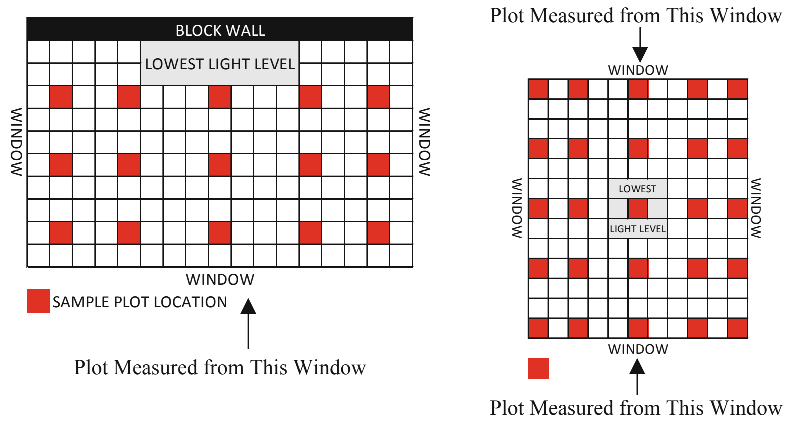


Fig. 3. Lux plot sample examples [8]

2.4 Energy Savings and Return on Investment

To determine the energy efficiency benefit of SDH, a simulation using 250 lights which represent a small office building, and 4500 light fittings (LFP) which represent a high-rise commercial building, was completed. To identify the total power (TP). Each lighting fitting power (LP) (where the wattage is constant) is compared to the rate of dimming required to maintain prescribed lux levels. To demonstrate, the data in Table 1 that was taken from AGI32 and used to simulate the financial benefit of installing SDH.

$$TP = \frac{\Sigma_{DimX=1}^{DimXn}}{DimXn} \times LP \times h\left(where\ LP = \Sigma_{LFP=1}^{LFPn}\right) \quad (5)$$

The external lux data was taken using an overcast sky and natural sky at latitude −27.38 and longitude −153.13 in summer. The power consumption estimate was calculated using 28 watts per light (LP), at a power cost of 21 cents per kWh (PC), with a total

Table 1. External lux values modelled from AGI (example)

Time	Overcast external lux values	Clear sky external lux values	Rate of dimming
8 am	13864	54425	75%
9 am	18787	78673	76%
10 am	22565	97498	77%

of 2,500 h (hrs) of operation which represents 1 year of use. To isolate the in-kind revenue from energy savings to calculate ROI, the difference in the non-daylight harvesting operating costs (NDH) is deducted from the operating costs (RC).

$$\begin{aligned} NDH &= \tfrac{LP}{1000} \times PC \times LFn \times hrs \\ RC &= \tfrac{LFP}{1000} \times PC \times hrs \end{aligned} \tag{6}$$

To estimate energy savings using dimming factors, external (EL) or internal (EL) lux level comparisons can be compared using overcast (OC) and clear sky (CS) simulations (Table 2). This prediction provides designers with data to include design variables such as the translucency settings on a blind that is used to minimises glare that causes occupant discomfort [21].

$$bt = 1 - \frac{osEL}{csEL} \approx 1 - \frac{osIL}{csIL} \tag{7}$$

To realise the benefit from the energy savings (TCS), the on/off state switching (NDH) is compared to the SDH (SDHCPY). These savings can then be used to provide a discounted (r) investment time (t) metric that the owner can evaluate using a net present value (NPV) calculation. In addition to establishing an ROI to compare the installation cost of SDH, the comparison can demonstrate a financial impact from applying a

Table 2. SDH return on investment comparison

Cost comparison based on number of fittings only				
	Overcast sky (OS)		Clear sky (CS)	
Number of lights	250	4500	250	4500
SDH power cost per year				
Year 4	$ 5,304.27	$ 95,476.81	$ 1,386.44	$ 24,956.01
Year 5	$ 6,630.33	$ 119,346.01	$ 1,733.06	$ 31,195.01
Power costs difference from on and off state and SDH switching				
Year4	$ 10,865.73	$ 195,583.19	$ 14,783.56	$ 266,103.99
Year5	$ 13,582.17	$ 244,478.99	$ 18,479.44	$ 332,629.99

translucency factor. Albeit, occupant comfort will need to be considered if the clear sky application is to be applied.

$$TCS = NDH - SDHCPY\left(whereCPY = \Sigma_{RC=1}^{RCn}\right) \tag{8}$$

3 Case Study

The sample data that was used to calculate SDH energy savings was taken from an iconic 39 floor high-rise building in the Brisbane CBD (Fig. 4). The simulation data included 100 light fittings per floor that operated for 10.5 h over 5-days per week, for a period of 48 weeks. This resulted in a non-DH energy cost of $88,353.72 when factoring a 28w light fitting at an energy cost of $0.29 per kWh. Using the SDH algorithm, the dimming value was calculated every 2 s for a period of 1 day in overcast conditions. The dimming rate was 56.36% and was used to conservatively estimate the energy savings for the entire year.

The total energy savings of 43.64% equates to approximately $39,000 per year which is consistent with the energy savings in existing literature authored by Doulos et al. [22], Delvaeye et al. [23], Li et al. [6], and Mayhoub [24]. However, unlike existing DH literature and practices that uses sensors to harvest daylight, SDH used mathematical modelling to eliminate the floor sensors. The difference in installation costs of $245,720 when comparing SDH to senor DH demonstrates a realistic ROI investment target can be achieved with SDH (Table 3).

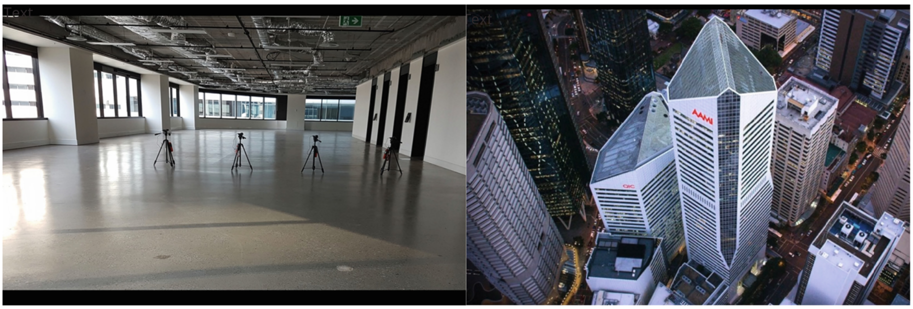


Fig. 4. Case study building

3.1 Practical Application of SDH and Prototype

A functioning model was developed to simulate the operational aspect of SDH. The system was scaled down to 1:50 lx of LED light to simulate the EL/sun. This light was housed in a light box that was connected to the LCS to visually display results. To summaries, the transfer of server commands includes; a) the data that is received from the light box, b) computation values from the MATLab compiler, and c) dimming values that are sent to the LCS. The LCS hardware and kWh meter was enclosed in a transparent box to show that standard products commonly installed in buildings can be integrated with the SDH algorithm, and that the actual power consumption of the LED lights was within 98% of the dimming percentage wattage (Fig. 5).

Table 3. Installation cost comparison between sensor operated dh and sdh

Sensor daylight harvesting	Unit cost per sensor	Sensors at 4 m intervals	Total
Preliminary Design	$ 15.00	1248	$ 18,720.00
Cost Per Sensor	$ 135.00	1248	$ 168,480.00
Electrical Installation (Inc Cable)	$ 95.00	1248	$ 118,560.00
Engineering/Programming	$ 10.00	1248	$ 12,480.00
Commissioning	$ 10.00	1248	$ 12,480.00
		Total DH cost	**$ 330,720.00**
Sensor-less daylight harvesting	**Hrs**	**Cost per hour**	**Total**
Preliminary Design	80	$ 250.00	$ 20,000.00
Light Control System OPC Mapping	40	$ 250.00	$ 10,000.00
Onsite Testing	160	$ 250.00	$ 40,000.00
Seasonal Inspections (1st Year)	20	$ 250.00	$ 5,000.00
Software (Site Specific)	1	$ 10,000.00	$ 10,000.00
		Total SDH cost	**$ 85,000.00**

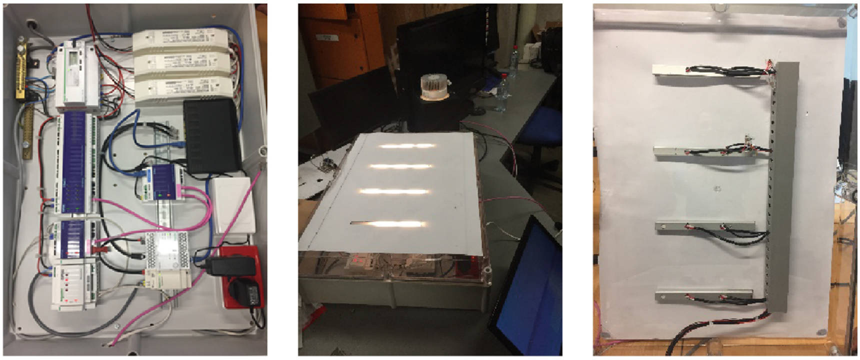


Fig. 5. SDH lighting control box with sun simulation by LED light

3.2 Future Research

The problem with current DH practices is that accuracy relies on complex computation models, or guesstimates with factors that significantly impact the overall efficiency of a

LCS. To provide a more robust business case for SDH, research regarding the efficiency of SDH and its financial benefits need to be expanded to include; a) asset whole of life benefits, b) technical integration requirements into lighting control designs that consider LED L70 factors, c) the use of machine learning techniques to predict daylight values, and d) the benefit that SDH brings to NABERs and other building schemes.

4 Conclusion

The case study and simulation proves that a sensor-less approach to DH provides owners with a satisfactory ROI. This research provides a financial metric that can be applied once the exponential equation to determine the natural lighting level at any location on a building floor is completed. The management of the 3-way relationship between internal, external, and artificial light to maintain a designed lux level can be completed using standard LCS software. The efficiencies from removing sensors reduces installation costs and provides a viable retrofit option for buildings that already have a LCS installed. Finally, this novel approach to DH will reduce energy consumption, extend the life of light fittings, and provide investment metrics to justify the implementation of a DH system.

References

1. Yu X, Su Y (2015) Daylight availability assessment and its potential energy saving estimation -a literature review. Renew Sustain Energy Rev 52:494–503
2. Kacprzak D, Tuleasca I (2019) Daylight savings calculation technique for smart buildings. 2013:102–107
3. Li DHW, Mak AHL, Chan WWH, Cheng CCK (2009) Predicting energy saving and life-cycle cost analysis for lighting and daylighting schemes. Int J Green Energy 6(4):359–370
4. de Rubeis T et al (2017) A first approach to universal daylight and occupancy control system for any lamps: simulated case in an academic classroom. Energy Build. 152:24–39
5. Ticleanu C (2015) Aiming to achieve net zero energy lighting in buildings*. Build Res Establ 6:63–78
6. Li DHW, Lam TNT, Wong SL (2006) Lighting and energy performance for an office using high frequency dimming controls. Energy Convers Manag 47(9–10):1133–1145
7. Van De Meugheuvel N, Pandharipande A, Caicedo D, Van Den Hof PPJ (2014) Distributed lighting control with daylight and occupancy adaptation. Energy Build. 75:321–329
8. Harris B, Montes J, Forbes N (2019) A novel approach to sensor-less daylight harvesting in commercial office buildings. In: Asset intelligence through integration and interoperability and contemporary vibration engineering technologies, pp 41–49
9. Saxena M (2008) A new & better way to calculate daylight, pp 12–15
10. Alshaibani K (2015) Average daylight factor for the ISO/CIE standard general sky. Light Res Technol 48:1–13
11. Ranasinghe S, Perera R (2016) Harnessing natural light for energy efficiency in buildings. Built-Environ Sri Lanka 1(1):12–18
12. Li DHW, Cheung ACK, Chow SKH, Lee EWM (2014) Study of daylight data and lighting energy savings for atrium corridors with lighting dimming controls. Energy Build 72:457–464
13. Pandharipande A, Caicedo D (2011) Daylight integrated illumination control of LED systems based on enhanced presence sensing. Energy Build 43(4):944–950

14. Caicedo D, Pandharipande A, Willems FMJ (2014) Daylight-adaptive lighting control using light sensor calibration prior-information. Energy Build 73:105–114
15. Peruffo A, Pandharipande A, Caicedo D, Schenato L (2015) Lighting control with distributed wireless sensing and actuation for daylight and occupancy adaptation. Energy Build 97:13–20
16. Caicedo D, Pandharipande A (2013) Daylight estimation in a faulty light sensor system for lighting control. In: Infornation Fusion (FUSION), pp 617–622
17. Yoo S, Kim J, Jang Cheol-Yong, Jeong H (2014) A sensor-less LED dimming system based on daylight harvesting with BIPV systems. Optics Exp 22(S1):A132. https://doi.org/10.1364/OE.22.00A132
18. Singhvi V, Krause A, Guestrin C (2005) Intelligent light control using sensor networks. In: Networked Sensing, pp 218–229
19. Bian Y, Ma Y (2017) Analysis of daylight metrics of side-lit room in Canton, South China: a comparison between daylight autonomy and daylight factor. Energy Build 138:347–354
20. Harris B (2021) Design to maintenance: a new construction design and building life cycle management approach
21. Suk JY, Schiler M, Kensek K (2016) Absolute glare factor and relative glare factor based metric: predicting and quantifying levels of daylight glare in office space. Energy Build 130:8–19
22. Doulos LT, Tsangrassoulis A, Kontaxis PA, Kontadakis A, Topalis FV (2017) Harvesting daylight with LED or T5 fluorescent lamps? the role of dimming. Energy Build 140:336–347
23. Delvaeye R, Ryckaert W, Stroobant L, Hanselaer P, Klein R, Breesch H (2016) Analysis of energy savings of three daylight control systems in a school building by means of monitoring. Energy Build 127:969–979
24. Mayhoub MS (2014) Innovative daylighting systems' challenges: a critical study. Energy Build 80:394–405

Resilience Rating System for Buildings Against Natural Hazards

Marta Duarte(✉), Nuno Almeida, Maria João Falcão, and Seyed M. H. S. Rezvani

Department of Civil Engineering, Architecture and Georesources,
Technical University of Lisbon – IST, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
marta.duarte@tecnico.ulisboa.pt

Abstract. In recent decades, there has been an increase in the frequency and intensity of natural disasters. The worldwide growth of population, and consequently of infrastructure, increases the exposure to risks of this type. The expectation that the frequency of such disasters will increase amplifies the need to act today, to minimize the associated economic risks and costs in the future. The ability of buildings to maintain or restore their functionality after disruptive events, within a certain period, has increasingly attracted the attention of academics and professionals. This work intends to study and develop a method to measure the resilience of built assets. Therefore, a resilience classification system is proposed, which assesses resilience according to 5 dimensions (environmental, economic, organizational, social, and technical), which are subdivided into 16 indicators and 75 parameters. This proposal is based on various existent systems such as REDi or Building Scorecard, and its applicability is tested with 11 buildings with varied uses. The results are analysed via SPSS using a Pearson correlation coefficient matrix and clustering techniques. These empirical cases allowed improvements in the system initially proposed. The proposed resilience classification system allows classifying and comparing the performance of buildings, identifying their vulnerabilities, essential information to establish investment priorities. Multiple stakeholders are involved in the life cycle of buildings that may benefit from the developed proposal. The work carried out is in its early stages of development and includes the identification of improvements to be developed in future work.

1 Introduction

The risks induced by natural and man-made disasters are inherently present throughout the entire lifecycle of buildings and engineering works. The built environment is thus vulnerable to risks that are impossible to eliminate, and this prompts the need for managing the resilience of constructed assets.

The impact of climate change on society, the construction sector, and individuals is widely debated. Various studies show that the frequency and intensity of natural disasters are increasing and that this, combined with high vulnerability and exposure, is also leading to increased economic and social losses [1, 2]. According to ISO/TR 22845 the frequency of these events is not expected to decrease, and this amplifies the need for action today to mitigate disaster risks in the future. For these reasons and given of

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 57–68, 2022.
https://doi.org/10.1007/978-3-030-96794-9_6

achieving the sustainable development goals, increasing attention is being drawn to the resilience of buildings and civil engineering works. Nowadays, the resilience of buildings is defined as "the ability to protect, maintain, or restore the functionality of, value of, and income generated by a building after a damaging event or circumstance within a prescribed time frame" [3]. The use of this concept is based on four pillars basic: i) resilience as trauma recovery and balance restoration; ii) resilience as synonymous with robustness; iii) resilience as the opposite of fragility; and iv) resilience as network architectures that can sustain the ability to adapt to future surprises as conditions evolve. For an infrastructure to have a high level of resilience it is necessary to be concerned with the four concepts, that is, pre-event drastic (preparation and mitigation) and post-event concerns (recovery and speed). It is worth noting the efforts currently involved in producing resilience-related international standards by both academics and various stakeholders such as construction project owners, managers, insurers, and municipalities. These standards are expected to enhance the understanding of resilience issues and allow comparison of pre and post-disaster measures of various infrastructures and building assets. In recent years, the development of building sustainability and resilience classification systems, such as ARMS [3], LiderA [4], REDI [5], RELi [6], Building Scorecard UN ARISE, [7], etc., has helped to establish concepts, indicators and metrics. However, the diversity of concepts is still quite prevalent, and the approaches and methods used to quantify resilience are not yet quite consensual [8]. The recently published international standard ISO/TR 22845 somehow helps to fill this gap within the context of buildings and civil engineering works. It establishes some core concepts and countermeasures and covers natural risks (e.g. earthquakes and climatic effects) and risks induced by man (e.g. terrorism). However, it does not solve all the difficulties in establishing resilience dimensions and metrics.

The overall goal of the present article is to: i) harmonize resilience metrics for buildings; ii) identify building vulnerabilities; and iii) streamline communication between various stakeholders. The interconnection of the objectives defined above is expected to be achieved by creating a multidisciplinary resilience rating system for buildings against natural hazards, thoroughly detailed bellow. The proposal builds upon ISO/TR 22845, ISO 31000 and the Sendai Framework for Disaster Risk Reduction.

2 Resilience Rating System

2.1 Design and Development

This chapter proposes a resilience classification system, with standardized metrics and simplified classification to understand and assess the resilience of buildings regarding natural disasters. The system is developed based on a holistic, comprehensive and systematic approach, allowing its application to different types of buildings (school, industrial, commercial, residential, hotel, etc.), at any stage of its life cycle (project, construction, use, etc.). This tool simplifies the identification of the building's resilience and weaknesses, allowing easy communication and comparison, either over time concerning the same building or others. It is intended for everyone involved in the construction, maintenance and building management processes, such as designers, contractors, project managers, construction owners and even insurance companies and municipalities, whose need to determine the resilience of the building and the community is high.

A deep approach at various levels is considered necessary, seeking to minimize interdependencies, and to achieve this, a three-level hierarchical structure was selected, consisting of dimensions, indicators and parameters. Each parameter is evaluated according to determined evaluation criteria. This work is recursive, with items and evaluation criteria calibrated and improved at each iteration.

The process is carried out with conceptual support from selected articles for this purpose. Their selection was based on the use of the keywords “resilience”, “buildings”, “natural hazard”, “indicators”, “seismic resilience”, “climate hazard” on the ScienceDirect platform whose results there were about 700. They are segregated based on publications and subjects, leaving only 50 for the literature review. Concepts were subsequently extracted from 9 selected documents: Almufti & Willford (2013), Asadzadeh & Kötter (2016), Burroughs (2017), Engle et al. (2013), Atrachali & Ghafory-ashtiany, et al., (2019), UN ARISE (2020), USGBC (2018), Verrucci et al. (2003) and World Economic Forum (2021). Their choice considers: i) their relevance to natural hazards; ii) their relevance to the built environment; iii) the justification given for the importance of the defined parameters; iv) recent documents. The definition of indicators and parameters aims at assessing resilience and facilitating communication and consultation procedures. The parameters subdivide indicators, and, in their turn, each set of indicators express in a more detailed manner each of the dimensions mentioned above. Their selection was substantiated through a literature review, bearing in mind that: (i) the selected parameters are possible to measure; (ii) information is available for their quantification and (iii) it is desirable to avoid overlaps or repetition of metrics. An initial list of more than 200 indicators was revised and reduced to 16 indicators, subdivided into 75 parameters that best fit the purpose of the intended rating system. The main drivers in the revision process were the elimination of repetitions of indicators and of those expressing a perspective at the level of urban and community concerns but that do not necessarily improve the resilience at the level of the constructed assets. Evaluation criteria were initially established based on thresholds for the different metrics.

The natural disasters covered by the proposed system correspond to disasters whose national exposure is high or medium, adapted from Union & Protection (2019): Earthquakes, floods (urban, river, sea), fires, tsunami.

2.2 Scoring

The proposed rating system builds upon existing resilience scoring systems [3, 5–7] and sustainability scoring systems, like LiderA (Portugal), Green Star (Australia), CASBEE (Japan), LEED (United States of America) and BREAM (United Kingdom) that are reasonably matured. A semi-quantitative scoring method is used. This allows grading progressive levels of performance for each indicator assuring: (i) accessible language, both in terms and concepts, that allows understanding by individuals who work or are qualified in the area of management of facilities and related built assets, (ii) criteria applicable to buildings with different types of use and (iii) identification of the level of attention needed for the analyses of indicators and dimensions. Following [9], the adopted scale complies with recommendations of ISO 11863 for considering five different levels expressed in whole numbers of a digit on a scale of 1,3,5,7 and 9, where 1 corresponds to the worst performance and 9 to the best. This scale allows even levels to be used when the correct assessment is between two levels. For reasons of simplicity, the weighting of each parameter in the pilot-test proposal is considered of equal importance. This allows the identification of the building's general performance and the performance of individual and unique aspects. For a clearer interpretation of the final score, the numeric scoring can be transposed into resilience classes from F to A++ (Fig. 1), allowing differentiation of resilience levels to be easily understood and intuitive.

Class	Average score
A++	[8,9]
A+	[7,8[
A	[6,7[
B	[5,6[
C	[4,5[
D	[3,4[
E	[2,3[
F	[1,2[

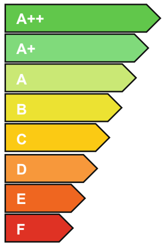


Fig. 1. Proposed rating scale

2.3 Dimensions

The output of the literature review was that the resilience of constructed assets in the face of natural risks could be structured into five dimensions: D1 - environmental; D2 - economic; D3 - organizational; D4 – social; D5 - technical. These dimensions are in line with: (i) the essential pillars for economic, social and environmentally sustainable development defined by the UN in EC0–92 Agenda XXI (United Nations Conference on Environment and Development); (ii) the four technical, organizational, social and economic dimensions defined by Bruneau et al. [10]; and (iii) the dimensions mentioned in various other selected documents (varying from 3 to 10 dimensions), taking into account that different terminologies were used to describe the same characteristic and the need to avoid overlapping concepts. Table 1 demonstrated the selected dimensions and indicators after the reduction was made.

In total, there are 5 dimensions, 16 indicators and 75 parameters in the system. The following sections provide descriptions of the rating system's contents for each of these dimensions.

2.3.1 Environment

The Environment dimension includes four indicators (I1 - Earthquake; I2 - Tsunami and tidal effect; I3 – Fire; I4 - Flood) and 25 parameters (P1 - Seismic zoning type 1 EC8; P2 - Seismic zoning type 2 EC8; P3 - Seismic vulnerability of the PDM soils; P4 - Slope of the terrain; P5 - Type of soil EC8; P6 - Distance to cliffs;P7 - Altitude of the terrain; P8 - Distance to the coast; P9 - Distance to the river; P10 - Natural barriers in the surroundings; P11 - Man-made barriers in the surroundings; P12 - Movable objects; P13

Table 1. System's dimensions and indicators

ID	Dimension	ID	Indicator
D1	Environment	I1	Earthquake
		I2	Tsunami and tidal effect
		I3	Flood
		I4	Fire
D2	Economic	I5	Insurance
		I6	Financial and strategic implications
D3	Organizational	I7	Internal organization
		I8	External organization
D4	Social	I9	Emergency infrastructures
		I10	Social responsibility
D5	Technical	I11	Conservation
		I12	Accessibility
		I13	Building seismic safety
		I14	Building security against fire
		I15	Building security against flooding
		I16	Building security against tsunamis

- Rows built between the coast and the building; P14 - Susceptibility to the direct tidal effect PDM; P15 - Relative location; P16 - Distance to the river; P17 - Natural barriers in the surroundings; P18 - Man-made barriers in the surroundings; P19 - Vulnerability to floods PDM; P20 - Distance to vegetation; P21 - Density of vegetation; P22 - State of maintenance of vegetation; P23 - Type of vegetation; P24 - Adjacent buildings; P25 - Proximity to the industrial zone).

This dimension seeks to foster a broad understanding of environmental issues, focusing on the vulnerability of the area to natural disasters of the upper and middle categories. The parameters were calibrated for the case of Portugal, providing an overview of potential threats as well as the determination of the intrinsic characteristics of the study area, such as altitude, distance to sea and river, slope, etc., since it can increase the risk propensity. The assessment related to natural disasters must be carried out for the present and future, considering that climate change modifies the frequency and intensity of disasters.

2.3.2 Economic

The Economic dimension includes two indicators (I5 – Insurance; I6 - Financial and strategic implications) and 3 parameters (P26 - Insurance against natural disasters; P27 - Financial plan; P28 - Economic assessment of downtime).

The economic aspects are crucial to make a building resilient and can greatly affect the quality of the building, especially during and after suffering the impacts of a natural disaster [11]. Studies show that good economic management and consistent financial availability improves the response to imposed natural disasters, and the recovery period is reduced. This dimension is related to the owner's monetary capacity in the face of imposed disturbances, including expenses on repairs, losses of assets and monetary losses from temporarily closed activities.

2.3.3 Organizational

The Organizational dimension includes two indicators (I7 - Internal organization; I8 - External organization) and 10 parameters (P29 - Business continuity plan; P30 - Risk management analysis; P31 - Post-disaster recovery plan; P32 - Routine; P33 - Plans and post-disaster exercises; P34 - Learning and updating; P35 - Destructive event data; P36 - Responsible; P37 - Compliance with the existing regulatory scenario; P38 - External standards for resilient construction).

The organizational capacity of buildings is related to the management capacity in emergency situations, that is, decision making by the owner regarding the identification, monitoring, and risk management. This dimension focuses on the pre-disaster, promoting preventive actions that reduce the impacts of natural disasters, guaranteeing a good performance of the building, minimizing the harmful consequences, and creating the least inconvenience for the users [12]. Topics outside of the owner's reach were also considered, like compliance with the existing regulatory scenario and the use of other standards of resilience. These indicators ensure construction safety and contribute to the preparation of buildings in the face of existing obstacles, helping to identify and prioritize problems.

2.3.4 Social

The Social dimension includes two indicators (I9 - Emergency infrastructures; I10 - Social responsibility) and 7 parameters (P39 - Access to police stations; P40 - Access to fire stations; P41 - Access to emergency infrastructure; P42 - Access to hospitals and health centers; P43 - Occupants; P44 - Disclosure; P45 - Social vulnerability).

The social dimension seeks to relate the building to society and the surrounding community, which are intrinsically related to each other, especially in times of stress whose individual response is difficult to identify and parameterize, but important to consider. Studies in resilient communities show that attentive and sensitive cities to individuals are better prepared for disasters, reducing its consequences [1], the same can be said for buildings. For this reason, factors like the building's social vulnerability, which corresponds to the number of elderly people, children, were considered. Additionally, it is intended to emphasize the role of citizens in response to disasters and the building's proximity to community infrastructures like fire stations, police stations, hospitals, etc.

2.3.5 Technical

The Technical dimension includes 6 indicators (I11 – Conservation; I12 - Accessibility: I13 - Building seismic safety; I14 - Building security against fire; I15 - Building security against floods; I16 - Building security against tsunamis) and 19 parameters (P46 - Year of construction; P47 - Structural system; P48 - Conservation status; P49 - Density of buildings; P50 - Alternative routes; P51 - Street characteristics; P52 - Plan irregularity; P53 - Height irregularity; P54 - Interaction with adjacent buildings; P55 - Slope difference; P56 - Expansion joint; P57 - Clearance between overlapping spans; P58 - Gas installations; P59 - Control and smoke evacuation systems; P60 - Intrinsic fighting means; P61 - Electrical installations; P62 - Fire compartment; P63 - Security team; P64 - Outdoor fire hydrants; P65 - Emergency lighting and signalling; P66 - Fire extinguishers; P67 - Fire detection and alarm; P68 - Escape routes; P69 - Barriers; P70 – Flood pumping systems; P71 - Exposure of the walls; P72 - Number of floors (flooding); P73 - Number of floors (tsunami); P74 - Orientation; P75 - Ground floor hydrodynamics).

This dimension focuses on all the technical and physical characteristics of both the building and its surroundings. The physical characteristics of the building are crucial to guarantee resistance to natural disasters and to minimize the damage caused by them [12]. This dimension derives from technical approaches and relates to the engineering component, which includes structural, security and the assessment of the building's physical vulnerabilities in the face of the natural disasters identified above. The building's redundancy and robustness strategies are included in this dimension, such as improvements beyond the building code or installing protection systems against natural disasters [11]. Intrinsic characteristics of the building like age, number of floors, irregularities, quality of construction, current condition, and state of conservation are considered in this dimension. Characteristics of the surrounding must also be analysed especially because of their impact on post-disaster recoveries [12] like the building accessibility that depends on multiple aspects like the existence of alternative routes, building density and street features.

3 Case Analysis of the System's Implementation

The choice of buildings used as case studies was made to test the feasibility of the proposed rating system in different situations. The sample of buildings covers new and old buildings, with or without rehabilitation interventions, higher and lower vulnerability to natural disasters, etc. Bellow are presented the results for 11 buildings: 2 residential buildings (C1 - single-family and C2 - multifamily), 2 schools (C3 - school 1 and C4 - school 2), 1 administrative building (C5 - research campus), 1 hospital (C6), 1 industrial building (C7 - carpentry factory), 2 commercial buildings (C8 - commercial building 1 and C9 - commercial building 2) and 2 hotels (C10 - hotel 1 and C11 - hotel 2). According to Portuguese regulations, this sample covers 7 out of 12 building use types. All the selected buildings are in the metropolitan area of Lisbon due to their ease of travel.

Table 2, indicates in a summarized way, the numerical score and corresponding resilience class for the five dimensions of each case study, derived from the arithmetic mean of the 75 parameters. In short, the hospital building obtained the best rating compared to the other case studies, A+. In the class below, A, there are both commercial buildings and school building 1. In class B, there are the hotels and school building 2. In class C there is multifamily residential building, industrial building and administrative building. The lowest class in this study proved to be D for the residential single-family building.

Table 2. Resilience rating of the case studies

Case studies	Dimensions					Total	
	D1	D2	D3	D4	D5		
C1	6,32	1,00	1,00	7,00	4,18	3,90	D
C2	5,95	2,00	3,75	6,00	6,72	4,89	C
C3	7,97	4,50	4,13	7,00	7,06	6,13	A
C4	7,31	4,00	2,88	6,33	5,64	5,23	B
C5	7,23	1,50	2,50	6,00	5,07	4,46	C
C6	6,73	8,00	7,00	7,67	8,72	7,62	A+
C7	6,11	2,00	2,00	6,42	6,89	4,68	C
C8	8,13	7,00	5,63	7,08	6,91	6,95	A
C9	5,81	7,00	5,63	6,33	7,13	6,38	A
C10	6,63	4,00	4,38	6,17	6,82	5,60	B
C11	5,27	4,00	4,38	5,42	7,44	5,30	B

After a first pilot-test application, the classification system was revised and purged from parameters that showed low applicability. An exhaustive analysis of the output results was made with the statistical software SPSS, performing a Pearson correlation analysis and Cluster analysis. The Pearson correlation matrix was made for every system's layer (dimensions, indicators and parameters), and in Table 3 it's possible to see the results for dimensions, which demonstrate a high correlation between D3 - Organizational and D2 - Economic, D2 - Economic and D5 - Technical, and D3 - Organizational and D5 - Technical, moderate correlation between D4 - Social and D1 - Environment, D4 - Social and D2 - Economic and D4 - Social and D3 - Organizational. The remaining showed no significant correlation. The dimensions with high correlation thus need revision in the future to ensure that each dimension has unique and distinctive items, avoiding repeated classifications or unnecessary criteria.

Table 3. Heat map of Pearson's correlation coefficients for dimensions

Pearson correlation	D1	D2	D3	D4	D5
D1		0,23	0,10	0,54	−0,11
D2			**0,93**	0,47	**0,73**
D3				0,31	**0,84**
D4					0,17
D5					

The clustering studies help us identify buildings with the same characteristics based on the clustered data, which is particularly helpful when comparing dimension's results (Fig. 2). Different clustering methods known as hierarchical (agglomerative) and non-hierarchical (K-mean) approaches were used for every system's layer. Results show that the k-means clustering approach narrowed down the outputs and provided the best results. The first cluster includes the residential single-family building, administrative building and the industrial building, the second cluster includes the hotels, schools and the residential multi-family building and the third one includes the commercial buildings and the hospital.

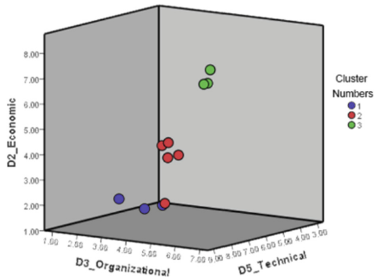


Fig. 2. K-mean (non-hierarchical) clustering allocation 3D representation of economic, organizational, and technical dimensions

The final scores, shown in Table 2, demonstrate that buildings with a high number of public users, such as hospital and commercial buildings, have higher levels of resilience, followed by hotels and school buildings and, finally, industrial, or residential buildings. In Fig. 3 it's possible to observe the score for each dimension from the first cluster, where the economic and organizational dimension show low scores. This information is in line with expectations, as buildings with no commercial activity or small activity typically have less financial capacity and less administrative resources [13]. The technical dimension shows high variance, thanks to different conservation states and other intrinsic

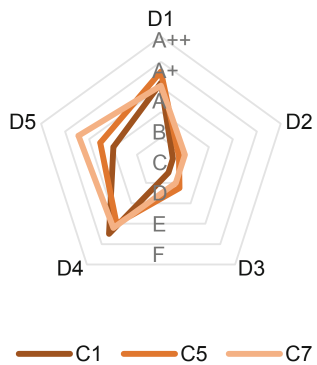


Fig. 3. Dimension's score for the first cluster

building's characteristics, fluctuating from class D to A, where social and environment dimensions show low variance, fluctuating from A to A+. The residential single-family building (C1) proved to be the worst classified on almost all indicators.

4 Conclusion

This paper contributes with a discussion about ways of measuring the resilience of constructed assets, namely based on a resilience rating system comprised of 3 layers, 5 dimensions, 16 indicators and 75 parameters. The proposed rating system covers not only the intrinsic building's qualities but also its interdependencies with the community, surroundings, and users in a post-disaster context. Eleven buildings are used as empirical case applications to test and calibrate the proposed system. The results were analysed with statistical techniques such as Pearson correlation coefficient and clustering via SPSS. The final case studies scores demonstrate that the system represents the building's resilience adequately, presenting values that are in line with expectations. Buildings open to the public, with intensive use and managed by large organizations, with greater administrative and financial capacity, present better scores, especially in the economic, organizational and social dimensions, demonstrating greater concern with the safety of users and the quality of the building. This is the case for commercial and hospital buildings (cluster 3). This was followed by hotels, schools and residential multifamily buildings (cluster 2) and, finally, buildings with fewer users and low or no economic activity, such as residential single-family building, industrial and administrative buildings (cluster 1). This last cluster is portrayed by the reduced concern with the economic and organizational theme.

It is concluded that the developed system is well dimensioned, since the results obtained allow the differentiation of groups with different degrees of importance. According to Almeida [9], each group is formed by constructions with similar technical risks and has a degree of relative importance. The application of the system allowed the differentiation of three clusters that are aligned with different groups of importance. The third cluster consists of buildings of high importance, essential for support in a catastrophe situation, vital for society and with high risks to human lives, economic and social issues in case of failure. The second cluster is made up of buildings with a beneficial influence on society and occupied by many people and the first cluster is made up of buildings with normal risk in terms of loss of life, economic or social issues in case of failure. Although the system presents satisfactory results, limitations were identified. Pearson's analysis allowed us to identify that the economic and organizational dimensions are highly correlated. In addition, a discrepancy was observed between the average scores of both dimensions compared to the others. These factors reveal that the dimensions are not completely uniform with each other, so the economic and organizational dimension should be reviewed in future studies. Indicators with reduced variance and high correlation between them should be revised in the future to ensure a uniform distribution score across the given scale and low proximity. Another limitation of the present study is the geographical limitation of the case studies carried out in the metropolitan area of Lisbon, which inhibited the proper analysis of 5 parameters, these being P1, P2, P20, P25 and P50. The proposed resilience rating system allows different stakeholders to efficiently identify which aspects should be improved and therefore establish investment priorities for enhancing the resilience of buildings. This information can be useful for all the stakeholders involved, i.e., owner, manager, insurers, and municipalities, enabling a better perception of the important contribution of constructed asset to resilient communities.

This methodology hopes to facilitate the operation, maintenance and construction phases of built assets, seeking to standardize recurring concerns. The main goal was attained, that is, to help translate natural hazards imposed on buildings to measurable resilient strategies and determine their classification that allows comparison both over concerning the same building, as with other buildings previously evaluated. Since the proposed system is in its initial stage of development, further work is needed to improve it. It is considered necessary to define thresholds corresponding to the minimum requirements for each group of degree of importance [9]. Different groups have different associated risks and different expected consequences, so the system should take this factor into account when determining the level of resilience. Finally, it is suggested to broaden the scope of the multidisciplinary rating system with regards to other types of disaster risks, like the man-induced risks. It should be noted that the developed items have an associated level of uncertainty, some being more developed than others. Thus, the system must be developed over time, improving items with a low level of development and or applicability such as the addition of new items emerging from future research opportunities.

Future work will determine the weight of each system's layer using the analytic hierarchy process (AHP) [14], which in this paper is considered equal for reasons of simplicity.

Acknowledgments. The authors gratefully acknowledge the contributions from Bruno Pacheco from Reify, João Couto from Pestana Group Hotel, Ana Filipa Salvado and Álvaro Vale e Azevedo from LNEC in the case studies.

References

1. World Economic Forum (2021) The Global Risks Report 2021 - 16th Edition
2. Intergovernmental Panel on Climate Change (2014) Climate Change 2014 Synthesis Report
3. Burroughs S (2017) Development of a tool for assessing commercial building resilience. Procedia Eng 180:1034–1043. https://doi.org/10.1016/j.proeng.2017.04.263
4. Pinheiro MD (2011) LiderA
5. Almufti I, Willford M (2013) REDi TM rating system resilience-based earthquake design initiative for the next generation of buildings
6. U.S. Green Building Council (2018) RELi 2.0 Rating Guidelines for Resilient Design + Construction
7. UN Private Sector Alliance for Disaster Resilient Societies (2020) Disaster Resilience Scorecard for Industrial and Commercial Buildings. For use by building owners, operators and managers, pp 1–103
8. Francis R, Bekera B (2014) A metric and frameworks for resilience analysis of engineered and infrastructure systems. Reliab Eng Syst Saf 121:90–103. https://doi.org/10.1016/j.ress.2013.07.004

9. Almeida N (2011) Performance and risk-based building technical management model - design, development and example of application to the structure. Instituto Superior Técnico. (in Portuguese)
10. Bruneau M et al (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. Earthq Spectra 19(4):733–752. https://doi.org/10.1193/1.1623497
11. Cerè G et al (2019) Urban-scale framework for assessing the resilience of buildings informed by a delphi expert consultation. Int J Disast Risk Reduct 36:101079. https://doi.org/10.1016/j.ijdrr.2019.101079
12. Atrachali M et al (2019) Toward quantification of seismic resilience in Iran : developing an integrated indicator system. Int J Disast Risk Reduct 39:101231. https://doi.org/10.1016/j.ijdrr.2019.101231
13. Meltzer R et al (2021) Regional science and urban economics localized commercial effects from natural disasters : the case of Hurricane Sandy and New York City. Reg Sci Urban Econ 86:103608. https://doi.org/10.1016/j.regsciurbeco.2020.103608
14. Rezvani SM, et al (2021) Urban resilience dimensions, indicators, and parameters for constructed assets evaluations by automate rational and consistent decision making using AHP (2021)

A Framework for Gamification to Encourage Environmentally Friendly Driving Habits

Turuna Seecharan(✉)

University of Minnesota Duluth, 1305 Ordean Court,
VKH 105, Duluth, MN 55812, USA
tseechar@d.umn.edu

Abstract. In recent years, modern society has been facing more traffic jams, higher fuel prices and an increase in Carbon Dioxide emissions. According to NASA Global Climate Change, the current warming trend is extremely likely (greater than 95% probability) to be the result of human activity since the mid-20th century. Although general awareness in sustainability issues has improved in recent years through mass media coverage, this knowledge is not always translated into actual sustainable practice. The transportation sector consumes more petroleum than any other sector, and that share has increased over time from about 50% in 1950 to about 70% in 2018. In 2016, light-duty vehicles accounted for 58.5% of transportation energy use while medium/heavy-duty trucks and buses accounted for 23.9%. Vehicle miles travelled was seven times higher in 2017 than in 1950. In the transportation sector, the Fourth Industrial Revolution (Industry 4.0) emphasizes advances in communication and connectivity with breakthroughs in emerging technologies in fields such as fully autonomous vehicles sector. Small to mid-sized cities are not always wealthy enough to adopt these infrastructure changes so sustainable transportation falls on the decision of commuters. This paper shows how gamification can be linked to the components of Industry 4.0 to encourage drivers to drive less aggressively and, thus, more environmentally friendly. The gamification approach is illustrated using a sample of nine college-aged drivers but can be extended to fleet drivers.

1 Introduction

Sustainable development focuses on meeting the needs of the present without compromising the ability of future generations to meet their needs [5]. Sustainable practices support ecological, human and economic health and vitality. This work defines sustainable driving as a set of driving techniques that saves fuel, is environmentally friendly and safe and hypothesizes that eco-driving is a sustainable driving technique.

The Fourth Industrial Revolution consists of many components including: mobile devices, location detection technologies, advanced human-machine interfaces, smart sensors, big analytics and advanced processes, augmented reality/wearables, and data visualization and triggered “live” training to name a few. Relative to the transportation sector, emerging technologies include self-driving cars, driver assistance systems or intelligent transportation systems [6].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 69–78, 2022.
https://doi.org/10.1007/978-3-030-96794-9_7

The definition of gamification varies between researchers such as "*The use of game elements in non-game contexts*" [8] or *"Incorporating game elements into non-gaming software applications to increase user experience and engagement"* [9]. Combining elements of Industry 4.0 and gamification to encourage environmentally friendly driving habits ("eco-driving"), in its simplest form, would require equipment to access data, cloud computing to store and analyze data and data visualization to communicate relevant information to the drivers. Effective data communication to the drivers is where gamification becomes relevant.

In 2018, there were 4,166,086 licensed drivers in Minnesota. Within the age groups 15–19, 20–24 and 25–29 who were involved in crashes, driving in a careless, negligent, or erratic manner was listed most often (10.9%. 3.5% and 15.2%) by officers at the scene [18].

In 2018, transportation activities were the largest source of emissions at 28% of total GHG emissions with 59% of CO2 emission from light-duty vehicles [25]. Improvements in vehicle fuel economy since 2005 has slowed the rate of increase of CO2 emissions; however, electric vehicles make up only a small percentage of vehicles sold at 1.9% in 2019 [20]. Although self-driving cars will help save lives, collecting data for self-driving cars is expensive and require billions of hours of footage of real driving for training Artificial Intelligence (AI) algorithms. Therefore, supporting the human driver in sustainable driving habits is still necessary [1, 15].

A variety of efforts have been undertaken to improve fuel economy and reduce emissions of on-road vehicles, including more stringent automotive emission standards, new engine and vehicle technologies, better fuel quality and renewable fuels. However, an important factor which is often overlooked and may improve vehicle fuel economy significantly is eco-driving technology. The investment for new vehicle technologies and fuels is usually significant and long-term, and an improvement of a few percentages may be considered significant. It was estimated that potential efficiency improvements of advanced engine and vehicle technologies were only about 4–10% and 2–8% respectively.

Behavioral approaches known as "eco-driving", can improve fuel economy by 15–25% along with promoting safety in road transport [3, 7, 14]. Eco-driving encourages drivers to anticipate traffic flow and maintain a steady speed with minimal braking and acceleration. Even with the benefits of eco-driving, studies have shown that drivers are still hesitant to change their driving style with motivation being the deciding factor when it comes to changing driver behavior [2, 17]. The objective of this paper is the development of tools to evaluate the behavior of drivers in order to make driving analysis to promote eco-friendly driving.

According to Self-Determination Theory, performing tasks that satisfy basic human needs for competence (feeling of proficiency), autonomy (control of one's actions), and relatedness (belonging or connectedness to others) enhances intrinsic motivation [21]. One way to engage and motivate people is through games. Elements of game design such as points, badges and leaderboards, have been applied to non-gaming contexts to enhance user engagement and experience [8].

The use of gamification has shown positive results in encouraging eco-drive habits [13, 17, 23]. Gamified eco-driving applications can be found in many cars including

Ford's SmartGauge with EcoGuide, the color-switching eco-gauge of the *Chevrolet Volt*, or *Kia's ECOdynamics* system. *Fiat's eco:Drive* is a computer application that Fiat promises will cut fuel costs and reduce a driver's carbon footprint. It requires a USB stick, a computer and either the Fiat Grande Punto, 500 or Bravo [11].

Recent studies in using gamification for driver behavior modification defines metrics based on the golden rules of eco-driving. The level of complexity of these metrics differ. The GamECAR consortium developed an eco-score that penalizes high level of engine rounds per minute (RPM) values during gear shift-up, cruising, abrupt braking and high acceleration, while the aggressiveness score penalizes high lateral acceleration, abrupt braking and high variances in throttle position and RPM [19]. This work also uses an eco-score and an Eco-Driver Grid based on the decision-making grid from maintenance literature.

In the maintenance of equipment, the allocation of maintenance strategies to various machines is analogous to prescribing drugs for different patients [13, 17, 22, 24]. The decision-making grid (DMG) acts as a map on which the performances of the worst machines are located according to downtime and failure frequency. Downtime is defined as the mean time to repair (MTTR) and failure frequency is the number of failure events (Fig. 1).

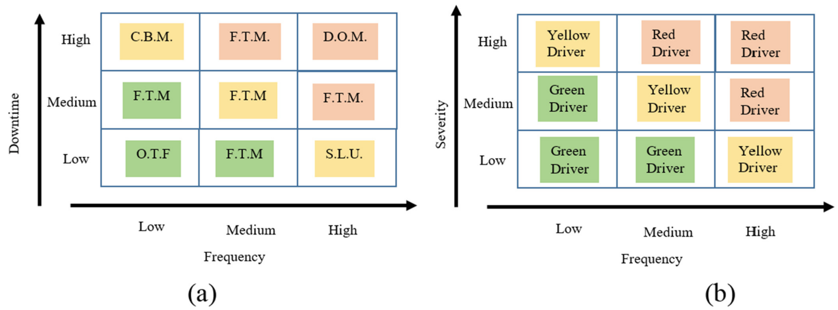


Fig. 1. (a) Decision making grid and (b). Proposed eco-driver grid

The DMG allocates one of five possible maintenance strategies for a failure event depending on its location in the DMG; Operate to failure (O.T.F.), Skills level upgrade (S.L.U.), Condition-based Maintenance (C.B.M.), Fixed-Time Maintenance (F.T.M.) and Design-Out Maintenance (D.O.M.). Assets that fall in the *O.T.F.,* section of the grid represent machines with low failure frequency and low downtime. This is the optimal state and is the area of the grid towards which all asset performance figures should aspire.

The goal of this work is to develop an eco-driver grid in a manner similar to the DMG where drivers are now identified as a green-, yellow- or red-driver. A long-term goal of this work would be to tailor feedback and training to "move" the driver toward the "green driver" section of the grid.

2 Methods

Participants were recruited using online postings at the University of Minnesota Duluth. Nine drivers (7 males and 2 females) between the ages of 18 to 30 years old were recruited to observe their driving habits in Duluth and develop the game elements. A telematics device that plugs in the on-board diagnostic post (OBD II) port of the study participants' personal vehicle collects engine and GPS data. The data collected is uploaded to a portal to which only the principal investigator and approved student researchers have access. Study participants do not have access to this portal in order to protect privacy. The goal of this study was to develop and test the eco-score and sustainable grid game elements.

2.1 Description of Data

The telematics device records GPS and engine data for each driver. From the engine data, acceleration forward and braking, acceleration side to side, acceleration up and down, GPS location, trip distance and number of times the driver "speeds" along with the distance and time spent speeding. For acceleration, this data is recording at small increments in time as shown in Table 1. Speeding, on the other hand, is a user-defined "rule" within the portal. A sample speeding report is shown in Table 2.

Table 1. Sample of acceleration data collected from the telematics device.

Vehicle	Date	Engine data	Value
CRV	Oct 27, 2020 6:14:03 PM	Acceleration forward or braking	0 m/s^2
CRV	Oct 27, 2020 6:14:03 PM	Acceleration side to side	0 m/s^2
CRV	Oct 27, 2020 6:14:03 PM	Acceleration up down	9.610517 m/s^2
CRV	Oct 27, 2020 6:14:05 PM	Acceleration side to side	−3.0204482 m/s^2
CRV	Oct 27, 2020 6:14:05 PM	Acceleration up down	10.483309 m/s^2
CRV	Oct 27, 2020 6:14:07 PM	Acceleration side to side	0 m/s^2
CRV	Oct 27, 2020 6:14:07 PM	Acceleration up down	9.7183895 m/s^2
CRV	Oct 27, 2020 6:14:34 PM	Acceleration forward or braking	0 m/s^2
CRV	Oct 27, 2020 6:14:34 PM	Acceleration side to side	0 m/s^2
CRV	Oct 27, 2020 6:14:34 PM	Acceleration up down	9.845877 m/s^2

2.2 Game Elements

Game elements include points, badges, leaderboards, performance graphs, meaningful stories, avatars and teammates [8]. Game elements in this pilot study included avatar names (chosen by drivers), the eco-score (points and leaderboard) and the eco-driver grid (badge and leaderboard).

Table 2. Sample speeding report.

Device	Location	Date	Start time	Duration	Distance (miles)
CRV	I-35, Pine City, MN 55063, USA	Nov 02, 2020	10:10:00 AM	0:01	1.98
CRV	I-35, Pine City, MN 55063, USA	Nov 02, 2020	10:11:33 AM	0:04	5.70
CRV	I-35, Rush City, MN 55069, USA	Nov 02, 2020	10:16:10 AM	0:09	12.65
CRV	5747 340th St, Stacy, MN 55079, USA	Nov 02, 2020	10:26:54 AM	0:02	3.75
CRV	30225 Fir Trail, Stacy, MN 55079, USA	Nov 02, 2020	10:29:42 AM	0:01	2.57
CRV	I-35W, Columbus, MN 55025, USA	Nov 02, 2020	10:40:14 AM	0:00	0.77
CRV	I-35W, Lino Lakes, MN 55038, USA	Nov 02, 2020	10:41:18 AM	0:02	2.72
CRV	5320 Main St NE Door, 70 Main St NE, Fridley, MN 55421, USA	Nov 02, 2020	10:56:29 AM	0:00	0.58
CRV	1701 James Cir N, Brooklyn Center, MN 55430, USA	Nov 02, 2020	10:58:01 AM	0:00	0.87
CRV	3319 67th Ave N, Minneapolis, MN 55429, USA	Nov 02, 2020	10:58:56 AM	0:00	0.53
CRV	12289 97th Ave N, Maple Grove, MN 55369, USA	Nov 05, 2020	9:09:56 AM	0:06	7.54

The eco-score was calculated based on four metrics – acceleration, braking, cornering and speeding. Thresholds were chosen based on previous research on the effect of hard acceleration on vehicle fuel economy and passenger safety [4]. Three levels were created for acceleration, braking and cornering to incorporate mid-range driving. Based on how data was collected from the telematics device, a mid-range level for speeding could not be designed. The scoring system for each metric is shown in Table 3. The relevant variables are defined as follows

AllAcc is the set of all AccX > 0; $N_{acc} = |AllAcc|$
AllBrk is the set of all AccX < 0; $N_{brk} = |AllBrk|$
AllCrn1is the set of all AccY > 0; $N_{crn1} = |AllCrn1|$
AllCrn2 is the set of all AccY < 0; $N_{crn2} = |AllCrn2|$
SpdFreq is the number of times a driver was recorded as speeding in one day

$L_S^{'}$ is the length of a trip not spent speeding (miles)
L is the length of a trip (miles)
g is the acceleration due to gravity

Table 3. Scoring system for eco-score

Range	Score	Metric	Level
$A_{CC}X_i > 0.4$ g	0	Acceleration ("AccScore")	Hard
0.29 g $< A_{CC}X_i \leq 0.4$ g	1		Medium
$0 < A_{CC}X_i \leq 0.29$ g	2		Soft
$A_{CC}X_i < -0.38$ g	0	Braking ("BrkScore")	Hard
-0.38 g $\leq AccX_i < -0.28$ g	1		Medium
-0.28 g $< AccX_i < 0$	2		Soft
$AccY_i < -0.38$ g	0	Left Cornering ("CrnScore1")	Hard
-0.38 g $\leq AccY_i < -0.19$ g	1		Medium
-0.19 g $\leq AccY_i \leq 0$	2		Soft
$AccY_i > 0.38$ g	0	Right Cornering ("CrnScore2")	Hard
0.19 g $< \mathbf{AccY}_i \leq 0.38$ g	1		Medium
$0 < \mathbf{AccY}_i \leq 0.19$ g	2		Soft
If Speed > 8 mph over speed limit for more than 30 s	0	Speeding ("SpdScore")	Speeding
If Speed < 8 mph over speed limit	2		Not Speeding

Table 4. The equations to calculate the score for each metric

Metric	Equation
AccScore	$\frac{\sum_{i=1}^{N} AccX_i}{2 \times N_{Acc}}$
BrkScore	$\frac{\sum_{i=1}^{N} AccX_i}{2 \times N_{brk}}$
CrnScore1	$\frac{\sum_{i=1}^{N} AccY_i}{2 \times N_{crn1}}$
CrnScore2	$\frac{\sum_{i=1}^{N} AccY_i}{2 \times N_{crn2}}$
SpdScore	$\frac{L_S^{'}}{L}$

$$\mathbf{EcoScore} = \left(\frac{\mathbf{AccScore} + \mathbf{SpdScore}}{2}\right) \tag{1}$$

In the preliminary study, the EcoScore was calculated for one day of driving. In future studies, the EcoScore will be calculated over a set distance (Table 4). Further research on the best distance to use is ongoing. Table 5 shows an example of the eco-scores obtained for each driver for one day.

Table 5. An example of the eco-Score for each driver from the study

Driver	Total
CRV	9.51
HW	9.34
Iron Man	9.50
Link	9.46
Liverpool1127	9.38
Pete	9.27
Universitario	9.33
Voldemort	9.20
Yung Modulus	9.40

The eco-driver grid was designed to display the relationship between severity and frequency of aggressive driving habits. The severity is calculated using Eq. (2).

$$\mathbf{Severity} = 1 - \mathbf{EcoScore} \tag{2}$$

Frequency is the number of times an aggressive driving habit was recorded.

$$\mathbf{Frequency} = |\{\mathbf{AccX_i} > 0.29\boldsymbol{g}\}| + |\{\mathbf{AccX_i} < -0.28\boldsymbol{g}\}| + |\{\mathbf{AccY_i} < -0.19\boldsymbol{g}\}| + |\{\mathbf{AccY_i} > 0.19\boldsymbol{g}\}| + \mathbf{SpdFreq} \tag{3}$$

In the grid, a lower frequency and lower severity is best. The goal is to move the driver towards the green section of the grid. Thresholds were defined based on the group driving performance and matches the decision making grid of the maintenance strategies literature. Let h be the highest value in the list and l the lowest value in the list.

$$\text{Medium/High boundary} = \boldsymbol{h} - \frac{\mathbf{1}}{\mathbf{3}}(\boldsymbol{h}) \tag{4}$$

$$\text{Medium/High boundary} = \boldsymbol{h} - \frac{\mathbf{2}}{\mathbf{3}}(\boldsymbol{h}) \tag{5}$$

$$\text{Low boundary} = \boldsymbol{l} \tag{6}$$

Figure 2 shows an example of the grid for a day of driving.

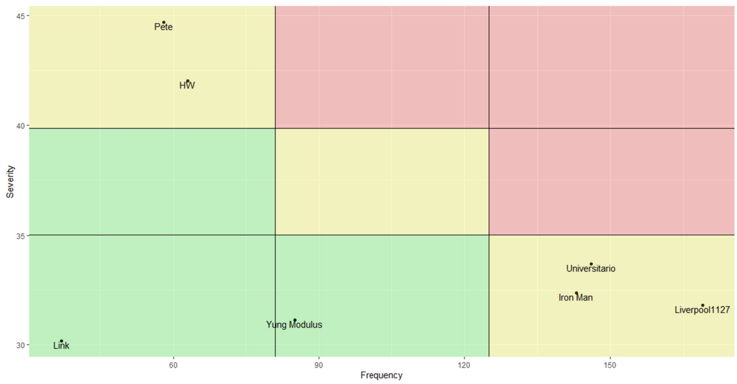


Fig. 2. Eco-driver grid.

This study developed the EcoScore and Eco-Driver Grid using data collected from a telematics device plugged into the OBD II port of the study participants' vehicles. Work is ongoing to optimize the EcoScore calculation and grid design.

2.3 Summary and Conclusion

This paper presents a framework of using game elements and components of Industry 4.0 to encourage drivers to adopt less aggressive and more environmentally friendly driving habits. The game elements include identification of drivers via avatars to protect privacy, a leaderboard is used to show the driver's position relative to others to encourage some friendly competition. The grid, developed based on the maintenance strategies literature, provides another data visualization technique to rank the drivers based on two dimensions of their driving metrics – severity and frequency of the non-eco-friendly driving habits. At the moment, drivers are informed of their performance at the end of the day via email. Work is ongoing to improve communication to the drivers, test peered driving and a longitudinal study to test the impact of gamification on over a longer period of time.

Acknowledgments. The author gratefully acknowledges the efforts of Karachi Ugwu and Francisco Arias in assisting with the analytics and communication with the study participants.

References

1. Adams T (2015). Self-driving cars: from 2020 you will become a permanent backset driver. Retrieved from The Guardian: https://www.theguardian.com/technology/2015/sep/13/self-driving-cars-bmw-google-2020-driving. Accessed 13 Sept 2015
2. Barkenbus J (2010) Eco-driving: an overlooked climate change initiative. Energy Policy 38(2):762–769
3. Barla P, Gilbert-Gonthier M, Castro MA, Miranda-Moreno L (2017) Eco-driving training and fuel consumption: impact, heterogeneity and sustainability. Energy Econ 62:187–194

4. Boodlal L, Chiang K-H (2014) Study of the Impact of a Telematics System on Safe and Fuel-Efficient Driving in Trucks. US Department of Transportation, Washington
5. Brundtland Commission (1987) Our Common Future. Oxford University Press, Oslo
6. Chan-Edmiston S, Fischer S, Sloan S, Wong M (2020) Intelligent Transportation Systems (ITS) Joing Program Office: Strategic Plan 2020–2025. U.S. Department of Transportation, Washington, DC. Retrieved from its.dot.gov
7. Cheng Q, Nouveliere L, Orfila O (2013) A new eco-driving assistance system for a light vehicle: energy management and speed optimization. In: Intelligent Vehicles Symposium (IV), pp 1434–1439. IEEE, Gold Cost, QLD
8. Deterding S, Dixon D, Khaled R, Nacke L (2011) From game design elements to gamefulness: defining gamification. In: Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, pp 9–15. Association for Computing Machinery, Tampere
9. Dominguez A, Saenz-De-Navarrete J, De-Marcos L, Fernandez-Sanz L, Pages C, Martinez-Herraiz J (2013) Gamifying learning experiences: practical implications and outcomes. Comput Educ 63:380–392
10. Entertainment Software Association (2020) 2020 Essential Facts About the Video Game Industry. Entertainment Software Association, Washington
11. Environment News (2009) Fleet technology: Fiat drives home the eco message. Retrieved from FleetNews. https://www.fleetnews.co.uk/news/2009/2/17/fleet-technology-fiat-drives-home-the-eco-message/29951/. Accessed 17 Feb (2009)
12. Fernandex O, Labib AW, Walmsley R, Petty DJ (2003) A decision support maintenance management system: development and implementation. Int J Qual Reliab Manage 20(8):965–979
13. Gunther M, Kacperski C, Krems JF (2020) Can electric vehicles be persuaded to eco-drive? A field study of feedback, gamification and financial rewards in Germany. Energy Res Soc Sci 63:101407
14. Hellstrom E, Ivarsson M, Aslund J, Nielsen L (2009) Look-ahead control for heavy trucks to minimize trip time and fuel consumption. Control Eng Pract 17:245–254
15. Hewitt C, Ioannis P, Amanatidis T, Sarkar A (2019) Assessing public perception of self-driving cars: the autonomous vehicle acceptance model. In: International Conference on Intelligent User Interfaces (IUI 19). ACM, Marina del Rey
16. Labib AW (1998) World class maintenance using a computerised maintenance management system. J Qual Maint Eng 4(1):66–75
17. Magana VC, Munoz-Organero M (2015) Gafu: using a gamification tool to save fuel. IEEE Intell Transp Syst Mag 7(2):58–70
18. Minnesota Department of Public Safety (2019) 2018 Crash Facts. Saint Paul: Minnesota Department of Public Safety. Retrieved from Minnesota Traffic Crashes in 2018. https://dps.mn.gov/divisions/ots/reports-statistics/Documents/2018-crash-facts.pdf. Accessed 24 July 2019
19. Nousias S, Tselios C, Bitzas D, Amazilatis D, Montesa J, Lalos AS, Chatzigiannakis I (2019) Exploiting gamification to improve eco-driving behaviour: the GamECAR approach. Electron Notes Theoret Comput Sci 343:103–116
20. Office of Energy Efficiency & Renewable Energy (2020) Plug-in Vehicle Sales Accounted for about 2% of all Light-Duty Vehicle Sales in the United States in 2019. Retrieved from energy.gov. https://www.energy.gov/eere/vehicles/articles/fotw-1136-june-1-2020-plug-vehicle-sales-accounted-about-2-all-light-duty. Accessed 1 June 2020
21. Ryan RM, Deci EL (2000) Intrinsic and extrinsic motivations: classic definitions and new directions. Contemp Educ Psychol 25(1):54–67
22. Seecharan TS, Ashraf L, Jardine AK (2018) Maintenance strategies: decision making grid vs Jack-Knife diagram. J Qual Maintenance Eng 24:61–78

23. Senova A (2018) Improving driver behavior using gamification. In: MobiWIS 2018, pp 10–20. ResearchGate, Barcelona
24. Shahin A, Labib A, Emami S, Karbasian M (2019) Improving decision-making grid based on interdependece among failures with a case study in the steel industry. TQM J 31(2):167–182
25. US EPA - Office of Transportation and Air Quality (2020) Fast Facts: U.S. Transportation Sector Greenhouse Gas Emissions 1990−2018. United States Environmental Protection Agency

Assessing the Economic and Environmental Effects of Gravel Recycling During Gravel Road Maintenance

Nea Svensson and Mirka Kans(✉)

Department of Mechanical Engineering, Linnaeus University, Växjö, Sweden
mirka.kans@lnu.se

Abstract. Approximately 300,000 km of the Swedish road network consists of gravel roads. These roads contribute to accessibility and accessibility throughout Sweden, which is especially important in rural areas. An annual operation and maintenance grant is paid to these roads to be maintained and kept open to public transport, but the grant covers only part of the total maintenance costs. Some of the costliest maintenance activities are planing and gravelling. When gravelling, natural resources in the form of rock and gravel are used, which is an energy-intensive process that has a negative impact on the environment. A couple of methods exist for recycling of gravel from the roads, but the utilization is rather limited. In order to promote and motivate recycling of gravel, it is important to highlight the environmental benefits of using recycled gravel, but also to be able to assess the economic impact as additional costs may arise. The overall purpose of the paper is to gain deeper understanding of the environmental and economic effects of recycling of gravel during gravel road maintenance. To achieve this, a calculation model is developed to estimate the environmental impact and economic effects of gravel road maintenance. The purpose of the calculation model is to be able to compare alternative methods for gravelling. The calculation model is evaluated through a test scenario with three alternative methods for gravelling; two where gravel recycling is performed by the means of two different methods and one in which new gravel is used. The test scenario shows that it is economically and environmentally beneficial, in a life cycle perspective, to use recycled gravel for road gravelling.

1 Introduction

The road network in Sweden consists of state, municipal and private roads [1]. The roads are either gravel roads or paved roads, and the traffic load determines whether the road should be paved or not [2]. In Sweden, there are approximately 300,000 km of gravel road of which approximately 280,000 km consist of private roads that are open for public access [3]. The formal responsibility for the public-private roads lies on the road association, and an annual operation and maintenance subsidiaries is paid by the government for the roads allowing them to be kept open for public transport. Normally, the grant covers only part of maintenance costs [4]. The maintenance work takes place e.g. in the form of gravelling and planning to promote one road safety [5].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 79–89, 2022.
https://doi.org/10.1007/978-3-030-96794-9_8

When gravelling, gravel or rock material that usually has a particle size of no more than 18 mm is used [2]. Either recycled gravel that is collected when grading the road or new gravel from a gravel pit or quarry is used. In 2015, 84 million tons of filling and ballast material in Sweden were produced: 56% within road construction and about 14% each to concrete production and filling material. Natural gravel is a non-renewable resource and therefore the outtake is regulated. In Sweden, the amount of gravel pits has considerably decreased: in 1990 there were about 5,000 gravel pits, in year 2000 about 3200, and 2012 about 1200 [6]. To supply the required amounts of material, quarries for macadam production and reuse of ballast material are replacing natural gravel. The decrease in gravel extraction locations means that the resources becomes more inaccessible, which leads to longer transport distances and thus increased transport costs [7]. Consequently, alternative strategies have been developed to solve the gravel shortage, such as paving of gravel roads with high traffic load, reduced maintenance work on gravel roads with low traffic load, and closure of gravel roads with non-existent traffic [8]. In addition, bottom ash that would otherwise be deposited has been suggested as a complement to natural gravel [7].

Other researchers advocate the importance of maintaining roads with low traffic loads, as these are important for the accessibility especially in rural areas and to enable leisure interests [9]. Yet other researchers promote finding new ways of reducing the use of natural gravel and conclude that the grain size of the gravel affects the loss of gravel and gravel can be saved by reusing the gravel that went down into the road ditch [10]. Unfortunately, this is a rather time-consuming process. It is unusual today for gravel to be recycled in Sweden [3]. To develop sustainable working methods for gravel recycling can, according to [11], provide environmental as well as financial benefits, because maintenance work is costly. However, a model for estimating environmental and economic effects during the maintenance activity of gravelling is currently lacking. More research is needed for understanding how gravel can be recycled, how to compare different available gravel recycling options, and how to estimate the economic and environmental effects of these options. Research regarding the estimation of environmental impact and costs of gravel recycling is, to high extent, missing today. It is e.g. unclear if gravel recycling, which is positive from an environmental point of view, is a more cost-effective option than using non-recycled gravel. The purpose of this study is to gain deeper understanding of the environmental and economic effects of raw material choices for the maintenance of gravel roads, i.e. the effects of using new or recycled gravel as material during the gravelling of the road. The goal is to develop a calculation model for the estimation of environmental and economic effects during gravel road maintenance, taking into consideration the choice of raw material source.

2 Maintenance of Gravel Roads, a Brief Introduction

Gravel roads are built up in layers, with the protective layer found closest to the ground and the wearing layer (or course) on the top. The protective layer prevents material from below to be pushed up into the roadway [12], while the wearing layer makes the road smooth, so it becomes safe and drivable [13]. Between these two layers the reinforcement and the bearing layer are found. The reinforcement layer helps to divert water from the road while the bearing layer distributes the load over the road surface, this to prevent deformation. The technical life of a gravel road is estimated to be 40 years for Swedish conditions [14]. The gravel road condition is affected by the thickness of the wearing layer and its composition. Inspections of the wearing course is made manually according to predefined assessment criteria and procedure [15]. General maintenance activities mainly consist of planning, gravelling and edge cutting [3]. The road needs to be shaped and leveled on a continuous basis, which is done by planning using a grader. It is recommended that the wearing layer is 7 cm, otherwise there is a risk of stones breaking through the bearing layer. If the course is too thin, the road must be graveled, which is done approximately every three years. Normally 14–25 m^3 of gravel is added per kilometer. Equipment required for gravelling is a truck and a spreader blade. Edge cutting is the activity of improving the road shape by pulling edge material into the roadway, which creates a slope required for better water runoff. This can be done with a grader.

There are few existing methods for gravel recycling, such as the Mähler method, the gravel recycler Saga, or to use a grid vibrator bucket. In all three methods, edges are first cut with a grader for the purpose of reusing discarded road material, and the drawn road material is put in a string about a meter from the roadside [16]. Sometimes, if the correct mix of gravel fractions are missing, supplementary gravel is added. A novel method for gravel recycling, the gravel recycler Rolf, is developed in the research project *Sustainable maintenance of gravel roads* [17]. However, this is still undergoing testing and is not available on the market.

3 Assessing the Economic and Environmental Effects of Gravel Recycling

A calculation model has been developed for being able to estimate the economic and environmental impact of different alternative approaches for gravelling; either adding new gravel to the road, or cycling gravel using one of the mentioned recycling methods. The model supports the calculation of the total cost and the environmental impact arising from the use of new and recycled gravel, respectively and is visualized in Fig. 1. The calculation model will be evaluated in the form of test scenarios, see Sect. 3.1 for the scenario descriptions and Sect. 4 for results. In [18] a model to calculate the cost of gravel when new gravel is used is available. The model is developed by practitioners and researchers within the area of wood and forest science (Skogforsk, LRF Skogsägarna and the Swedish Forest Agency), and is therefore seen as a reliable source. Information required for the calculations are the length of the road, traffic intensity, and the climate zone of the road. Traffic intensity is divided into the categories low (annual daily average 0–24 vehicles), medium (annual daily average 25–124 vehicles), and high (annual daily average <124 vehicles). The climate zones are divided into southern Sweden, central Sweden, and northern Sweden. The cost of recycling gravel is between 7–12 Swedish krona (SEK) per meter for one grader with a gravel recycling machine such as Mähler [16]. During gravel road maintenance costs for e.g. gravel and for the machinery required for execution arise [3]. The costs can be divided into direct and indirect costs [19]. A direct cost is attributable to a specific calculation object while an indirect cost arises from general activities shared by several objects, i.e. not attributable to a specific calculation object. Direct costs for the maintenance activity of gravelling are material, personnel, transport, and fuel costs. For calculating the costs, only direct costs are included in the form of transport, material, and personnel costs.

The Network for Transport (NTM) method will be used for calculating the environmental impact. The NTM method is an approach for environmental calculation of vehicle transport. Using the NTM method requires knowledge of the vehicle's load capacity to calculate the degree of filling, traffic work and fuel consumption. Allocation means that the environmental impact is distributed e.g. if a transport of two goods is shared environmental impact needs to be divided among the products. Environmental impact is calculated by multiplying the vehicle's fuel consumption by vehicle distance [20]. Estimating a systems' total life cycle environmental impact is a complex task, and therefore delimitations should be made [20]. A natural delimitation is to include only the activities and parameters that are directly affected by the object of study. In this case, only direct transportation activities and material use are included, i.e. impact connected with operations, while impact that arise from vehicle manufacturing and maintenance, as well as gravel extraction, are excluded. The environmental impact is calculated as the fuel consumption required for the execution of the specific maintenance activity. Transportation of vehicles to and from the gravel road is not considered in the study because the test is limited to direct costs and environmental impact to obtain the gravel required to carry out the gravel maintenance activity. Moreover, as all scenarios require a grader the grader cost and fuel consumption are excluded in the test. The transport cost and environmental impact are calculated based on the distance traveled. When calculating the mileage for the transport of new gravel maximum load level of the truck is used. Transportation costs are by Swedish Petroleum Institute's annual average price for diesel (8,21 SEK per liter without tax) multiplied by the fuel consumption for the driven distance. The personnel cost is calculated based on the time required to carry out the activity. Costs derived during the use stage of the item are often higher than the investment costs, and therefore it is advisable to make a total cost analysis to get an overview of the actual cost [21]. A total cost analysis reflects the total cost during an assets' or product's lifetime, including the purchase price and sub-costs such as warehousing, transport, materials, quality control and maintenance costs. To achieve a fair comparison of different alternatives the lifetime cost of 40 years is calculated, taking into account supplementary gravelling.

3.1 Test Scenarios

The calculation model is evaluated using three test scenarios (see Table 1); one based on using new gravel, while two are based on recycling gravel.

- Scenario one includes a grader and a truck with a trailer. This requires two persons in total: one person who operates the grader and cuts the road and one person who transports new gravel and add it on the road. Two travel distances occur in this scenario: the distance travelled while conducting maintenance work equal to the length of the maintained road, and the material transport distance between the quarry and gravel road that undergoes maintenance.

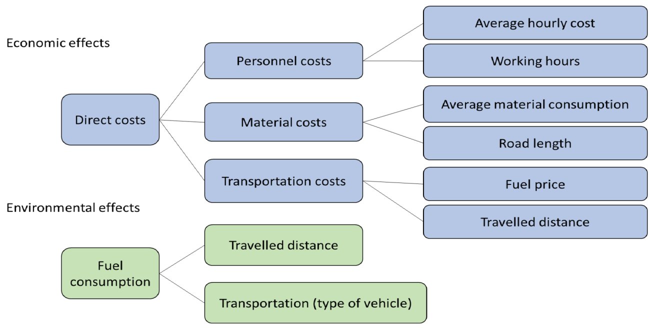


Fig. 1. Conceptual model

Table 1. Test scenarios

Scenario	Vehicles	Personnel	Material	Distance
1. No gravel recycling	Grader, truck with a trailer	2 persons	New gravel is added, average 28.7 tons/km	Material transport and maintenance work
2. Gravel recycling with Saga	Grader, wheel loader pulling Saga	2 persons	No new gravel is added	Maintenance work
3. Gravel recycling with Rolf	Grader pulling Rolf	1 person	No new gravel is added	Maintenance work

- Scenario two uses a grader and a wheel loader. It takes one person driving the grader and one person driving the wheel loader pulling the gravel recycler denoted "Saga". The distance travelled while conducting maintenance work, equal to the length of the maintained road, is taken into consideration in this scenario. The gravel recycler can only manage one side of the road at a time, which means that the distance traveled must be doubled.
- The third scenario includes the gravel recycler denoted "Rolf". Rolf is connected to a grader and accounts as one vehicle, which means that only one person is required. Like Saga, Rolf can only manage one side of the road at a time, which means that the distance traveled must be doubled. Unwanted materials such as rocks and tufts of grass are sorted out and discarded by a conveyor belt located on the side of the machine. The unwanted material ends up in the ditch edge while desired material, i.e., gravel, ends up on the road again.

The cost and fuel consumption of the grader are excluded from the calculations presented in Sect. 3.2 due to following reasons: a) all scenarios include the use of a grader, b) the grader relates to the activity of "planing", while the assessment is focused on the activity of "gravelling". Even though both activities in practice are tightly interconnected and carried out in coordination, the calculation model is focused on the effects of recycling versus non-recycling of gravel.

3.2 Input Data

Empirical data have been collected through interviews and company documents. An interview was conducted via Zoom with the CEO of a contractor that provides gravel road maintenance services, and a telephone interview was held with one of the constructors of the gravel recycling machine Rolf. Both interviewees have also provided documentation of technical and administrative nature. In case empirical data have been missing, assumptions or averages have been used based on the experiences of the interviewees. In addition, distances have been extracted based on commercial map applications. The costs are based on information from the maintenance contractor. According to the interviewee, the personnel cost is in average 460 SEK/h, and material cost for gravel is approximately 77 SEK/ton. Fuel consumption/km for a truck with a trailer with maximum load of 35 tons has been estimated to 0,6 L/km. This value is based on a daily environmental report, with the average distance of 445,6 km and an average fuel consumption of 276,2 L, and the average fuel cost of 8,21 SEK per liter without tax. To obtain the amount of gravel required, an average value is calculated using data from an agreement between the contractor and the major customer, which is a municipality.

The agreement contains information about the road number, length, and estimated amount of gravel in tons required for gravelling for 15 different roads. The material consumption for the 15 roads as well as the average consumption is presented in Table 2. In the test scenario an average length of the gravel road of 3.167 km (3,2 km rounded up) is used which corresponds with the average length of the 15 roads according to Table 2.

The average distance between the gravel extraction location and the gravel road to be maintained is estimated using three different map applications: Apple map, Hitta, and Eniro. A place close to the gravel road (see Table 3) was chosen as the destination point while the starting destination was a gravel pit located in Urshult. The distance and driving time for each of the 15 roads were extracted from each of the three map applications, and thereafter the average distance and total time was calculated. The average distance according to *Apple map* is 153,6 km, and the total driving time is 234 min. For *Hitta*, the average distance is 163 km and the total driving time is 209 min. Using *Eniro*, the average distance is 200,6 m and the total time is 264 min. The average total distance of 172,4 km was gained taking the average distance from the three map applications, which corresponds to an average of 11,5 km. *The average distance of 11,5 km between gravel road and gravel extraction will therefore be used in the test scenarios.* Similarly, the time required was calculated between gravel extraction and gravel road, which gave an average time consumption of 0,26 h. Based on the calculated distance and time required, an average driving speed was obtained of 44,2 km/h. When recycled gravel is used, only personnel and fuel costs arise. Personnel costs are calculated by multiplying the time required for performing the maintenance action by the hourly cost. According to the constructor, the average speed of gravel recycler Saga is 1,4 km/h, while the average speed of Rolf is 2 km/h. The fuel cost is calculated by multiplying the fuel consumption by the price of fuel. The fuel consumption of the grader is estimated to 0,5 l/km. For Saga, the average fuel consumption is 1,1 l/km and for Rolf, the average fuel consumption is 0,5 l/km.

Table 2. Material costs

Road number	Distance (km)	Material consumption (tons)	Material consumption per unit of length
17529	6,960	246	28,3
17530	5,625	200	28,1
17582,2	4,220	147	28,7
17746	4,434	160	27,7
17794	3,520	119	29,6
17806	2,000	68	29,4
17806,2	3,410	131	26,0
17916	1,530	51	30,0
18020	2,750	93	29,6
18030	1,900	66	28,8
18165	3,500	127	27,6
18205	2,100	73	28,8
18226	1,300	40	32,5
18327	1,600	58	27,6
17661,2	2,662	94	28,3
Average	3,167	111,53	28,7

4 Results

Table 3 accounts for the calculations made for scenario 1, 2 and 3. Scenario 1 includes material costs as new gravel is used when gravelling while scenarios 2 and 3 only include personnel and transportation costs. Moreover, calculation of transportation costs differs depending on whether new gravel is added or not. In scenario 1 the distance driven, and the material tonnage are taken into account while fuel costs are calculated for scenarios 2 and 3.

The highest cost for scenario 1 is material costs, which constitutes about 2/3 of the total cost. Transport costs for scenario 1 accounts for around 25% of the total costs, while the transportation costs are 21% and 15% respectively for scenario 2 and 3, where the largest cost factor is personnel costs. According to the calculations, the total cost for scenario 1 is 8 169 SEK, 2 694 SEK for scenario 2, and 1 735 SEK for scenario 3. When comparing the fuel consumption for the different alternatives, 43,3 L of diesel is required for scenario 1, 70 L for scenario 2, and 32 L for scenario 3. According to the test, it is most cost-efficient and environmentally friendly to use the gravel recycler Rolf. However, to be able to make a fair comparison, the total life cycle impact should be assessed.

The technical life of a gravel road is estimated to be 40 years for Swedish conditions [14]. According to [3] gravelling should be done every three years, which means 13

Table 3. Calculations for scenario 1 to 3

Cost factor	Formula	Calculation
Scenario 1		
Amount of new gravel	Length of the road (km) × material consumption (ton/km)	3,2 × 28,7 = 91,8 tons
Number of distances to drive	Total amount of material (tons) / maximum filling weight (tons)	91,8 / 35 = 2,6 (3 distances)
Total distance	Number of distances × distance (km) × 2[1]	2(3 × 11,5) = 69 km
Total time	Total distance (km) / speed (km/h)	(69 + 3,2) / 44,2 = 1,6 h
Transportation cost	Fuel consumption (l) × fuel cost (SEK)	8,21 × 43,3 = 355 SEK
Material cost	Average material cost (SEK) × Average material consumption (tons)	77 × 91,8 = 7 069 SEK
Personnel cost	Average hourly cost (SEK) × driving time (h)	460 × 1,6 = 745 SEK
Total cost	Transportation cost + material cost + personnel cost	355 + 7 069 + 745 = 8 169 SEK
Fuel consumption	Fuel consumption per distance (l/km) × (Distance road to quarry + length of the road) (km)	0,6 × (69 + 3,2) = 43,3 l
Scenario 2		
Total time	Road length (km) / speed (km/h)	6,4 / 1,4 = 4,6 h
Transportation cost	Fuel consumption (l) × fuel cost (SEK)	8,21 × 70,4 = 578 SEK
Personnel cost	Average hourly cost (SEK) × driving time (h)	460 × 1,4 = 2 116 SEK
Total cost	Transportation cost + personnel cost	578 + 2 116 = 2 694 SEK
Fuel consumption	Fuel consumption per distance (l/km) × length of the road (km)	11 × 2(3,2) = 70,4 l
Scenario 3		
Total time	Road length (km) / speed (km/h)	6,4 / 2 = 3,2 h
Transportation cost	Fuel consumption (l) × fuel cost (SEK)	8,21 × 32 = 263 SEK
Personnel cost	Average hourly cost (SEK) × driving time (h)	460 × 3,2 = 1 472 SEK
Total cost	Transportation cost + personnel cost	263 + 1472 = 1 735 SEK
Fuel consumption	Fuel consumption per distance (l/km) × length of the road (km)	5 × 2(3,2) = 32 l

1 The total length is multiplied by two as both sides of the road must be maintained

times over a 40-year period. Recycled gravel is gradually broken down, which affects the composition of the course. Supplementary gravel must therefore be added regularly approximately every third time gravelling is carried out in the case of gravel recycling. As the required amount of new gravel is unknown, and, in addition, varying depending on factors such as road structure and composition, the assumed estimation of 25% new gravel and 75% recycled gravel are used in the test scenarios, based on practitioners' experience. The lifecycle cost is calculated with the same assumptions as in the test scenarios. The only difference is that supplementary gravel is taken into account. The lifecycle cost of 40 years is 106 197 SEK for scenario 1, 45 694 SEK for scenario 2, and 33 227 SEK for scenario 3, see Fig. 2. The fuel consumption of 40 years is 563 L for scenario 1, 965 L for scenario 2, and 471 L for scenario 3.

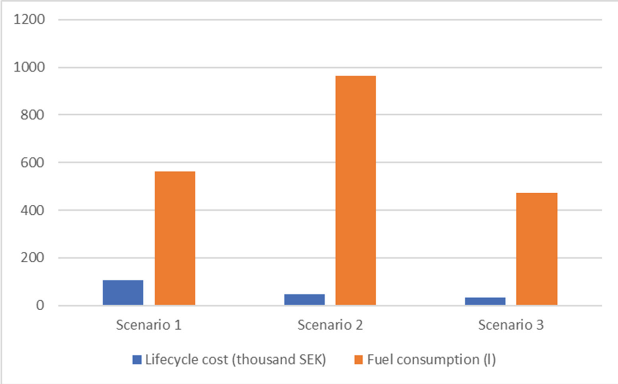


Fig. 2. Lifecycle cost and fuel consumption

5 Discussion and Conclusions

There are several factors that affect the result of the model test. It is not optimal to calculate the fuel consumption based on distance traveled because the fuel consumption is affected e.g. of the terrain in which the vehicle is driven. Slopes and high speeds for instance require more fuel than a straight road. Was used in the test, but the distances differed quite a lot between applications. Real milage and fuel consumption of the vehicle should be used for gaining higher accuracy. Moreover, it is a bit misleading to exclude the cost and environmental impact of the grader in the test scenario 3. In scenario 1 and 2, an additional vehicle and one more person is required, while when using the gravel recycler according to scenario 3, only additional fuel is required. The model should therefore include both the gravelling and planning into one activity, and not just focus on graveling. The model was delimited with respect to variables reflecting direct impact on the economy and environment. If a full LCC and LCA on the maintenance activity of gravelling had been performed, a greater understanding of which aspects affect cost and environmental impact might have been identified. The possibility to expand the model with more variables is therefore a suggested area of future research. The calculation model could for instance be extended with investment costs for machinery and vehicles as well as their maintenance costs. From an environmental point of view, the model could consider the environmental impact when producing natural gravel and the vehicle's environmental impact. In the model, the environmental impact is calculated based on the fuel consumption. The fuel consumption could be recalculated to e.g. carbon dioxide emissions, or carbon dioxide equivalents, for reflecting the total environmental impact in a better way. In this study, calculations on the full technical life length of a gravel road was performed. Calculating the cost and fuel consumption over 40 years and taking into account supplementary gravel creates more accurate results than focusing on single activities or outcomes in shorter term. The disadvantage is that, due to limited scientific and practical data, there is no exact answer to how often supplementary gravel need to be added. If supplementary graveling occurs every fourth year instead of every third year, the cost of recycling will consequently decrease. Long term studies on the gravel degradation when using recycled gravel is therefore suggested.

The purpose of the study was to create an understanding of the environmental and economic effects of using new and recycled gravel during gravel road maintenance. The results of the study show that environmental and economic effects can be estimated based on the model developed in this paper. The model was tested using a test scenario where three different options for the maintenance activity of gravelling were evaluated. Usually, the most environmentally friendly alternative is the more expensive, but the results show that alternative three, which involves the use of the gravel recycler Rolf, was best both from an economic and environmental perspective. The result for fuel consumption in the test scenario shows that it is more environmentally friendly to use recycled gravel than to use new gravel. Even when the total cost and fuel consumption is calculated with supplementary gravel, alternative three was most advantageous. This shows that it does not necessarily have to be more expensive to choose the most environmentally friendly alternative. The test scenario also shows that the cost and environmental impact differ depending on chosen maintenance method. Therefore, using a calculation model is advantageous, as it can help maintenance contractors to assess whether it is appropriate to use recycled gravel or not. The study showed that recycling can be both economically and environmentally beneficial, which might lead to an increased interest in recycling gravel and to the development of new gravel recycling methods. A lower consumption of natural resources is advantageous because it is an energy-intensive process to produce new gravel. To ensure which alternative is best from an environmental and economic point of view, the study should be repeated and the possibilities of extending the model should be investigated.

Acknowledgments. The research has been conducted as part of the project Sustainable maintenance of gravel roads funded by the Kamprad Family Foundation. The project develops new methods and technologies for gravel road maintenance. The authors are grateful for the valuable support from RJ Schakt and Transport AB and Bryne AB.

References

1. Swedish Transport Administration (2017) Sveriges vägnät. https://www.trafikverket.se/resa-och-trafik/vag/Sveriges-vagnat/. Accessed 26 Jan 2020
2. Swedish Transport Administration (2019) Så sköter vi grusvägar. https://www.trafikverket.se/resa-och-trafik/underhall-av-vag-ochjarnvag/Sa-skoter-vi-vagar/Grusvagar/. Accessed 06 Feb 2020
3. Edvardsson K (2015) Mer grus under maskineriet: handbok för tillståndsbedömning och underhåll av grusvägar. Sveriges kommuner och landsting, Stockholm
4. Swedish Transport Administration (2014) Ny beräkningsmodell för årligt driftbidrag till enskilda vägar. https://www.trafikverket.se/contentassets/432a6b0ba0014eaf93256e7df8ed05f5/enskilda_vagar_informationsblad.pdf. Accessed 13 Mar 2020
5. Swedish Transport Administration (2019) Barmarksunderhåll på grusväg. https://www.trafikverket.se/resa-och-trafik/vag/Enskilda-vagar/Underhallav-enskild-vag/Barmarksunderhall-pa-grusvag/. Accessed 26 Jan 2020
6. Nationalencyklopedin (n.d.) Grustäkt. http://www.ne.se/uppslagsverk/encyklopedi/lång/grustäkt. Accessed 26 Jan 2020
7. Carpenter AC, Gardner KH, Fopiano J, Benson CH, Edil TB (2007) Life cycle based risk assessment of recycled materials in roadways. Waste Manage 27(10):1458–1464

8. Smadi A, Hough J, Schulz L, Birst S (2009) North Dakota gravel road management: alternative strategies. Trans Res Rec J Trans Res Board 1652(1):16–22
9. Marinello F, Rosario Proto A, Zimbalatti G, Pezzuolo A, Cavalli R, Grigolato S (2017) Determination of forest road surface roughness by Kinect depth imaging. Ann Forest Res 60(2):217–226
10. Berthelot C, Carpentier A (2003) Gravel loss characterization and innovative preservation treatments of gravel roads: Saskatchewan, Canada. Trans Res Rec J Trans Res Board 1819(1):180–184
11. Ullberg J (2006) Riktlinjer för ett miljöanpassat arbete. http://www.roadex.org/wpcontent/uploads/2014/01/Environmental_Swedish.pdf. Accessed 06 Nov 2019
12. Johansson K (2005) Grusvägars uppbyggnad, nedbrytning och underhåll – En litteraturstudie med inriktning mot tillämplig av planeringsverktyget HDM-4. Research report. Luleå: Luleå Technical University
13. Tholén O, Wiman GL (2018) Asfaltsboken, Vägens uppbyggnad. http://www.asfaltboken.se/vagens-uppbyggnad/. Accessed 25 Feb 2020
14. Alzubaidi H (1999) Operation and Maintenance of Gravel Roads - A Literature Study, vol 852A. VTI meddelande
15. Swedish Transport Agency (2014) Bedömning av grusväglag. TDOK 2014:0135 Version 1.0
16. Skogskunskap (2016) Grusåtervinning. https://www.skogskunskap.se/vagar-i-skogen/drift-och-underhall/sommarunderhall/grusatervinning/. Accessed 16 May 2020
17. Linnaeus University (2018) Project: Sustainable maintenance of gravel road. https://lnu.se/en/research/searchresearch/research-projects/project-sustainable-maintenance-of-gravel-road/. Accessed 01 May 2021
18. Skogskunskap (2016) Kostnad och prestationer vid vägarbete. https://www.skogskunskap.se/vagar-iskogen/vagnatsplanering/projektering-och-konstruktion/kostnader-ochprestationer-vid-vagarbete/. Accessed 05 June 2020
19. Lantz B, Isaksson A, Löfsten H (2014) Industriell ekonomi: grundläggande ekonomisk analys, 1st edn. Studentlitteratur, Lund
20. Björklund M (2018) Hållbara logistiksystem, 2d edn. Studentlitteratur, Lund
21. Jonsson P (2008) Logistics and Supply Chain Management. McGraw-Hill, Maidenhead

Revisiting Agricultural Technologies in the 4IR Era

Nonceba Ntoyanto-Tyatyantsi(✉) and Anthea Amadi-Echendu

Preller Street, Muckleneuk, Pretoria, South Africa
ntoyan@unisa.ac.za

Abstract. Food security is becoming a growing problem worldwide. Much focus should therefore be placed on the agricultural sector with a view of equipping the sector for increased food production capabilities. The agricultural sector is changing rapidly globally due to the fourth industrial revolution (4IR) mega technologies, resulting in smarter ways to farm. These technologies allow farmers to maximise production remotely while controlling every aspect of crop farming such as pest control, soil conditions, crop monitoring, and soil moisture. These advances will allow farmers to be more profitable, efficient and environmentally friendly. However, evidence suggests that small-scale farmers are left behind in the use of 4IR mega technologies in South Africa.

The purpose of this paper is to highlight the use of various 4IR technologies in the agricultural sector by using a desktop review of current literature. The paper recommends for a government-driven entity to be established that will focus on building capacity for small-scale farmers to build more sustainable and bigger businesses to assist in increased food production through the introduction of 4IR technologies.

1 Introduction

Food security is a global phenomenon that also forms part of the 2030 Agenda for Sustainable Development in Sustainable Development Goals. Failure to grow enough food may lead to food insecurity. Food may need to be imported, thus inflating food prices which may further lead to poverty as households may not be able to afford the higher food prices. According to the 2001 policy brief of the Food and Agriculture Organisation (FAO) of the United Nations, "food security exists when all people, at all times, have physical and economic access to sufficient safe and nutritious food that meet their dietary needs and foods preferences for an active and healthy life" Policy Brief [1].

Du Toit (2011) clarifies the meaning of food security at three levels:

1) **At national level:** food security "refers to the condition whereby the nation is able to manufacture, import, retain and sustain food needed to support its population with minimum per capita nutritional standards".
2) **At community level:** food security means "the condition whereby the residents in a community can obtain safe, culturally accepted, nutritionally adequate diets through sustainable systems that maximises community self-reliance".

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 90–100, 2022.
https://doi.org/10.1007/978-3-030-96794-9_9

3) **At household level:** "food security refers to the availability of food in one's home which one has access to and when the member of the family does not live in hunger or fear of starvation".

United Nations (UN) organisations forecast population growth to be between 8.3 and 10.9 billion people by 2050. Such growth rates will require a 50%–75% increase in food supply. At the same time, the needs of developing countries will double, including a 60% growth of rice-consuming countries and Sub-Saharan African growth by 250% Prosekov and Ivanova [3]. According to the UN data, agricultural production will have to grow with 75% by 2050 Godfray et al. [4]. Modern technologies may assist to increase food production through enhanced fertility of land, using organic ocean resources, switching to renewable energy sources, advances in genetics, and breeding more productive breeds of animals Prosekov and Ivanova [3].

In addition, many people have adopted a healthy eating lifestyle whereby emphasis is placed on healthy food habits, public health, quality of food, environmental protection and empowerment of communities. The role of local food production has thus become more important. Aquaponics, gardening and sustainable customised agriculture are emerging as practices in cities. Digital platforms provide new marketplaces to match supply and demand. In addition, a diversified range of alternative food production and supply chains has emerged. Considering the anticipated food shortages, it stands to reason that strategies need to be pursued now to mitigate these shortages. This paper will explore various 4IR technologies that may be used within the agricultural sector to increase food production. It further will look at SMEs as a vehicle to increase food production sources. This paper will focus on technologies that can be introduced as assets into subsistence farming practices to enable small-scale farmers to increase their business yields. In addition, the circular economy promotes greater land productivity, reduction of waste in the food value chain, improvement of soil nutrients to increase the value of land and the indorsement of soil as assets.

2 The Ecological Imperative

Carbon dioxide produced by human activities is one of the main greenhouse gases (GHG) responsible for climate change. Managing the carbon emissions has become the biggest social, economic and political challenge in recent years. In December 2015, the world's nations agreed in Paris to a goal of "holding the increase in the global average temperature to well below 2 °C above preindustrial levels and pursuing efforts to limit the temperature increase to 1.5 °C above preindustrial levels" UNFCCC [5]. This requires public and private entities' pursuance to curb greenhouse gas emissions from all sectors. The global food system is responsible for ~21–37% of annual emissions Mbow et al. [6].

Agriculture is particularly susceptible to changes in climate. South Africa is a dry country and farming is highly dependent on weather patterns. Ofoegbu et al. [7] opine that, for South African small-scale farmers, the changes in climate are resulting in higher temperatures, unpredictable rainfall patterns and droughts becoming more frequent. That, coupled with the fact that South Africa is already a water scarce country, means that planning and practicing small scale agriculture is becoming more tenuous. It is therefore imperative to review strategies to optimise agricultural processes and the optimisation of resources. The circular economy is one such strategy that has been gaining support across various industries.

The circular economy (CE) entails the end-of-life management of a product whereby a product's functional life is extended through reintroduction into the supply chain as non-waste Govindan and Soleimani [8]. No production in a circular economy is wasted, thus advancing from the current linear 'take-make-dispose' economy. A CE is planned and designed to be a regenerative industrial system and restores used products, uses renewable energy, removes toxic chemicals and aims for the eradication of waste by using environmentally friendly designs, resources systems and business models. CE promotes reuse, remanufacturing, rebuilding, recovery, restoration and recycling. The agricultural sector is reliant on natural resources and cycles as primary inputs into crop and livestock production. Regenerative agricultural systems like crop rotation, minimum till and cover cropping may be used to preserve the sustainability of the natural systems that support agricultural production Shrestha et al. [9]. Using resources efficiently and the reuse of consumer products may encourage more sustainable agricultural business models.

3 Technologies

Droughts causes water scarcity that have a direct impact on the ecosystem, livestock, wildlife and humanity, which in turn negatively affect food production. Factory farming or automated indoor farming has gained prominence, particularly in areas that have water scarcity. Produce is grown in vertical stacks under LED lights. Plants are grown in a water-based solution instead of soil through a hydroponic indoor growing system that allows for vegetables and fruit to be grown throughout the year Firbank et al. [10]. In Pretoria, South Africa, CAN-Agri is a business that provides a vertical farming solution to produce leafy greens in a controlled environment (www.can-agri.com). Kranji farms in Singapore grow vegetables on 9 m-tall towers made of tiers of planting troughs that rotate around an aluminium frame that caters for 10% of the vegetable market in Singapore. Valcent products uses multi-level rotatable, stacked plastic trays in a climate-controlled glasshouse enclosure to produce tomatoes that are exported globally. The vegetables require only 8% of the water required by traditional farming. Mirai in Japan produces 10,000 heads of lettuce per day which is 100 times more per square foot than traditional farming; produced with 40% less energy, 80% less food waste, and 99% less water usage than traditional outdoor farming Benke and Tomkins [11].

Automation of outdoor farming is also advancing well. Artificial intelligence systems and machines perform tasks like raising seedlings, replanting and harvesting Bannerjee et al. [12]; Han et al. [13]. Precision farming uses information and technologies (GPS, satellite imagery, control systems, sensors, robots, variable rate technology, telematics, software, etc.) to improve crops and increase yields, reduce harvest times, and reduce costs and environmental impact by focusing on soil preparation, seeding, and harvesting. FruitLook is an information technology that is used by farmers in the Western Cape in South Africa that helps deciduous fruit and grape farmers to be water efficient and climate-smart. The Chameleon and Wetting Front Detector Sensors have enabled small scale farmers in Mozambique, Tanzania, and Zimbabwe to cut down irrigation frequency fifty times and double productivity Ncube et al. [14].

Speed of driverless machines and tractors can be adjusted through cloud-based platforms to fertilise, plow or harvest fields, while locations are monitored Baskerville [15]. Agrorobots can be used to identify and kill weeds in targeted areas by using cameras and sensors, thus reducing the use of pesticides where they are not needed. In addition to cost reductions, a reduction of the environmental footprint of the farm can be achieved. Replacing tasks that people previously performed implies that people may eventually not be needed to grow crops at all. All functions in future will be controlled by a central artificial intelligence mechanism Ponnambalam et al. [16].

DNA testing of crops can be invaluable to adjusting herbicides and other products used in agricultural processes. GPS locations can be stored in the cloud to provide historic information of how crops change over time in specific locations. The farmer's attention could be centred to areas where specific interventions are needed instead of applying techniques and products to the entire crop. The result may yield bigger crops (bigger quantities), less chemicals being used, and better harvests (enhanced quality) at a reduced overall cost Peace [17].

Food Circles Network is a new way of organising food systems and agriculture that links the different role-players of food production (www.showmefood.org). It is concerned with the safe production and consumption of food. Rural farmers are promoted in the region where their farms are situated. These community-supported agricultural programmes may counter many criticisms raised over food security, food safety, health, as well as ecological and social sustainability associated with the industrialisation of food chains. In addition, community gardening can also help to connect food producers and food consumers in the same geographic regions. The result could build more self-reliant and resilient food networks whereby food is grown on unused land that are not being used Fox-Kämper et al. [18]. Permaculture (permanent agriculture) revisits traditional agricultural methods of not ploughing soil to retain moist, rotating cultures and using complementary crops in order to replenish depleted soils. Small scale farming is potentially more intense and productive as it is presumed that smaller farms produce more food per area. Nutrient recovery (nitrogen and phosphorus) in wastewater can be converted into fertiliser that can be used for agriculture and feedstock Krebs and Bach [19].

Blockchain technologies may in turn be used to link the agricultural supply chain for easier traceability of food to provide for shorter delivery time to market. It is important to be able to trace the origin of food products. Regulators should be able to identify the source and extent of food contamination for example Aldag and Eker [20]. It can also be used to link global networks to enable peer-to-peer transactions in self-organising communities in the economy. A peer-to-peer operating platform can replace complex contracting models as no mediating financial institution needs to be involved. Blockchain also improves transparency as information cannot be changed once it has been entered into the system and it can be viewed by the entire value chain. Private blockchains can be put together for a ringfenced supply chain. A blockchain based supply chain traceability system using RFID and Blockchain technology was proposed to ensure the traceability of wine from the grape to the bottle distribution Biswas et al. [21]. AgriDigital is an Australian agricultural commodity management platform provider that uses blockchain technology in a verification system to record data such as pharmaceuticals administered to animals Rogerson and Parry [22].

Technologies that reduce carbon emissions by providing sustainable energy supplies, storage and demand solutions, as well as waste reduction and the preservation of clean water are preferred. The wireless sensor network nodes use batteries that consume limited energy within days, depending on the duty cycle of the application. Renewable energy sources like solar photovoltaic energy can increase the sensor's lifespan. Harvesting energy by converting energy from the environment into usable electricity by drawing energy from the sun, the wind, solar, geothermal heat, hydropower, biomass, waste burning safe nuclear power and water (oceans tides) has become a major focus area globally Sharma et al. [23]. The use of data is a core interface between social and technological innovations. Big data analytics helps make sense of vast amounts of data and provide deeper insights into production processes, quality and risk management Lohmer and Lasch [24]. Using the analysis of data, farmers can predict the amount of rainfall that might occur in a given season and even the incidence of flooding. This will assist the farmers in terms of planning and deciding what to grow and when to grow it. In Southern Africa the weather patterns have become unpredictable and with descriptive and predictive analytics farmers can determine the optimal solution Chen et al. [25].

Drones are used for the spraying of pesticides on crops as well as crop monitoring and this has allowed for a larger surface area to be covered than the traditional knapsack sprayer would have covered Veroustraete [26]. In irrigation, drones with hyper-spectral, multispectral, or thermal sensors can recognise parts of a field that are dry or need improvements. They are also able to calculate the vegetation index, which describes the comparative density and health of the crop and show the heat signature Ahirwar et al. [27].

4 The 4IR Technologies and Sustainable Development

Climate change affects agricultural yield, which may interfere with food supply chains and progress towards a zero-hunger future. Climate action shows that progress by AI will encourage knowledge and modelling of potential weather and climate prediction and response analytics. Precision agriculture will enable farmers to detect crop diseases early and enhance productivity and returns from rural poverty Silvestre and Ţîrcă [28]. Climate control systems and hydronics technologies help minimise water, land and nutrient usage in vertical farming. Technologies such as sensor, AI, robotics and drones reduce greenhouse gas emissions, create critical eco-systems, and promote nature and water conservation Mishra and Maheshwari [29]. Another main area in which sustainable development can be achieved through the fourth industrial revolution is clean energy and less carbon emissions. Energy for development and sustainable energy solutions are required for long-term sustainability. Big data platforms will enable earth management by monitoring carbon emissions and provide precision analytics for agricultural management Østergaard et al. [30]. Table 1 provides a summary of the 4IR technologies highlighted in this paper that may assist to enhance the agricultural sector.

Table 1. Summary of the 4IR technologies proposed to solve problems in agriculture

Problems in agriculture	Technologies proposed	Solution
Droughts that cause water scarcity	Climate control system, artificial lighting (LED's) and hydroponics	Automated indoor farming
Automation of outdoor farming	Artificial intelligence systems	Precision farming
High cost of labour and energy usage	Agrorobots	Reduction in cost of labour and of the environmental footprint
Lack of supply chain traceability	Blockchain technologies	Improves traceability and transparency
Carbon emissions	Wireless sensor network nodes	Sustainable energy efficiency
Unpredicted rainfalls	Big data analytics	Descriptive and predictive analytics assist in determining the optimal solution
Pesticides	Drone-based sprayers	Crop monitoring and management
Lack of agricultural space	Vertical farming	Using smaller spaces by leveraging height and not distance to grow crops

5 Mitigating Environmental Changes/Challenges

Circular economy as a resource, offers a solution to some of the most demanding issues in sustainable development, and has gained popularity. The circular economy seeks to redress waste pollution and overexploitation of mineral and other resources. Materials that are being used in processes are now used with their end-of-life possibilities in mind. Because of reuse and recycling capabilities, production is now less dependent on virgin raw materials. Recycling should be approached with caution as some materials become toxic if they were recycled too many times. Leakage of plastic in water must be prevented as it may affect the water resources used in agriculture which in turn may affect food production. Bioplastics that have wide use in food and beverage packaging, health care, textiles, agriculture, automotive or electronics need to be biodegradable to reduce the ecological footprint.

Undesirable changes in water systems, caused by climate change, are addressed with collaborative geo-engineering efforts by scientists, companies and private actors. The plastic and chemicals released into the oceans and water systems can cause much harm in the form of ocean pollution and acidification. New water-harvesting solutions provide water for agriculture or drinking in areas that suffer from low rainfall; and they help to mitigate floods Lazurko and Venema [31].

6 The Small-Scale Farmer/Entrepreneur

Small medium and micro enterprises (SMME's) play a significant role in the economies of developing countries and play a greater role in the socio-economic stimulation of South Africa. SMME's contribute around 50% to the national Gross Domestic Product (GDP) and believed to be employing up to 60% of the labour force Bruwer [32]. Small-scale farmers have the potential to create employment, provide for food security through increased food supply and contribute significantly to poverty alleviation Oluwatayo [33].

Challenges that small-scale farmers are facing include low profitability, problems in marketing their produce, poor infrastructure and a paucity of policies that support small-scale farmers. The small farmers are disadvantaged as they compete with big farmers who benefit from economies of scale and own a sizeable chunk of the value chain. Land sizes that are becoming smaller coupled with the fact that urbanisation is reducing agricultural land Ferris [34]. Reducing land sizes can however work to the benefit of small farmers. According to Krebs and Bach [19], small farms (1–2 ha) cultivate 12% and even smaller family farms (less than 1 ha) cultivate 72% of the word's agricultural land, and therefore secure nutrition for more of the world's population.

Chisasa [35] states that financial institutions consider a number of factors which include: security, size of the farm, the income of the household, the net worth of the family as well as the demographic characteristics when assessing credit applications. Chauke et al. [36] note that about 77% of the farmers use land as collateral for loans advanced to them. It stands to reason that farmers with a positive net worth are more likely to obtain loans. Small scale farmers are at a disadvantage because they do not own high value assets and are unable to secure finance necessary to grow their farming operations. They therefore mostly participate in subsistence farming.

7 The South African Agricultural Landscape

The South African agricultural landscape is one that is dominated by large scale farmers who are predominantly white. South Africa therefore has a de facto bimodal agricultural system which prior to 1994 was epitomised by the interference of the state in the sector through the giving of subsidies to commercial white farmers Lepheane [37]. There is a plethora of laws that were passed before 1994 that gave rise to the dual nature of South African agriculture which include inter alia the 1913 Natives Land Act, the Development Land and Trust Act 18 of 1936, Group Areas Act 36 of 1966, and the Black Homelands Citizenship Act 26 of 1970 Coles [38]. Land distribution in South Africa has created a scenario in which the rural areas are dominated by large scale farms which occupy about 88% of the available agricultural land. This has left only 12% of the land for the 72% of the rural population which is largely black without the necessary infrastructure needed for successful agricultural practices Ministry for Agriculture and Land Affairs South Africa [39].

Despite the intended consequences of policy instruments there are inconsistencies that have emerged. Gates [40] contends that small-scale farming in South Africa, as with rest of Southern Africa, is currently not sustainable given the extent of the damage caused to the environment. Unsustainable farming practices is further compounded by insufficient financial resources to acquire technology that will align practices, products and processes that will result in sufficient and profitable yields that preserves the environment for future generations Khwidzhili and Worth [41].

8 A Case Study Example from the Automotive Sector in South Africa

A few years ago, the automotive sector in South Africa faced a dilemma when manufacturers wanted to close their operations in South Africa. The closing of automotive manufacturing facilities would have resulted in a significant reduction in GDP and numerous job losses.

The Automotive Industry Development Centre (AIDC) was established to strengthen global competitiveness, stabilise jobs in the automotive sector as well as to enhance Gauteng Province as the automotive industry destination of choice. The AIDC has accelerated economic growth within the automotive industry through strategic partnerships with government, non-governmental agencies and industry leaders by focusing on skills development and training, enterprise development, incubation programmes, management of incentive programmes and facility management. Since their incorporation, vehicle manufacturing in South Africa has soared and infrastructure development has grown.

The AIDC established incubation programmes to support black-owned entrepreneurs during their business' start-up phases. These start-up businesses benefitted by receiving subsidised rental space within the incubation facilities, mentorship and training by professionals with business development and financial support. The resultant benefits include nurturing aspirant entrepreneurs into successful businesses; providing business support, mentoring and training to the incubatees, and identifying new opportunities for the SMMEs in the automotive sector to enhance local content for the sector.

9 The Way Forward

Although many initiatives have been formulated in the agricultural space to advance small scale farmers (Agricultural Black Economic Empowerment; restructuring of land reform), and to particularly assist black farmers, more can be done to assist with skills transfers and mentorship in addition to making money and land available to small-scale farmers. A similar intervention to what was done in the automotive space with the AIDC is required in the agricultural sector, whereby government works on a long-term plan to grow the agricultural infrastructure, upskill farmers and assist to equip farmers with technological advancements to enhance agricultural practices and yields. In addition, government can provide security for making funding available to small scale farmers as part of the programme so that modern technologies and much needed training can be financed.

Strategic partnerships can assist to make much needed technologies available to small-scale farmers to use in their businesses. Servitisation collaborations with key partners may be concluded whereby these partners can provide equipment and technologies as part of a service agreement to small scale farmers while they are being incubated. As part of the incubation process, these small-scale farmers can build their businesses to a point where they become self-reliant and independent within the agricultural economy. Warren [42] points out that farmers could invest in producing food products that target specific niche markets, transport and also opening up their farms to domestic tourism. For this to be effective there is need for farmers to have access to the internet and other technologies. Government and the private telecommunications companies can play a part by making sure that rural areas have access to fast and reliable internet. A study in Hungary concluded that, if there is proper access to technology, the gap between the small farmers and the big farmers will be reduced Boros-Papp and Várallyai [43].

There is also scope for investing in areas and technologies such as hydroponics and urban farming, but farmers need assistance with soil analysis, help with harvesting as well as post-harvest technologies and financing Venter [44]. The shift to commercial farming by the rural and small-scale farmers should be measured. Economies of scale could be achieved faster if the farmers work together as consortiums. For example, the waste from crop farming can be used to feed the animals while the waste from the animals can be used to generate biogas to produce vegetables. Given that most of the small-scale farmers are farming in areas that have poor soils, this could be a solution to minimising waste and making the soils more suitable for farming in order to yield larger crops. Lastly, it is crucial for small-scale farmers to be supported in their local communities. A collaborative approach among the farmers may allow for community needs to be met and for a more distributed approach to fulfilling their community's needs. Farmers can decide to diversify the products they grow and farm to avoid oversupply of certain products while there are shortages of others which will also enable all parties to benefit in the value chain. Currently, small-scale farmers compete with larger farming businesses to sell their products and end up incurring expenses in trying to sell their products that often are not prioritised at markets. The supply chain needs to be shortened to make the products less expensive to the end users and allowing fresher products to reach consumers earlier on in the process. This can be achieved by making products available within the community.

10 Conclusion

This paper reviewed literature that pertains to the agricultural sector with specific focus on the South African agricultural environment. It is a known fact that the anticipated food insecurity necessitates the implementation of technologies and mechanisms to enhance food production and yields. It is also important to grow food in a sustainable manner that does not cause harm to the environment. Consumers have also become more health conscious and as such also demand healthier food options. Although policies have been enhanced to incorporate small scale black farmers, changes are not occurring at a fast-enough pace, and the value of the programmes put into place are therefore diminished. There is a need for a more structured, long-term, and large-scale project to be put into place to drive growth in the agricultural system, with particular assistance and support to be provided for small scale famers. By planning towards infrastructural growth in agriculture, Africa and in particular South Africa, can be placed on a trajectory to supply food to other continents by 2050 and beyond.

References

1. Policy Brief (2006) Food security. FAO Agricultural and Development Economics Division. Available at: http://www.fao.org/es/esal, accessed on 24 October 2019
2. Du Toit DC (2011) Food security. Discussion notes on food security. Available at: http://www.nda.agric.za/docs/genreports/foodsecurity.pdf, accessed on 28 October 2019
3. Prosekov AY, Ivanova SA (2018) Food security: The challenge of the present. Geoforum 91:73–77
4. Godfray HCJ et al (2010) Food security: the challenge of feeding 9 billion people. Science 327:812–818
5. UNFCCC (2015) Paris Agreement. Available at: https://unfccc.int/process-and-meetings/the-paris-agreement/the-paris-agreement, accessed on 20 May 2021
6. Mbow HOP, Reisinger A, Canadell J, O'Brien P (2017) Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems (SR2). Ginevra IPCC
7. Ofoegbu C, Chirwa PW, Francis J, Babalola FD (2016) Assessing forest-based rural communities' adaptive capacity and coping strategies for climate variability and change: The case of Vhembe district in south Africa. Environmental Development 18:36–51
8. Govindan K, Soleimani H (2017) A review of reverse logistics and closed-loop supply chains: a Journal of Cleaner Production focus. J Clean Prod 142:371–384
9. Shrestha J, Subedi S, Timsina K, Chaudhary A, Kandel M, Tripathi S (2020) Conservation agriculture as an approach towards sustainable crop production: a review. Farm Manage 5:7–15
10. Firbank LG et al (2018) Grand challenges in sustainable intensification and ecosystem services. Frontiers in Sustainable Food Systems 2:7
11. Benke K, Tomkins B (2017) Future food-production systems: vertical farming and controlled-environment agriculture. Sustainability: Science. Practice and Policy 131:13–26
12. Bannerjee G, Sarkar U, Das S, Ghosh I (2018) Artificial intelligence in agriculture: A literature survey. International Journal of Scientific Research in Computer Science Applications and Management Studies 7:1–6
13. Han C, Hu X, Zhang J, You J, Li H (2021) Design and testing of the mechanical picking function of a high-speed seedling auto-transplanter. Artificial Intelligence in Agriculture 5:64–71
14. Ncube B, Mupangwa W, French A (2018) Precision agriculture and food security in Africa. In Systems analysis approach for complex global challenges, Springer 159–178
15. Baskerville R (2018) Information Security: Going Digital. In Annual Conference of the Southern African Computer Lecturers' Association 3–14
16. Ponnambalam VR, Fentanes JP, Das G, Cielniak G, Gjevestad JGO, From PJ (2020) Agri-Cost-Maps-Integration of Environmental Constraints into Navigation Systems for Agricultural Robots. In 2020 6th International Conference on Control, Automation and Robotics (ICCAR) 214–220
17. Peace CP (2017) DNA-informed breeding of rosaceous crops: promises, progress and prospects. Horticulture research 411–413
18. Fox-Kämper R, Wesener A, Münderlein D, Sondermann M, McWilliam W, Kirk N (2018) Urban community gardens: An evaluation of governance approaches and related enablers and barriers at different development stages. Landsc Urban Plan 170:59–68
19. Krebs J, Bach S (2018) Permaculture—Scientific evidence of principles for the agroecological design of farming systems. Sustainability 10:3218
20. Aldag MC, Eker E (2018) How to use blockchain technology in agriculture. 23rd International Scientific Conference "Restructuring Management. Development and Efficiency in the Face of Changes"

21. Biswas K, Muthukkumarasamy V, Tan WL (2017) Blockchain based wine supply chain traceability system. In: Future Technologies Conference (FTC) 56–62
22. Rogerson M, Parry GC (2020) Blockchain: case studies in food supply chain visibility. Supply Chain Management: An International Journal 25:601–614
23. Sharma H, Haque A, Jaffery ZA (2019) Maximization of wireless sensor network lifetime using solar energy harvesting for smart agriculture monitoring. Ad Hoc Networks 94:101966
24. Lohmer J, Lasch R (2020) Blockchain in operations management and manufacturing: potential and barriers. Computers and Industrial Engineering
25. Chen M, Mao S, Liu Y (2014) Big data: a survey. Mobile Network and Applications 19:171–209
26. Veroustraete F (2015) The Rise of the Drones in Agriculture. EC Agriculture 2(2):325–327
27. Ahirwar S, Swarnkar R, Bhukya S, Namwade G (2019) Application of Drone in Agriculture. Int J Curr Microbiol App Sci 8:1
28. Silvestre BS, Ţîrcă DM (2019) Innovations for sustainable development: Moving toward a sustainable future. J Clean Prod 208:325–332
29. Mishra H, Maheshwari P (2020) Achieving sustainable development goals through Fourth Industrial Revolution: an Indian perspective. Indian Journal of Commerce and Management Studies 11(2):63–75
30. Østergaard PA, Duic N, Noorollahi Y, Mikulcic H, Kalogirou S (2020) Sustainable development using renewable energy technology. Renewable Energy 146:2430–2437
31. Lazurko A, Venema HD (2017) Financing high performance climate adaptation in agriculture: Climate bonds for multi-functional water harvesting infrastructure on the Canadian Prairies. Sustainability 9:1237
32. Bruwer JP (2020) The sustainability of South African Small Medium and Micro Enterprises (SMMEs) operating in the retail industry amidst the ever-increasing excise taxation on tobacco products, alcohol products and plastic bags: A literature review
33. Oluwatayo IB (2019) Towards assuring food security in South Africa: Smallholder farmers as drivers. AIMS Agriculture and Food 4:485–500
34. Ferris S (2016) Module 11: Agricultural Entrepreneurship. Global Forum for Rural Advisory Services: Lindau
35. Chisasa J (2019) Determinants of access to bank credit by smallholder farmers: evidence from South Africa. Academy of Accounting and Financial Studies Journal 23:1–10
36. Chauke PK, Motlhatlhana ML, Anim PTK, FDK, (2013) Factors influencing access to credit: a case study of smallholder farmers in the Capricorn district of South Africa. Afr J Agric Res 8:582–585
37. Lepheane TJ (2007) Land Redistribution and its impact on Agri-BEE: A case study of Calcom PTY, Barbeton, Mpumalanga, South Africa. MA Dissertation Wits University
38. Coles CM (1993) Land reform for post-apartheid South Africa. Boston College Environmental Affairs Law Review 20:699
39. Ministry for Agriculture and Land Affairs South Africa (1998) Agricultural policy in South Africa: a discussion document
40. Gates B (2017) Developing sustainable agriculture in Southern Africa
41. Khwidzhili RH, Worth SH (2016) The sustainable agriculture imperative: implications for South African agricultural extension. South African Journal of Agricultural Extension 44:19–29
42. Warren M (2004) Farmers online: drivers and impediments in adoption of Internet in UK agricultural businesses. J Small Bus Enterp Dev 11:371–381
43. Boros-Papp S, Várallyai L (2019) The opinion of farmers and small and medium-sized enterprises on the importance of ICT in Hajdú-Bihar County. Hungary. Journal of Agricultural Informatics 10:2
44. Venter C (2019) The future of farming in South Africa

Speed of Innovation Diffusion in Green Hydrogen Technologies

Lourenço Correia(✉), Oliver Schwabe(✉), and Nuno Almeida(✉)

IST – University of Lisbon, Avenue Rovisco Pais 1, 1049-001 Lisboa, Portugal
{lourenco.correia,oliver.schwabe,
nunomarquesalmeida}@tecnico.ulisboa.pt

Abstract. In face of increasing pressure regarding climate change and greenhouse gases emissions, various countries and economic regions are setting ambitious action plans for a systematic transition towards a carbon-neutral economy. For example, the European Union aims at achieving this goal by 2050 through actions such as the investment and adoption of hydrogen technologies. Green Hydrogen – meaning hydrogen produced from renewable energies – is expected to play an important role towards a sustainable energy transition as an energy carrier with numerous use applications in transportation, industry, heating, and energy storage. Several experts see it as a viable solution to decarbonize different sectors over time. However, this technology is still in its early stages of development at various points of the supply chain. This paper discusses existing Green Hydrogen technologies and focuses on the diffusion of innovations through the extended Green Hydrogen Supply Chain. It uses a dependency model based on available Case Studies and semi-structured interviews to assess how these innovations will diffuse in the current market conditions until market saturation.

1 Introduction

Green Hydrogen – hydrogen produced from water and renewable sources [1] – and used in fuel cells, both for stationary and mobile applications, constitutes a very promising energy carrier in the context of sustainable development. Hydrogen technologies have significant potential to improve energy security and mitigate the effects of climate change, hence creating a path to a clean, sustainable energy system. Currently, the cost of green hydrogen is not competitive compared with fossil fuel-based hydrogen, yet with the development of hydrogen and fuel cell technologies, with increasing fossil fuel prices and taxation, have resulted in greater competitiveness recently which is expected to improve even more in upcoming years. Several European countries have presented plans to deploy hydrogen infrastructures and to accelerate the deployment of the hydrogen economy [2]. The required technology is already available, but the deployment of a hydrogen infrastructure constitutes a challenging task.

Hydrogen infrastructure and technologies are seen as an important part of the future energy mix, due to their advantages in terms of reducing GHG emissions in various sectors, from transportation, to industry, and the energy sector itself [3]. There is a need for appropriate methodologies for accelerating the deployment of extended hydrogen

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 101–111, 2022.
https://doi.org/10.1007/978-3-030-96794-9_10

supply, enhancing both supply and demand towards the realization of the expected environmental, economic, and social benefits. Research and implementation projects in the field have been growing at an increasing rate in recent years [4, 5], yet many concepts still remain undeveloped, in particular the diffusion process for innovations and variables influencing the adoption of Green Hydrogen technological innovations.

This paper discusses the acceleration of diffusion of technological innovations in the Green Hydrogen Supply Chain (GHSC). Section 2 provides an overview of Green Hydrogen technologies and a proposed description of the GHSC. Section 3 presents a maturity model that forecasts the speed of innovation diffusion from ideation to market saturation. Section 4 presents a test application of this maturity model to six case studies related with the production by PEM Water Electrolysis (PEMWE) recurring to renewable energy sources. Section 5 includes the preliminary results and the insights deriving from semi-structured interviews with regards to the factors enabling the diffusion of the mentioned innovations, the variables influencing their acceleration and a forecast of how they will actually diffuse, namely in terms of the time needed to achieve market saturation. The concluding remarks are presented in Sect. 6.

2 Green Hydrogen and Related Technologies

The decarbonisation of the world economy is expected to enhance hydrogen-related technologies within the global energetic mix and specially in hard to electrify sectors, such as industry (feedstock to petrochemical and fertilizer sectors), heavy transportation, heating, and energy storage [3].

Green hydrogen is an energy carrier characterized as hydrogen (H2) produced from renewable sources, energy, and feedstock. Currently, it represents 3.9% of the global hydrogen production [6]. The most established technology to produce Green Hydrogen is water electrolysis using renewable electricity in an electrolyser [7].

A major part of the global hydrogen consumption is dominated by two industries: oil refineries 52% and ammonia production 43%. The remaining consumption lies in other industrial uses. In Europe, ammonia and oil production account for 50% and 30% respectively, methanol production represents 5% and metal industries around 3% [1]. Most of this consumption derives from fossil fuel-based hydrogen. Iberdrola is currently developing an innovative project with the largest Green Hydrogen project for industrial use in Europe as an off grid hydrogen production to supply an ammonia factory in Spain with an electrolysing capacity of 20 MW [8]. Adding to the already developed utilizations, where hydrogen is used as feedstock in refining, chemical and fertiliser industries, the hydrogen potential applications are numerous and still mainly unexplored [9].

According to the International Energy Agency (IEA), renewable electricity production increased 45% to 280GW in 2020. This was the only energy source to increase this year and it is expected that the share of renewable energy sources in the global energy mix will also increase in the near future [10]. By 2022 solar PV production increased 162 GW (50% higher than in 2019), while wind energy production increased a record breaking of 114 GW in 2020 (an yearly increase of 90%) [10]. Bearing in mind the fluctuating nature of renewable energy production compared to the energetic supply and demand, renewable hydrogen has been considered a good solution as an energy storage method, particularly with large amounts of energy needed for an extended period of time, through the electricity-hydrogen-electricity cycle (Power-to-Power). The production of hydrogen from renewables through electrolysis, storage and the reconversion into electricity for grid supply, by fuel cells or gas turbines, presents a favourable off-grid application, for instance in isolated areas or as back-up power [3, 9, 11]. However, it does not yet seem viable due to the low full-cycle efficiency currently between 30% to 40% [5].

Another possible pathway lies on supplying hydrogen through the existing natural gas grid, by blending both together to generate hydrogen enriched natural gas (HENG), which can be used in buildings or industrial complexes in combined heat and power systems [12]. This represents a viable transition solution while dedicated infrastructure and hydrogen grids are not installed.

Hydrogen can also be distributed through refuelling stations for hydrogen powered vehicles or fuel cell electric vehicles (FCEV). This is considered as a clean energy source, as compared to other types of vehicles, due to leaving no bi-products other than water. FCEVs are also considered superior in terms of operating range and refuelling time when compared to battery electric vehicles (BEV) [13], while their energetic efficiency is significantly lower than that of BEV (electrolysis alone represents an energetic loss of approximately 30%). As with common electric vehicles, a substantial infrastructure is necessary to supply hydrogen-powered vehicles [13]. The FCEV is applicable to cars and trains. Fuel cell technologies are most suited in hard-to-electrify, long-haul, and heavy-duty vehicle markets, such as trucks, buses, maritime shipping, and aircrafts [13]. These represent means of transport where electrifying through batteries is inefficient, with current technologies, and a different fuelling method is needed to change from fossil fuel based transports and decarbonise each sector [14].

The designing of a supply chain is related with the desired goal, hence there is no unique solution for it. The Hydrogen Supply Chain (HSC) suggested in Fig. 1 highlights the fact that innovations and their impact diffuse along different paths. This proposal

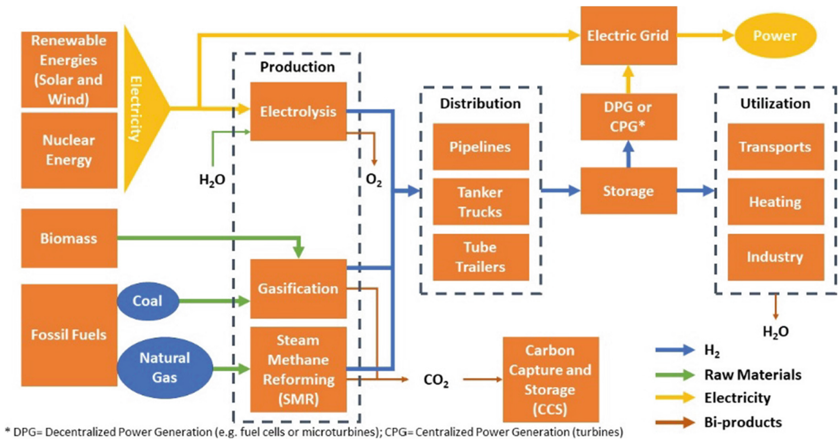


Fig. 1. Hydrogen supply chian, adapted from [7]

considers multiple energy sources, production processes, means of distribution, storage modes and end uses. It is nevertheless a simplification of the complexities of the various pathways involved in the HSC.

3 The Innovation Diffusion Model – Litmus Test

The Innovation Diffusion Litmus test [15] was developed to help understand how fast technological innovations can diffuse from ideation to market saturation. This test enables the identification of what variables influence the speed of innovation and and what is their interdependence in innovation ecosystems. The test is aligned with previous literature in Diffusion of Innovations [16], Value Networks [17, 18], as well as the Bass Diffusion Model [19]. This test is embedded with a Excel™ tool that captures the responses of semi-structured surveys and provides a high-level assessment of the underlying innovation web [20, 21].

The rational of this test is that an innovation needs a network/ecosystem (innovation web) in order to be successful. The test's objective is to assist those involved in the innovation web to enable ideas to progress and travel as fast as possible through that same network [20]. The key elements of the innovation web are (1) roles of (2) individual participants, who exchange (3) tangible and (4) intangible deliverables.

The innovation web shown in Fig. 2 consists in a series of archetypical relations and exchanges that begin when the Inventor(s) receiving an intangible challenge from the Key User(s). Inventor(s) investigate the matter with Super User(s) and develop the possible solution for the problem as a tangible prototype that is presented to the Product Owner(s). The Product Owner(s) take the prototype and transform it into a potential tangible solution with the clarification of the ways it can generate value to the consumer.

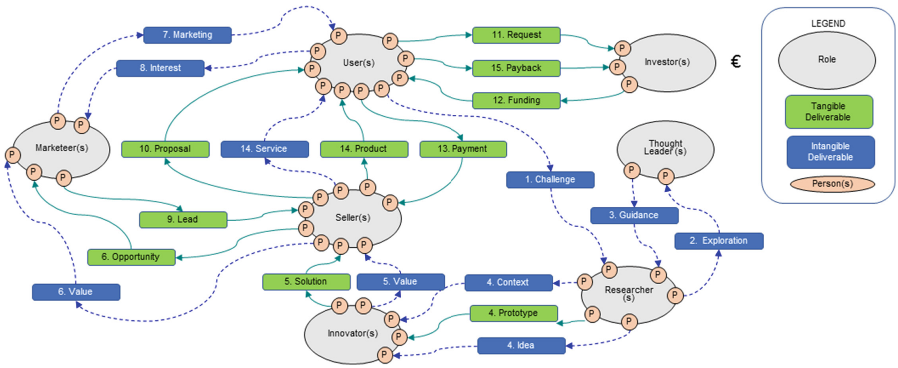


Fig. 2. Generic diffusion of the innovation web [20]

Afterwards, the solution and possible value creation is shared with the Business Sponsor, who reshapes the solution into a tangible opportunity and presents it to the Influencer(s).

The Influencer(s) commercialize the solution and provide value proposition back to the Key User(s), to create an intangible expression of interest, which the Influencer(s) convert in a tangible asset passed through to the Business Sponsor(s). The Business Sponsor(s) provide a tangible commercial proposal for selling the solution to the Key User(s), after receiving the proposal, they request funding from the Investor(s), and ideally it is provided that is used as payment for the tangible solution to the Business Sponsor(s). After receiving the required payment, the Business Sponsor(s) provide the developed solution and needed intangible services back to the Key User(s). Key User(s) eventually use the product to solve the problem/need initially announced and then provide the feedback and assets, necessary for relevant changes in return of the funding from the Investor(s) [20].

The archetypical roles in the web must be populated by at least one participant for each of the core and accelerator roles. The core roles responsible for triggering the initial innovation development are: Key User, Inventor, Product Owner and Business Sponsor. Accelerator roles provide the momentum needed to reach 84% of adoption in the proposed schedule. These are: Investor, Influencer, Super User and Moderator.

The high levels of uncertainty that an innovation suffers, mainly result from complex adaptations of highly regulated design and technical production solutions with highly complex product life cycles, also these processes frequently happen in worn out global supply chains and ecosystems. An important factor to consider is the shift from a linear and modular perspective to considering the whole ecosystem as a living and continuous interchange of tangible and intangible assets between the many stakeholders present in the diffusion process. Precisely, the idea presented in the Litmus Test is as a "virus" spreading as quickly as possible through the living ecosystem of the innovation web.

Moreover, it is considered that the "success", or market saturation, of the diffusion of an innovation is the sustained use of the idea or product by the late majority of adopters, accordingly aiming to reach late adopters in the total market share to achieve sustainable use and enough value created for continued stakeholder investment, whereby this equates to approximately 84% of the total market [16, 22].

The Litmus Test is designed as a semi-structured interview assessing the factors that influence the diffusion of the idea, as well as the degree of commitment that all the roles and participants present in the generic diffusion of the innovation web, while at the same time, intaking the level of confidence for each factor assessed. The evaluation of the aspects will be performed through a qualitative analysis based on a Likert scale from 0 (Not at all) to 5 (Very High). The aspects analysed for the idea are degree of innovativeness, technical readiness level, budget and resources, number of competitors, degree of complexity, compatibility with existing technologies, ease of understanding, ease of use and ease of adoption. Regarding the participant aspect the main attributes are if its behaviour in the web respects: urgency, priority, motivation, expertise, collaboration, and if they are voluntarily engaging. The score for the idea and participant aspects is weighted as an average of all answers to reach a relative score while considering the degree of confidence.

Two important factors to consider when modelling the diffusion are the total market size and the time forecast of the project. The level of maturity reached, originates from the case study assessment tool, and evaluates the ability of the case study to diffuse to the late majority within the expected timeframe.

The referred maturity levels are: Level 5 (Maturity: 80%–100%) the idea is successful and should be launched, Level 4 (Maturity 60%–79%) relatively high diffusion where quotation is required, Level 3 (Maturity: 40%–59%) intermediate diffusion rate and proposal is required, Level 2 (Maturity: 20%–39%) lower diffusion and more information is required, Level 1 (Maturity: 1%–19%) very low diffusion where the recommendation is to explore the strategy and Level 0 (Maturity: 0%) when the innovation does not diffuse and should not be launched. The band-widths are chosen based on an equal distribution of probability and considered by the authors as sufficient for drawing robust conclusions.

The amount of new adopters over time (sa(t)) is determined using the two coefficients of innovation (p) and imitation (q), Total market size (m) and the Cumulative number of adopters (S(t)) [22], through the following equations adapted from the Bass Diffusion base Eqs. (1) to (3).

$$s_a(t) = \left(p + \left(\frac{q}{m}\right) \times S(t)\right) \times (m - S(t)) \tag{1}$$

$$p = m \times s_r(t) \tag{2}$$

$$q = p \times s_r(t) \tag{3}$$

Presuming that each phase of adoption only starts when reaching the 84% of the adopter category, this is applied to each adopter segment separately and then aggregated.

The main outcomes of the test can be resumed as:

1. Evaluate the Innovation's maturity level (adequate for Innovators, Early Adopters or Late Adopters)
2. Evaluate the participant's maturity level (Forming, Exploring, Educating/Training and Performing)
3. Assess the overall maturity level of the project (Do not Launch, Improve or Launch)
4. Forecast how long the innovation takes to reach the late majority share of the total population relatively to the initial timeframe.
5. Identify the aspects that need improvement to accelerate diffusion and reach sustainable market growth.

4 Case Studies

In this paper, the focus is the Green Hydrogen production (see GH Supply Chain in Fig. 1) by Proton Exchange Membrane Water Electrolysis (PEMWE) recurring to renewable energy sources. The Case Studies presented in this section are a selection of projects being implemented in the European Union, a region that is specially focused on the implementation of a Hydrogen economy as one of the solutions to reach the carbon-neutrality by the year of 2050. The selection of the Case Studies considers the innovation itself and its characteristics on solving the pretended problem and the ecosystem members involved in each project.

The Green Hydrogen technologies are still on the early stages of development worldwide. The projects studied are in its majority part of the Horizon 2020 from the EU funded Fuel Cells and Hydrogen Joint Undertaking (FCH – JU), with the exception of the NortH2 project which is orchestrated and funded by the local government of Groningen in the Netherlands. The 6 case studies from the Supply Chain are presented in Table 1.

5 Results and Key Findings

The key findings are based on the analysis of the results of the Litmus test, with a focus on innovation maturity, population maturity, overall maturity, the degree of schedule adherence and confidence in the forecast.

The Litmus Test provides a forecast of adopter diffusion for each idea (Fig. 3), from ideation to market saturation, considering the inputs from the questionnaires with regards to the innovation itself and the inherent innovation web. Additionally, the model provides the cumulative total share of adopters throughout time resulting in different maturity levels as presented in Table 2.

Table 1. Case study definition and results

Case	Name	Context	Idea maturity
1	H2 Future	Generation of green hydrogen with the purpose of supplying a steel production plant in Austria	3
2	NortH2	Large-scale green hydrogen production resourcing to offshore wind power in the Netherlands	4
3	BIG Hit	Production of green hydrogen in isolated territories in the Scottish islands	3
4	H2 Ref	Develop cost effective and reliable FCEV refueling systems	3
5	Hy STOC	Supply and transportation using liquid organic hydrogen carriers (LOHC), to a commercially operated hydrogen refueling station in Finland	3
6	HPem2Gas	Develop, validate, and demonstrate robust, flexible, and rapid response PEM electrolysis	3

The diffusion forecast presented in Fig. 3, shows the results of the 6 case studies and the reference model which is the basis of comparison of the ideal diffusion pattern known as the bell-shaped curve presented by Everett M. Rogers [16]. The results of the practical Case Studies in general, show a very similar rate of adoption throughout time with small if any changes, with exception of the Case Study 2 – NortH2. NortH2 differs from the remaining cases as the largest and most ambitious one to be developed, although only fully operational in 2030, it is a very promising innovation, with a relatively high maturity level of 4, and presents very specialized and interested stakeholders in the consortium. For instance, Shell Nederland one of the world's major energy companies is highly invested in reducing its carbon footprint and become a net-zero emission energy company by 2050 or sooner, by finding new ways of sustainable income just where the Green Hydrogen comes in. The innovation itself also shows promising growth since the launch of offshore wind power production and technologies are increasing at a fast pace in recent years.

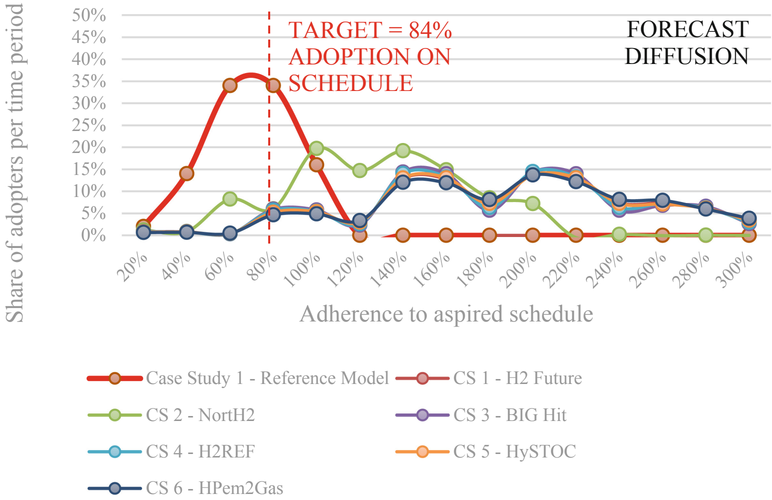


Fig. 3. Diffusion forecaset

The remaining Case Studies, the Horizon 2020 projects, also show promising results with a maturity level of 3, an intermediate diffusion rate where proposals and enhancements are still needed. However, these cases seem to present lower rates of adoption and reach market saturation much later in the aspired schedule, as mainly these projects are of small-scale and specified in the research and development of new technologies, that still require a lot of investment and legislation to take off. Moreover, all these cases depended on the orchestration and funding of the same entity, the FCH-JU, making them extremely dependent on the EU.

Recurring to a pareto analysis of the inputs on each case, the different variables influencing the maturity level of each Case Study present on Table 2 disclose the factors that show higher needs of improvement where the maturity levels are lower, as well as the small differences in each case's maturity level. The maturity of each factor considers the assessment value added in the questionnaire multiplied with the associated confidence level and divided by the maximum possible value of the product. The average maturity of each case results from the average of all the factor's maturity, where NortH2 presents the highest overall score as previously discussed, due to high maturity levels in factors such as degree of innovativeness, number of competitors and observability of impact. On the other hand, the worse maturity score is at the HPem2Gas, where factors such as compatibility with existing ways of work, ease of adaptation and the presence of the Super User disclose lower scores of maturity and need special attention for improvement. These factors can be improved through enhancing the confidence in a score by learning more about what it takes to score highly this is itself a strong way of improving maturity.

Table 2. Maturity level and factors of each case study

CS ranked by overall factor average maturity

Factor	NortH2	H2Future	BIG Hit	H2Ref	HySTOC	HPem2Gas	Average Maturity
Innovation - Degree of Certification (Legal/Policy)	64%	60%	36%	48%	36%	48%	**49%**
Innovation - Degree of Complexity	64%	64%	48%	48%	48%	48%	**53%**
Population - Super User (Identified)	75%	61%	50%	44%	69%	44%	**57%**
Innovation - Ease of Use	64%	60%	48%	64%	64%	48%	**58%**
Innovation - Technical Readiness Level	80%	48%	64%	48%	48%	60%	**58%**
Innovation - Ease of Adaptation	64%	80%	60%	48%	64%	36%	**59%**
Population - Influencer (Identified)	72%	67%	67%	67%	46%	48%	**61%**
Innovation - Budget and Resources	64%	64%	64%	80%	48%	48%	**61%**
Population - Key User (Identified)	72%	64%	48%	56%	69%	61%	**62%**
Population - Inventor (Identified)	72%	56%	72%	59%	52%	64%	**62%**
Innovation - Compatibility with Existing Ways of Work	80%	60%	64%	80%	60%	32%	**63%**
Innovation - Ease of Trialing	80%	80%	80%	64%	48%	36%	**65%**
Innovation - Observability of Impact	100%	64%	60%	64%	60%	48%	**66%**
Population - Business Sponsor (Identified)	90%	56%	75%	72%	64%	52%	**68%**
Innovation - Urgency of Need	80%	80%	48%	80%	64%	60%	**69%**
Population - Product Owner (Identified)	72%	69%	90%	48%	72%	69%	**70%**
Innovation - Ease of Understanding	64%	100%	64%	100%	36%	64%	**71%**
Population - Moderator (Identified)	72%	60%	75%	75%	75%	75%	**72%**
Population - Investor (Identified)	90%	72%	75%	77%	75%	56%	**74%**
Innovation - Number of Competitors	100%	60%	80%	60%	60%	100%	**77%**
Innovation - Degree of Innovativeness	100%	80%	80%	60%	100%	80%	**83%**
Case Study Average Maturity	**77%**	**67%**	**64%**	**64%**	**60%**	**56%**	**63%**

Moreover, the factors influencing more negatively the diffusion on the average of the 6 case studies are the degree of certification, degree of complexity on the innovation itself and regarding the contribution of the ecosystem members, the Super User is the entity that presents lower maturity. These are some of the factors that in general must be improved in order to reach a higher level of diffusion for Green Hydrogen technologies, where there are still some issues certifying and legislating the referred technologies, since they present high complexity on energetic/gas production and additionally there is a low presence of a Super User in the ecosystem due to early stages of development in this field.

6 Conclusion

In this paper, Case Studies related with different Green Hydrogen technologies were assessed in their perspective of diffusion and adoption throughout time. The need to study this technology and its diffusion is increasing as the EU climate goals as well as the lack of infrastructure, legislation and funding, force companies and governments to act fast on applying these technologies to our current energetic framework by reducing costs of production, supply and use of hydrogen and complementary technologies. Insights from the research effort suggest that the factors influencing diffusion are mainly the degree of certification and complexity, interdependent availability of budget and resources to achieve market saturation, all these projects are funded by governmental entities that intend research and deployment of the hydrogen economy making them extremely dependent on these entities. Additionally, the urgency of need, interest and expertise must be shared by all participants to reach higher adoption rates.

Further research is recommended in understanding external factors influencing the adoption of such technologies, by extending the application of the model into a higher number of cases and the validation of results with experts in the field of hydrogen is still required.

Acknowledgments. This work is supported by the Innovation Capacity Building for Higher Education (INCORE) project funded by European Institute of Innovation and Technology (EIT) in collaboration with the Open European Network for Enterprise Innovation.

References

1. Kakoulaki G, Kougias I, Taylor N, Dolci F, Moya J, Jäger-Waldau A (2021) Green hydrogen in Europe – a regional assessment: Substituting existing production with electrolysis powered by renewables. Energy Convers Manage 228:113649. https://doi.org/10.1016/j.enconman.2020.113649
2. European Comission, A hydrogen strategy for a climate-neutral Europe, Brussels (2020). Accessed 20 May 2021. https://www.eu2018.at/calendar-events/political-events/BMNT-
3. IEA, The Future of Hydrogen, Futur. Hydrog., June 2019. https://doi.org/10.1787/1e0514c4-en
4. Fuel Cells and Hydrogen EU, All H2020 projects, www.fch.europa.eu 2020. https://www.fch.europa.eu/fchju-projects/h2020 Accessed 03 Jun 2021
5. Chapman A et al (2019) A review of four case studies assessing the potential for hydrogen penetration of the future energy system. Int J Hydrogen Energy 44(13):6371–6382. https://doi.org/10.1016/j.ijhydene.2019.01.168
6. Dincer I (2012) Green methods for hydrogen production. Int J Hydrogen Energy 37(2):1954–1971. https://doi.org/10.1016/J.IJHYDENE.2011.03.173
7. Azzaro-Pantel C (2018) Preface hydrogen supply chains: design, deployment, and operation. Hydrogen Supply Chains. Elsevier, pp xix–xxiv
8. Iberdrola F (2020) Puertollano green hydrogen plant - Iberdrola. https://www.iberdrola.com/about-us/lines-business/flagship-projects/puertollano-green-hydrogen-plant Accessed 05 Jun 2021
9. Maggio G, Nicita A, Squadrito G (2019) How the hydrogen production from RES could change energy and fuel markets: a review of recent literature. Int J Hydrogen Energy 44(23):11371–11384
10. I. International Energy Agency, Renewable Energy Market Update - Outlook for 2021 and 2022 (2021). Accessed 05 Jun 2021. www.iea.org/t&c/
11. McKinsey and Hydrogen Council, Hydrogen Insights, February 2021
12. IRENA, Green hydrogen: a guide to policy making, vol. 2 (2020)
13. IRENA, Hydrogen from renewable power: technology outlook for the energy transition, September 2018
14. IEA, The Future of Hydrogen (2019)
15. Schwabe O (2020) The Innovation Diffusion Litmus Test. https://open-european-innovation-network.blogspot.com/2020/04/the-innovation-diffusion-litmus-test.html Accessed 02 Jun 2021
16. Rogers EM, Singhal A, Quinlan MM (1983) Diffusion of innovations
17. Allee V, Scwabe O (2011) Value Networks and the true nature of collaboration, ValueNet Works and Verna Allee Associates. https://www.goodreads.com/book/show/26216692-value-networks-and-the-true-nature-of-collaboration Accessed 27 May 2021

18. Homans GC (1958) Social behavior as exchange. Am J Soc 63(6):597–606. https://doi.org/10.1086/222355
19. Bass FM (1969) A new product growth for model consumer durables. Manage Sci 15(5):215–227
20. Schwabe O (2020) The Narrative for Generic Diffusion of Innovation. https://open-european-innovation-network.blogspot.com/2020/01/the-narrative-for-generic-diffusion-of.html Accessed 31 May 2021
21. Schwabe O, Schneider L, Marques N, de Almeida A, Salvado F (2021) A framework for accelerating innovation through innovation webs. In: Rodrigues H, Gaspar F, Fernandes P, Mateus A (eds) Sustainability and Automation in Smart Constructions: Proceedings of the International Conference on Automation Innovation in Construction (CIAC-2019), Leiria, Portugal. Springer, Cham, pp 205–210. https://doi.org/10.1007/978-3-030-35533-3_23
22. Schwabe O et al (2020) A maturity model for rapid diffusion of innovation in high value manufacturing. Procedia CIRP 96:195–200. https://doi.org/10.1016/j.procir.2021.01.074

Servitization and Industry 4.0

Overview for Leasing or Buying Decisions in Industrial Asset Management

Vicente González-Prida[1,2](✉), Carlos Parra[1], Fredy Kristjanpoller[3], Pablo Viveros[3], Antonio Guillén[1], and Adolfo Crespo[1]

[1] University of Seville, 41092 Seville, Spain
{vgonzalezprida,ajguillen}@us.es, parrac@ingecon.net.in, adolfo@etsi.us.es
[2] UNED, 28040 Madrid, Spain
[3] Universidad Federico Santa María, Valparaíso, Chile
{fredy.kristjanpoller,pablo.viveros}@usm.cl

Abstract. One of the main interests for today's companies is to reduce the life cycle cost of those assets included in their productive processes. With that purpose, new business models are considering the acquisition of services provided by an asset, instead of the direct ownership of such asset. In other words, the focus is that the asset's ownership remains in the service supplier, since the company just purchases the results obtained using the asset during a specific period. This asset utilization is at the end of the day the value added by the asset to the productive process. With that idea, this contribution is intended to summarize an overview of factors in order to decide between "lease or buy" an industrial asset. The choice between buying or leasing an asset must take into account different aspects, all of them with advantages and disadvantages, that must provide finally a positive influence in the profit and loss statement. Together with this discussion, one of the factor to take into account is nowadays the new digital tools that come from the so-called Fourth Industrial Revolution (4ID). These new tools (i.e. the IoT) support the asset management in a servitization context, contributing with big data analytics to facilitate the decision making in terms of criticality and reliability assessments as well as other decisions involved in the industrial asset management. In fact, the Digital Transformation may simplify the connection among system, process, asset and service, dealing with massive information interconnected among different assets, as well as different organizations like the asset user and the service supplier.

1 Introduction

This document is intended to briefly overview the possibilities of leasing or buying an asset from a cost-risk-benefit point of view. With this purpose, it is important to start mentioning the Life Cycle Costs Analysis (LCCA) as a methodology that helps to assess the variation of asset costs during its useful life. Such an analysis requires exploring aspects related to the impact of reliability in the total costs of life cycle among other relevant cost drivers. In the last few years Value Engineering, Design and Production Organization field specialists have improved the costs quantification assessment process,

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 115–125, 2022.
https://doi.org/10.1007/978-3-030-96794-9_11

including the use of techniques that quantify the reliability factor and the impact of the failure events on the total costs of a production system throughout its life cycle [1]. These improvements have allowed decreasing uncertainty in the decision-making process in areas of vital importance such as design, development, substitution and acquisition of production assets. It is important to clarify that in all the LCCA process there are many decisions, both technical and non-technical actions, which must be adopted throughout the period of use of an industrial asset. There are particular interests in improving the "reliability" factor, since it has a great influence on the asset life cycle total cost, and a greatly influence in the expected asset function. On the other hand, today's companies are exploring the possibility to reduce costs by contracting just those services or functions provided by the asset, instead of owning it. In other words, there is a tendency to reduce the ownership of assets at the expense of a shared responsibility for the proper use and conditioning of a specific asset [2], via pay-per-use instead of sales transactions for an asset ownership. This new provider-user relationship model is usually known as "servitization" and consist in a partnership with continuous interaction between both parties where the supplier of goods assists the user in the asset operations and maintenance, sharing benefits and risks between the two parts throughout the defined asset life cycle. Of course, services are an elemental part of the manufacturing and industrial world and, consequently, they require a specific contractual definition between the interested parties. As commented, servitization is a quite extended word that deserves some clarifications. A couple of definitions can be found in references like Vandermerwe and Rada (1988) [3], where servitization is the tendency among companies to base their businesses on services in order to gain competitiveness in the market. In a similar way, Baines and Lightfoot (2014) [4] indicate that servitization refers to the development by manufacturing companies of those skills needed to provide services and solutions to their customers (beyond or instead of their ability to manufacture a product). When companies decide to lease an asset instead of buying it, they are more focused in the value creation provided by the asset function (integrating such asset within a specific manufacturing process), than in increasing the fixed assets value in the balance sheet. Nevertheless, the decision has to be analyzed and quantified in an objective manner, since both (leasing and buying) provide strengths and limitations. Leasing an asset refers to the act of contracting an element where one part (the supplier) allows to the other part (the user), the use of an asset for a certain period of time, in exchange for rent or other payment. Therefore, the asset ownership stays in the supplier, and the user is the beneficiary of the services provided by the asset. Together with the services provided by the asset to the user, the leasing contract may include other services to be under the responsibility of the supplier. For instance, such a leasing contract may include also the asset installation, commissioning, maintenance and, finally, the asset dismantling. As an example, when the leasing contract refers to a car, it usually includes the preventive maintenance and, at the end of the leasing period, the user has the opportunity either to extend the leasing contract, to finalize it, or to buy the asset by a previously stipulated remaining cost. When the asset is not a car, but a physical asset installed within the layout of a specific industry, the relationship between the companies and the services to be provided from the supplier to the user can be taylor-made at the best interest of both parts. These services usually receive nowadays the support of the so-called Industry 4.0

and all the improvement techniques that can be applied throughout the asset lifecycle [5]. In order to deal with all these topics, the paper is structured in sections that address to the commented matters.

2 Cost Breakdown Structure Considering Servitization

According to the traditional ownership concept of an asset, the user (and owner in this case) supports the total costs of the asset, from its acquisition to its disposal [6]. Therefore, the total life cycle costs can be decomposed in different categories as: Acquisition Cost, Operation Cost, Maintenance and Repair Cost and Disposal Cost. This breakdown structure gathers the main types of costs associated with the process of use of the asset. The breakdown level and the different categories of costs will depend on different factors such as the nature of the asset, the type of information available, the process of asset utilization, the economic variables, the human resource, the existing technology, among others [7]. The acquisition cost is linked to the asset development, its manufacturing (including direct material and direct labor), marketing, distribution… and all those stages of the asset value chain until the asset ownership changes from the manufacturer to the user. In other words, the user pays proportionally all these aspects considered in the asset value chain until this is purchased. Once the asset has been acquired by the user, the new owner must face costs that specially affect to the use of the mentioned asset. Such costs are associated to the operation itself like energy, inputs and raw material, training of the operators, the required documentation etc. The operating costs are one of the most significant during the life cycle of an asset. With this, it is important to add those costs related to preventive and corrective maintenance, which may include: engineering design and modifications, spare parts logistics, training to the maintainers and any other cost that may be generated throughout the asset life cycle [7]. These support costs are the most difficult to predict. In some cases, operational and maintenance cost may exceed up to 10 times more than the initial costs of acquisition [6]. An important aspect in this stage is connected to the availability that the asset must have in order to meet the user´s production expectations. In reference to availability, the asset must be designed in such a way that it can be maintained in the shortest time and cost without negatively effects on the characteristics or reliability and safety of the asset. In most of the production processes, every minute in which the asset is out of service represents financial losses for the user. In that scenario, time is "cost", in such a way that maintenance activities must be executed quickly (short repair times) and efficiently, in order to be able to meet the availability levels required by the user [8, 9]. At the end of the asset lifecycle, it is when is needed to dismantle the asset installation. The costs at this stage are specifically related to activities for conversion of materials, waste management, consumption of energy, etc. All of these activities must be developed considering that are environmentally clean [10]. Summarizing, all the above-mentioned cost can be compiled in the following formula:

$$\begin{aligned} LCC_{Ownership} &= \sum_{T=1}^{T} [Ownership\ Cost]_T \\ &= IC + OC + PMC + \mathrm{CM}C + MMC - RV \end{aligned}$$

Where:

- LCC: Life cycle costs (in this regards, the total ownership cost).
- T: expected useful life (in years).
- IC: Acquisition and installation costs.
- OC: Operation costs.
- PMC: Preventive maintenance costs (planned tasks).
- CMC: Corrective maintenance costs (due to failures, costs produced by the effects of low reliability)
- MC: Major – specials maintenance Costs (e.g. overhauls).
- RV: Remaining Value (or –[Disposal Cost])

On the other hand, considering just the leasing contract of the asset (that means, it is contracted just the service provided by the asset, and not the ownership), there are many possibilities to represent that relationship between the user and the service provider. One of the simplest way is to consider that, besides the stipulated leasing cost, the user is responsible just for the Operating costs. This is the case illustrated in Fig. 1 and in the following formula:

$$LCC_{Service} = \sum_{T=1}^{T} [Service\ Cost]_T = LC + OC$$

Where:

- LCC: Life cycle costs (in this regards, the total service cost).
- T: expected useful life (in years).
- LC: Leasing costs.
- OC: Operation costs.

According to this case, the user is just responsible for the asset operation. In other words, many of the total ownership costs are removed when the user just considers leasing the asset instead of buying it. Nevertheless, in that situation, the user has to face a new cost, which is the cost of the leasing.

When is more convenient to lease than to buy an asset? Of course, when the total ownership cost is higher than the total service cost. It can be expressed by the following formulas:

$$LCC_{Ownership} > LCC_{Service}$$

$$\sum_{T=1}^{T} \left[Ownership\ Cost\right]_T > \sum_{T=1}^{T} [Service\ Cost]_T$$

$$IC + OC + PMC + \mathrm{CMC} + MMC - RV > LC + OC$$

Therefore, in order to take the decision of leasing the asset, the leasing cost has to be:

$$LC > IC + PMC + \mathrm{CMC} + MMC - RV$$

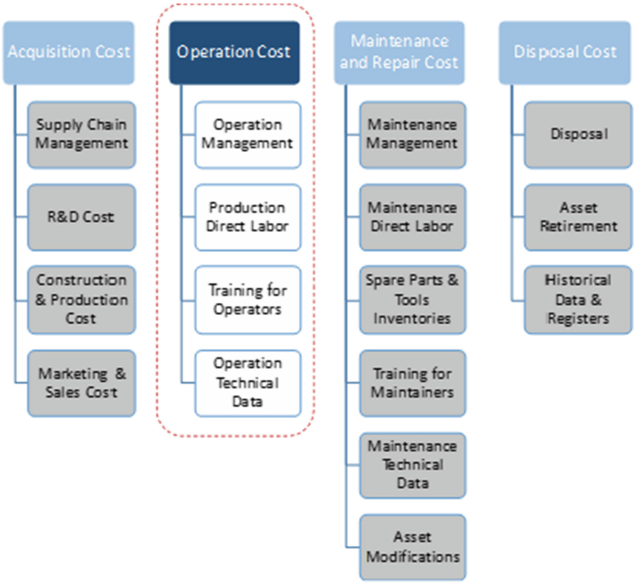


Fig. 1. Possible costs with leasing (adapted from [7])

An interesting aspect in the leasing process, would be to include a procedure in the design phase, which allows estimating the economic impact of failures (low reliability) on life cycle costs (LCC) [11, 12]. Once the design has been completed, it is substantially difficult to modify the economic results. Also, the economic considerations related with the life cycle should be specifically outlined during the phases previously mentioned, if it is that one wants to totally exploit the possibilities of an effective economic engineering. It is necessary to keep in mind that almost two thirds of the life cycle cost of an asset or system are already determined in the preliminary conceptual and design phase (70–85% of value creation and costs reduction opportunities), according to Dowlatshahi (1992) [12].

3 Supplier's Viewpoints on Servitization

Many suppliers (mainly manufacturing companies) still hold a high importance to product sales in comparison to the sale of services. These suppliers consider that the customer chooses a supplier based on the quality, price and delivery time of the tangible product (asset) they supply. This represents a product-based logic, which has been until now the dominant way of thinking and business understanding in many manufacturing companies [2]. Nevertheless, it is growing the awareness that customers do not necessarily require to own an asset, since that asset is used to perform operations or functions within a manufacturing or testing process. This new way of thinking relativizes the will of the users to own the supplied asset with additional advantage due to fractional payments according to the asset's performance when it is used in the user's own value creation process. In general terms [13], product-service systems are usually classified according to their product-service content [14] or levels of servitization:

- Product-oriented: selling the product in a traditional manner while including within the sale additional after-sales services, such as maintenance, warranty periods, training, etc.
- Use-oriented: selling the use or availability of a product while it is owned by the asset producer.
- Result-oriented: selling the outcome or capability provided by an asset while it is owned by the asset producer.

Each of these levels of servitization will have a different cost breakdown structure that will impact in the service sales to the asset user. Hence, the asset user will consider if the comparison between such service cost (plus those services not included, depending on the contracted level of servitization) with the total ownership cost, in order to check which possibility is more affordable and profitable for the company. In reference to the possibilities of servitization within an industrial context, there are already sectors where the sales of new products are very small in comparison to the so-called past installed sales, which refer to those sales accumulated over the years. In these sectors, it is very important not to lose business opportunities by offering services around the company installed base. Losing this opportunity opens the possibility to competitors to take the company customers portfolio to provide the initiative of offering services around others' products. Hence, the company that sold the product, loses possibilities to learn from customers by accompanying them in the use of the product, learning how they are really using it, how they value its use and what services or added products may interest them. On the contrary, a company that implements servitization, increases its chances of customer loyalty improving its differentiation capacity and opening new business opportunities [2]. Due to that reasons, companies in practice are nowadays paying more attention to their service sales and their related management processes. Consequently, there are opportunities of significant cost savings as well as the possibility of increasing the business' value through higher levels of customer satisfaction, not through better quality assets, but for better quality functions. Under this situation, it is mandatory for companies with focus on service, to take into consideration the interactions between engineering, marketing and customer support from an operational and strategic level [15]. Therefore, customer relationship management should be handled as an aspect that can create more value in the performance of the business. The result will be to improve services, gaining consequently higher levels of customer retention. Servitization, viewed from a strategic perspective, begins with a strategy linked to technical and commercial planning from the very start of the asset development process, although the execution of such services is performed during the asset life cycle. Therefore, most of the information required to provide services has to be strategically determined from the very beginning of the process. As a result of such analysis, a deep knowledge about the asset capacity has to be obtained in order to perform properly the functions expected from it, as well as the cost of spare parts, reparation times, etc. among other data needed to elaborate a consistent portfolio of services. For instance, with the results obtained from the RAMS and RCM analysis, an initial maintenance plan, applied to the life cycle horizon, can be a good first approach for the service capacity planning, the spare parts provisioning, the service task schedule, skill level of the technicians, etc., as well as an analysis on service cost-risk-benefit. In this stage is also of importance to start implementing a Customer Relationship

Management (CRM), which will be used, in general, for the definition of a business strategy centered in the customer. Additional analysis will deal mainly with aspects more related to the service logistics and the achievement of a balance between costs and customer satisfaction. In general terms, it is possible to build management frameworks that help asset suppliers in this process of servitization [16]. This framework covers from data analysis, to advance simulation and optimization mechanisms that facilitate the final decision-making process of product-service systems. Together with the decision on what is the most appropriate level of servitization (and with which specific level) in terms of cost and business strategy, it is important to stablish the contractual frame that links the asset supplier with the asset user [15]. Service contracts must encompass the conditions for such services. A service contract can include and/or exclude different topics. In legal terms, service provision is a concept expressed through a contract between the parties for specific assurances and applied to customers during a determined period. Service contract can give different conditions of the technical service to be provided by the manufacturer (service provider), which must be described in the fundamental terms of the contract. These agreement clauses give consequently a legal obligation to the manufacturer (service provider) in front of the customer (asset user). Usually, the asset function (which is ultimately what is buying the customer), must be reasonably conformed to an ordinary user's expectations, and the supplier must be sure that the asset is fit for that particular purpose. In general terms, assets must comply the trade standards applicable to the service contract, being asset functionalities uniform in quality. It is interesting to add here that, in connection with suspension periods, they should start for all the affected elements from the moment when the appearance of the defect is communicated to the service provider, until the repair or substitution is performed. It is not considered affected by this suspension those elements which stay operative despite the failure [17]. This issue is frequently not clear in contracts and supposes usually a confrontation topic between user and supplier. Apart from deciding what kind of service is more interesting for the parts, it is important to include in the service contract a definition as clear as possible of all those requirements, conditions, works scope etc. by written for the mentioned services.

4 Influence of the New Technologies on Servitization

Servitization requires improving the strategies with the use of more efficient tools and integrated systems, as well as through an organizational structure in order to link service supplier with asset user. The application of new technologies represents higher levels of knowledge, experience and personnel training in the technical service. Logically, one of the most important parts within a service management program is the relation with the customer. The techniques used in these relations generally considered as a part of an ERP (Enterprise Resource Planning) system, or as a complementary part of them. The CRM (Customer Relation Management) is not exclusively used to define a business strategy centered on the customer, but also to include a group of information technology applications useful to sort and manage data related with customer, reclamations, and in general with the commercial activity of the company [15]. As it has been mentioned before, the CRM is often used as a module for customer management including ERP software. Originally, the CRM was once focused mainly in marketing and commercial aspects. However, the CRM module can be adapted to service management, allowing the company to:

- Identify products and services requiring assistance.
- Reducing the time of the assistance and optimize channels of information.
- Identify groups of customers in order to develop common strategies.
- Be conscious of customer's real needs.
- Increase sales on the same level as customer satisfaction, etc.

Customer relationship management (as part of an ERP or as a complementary system), includes operative areas that complement tasks directly related to the customer (front office), and other analytical more closely associated to different internal parts of the company (back office). The continuous improvement in the service management of a supplier or a user requires the application of emerging techniques and technologies in those areas considered with more impact. New and emerging technologies on information and communications, facilitate the work inside a multiple-users environment by a cooperative and distributed way. Therefore, these new tools should be defined as the support to a service program which includes those resources, assistances and management needed to enable proactive decision-making. Included in the digital technologies, there must be activities which are typically associated to service provision like monitoring, diagnostics, prognosis, etc. Together with the application of new technologies for the service management, is essential the involvement of the technicians from the after-sales service in the continuous improvement process in order to achieve a success goal which supposes higher levels in the quality of the asset function, as well as service effectiveness of the technical assistance. Consequently, it will be required higher levels of knowledge, experience and training [18]. Considering the implementation of digital tools, they propose a new way of managing a service program by which assets function is controlled via Internet by the service supplier, e.g. information can be gathered or is possible to manage alarms. Thus, the service supplier obtains real-time data from the asset through digital technologies. The data obtained by sensors pre-installed on the asset, provide information about its status (temperature, run time, pressure, etc.) and allow a continuous diagnostic, which enables the failure prediction due to the malfunction of equipment or due to a bad use. This application may also be a useful strategic marketing tool when seeking new business, offering to the users the possibility to contact services as the remote tracking of asset performance and, consequently, being able to foresee possible failures. In any case, this new advanced model of customer service management breaks the physical distances between supplier and customer using ICT and transforming the company into a manufacturing services business that provides support to its customers anywhere, anytime. All these improvements can be carried out based on a computerized information system for the service programs, which enables the cyclical processing of data management as well as the analysis of results. Digital tools facilitate the decision-making, processing for that purpose large volumes of data and information from multiple sources and with a great variety of types. Therefore, servitization requires to be based on the monitoring, diagnosis, prediction and real-time control of assets and the detection of abnormal conditions. The diagnosis identifies possible causes of early degradation or failure in asset (important when doing the RCA of the failure) and the prediction analyzes the impact of the failure on the asset itself (essential for the Criticality Analysis). Thus, the manufacturer is able to obtain in real-time whether a failure has occurred on a specific asset and, consequently, the failure (in terms of the causes that produced it) is included in the coverage of the service contract, or it is a result of a misuse or an accident with the asset. With this information, the manufacturer can anticipate the failure, warning the user (such as an added value in the Customer Relationship Management) about the existence of anomalies in the control parameters and recommending, for instance, the preventive shutdown of the asset, reducing thereby the repair time and cost (avoiding further damage with the shutdown), achieving at the same time a mutual benefit for both, manufacturer and customer. The collected information also allows the service supplier to make decisions and to carry out maintenance actions if necessary, generating automatically work orders and speeding up the administrative procedures. Sometimes the historical information from the asset may be enough to verify the “natural” cause of the failure, or an abnormal situation, thus it is not required a previous physical assessment to start planning the repair.

The application of digital technologies in servitization allows a fully knowledge on how assets behave by receiving signals from sensors on remote. Thus, during the specified service contract period, possible failures can be predicted reducing substantially the administrative procedures, increasing functioning time and optimizing the asset performance and productivity. Thus, the customer takes advantage of the sensors network pre-installed on the assets in order to predict failures which may affect the function impacting on the production line where the asset is installed. All this is translated into a fast, effective and efficient service to the user, increasing the operating time of the asset under the service contract, its productivity and finally the customer satisfaction. Obviously, the application must be clearly and explicitly agreed in the service contract of industrial assets, in order to avoid conflicts or problems regarding to the privacy of customer activities [14]. The digital servitization is a proactive system, where information is automatically and instantly received from the asset, being able to configure what kind of information is desired depending on the asset under analysis, for example: failures, asset performance, anomalies in measurable parameters (temperature, pressure, etc.) workloads peaks, asset status, among others. The transmitted data can only provide information about the asset. In addition to the above mentioned, digital servitization has other advantages:

- It eliminates spatial barriers between asset user and service supplier.
- It eliminates the time barriers.
- It reduces the cost per failure to the customer and supplier.
- It enables permanent access and control in real-time to all the information on the asset under the scope of the service contract.
- It facilitates fast communication between customer and supplier.

It eliminates paperwork for the supplier in case of equipment failure (cause verification, repair record, etc.). The relevance of well-established decision making processes in this context is emphasized in reference [19].

5 Summary and Conclusion

As commented throughout this document, servitization refers to a new focus on inter-firm relationships. This focus tends more towards cooperation, joint operations, risk/reward sharing and, of course, to innovate in earnings and payment models. From this overview, it is easy to understand the economic impact that can bring the decision between leasing or buying an asset, in order to integrate it within a production system. Therefore, a cost model for the analysis of the alternatives between buying or leasing an asset, seems to be necessary as a support instrument in the company's decisions (Jiménez 2008). This analysis can be performed using models that estimate the life cycle cost (LCC). In particular, this document has shown its usefulness calculating the summation of acquisition cost, maintenance cost and disposal cost (that is the total ownership cost) in comparison to the expected leasing cost. The application of LCC techniques allow organizations to efficiently select their physical assets with a lower level of uncertainty and, according to this document, to select them according to a specific ownership or service contract. Together with some suggestion related to cost structure of servitization, this paper has dealt with the suppliers' point of view on this new business line, as well as the consideration of a digital servitization where the application of new technologies appear as a mandatory requirement for the fulfilment of this kind of contracts. Besides this general overview, future research may deepen on the risk analysis, since the leasing assumes that part of the company value chain depends on an external part. In fact, further researches on this topic would be very interesting and welcome, in order to contribute to the development of the after-sales service area. Other special areas to deepen in regard to servitization can be for instance: (i) Analysis of customers and complaints; (ii) Modernization of technical assistance plans; (iii) Processing of large volumes of information; (iv) Data feedback on the asset life cycle cost; among others.

References

1. Woodhouse J (1993) Managing Industrial Risk. Chapman Hill Inc, London, pp 200–241
2. Kamp B, Alcalde H (2014) Servitization in the basque economy: servitization in the Basque economy. Strateg Change 23(5–6):359–374. https://doi.org/10.1002/jsc.1982
3. Vandermerwe S, Rada J (1988) Servitization of business: adding value by adding services. Eur Manage J 6(4):314–324
4. Baines T, Lightfoot H (2014) Servitization of the manufacturing firm: exploring the operations practices and technologies that deliver advanced services. Int J Oper Prod Manage 34(1): ISSN: 0144–3577
5. Crespo Márquez A, González-Prida Díaz V, Gómez Fernández JF (eds) (2018) Advanced maintenance modelling for asset management. Springer, Cham. https://doi.org/10.1007/978-3-319-58045-6
6. Wilson RL (1986) Operations and support cost model for new product concept development. In: Proceedings of the 8th Annual Conference on Components and Industrial Engineering, pp 128–131
7. Fabrycky WJ, Blanchard BS (1991) Life Cycle Costing and Economic Analysis. Prentice Hall Inc, New Jersey
8. Smith DJ, Babb AH (1973) Maintainability Engineering. Halsted Press, Sydney

9. Ruff DN, Paasch RK (1993) Consideration of failure diagnosis in conceptual design of mechanical systems. Des Theor Methodol ASME DE 53:175–187
10. Weule H (1993) Life cycle analysis a strategic element for future products and manufacturing technologies. Ann CIRP 42(1):181–184
11. Parra C, Crespo A (2015) Ingeniería de Mantenimiento y Fiabilidad aplicada en la Gestión de Activos. INGEMAN, Seville, Spain, pp 203–226
12. Dowlatshahi S (1992) Product design in a concurrent engineering environment: an optimization approach. J Prod Res 30(8):1803–1818
13. Morelli N (2003) Product-service systems, a perspective shift for designers: a case study - the design of a telecentre. Des Stud 24(1):73–99
14. Tukker A (2004) Eight types of product–service system: eight ways to sustainability? experiences from SusProNet. Bus Strateg Environ 13(4):246–260. https://doi.org/10.1002/bse.414
15. González-Prida V, Crespo A (2012) A framework for warranty management in industrial assets. Comput Ind 63(9):960–971
16. Erguido A, Crespo A (2019) A management framework for enhancing asset management decisions in service-oriented business models. PhD Thesis, University of Seville
17. Jiménez NS (2018) Decisión de Comprar o Alquilar Activos Fijos. Revista de la Facultad de Ingeniería Industrial. 11(2):9–17. UNMSM. ISSN: 1560–9146
18. Crespo A, Parra C (2017) The Maintenance Management Framework. Models and Methods for Complex Systems Maintenance, Springer, London https://doi.org/10.1007/978-1-84628-821-0
19. Trindade M, Almeida N, Finger M, Ferreira D (2019) Design and development of a value-based decision making process for asset intensive organizations. In: Mathew J, Lim CW, Ma L, Sands D, Cholette ME, Borghesani P (eds) Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies. LNME. Springer, Cham, pp 605–623. https://doi.org/10.1007/978-3-319-95711-1_60

Creating Value and Business Benefits from Joint Offerings of Asset Performance Management Tools in the Capital-Intensive Industries

Minna Räikkönen[1(✉)], Leila Saari[2], Katri Valkokari[1], Antti Rantala[1], and Helena Kortelainen[1]

[1] VTT Technical Research Centre of Finland Ltd, Tampere, Finland
{Minna.Raikkonen,Katri.Valkokari,Antti.Rantala, Helena.Kortelainen}@vtt.fi
[2] VTT Technical Research Centre of Finland Ltd, Oulu, Finland
Leila.Saari@vtt.fi

Abstract. In capital intensive industries, the purchasing of IoT, predictive maintenance or other digital solutions is very complex as there already exists several separate legacy systems. Also for a single solution provider, the customer needs and expectations can be huge, and they might not have all resources or competences needed for the delivery. In our research, we examined various aspects related to joint offering development and value creation. We define the joint offering as a concept or solution that is co-created in collaboration with two or more actors that usually have complementary technological skills or value creation logics. Based on our study, there is still relatively little joint offerings deployed supporting asset performance management and execution of asset management operations. To discover the interaction between the actors able to provide a joint-solution, we further sketched value network models from the anonymised industrial challenges. Finally, the developed conceptual framework clarifies the scene of both the business opportunities and value assessment. The first part of the framework, business opportunities, considers the business models, value networks and analysis of business opportunities and risks. The later part, business value assessment, is built on capital and operating expenditure and revenues summarising the business value.

1 Introduction

The digital transformation is currently accelerated by exponentially growing availability of various technologies and solutions (e.g., digital twins, advanced sensor technologies, predictive maintenance solutions, Internet of Things IoT related products and services, artificial intelligence) [1]. Thus, the opportunity for companies in capital-intensive industries to renew their business and operating models through application of digital technologies has become evident. Over past decades, there has also been a trend for focusing on the most important functions of core business and outsourcing the non-core operations [2]. We argue that this direction of strategic business development in capital-intensive industries combined with digital transformation provides IT sector with significant opportunities to open up new markets, to acquire new customers and to develop new products and services.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 126–136, 2022.
https://doi.org/10.1007/978-3-030-96794-9_12

The disruptive aspects of IoT technologies dominate the current developments and drive innovation in the IoT space [1]. Increasingly, IT companies are realizing that they can no longer compete through their own product and service excellence alone [3]. Novel and more specific customer needs can lead IT companies to use external skills in the development by employing more domain competence and by acquiring tailored solutions from other companies. In addition, there is a growing need to examine models for collaboration and co-development to share the costs, benefits and risks of the product and service development with other companies. In this paper, the joint offering is defined as a concept or solution that is co-created in collaboration with two or more actors that usually have complementary technological skills or value creation logics. Further, they aim to solve a challenge that is identified by their customers and they do not have resources nor competences to implement it alone.

However, digital product and servitization business and related joint offerings differ quite fundamentally from the traditional ways of doing business [4]. It is strongly related to the capacity to integrate different resources and data flows [5] and to orchestrate meaningful collaboration [6]. Collaborative business settings have been growing and increasing lately, and the concept of a software ecosystem has also made a large impact on the platform business and research [3]. The ecosystems that surround platforms are undeniably responsible for a large part of a platform's success. Hence, companies should carefully analyze the options for the collaboration. In general, the options may range from joint ventures to short-term contractual agreements on specific tasks. Examining and choosing between the different strategic options, however, may be challenging because of the nature of novel business opportunities, parties involved, customer needs, changing business environment, and related risks [7]. Every company has several processes and activities through which they create value [8, 9]. The traditional perspective highlights typically only the company's role in value creation, whereas the new perspective considers customers and other business partners to also be important players [10]. As the concept of value creation is different from different perspectives and for different parties [8], lifecycle value and business benefits of joint offerings needs to be considered in a manner that involves relationships, exchanges and interactions between parties. In this context, the understanding on the collaboration potential in value creation is important. For example, how the co-creation and data sharing can lead to better solutions. The challenge here is to understand how companies' business models are able to interact and to find a common ground and a win-win situation in the value network [6, 11]. In addition, there is still relatively little innovation in building joint offerings in the capital-intensive industry.

2 Research Objective and Methods

The aim of this paper is to examine various aspects related to joint offering development and value creation in the capital-intensive industries. The paper is based on the preliminary results of the SEED project[1]. It is one of the forerunners ecosystemic joint actions in Finland bringing together the pulp, paper, board and tissue manufacturers, engineering companies, technology suppliers and IT sector - over production line lifecycle from

[1] https://seedecosystem.fi/.

engineering to operations and asset management. In the pulp and paper industry, even automation is already widely applied, there is still various systems, processes and assets where digital solutions can be deployed and jointly developed.

The research is challenge driven and interdisciplinary by nature and applies scientific methods to find solutions to the practical problems. The approach aligns with the principles of design science [12] and empathize the problem-solving as a source of innovation. In this joint problem solving process between the practitioners and the researchers, we follow recent ideas of design thinking methods [13, 14] that constructs (or concepts) form a process for the joint problem solving. This kind of approach also suggests that research is not just about understanding and explaining issues and phenomena but also about changing them [15] and affecting creation of new ideas and innovation. Therefore, a joint project (or process) of researchers and practitioners – companies or other organisations – and close cooperation throughout the problem solving are necessary in such a research setting. Through a design science approach, we explored the latest academic literature and companies' viewpoints and experiences on joint offerings, and value creation mechanisms, and ended up to build a conceptual model for joint offering development. Firstly, we used a content analysis to research collaboration and value creation mechanisms for joint offerings, and to examine revenue, cost and risk sharing models and assessment methods. Secondly, to set the scene and understand the practical problems the researchers conducted a web survey and semi-structured interviews (typically more like discussions) with the key persons of the SEED companies participating in the industrial cases.

The survey was implemented in the beginning of the research to gather basic information of the technological solutions and services that the companies in the SEED ecosystem could provide. Altogether, there were 13 respondents operating in the IT sector and providing products and services for pulp and paper industry. *The interviews* were conducted in one (out of seven) industrial use case (UC) of the SEED project. The use case in question targets to develop a production digital twin [16] to improve process control in de-inking, paper machines and converting in terms of productivity, quality and cost. Regarding the business requirements in the use case, six of the interviewed companies have a main business model linked to IT domains (and are later called solvers), whereas one company was in the role of use case (problem) owners in pulp and paper industry.

3 Background for the Joint Offering Development

In this chapter, the reasoning for the joint offering development is briefly discussed. Firstly, the risks and opportunities of the collaboration are considered, and after that the characteristics of the value flows and interaction among actors is addressed. The last chapter emphasizes the role of business value assessment and demonstration in the context of joint offering development in capital-intensive industries.

3.1 Risks and Opportunities in Multi-actor Collaboration

Collaboration between multiple parties with diverse technical capabilities is common in high-technology industries [17]. By deploying collaborative business models, partners can jointly create, capture and deliver value. Multi-firm collaboration includes various benefits and risks that should be considered before participating in multi-firm arrangements.

Partnering with other companies may enable access to new markets and help to gain new customers. Companies may also increase their reputation by collaborating with forerunners [18]. Companies that engage in collaborative product development have the opportunity to combine their complementary knowledge sources, facilitating the generation of technological inventions that the partners could not achieve on their own [19]. Because collaboration requires the combined effort of multiple companies, they entail coordination complexities, often resulting in conflicts and frustrations. Collaboration may also cause financial and organizational risks such as the risk of becoming overly dependent of the partner [18]. Also, external changes will affect partners differently, requiring continuous calibration of the collaborative business model [20].

3.2 Value Network View on Joint Offering Development

Understanding the value flows and different dimensions of value are key to understanding the dynamics between the actors. In addition to monetary value, tangible and intangible forms of value include quality, goods, knowledge, benefit or service [21]. In the capital-intensive industries, high-volume processes and large-scale capital investments have been the key drivers. However, in a development process aiming to joint offerings the challenge is that the content of some – if not all – of these core building blocks is unknown. Thus, an open approach is needed to broaden the view of the current value network. Further, the traditional approach to value stream mapping (VSM), which is suited to linear systems, must be connected to other tools like system dynamic modelling [22, 23].

In literature, the joint offering is discussed within wind energy sector, brand alliances (ingredient branding) or service bundles at a marketplace. In a dictionary, a joint venture is defined as a business or business activity that two or more people or companies work together. In this paper, we define the joint offering as a *concept or solution that is co-created in collaboration with two or more companies and where parties usually have complementary technological skills and value creation logics*. They aim to solve a challenge that is identified by their industrial customers and do not have resources nor competences to implement it alone. This goes well with the idea of co-creation, where the mission oriented viewpoint has three optional processes: initiation, co-design and co-implementation [24].

To discover the business opportunities and interaction between the actors willing to provide a joint-solution we sketched value network models from the anonymised industrial challenges in SEED based on the interview findings. We exploited the notation of Verna Allee's Value Network Analysis (VNA) [21]. The notation has: i) Entities that are either individuals, teams, groups or companies, and ii) Transactions are exchanges between entities (in a value network). Transactions are either tangible or intangible.

3.3 Business Value of Joint Offerings

Business value of joint offerings can be considered both from the value network and from the customer viewpoints. Perceived use value is based on customer's perception of the joint offering's use and on the other hand, exchange value is the amount paid for the joint offering by the customer to the value network [9]. The demonstration and verification of the joint offering's business value to customer can represent a critical area in the marketing and in the business negotiations. However, the increasing complexity of both digital technologies and value networks has an impact on the monetization models to create value. For example, the solutions for predictive maintenance in the capital-intensive industry can reduce the life-cycle cost and the cash tied-up to production assets by minimizing the asset downtime and the need for costly corrective and preventive maintenance activities, without implying loss of benefits [25]. However, the most current economic models regarding the value creation and assessment of digital technologies focus largely only on short-term cost reduction rather than valuation of life-cycle revenues, costs and risks for value network parties.

Typically, methods for evaluating the business value of joint offerings for various parties and for the entire value network can be divided into financial assessment, alignment with the business strategy, scoring models and checklists [7]. Financial assessment is the most widely used, and strategic approaches are also quite popular. However, regarding the joint offering development, a variety of above mentioned approaches should be used to be able to cover all the relevant aspects affecting the business value for various value network partners.

4 Insights on Supporting the Joint Offering Development

4.1 Defining Business Requirements

During the interviews, the expert insights on the development of joint offerings in the Finnish pulp and paper industry were gathered. The joint offerings on digital solutions and related business requirements were discussed as widely as possible, also taking into account the perspectives of various partners in the SEED project, namely the IoT technology providers and pulp and paper mills.

From the business point of view, joint offerings are gradually increasing also within the pulp and paper industry. However, there is still relatively little joint offerings deployed and innovations regarding novel business models realized. It should also be noted that in addition to benefits and opportunities, there are also risks and challenges that have an effect on the implementation of digital solutions and hindering digitalization in the pulp and paper industry. According to the interviews, in the case of developing, acquiring and implementing novel technologies, willingness to take risk is always needed. The main interview findings on the business requirements are summarized in the following table (Table 1).

Table 1. Insights from the interviews.

+ Positive	"Business as usual is history": Processes also in pulp and paper industry will increasingly include digital solutions. Different application targets possible **"Win-win situation":** Clear business benefits for pulp and paper industry if the expected goals set for joint offerings is achieved; increased business and business opportunities for IoT technology providers **Joint offerings** can be quite **reasonable in price** (min. ten(s) of thousands to max. hundred of thousand(s) Technology development is already on **a high readiness level** **Common understanding** on co-operation potential and possibilities in value creation i.e. joint offerings could increase and maximise value; co-creation and data sharing can lead to better decisions Clear signs of market change, moving **from CAPEX (capital expenditure) towards OPEX thinking (operating expenditure)** Increasing discussion and demand for **performance based contracts** **Pricing and business models changes**: shared risk models, risk-reward models, success fee e.g. yearly milestones, outcome economy based benefit sharing business model **Expertise and competence increase** both in pulp and paper industry and in IT sector
- Negative	Concept of the joint offering is unclear, which can cause **misunderstandings** also in business-wise **"Build and forget**" type of projects can cause challenges. Continuity of collaboration and continuous improvements after the project would lead to better results **System integration challenges** might have an impact on business values **Complexity** increases **Different IT knowledge and competence level** in various plants and companies **More reference cases are needed** to demonstrate the business benefits of joint offerings **No clear costing and investment appraisal** model **Unclear value of the collected data**. Price of the data needed when developing joint offerings

4.2 Cost and Revenue Drivers of Digital Technologies in Capital-Intensive Industries

From the pulp and paper industry point of view, the interviews included also the discussion on the main cost and revenue drivers of digital solutions (Fig. 1). For example, regarding the scheduled maintenance (both spare parts and service work) and unavailability, there is a clear linkage between economic aspects, risk and reliability issues.

For example, for an operator at pulp and paper plant, it is important to maximize the uptime of the production assets and to keep the production lines constantly in service. Reliability issues and scheduled maintenance periods interrupt the operation and increase the cost. Moreover, due to optimized processes, also the scheduled maintenance can more costly than expected and should therefore be minimized. The joint offerings can be created to focus for solving these challenges e.g. on the development of predictive maintenance solution to increase the predictability and thus to reduce the need for scheduled maintenances and downtime due to failures.

Fig. 1. Cost and revenue drivers of digital technologies deployed in the pulp and paper industry.

4.3 Value Network Analysis Identified Two Collaboration Models

In literature, two collaboration models identified are i) traditional bilateral supplier–customer relationships, and ii) multi-actor models based on system integration or open boundaries of the digital platform [5]. Both models have strengths and weaknesses; the traditional bilateral supplier–customer relationships ensure effective co-operation, whereas multi-actor models could boost novel business models or even solve the challenges regarding so called vendor lock-ins.

Within *the bilateral model* [5], the Use Case (UC) owner of the SEED project provides both requirements and data to the solver company (Fig. 2). The selected solver company implements and delivers a digital solution and will receive a fee. The questions related to this model concern the sustainable earning logic of the solver and the return of investment for the use case owner. Traditionally this means a closed contract where the requirements are specified in detail and the reward is fixed. In many cases the requirements will change and the software modifications will cause extra work to solver and extra costs to the UC owner. For the UC owner the total cost cannot be foreseen. Another invoicing model is to agree on the development phase reward (fee) and monthly service maintenance fee. These two models include reward for a separate development phase unlike a monthly or bi-annual service fee that is based on a Service Level Agreement (SLA). In addition to those the invoicing could be periodic and based on the business impact of the solution. This is often challenging as the business impact of the solution may not be transparent. Also the mutual understanding and agreeing on the Key Performance Indicators (KPIs) is required.

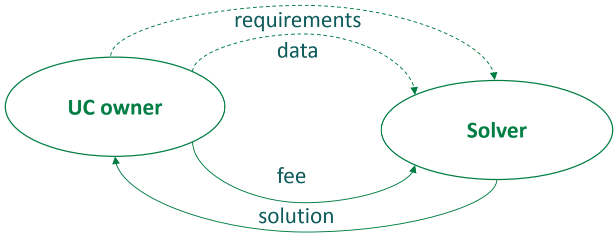


Fig. 2. Bilateral model with single solver does not create any joint offering. The notation is based on Verna Allee's Value Network analysis (VNA) [21].

The collaborative multi-actor model [5] in the SEED project is a case with *several solvers having complementary technological (*software implementation) *skills* (Fig. 3). Skills can vary from capability to provide data interfaces via analytics to user interfaces. In this model each solver has their own niche competences and they are not competing with each other. How this could happen?

Is it possible for an UC owner to pick a group of interacting solvers without any existing network or ecosystemic project where the solvers are in principle committed to solve the challenges of an UC company? How to interact with the group of selected solvers? How to share costs and profit among solvers? Shall one of the solvers become a coordinator intermediating both the requirements and rewards? In case the UC owner agrees bilateral with each solver company, who will grant the interoperability of separate sub-solutions? Usually an integrator is needed. Does the integrator act as the contact point to UC owner both delivering solution and sharing fees? What is the basis for fees of the sub-solution solvers as each sub-solution is necessary to deliver the total solution? Does the integrator carry total responsibility (with potential risks and profits) and purchase the sub-solutions from each solver? In this case it is just hierarchical bilateral subcontracting. Still, a total joint solution is delivered. Did it also create joint offering among the solvers? This depends mainly on the scalability and modularity of the solution as well as the agreement of ownership of the jointly created solution.

In case the sub-solvers deliver the data interfaces and intelligence (analytics etc.) the integrator does not to have access on the raw data (grey dotted line). Integrator creates the user interface to display the result or status. In case the solution includes a control loop, the feedback signal goes via the data interface included in the solution.

It is clear that the bilateral model does not create any joint offering between solvers as there is only one deliverer. It might still bring the solver company to a new domain or even to international markets if the solution is scalable and general enough with modular interfaces. The multi-solver model delivers a total solution. Will it also became a joint offering and create new business still is open. Is the solution scalable? Further, does the contracts allow to scale a solution to another company or another domain? Is a common (technical) platform the best solution to grant common interfaces, modularity and scalability for joint offerings? How to grant sustainable business in a multi-solver case? Which are the earning logics of solvers and ROI of UC owner? Further, can an ecosystem boost shaping the (international) markets for joint offerings?

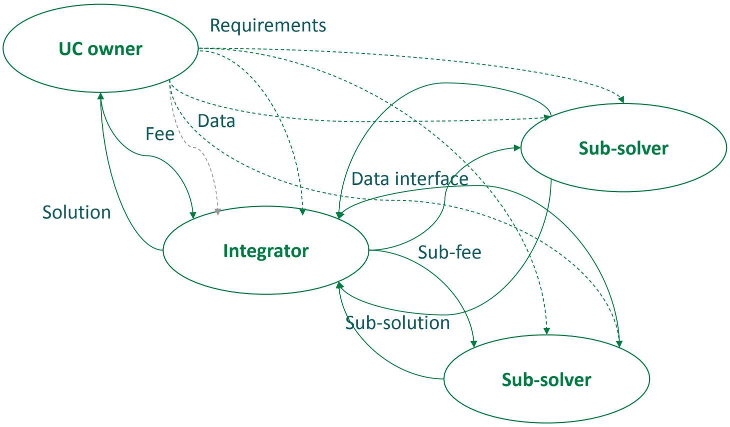


Fig. 3. Multi-solver model with complementary technical skills.

5 Concept for Joint Offering Development - Business and Value Potential Assessment

In order to guide the research, a conceptual framework indicating the different analyses associated with the joint offering development, was co-created in the SEED project (Fig. 4). The starting point is an intention for a joint offering among technology developers and a perceived customer need for improvements that can be fulfilled by the joint offering. The purpose of the framework is to give a high-level understanding of the business opportunities and risks of joint offerings and of the mechanisms and benefits of value networks in the joint offering development. The framework also includes the analyses and the estimations of revenue potential and cost impacts of joint offerings as well as the identification of the most promising business models for joint offerings from the business perspective.

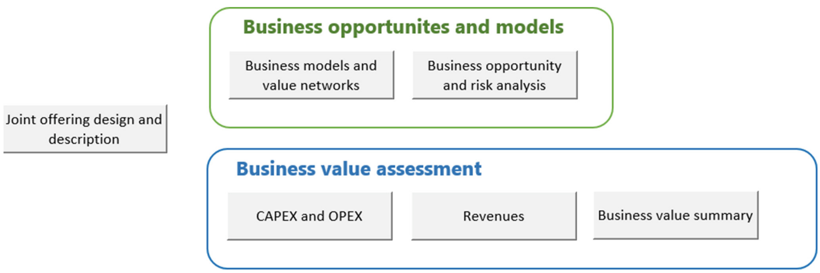


Fig. 4. The conceptual framework for the joint offering development.

The framework contributes to asset performance management research by providing a new conceptual understanding of joint offerings and the value creation that occurs in the interplay of parties in value networks. In the next phases of the SEED project, we will further research the value dimensions of joint offerings and provide empirical insight in respect of metrics to illustrate the advantages of joint offerings for the pulp and paper industry. Moreover, the conceptual framework (Fig. 4) will be further evaluated and tested with the SEED solver companies.

It is also evident that companies need to devise and reinvent asset performance management tools and strategies that are suitable for the joint offering development and for the value creation in the collaboration network. As a practical implication of the research in the SEED project, we will develop a practical toolset supporting the joint offering development and including various analyses depicted in the figure. The toolset is especially aimed at IoT companies and supporting them in creating joint offerings for the pulp and paper industry. Further, we will research how an ecosystem can boost shaping the (international) markets for joint offerings.

Acknowledgments. The research is part of the Business Finland ecosystem project SEED "Solid value of digitalisation in forest industry" which is jointly funded by Business Finland, participating companies and research parties. The authors want to thank all the interviewed companies and key persons for insightful insights, all the information and discussions during the interviews. We also give our sincere thanks to Essity, UPM, Metsä Group for providing the use cases for the research.

References

1. Schlaepler RC, Koch M (2015) Industry 4.0. challlenges and solutions for the digital transformation and use of exponential technologies. Zurich
2. Kakabadse A, Kakabadse N (2002) Trends in outsourcing. Eur Manag J 20(2):189–198
3. Jansen S (2020) A focus area maturity model for software ecosystem governance. Inf Softw Technol 118:106219
4. Nicolescu R, Huth M, Radanliev P, De Roure D (2018) State of the art in IoT - beyond economic Value, pp 1–16
5. Valkokari K, Valkokari P, Kortelainen H, Nyblom J (2020) Building Business impacts of an industry 4.0 ecosystem through collaborative network settings between IT and forest companies. In: Camarinha-Matos LM, Afsarmanesh H, Ortiz A (eds) Boosting Collaborative Networks 4.0: 21st IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2020, Valencia, Spain, November 23–25, 2020, Proceedings. Springer, Cham, pp 463–474. https://doi.org/10.1007/978-3-030-62412-5_38
6. Sorri K, Seppänen M, Still K, Valkokari K (2019) Business model innovation with platform canvas. J Bus Model 7(2):1–13
7. Hassett M, Räikkönen M, Rantala T (eds) (2011) M&A as a Strategic Option: From Opportunities to New Business Creation. Teknologiainfo Teknova oy, Helsinki
8. Grönroos C, Voima P (2013) Critical service logic: making sense of value creation and co-creation. J Acad Mark Sci 41(2):133–150
9. Bowman C, Ambrosini V (2007) Firm value creation and levels of strategy. Manag Decis 45(3):360–371
10. Meyer C, Schwager A (2007) Understanding customer experience. Harv Bus Rev 85(2):116–128
11. Viljakainen A, Hemilä J (2016) Theme digitalisation: How do business models change with digitalisation? VTT blog https://www.vttresearch.com/en/news-and-ideas/theme-digitalisation-how-do-business-models-change-digitalisation
12. Simon H (1994) The Sciences of the Artificial, 3rd edn. The MIT Press, Cambridge

13. Valkokari K, Amitrano CC, Bifulco F, Valjakka T (2016) Managing actors, resources, and activities in innovation ecosystems – a design science approach. In: Afsarmanesh H, Camarinha-Matos LM, Soares AL (eds) Collaboration in a Hyperconnected World: 17th IFIP WG 5.5 Working Conference on Virtual Enterprises, PRO-VE 2016, Porto, Portugal, October 3-5, 2016, Proceedings. Springer, Cham, pp 521–530. https://doi.org/10.1007/978-3-319-45390-3_44
14. Vesalainen J, Hellström M, Valkokari K (2017) Managerial tools and the network-as-practice perspective. In: Vesalainen J, Valkokari K, Hellström M (eds) Practices for Network Management. Springer, Cham, pp 323–339. https://doi.org/10.1007/978-3-319-49649-8_22
15. Burke, W.: Organization Change. Theory and Practice. SAGE Publications, Inc., Los Angeles (2002)
16. Thomasson T, Sihvonen T, Rantanen T, Hytönen E (2021) Building a Digital Twin: a case study for a Finnish fibre process research environment. In: PWC 2021 PaperWeek Canada
17. Van der Meer-Kooistra J, Scapens RW (2015) Governing product co-development projects: the role of minimal structures. Manage. Acc Res 28:68–91
18. Tjemkes B, Burgers K, Vos P (2012) Strategic Alliance Management. Taylor & Francis Group, New York
19. Belderbos R, Cassiman B, Faems D, Leten B, Van Looy B (2014) Co-ownership of intellectual property: exploring the value-appropriation and value-creation implications of co-patenting with different partners. Res Policy 43(5):841–852
20. de Man AP, Luvison D (2019) Collaborative business models: aligning and operationalizing alliances. Bus Horiz 62(4):473–482
21. Allee V (2002) The Future of Knowledge: Increasing Propsperity though Value Networks. Butterworth-Heinemann, Oxford (2002)
22. Nathan AJ, Scobell A (2012) How China sees America. Foreign Aff 91(5):1689–1699
23. Parv L, Deaky B, Nasulea M, Oancea G (2019) Agent-based simulation of value flow in an industrial production process. Processes 7(2):82
24. Eckhardt J, Kaletka C, Krüger D, Maldonado-Mariscal K, Schulz AC (2021) Ecosystems of co-creation. Front Sociol 6:1–11
25. Bokrantz, J.: Smart Maintenance-Maintenance in Digitalised Manufacturing. Chalmers University of Technology, Gothenburg (2019)

The Journey Towards Successful Application of Maintenance 4.0 and Service Management 4.0

Mirka Kans(✉)

Department of Mechanical Engineering, Linnaeus University, 351 95 Växjö, Sweden
mirka.kans@lnu.se

Abstract. The paper discusses the role of maintenance in the digital era and propose directions for supporting the technology and business transformation towards Maintenance 4.0 and Service Management 4.0. For achieving this, the paper summaries previous research in the area and conceptualizes the transformation of maintenance towards Maintenance 4.0 and Service Management 4.0. The concept applies a systems approach on digitalization and recognizes the need for combining several working areas (production and maintenance management, information systems management, improvement processes and business management) for successful digital transformation within maintenance. Service Management 4.0 describes how maintenance becomes a business opportunity and modular maintenance offerings a way to approach the new opportunities. The concepts proposed in the paper provide support for industries in the digital transformation process towards Maintenance 4.0 and Service management 4.0.

1 Introduction

Ever since we entered the digital era, rapid changes in industrial business as well as daily life have been a reality. New customer behaviour affects industrial production; individualized products delivered with precision from all around the world is expected, which puts demands on efficiency and adaptivity in the production systems, and demands new kind of production processes where for instance the customer is a part of the development process, so called co-creation [1–3]. In addition, the focus is switching from products and their technical specifications towards value and utility that a product, or a product combined with additional services, could offer [2]. The quality and variety of services becomes important, as it is hard to compete solely with product quality. Above mentioned requirements must be reached in a sustainable way, which complicates the situation even more. Maintenance is a service that gained a lot of attention the past decades, as it has large effect on production output as well as on asset sustainability.

Maintenance is affecting the asset's full lifetime, from the conceptualization to scrapping [4], in the form of design choices, restoring or conserving activities until decisions regarding whether to prolong the lifetime with additional maintenance or refurbishing activities or not. The digitalization of the industry gives great possibilities to support the maintenance activities with reliable and relevant data and decision-making capabilities [5–8]. However, this development seems to be slow. The reason behind low utilization of smart technology in maintenance are technical as well as organizational and managerial

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 137–147, 2022.
https://doi.org/10.1007/978-3-030-96794-9_13

[8]. This paper will discuss the role of maintenance in the past and in the digital era and propose directions for supporting the shift of maintenance practices towards smart utilization of maintenance in the form of Maintenance 4.0 and Service Management 4.0.

2 Maintenance Transformation from an Industry 4.0 Perspective

The physical assets used in production have undergone drastic changes over the years, see Fig. 1. The role of maintenance could be described in the view of this industrial development, providing a framework in which the maturity of an organization or enterprise could be understood. This supports the transformation of maintenance, especially as many organizations still rely on maintenance practices related with Industry 1.0 and 2.0.

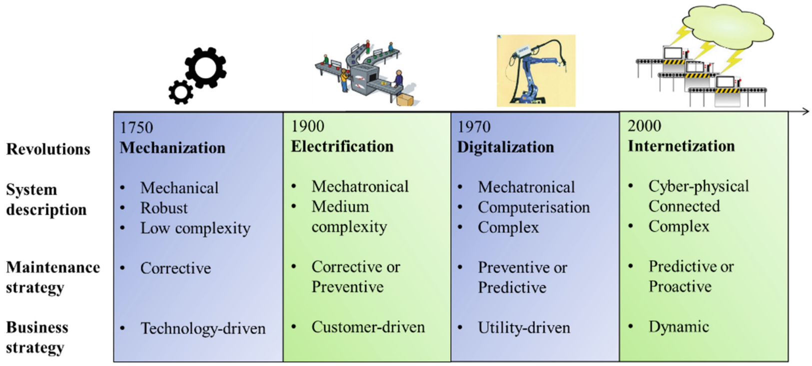


Fig. 1. Maintenance transformation from an Industry 4.0 perspective

The introduction of the steam engine, which made localization of factories possible everywhere is referred to as the first industrial revolution. Industry 1.0 was characterized by mechanical equipment with robust construction and quite low complexity, compared with today's equipment. Main maintenance strategy used was "run to failure", i.e. a corrective strategy. This is partly due to lack of knowledge regarding failure mechanisms as statistical data on failures were missing. The second revolution happened around 1900 when the production was electrified and standardized. With the electrification and standardization of production, the dependencies between different production assets became higher, requiring a better planning of maintenance for keeping availability high. Especially after World War II a scientific approach on maintenance was applied, such as statistical modelling of failure modes [9]. The digitalization of industry is referred to as the third revolution, characterized by seamless digital information flows, decision support, production automation, and focus on distribution rather than on production [10]. Highly automated equipment required new maintenance strategies as well, as the traditional corrective and preventive strategies were not reliable enough. A risk-based approach (Reliability-Centered Maintenance) was proposed around 1970 [11]. For critical assets condition monitoring was seen as the best basis for maintenance planning and scheduling, and new technologies for detecting and assessing severity of failures as well as predicting the future state were developed. Industry 4.0 is the extension of the digitalization in the form of interconnectivity and smart equipment [12]. One could describe it as the internetization of industry. A key concept is Internet of Things, i.e. physical assets that are interconnected and coordinated through Internet technology [13]. A key concept of Industry 4.0 are cyber-physical systems, i.e. assets with built-in intelligence, allowing the physical asset to be represented as a virtual twin where e.g. simulations could be run [14]. With the emergency of Industry 4.0 maintenance planning will increasingly become an integrated part of production planning, requiring dynamics in when to do maintenance and what actions to do, to optimize the production output, see e.g. [15].

Just as the technical advances and increased digitization will change the way we produce it will radically change the way we understand and manage value creation. Access to large amount of data and increased number of stakeholders involved in value creation enables more complex business models and collaboration patterns. Business modelling, i.e. development of new business models or redesign of current models has received increased attention as a way of developing and maintaining competitiveness. Business models define how the business is to be conducted, for example in terms of strategies, market segments, product platforms, resource requirements, customer relations and value creation mechanisms [16]. In Fig. 1 the main service business strategy for each development stage is found based on the concept of Service Management 4.0 [1, 2, 17]. The business model of Industry 1.0 is driven by technical competence manifested in the physical product. The main maintenance strategy is corrective; therefore, no maintenance planning or management is necessary. Maintenance is solely the responsibility of the buyer, conducted by the buyer or bought when necessary. This technology-driven approach is found for instance in the form of manufacturers supplying original manufacturer spare parts and specialized technical competence regarding the machine. The product could also be designed with locked-in effect, making it impossible to change spare parts without contacting the original manufacturer, such as the Iphone battery. Industry 2.0 is characterized by standardization, mass-production, and straightforward value chains: customers are classified according to the anticipated needs they would have. This is a customer-centered perspective, i.e. the manufacturer tries to define what the customer wants and thereafter acts according to this. The maintenance provider, either the original manufacturer or a service provider, offers preventive maintenance according to the standardized maintenance plan of each specific asset type. The relationship to customers and distribution channels could be quite impersonal because of the standardization. This business model is quite common today for many types of assets.

The maintenance service focus of Industry 3.0 is to maintain the functionality of the asset. The complexity of the assets in Industry 3.0 is reflected in the business models in the form of combined offerings where the product and aftersales services often are sold together as a utility-driven offering, i.e. focus is on the explicit as well as the implicit needs of the customer in form of asset functionality. For Industry 4.0, the dynamics in the service business model will become increasingly important. The business model of Industry 4.0 is a value proposition that goes beyond the physical appearance and the requirements on functionality. The focus is instead on supportability for the production, and therefore dynamics in the maintenance activities are important as well [17]. The dynamics could be in time; to find the most optimal time or frequency for maintenance tasks to maximize production performance, but also in scope, i.e. to what and how much to do when a maintenance intervention is made. This is enabled by Maintenance 4.0 and proactive maintenance in the form of Service Management 4.0.

3 The Future of Maintenance

Until now, maintenance has mainly been treated as a separate function and as a cost driver, which might result in a reactive approach to maintenance. Instead, proactivity and value will be catchword for maintenance of the future. A value-driven maintenance approach provides availability, quality, safety, or other production-related objectives [18]. Proactivity means avoiding problems and disturbances in the production by appropriate maintenance [19], but it could be extended to mean sound decisions regarding design or prolongation of equipment lifetime as well. This requires a holistic and integrated perspective on technical assets, switching focus from maintenance management to asset management [20] and coordination of stakeholders within ecosystems rather than focusing on internal company business [1, 2].

3.1 Maintenance 4.0

Maintenance 4.0 is a subset of industry 4.0 with emphasis on the prospects of maintenance that involves automatic data collection, analysis, visualization, and decision making for assets [21, 22]. Maintenance 4.0 supports predictive and proactive maintenance strategies and a holistic view on assets in terms of asset management. The core is found in the possibilities to apply advanced technology for condition monitoring and diagnostics, thus supporting the process of predictive maintenance as described in the standard ISO 13374 [23], see Fig. 2. *Internet of Things* allows health data to be collected simultaneously from several production assets, which creates a basis for advanced data analysis, predictive maintenance, and coordination of resources [7]. *Big data* sets are suitable for advanced statistical and mathematical modelling, for instance by utilizing machine learning algorithms and methods [8]. *Big data analytics* enables the decision making based on understanding of the assets and their physical behavior, but also on yet unrevealed correlations between different system parameters and environmental parameters [24, 25]. *Visualization* is a way to present large data set in a comprehensive manner [26]. The enhanced data analysis capabilities are directly applicable in the area of prognostics diagnostics, but also for planning, scheduling and optimization [6]. In particular, the objective for *smart maintenance* is minimizing time for maintenance and maintenance costs as well as risks and accidents, while maximizing device life and reliability, and optimizing the maintenance strategy and maintenance plan. The smart data analytics will provide a holistic picture of the operations and the machinery, maintenance planning becomes more dynamic and adapted to machine operating conditions and age, automatic warnings occur when something abnormally occurs, otherwise business as usual, and business-effective decisions could be reached: not over maintaining while avoiding unplanned stoppages, production disturbances and excessive spare parts inventories. *Cyber-physical systems* that automatically monitors and predicts their degradation is an example of the smart assets of the future [14].

Maintenance 4.0 is characterized by self-adapting and self-healing CPS [14, 27]. The *self-adapting* capabilities could for instance be in the form of artificial intelligence that utilizes health information for adapting the production speed or regulating functionality to avoid stoppages and breakdowns. *Self-healing* could be in form of automatic maintenance conducted by the machine, for instance increasing the lubrication, based on condition data. The smart assets can automatically detect changes, analyze the current state, repair or perform maintenance, adjust performance, alerts about failures, and generates maintenance action. The impact of Maintenance 4.0 is not only to be seen on managerial level. Smart technologies will also ease the daily work of the maintenance technician. Digital and automated inspection rounds and reporting procedures reduce time and competence required, real-time data provides better support for task prioritization, and automatic failure reporting and work order generation eases the daily and weekly planning [6]. The more automated the maintenance process becomes the less failures are to be expected. This is of course a huge advantage, as less resources are spent on maintenance, but it also leads to less practice for the maintainers. Learning and training will thus become more important in the future. *Virtual and augmented reality* (VR/AR) provides good opportunities for the technicians to practice complicated situations or such situations that rarely happens in real life [28]. In addition, VR/AR is a good support during actual execution of maintenance activities [29].

Increased digitalization, automation, and interconnectivity brings new threats to the business as well, e.g. in the form of technology heterogeneity and information security problems [8, 30]. The technical core for realizing Maintenance 4.0 exists today, but not as a “one-system-fits-all” package, like the Enterprise Resources Planning system. Instead, it is up to every enterprise to define and create their own solutions, or to wait until such off-the-shelf-systems emerge. In either case, the successful digitalization is dependent on the capabilities of the company, and the ability to utilize information technology (IT) for reaching business goals, i.e. the IT maturity of an organization will set the baseline. Technology investments should always be aligned with the needs of the business, and not vice versa. Thus, it is important to stop focusing on digitalization as “a technology solution for everything”. Instead, the goal is finding the digitization solution that best leads to highest benefits based on core competencies and conditions of the enterprise [31]. Moreover, large technology projects have an increased risk of failing, so prioritization between different projects is important. Smart technology solutions must be combined with smart people and a smart approach to digitizing. This is the basic idea behind the approach “The thinking industry” [31]. The motto is “think hard, think right, think system(atically)”, i.e. a systems perspective on Maintenance 4.0. The thinking industry combines smart people, smart processes, and smart technology solutions, to reach smart solutions that address real problems or opportunities at hand. The approach proposes combining theories, methods, and tools from areas of information systems and project management, maintenance and production management, systematic improvement work, and strategic management and business modelling.

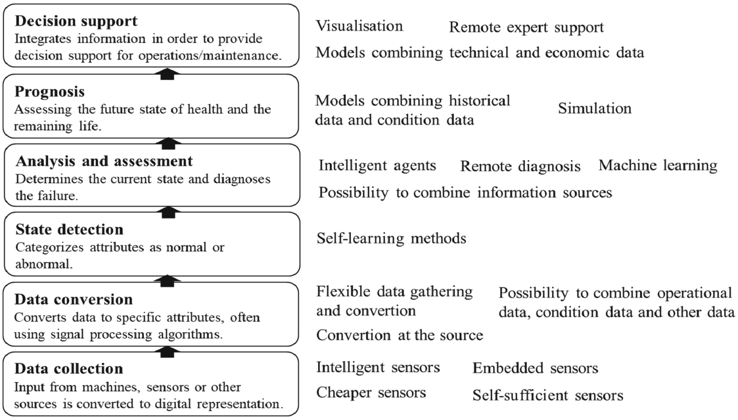


Fig. 2. Key technologies of maintenance 4.0, inspired by [23]

3.2 Service Management 4.0

Service management 4.0 is all about delivering value to the customer, where the physical product and the maintenance are fully integrated in the value proposition. The value creation is a result of several actors involved, each getting a share of the profit; sometimes even the customer gets a part of the profit, if profit-sharing mechanisms are used for regulating contracts. The journey towards Service Management 4.0 has already started, but there is still a long way to go. A trend is to combine or integrate products and services into new types of value offerings [1, 32–34]. Another trend is to sell after-market services based on performance rather than function [35, 36]. This means, for example, selling service and maintenance based on specified technical performance or production instead of number of maintenance hours or predefined maintenance intervals. However, to achieve business opportunities that could be achieved by providing after-market services, the view of the business must change [32], and digitalization initiatives must be coordinated with and reflected in the strategic business models [2]. A holistic view on the value creation process, as well as on the value proposition, is required [34]. The value could be delivered in the form of performance-based and collaborative contracts, regulating not only the performance of maintenance, but also the production output, the long-term relationships as well as the information sharing principles, see e.g. [35]. Today, companies interact in complex patterns, and the border between the roles could be diffuse; one actor could act as supplier and producer in the same business environment and new actors enters rapidly while others disappear. For better capturing the complex and dynamic business environments the term business ecosystem could be used [37]. The business ecosystem describes all actors involved in the value generating process, but also other stakeholders that possess power and could influence the value creation, such as competitors, companies producing substitutes, standardization organizations, public authorities, and customer groups [38]. New business models are knowledge and information intensive; information technology is therefore a prerequisite for its implementation. The technology development in form of Industry 4.0 brings more opportunities to exchange information within and between organizations, such as in the form of open data sources. Open data sources also imply potential security risks and therefore it is important to find ways for safe information management in the value chain [39]. Asset as a Service (AaaS), a virtual representation, or ontology, of tangible and intangible assets is a way facilitate communication and collaboration between actors in the business ecosystem and could be a solution to the data sharing problem. AaaS creates a holistic view on asset information, and with a neutral third party that manages and governs the AaaS, there are possibilities to create a shared common information platform without risking privacy and IT security problems [39].

Modular maintenance offerings apply ecosystem reasoning as well as an integrated approach to service contacts, but adds the dimensions of standardization, progression, and dynamics [17]. Different types of service contracts should be offered to customers, but instead of great diversity in contract forms and individualized contracts, leading to almost one contract type per customer and great administrative inefficiency, the offerings should be based on the core competencies of the service provider, and the needs and behavior of the customer. Modular maintenance offerings, see Fig. 3, classify maintenance services offerings with increasing integration of the offering, from a traditional technology-driven approach to a utility-driven approach.

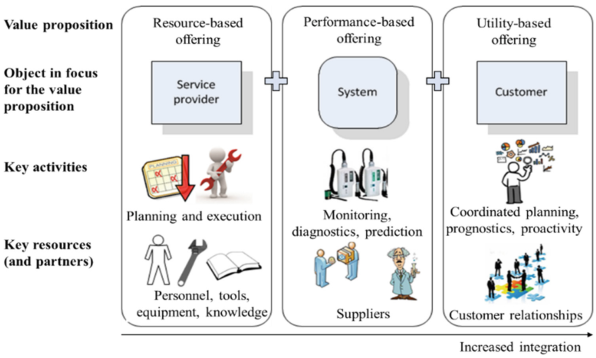


Fig. 3. Modular maintenance offerings

Working with only a few variations of these three basic contract types, the administration of contracts as well as the maintenance activities could be streamlined. The basic value proposition is maintenance as a resource, such as corrective maintenance, predetermined scheduled maintenance or providing spare parts. This reflects the core competencies of the service provider in terms of key resources and activities. The customer relationship is not regular, and many customers might buy services only once. This allows for simplicity in the value distribution as well, for instance a web shop offering original spare parts. The offering is considered as "bulk" business available for all customers. In the next level maintenance is considered from the perspective of the asset's or system's performance. Key competencies for performance-based offerings are the ability to monitor and predict system behavior, i.e. predictive maintenance capabilities. The contract is regulated based on system performance measures, such as lifted tons, kilometers run or number of products per time unit. Leasing-based contracts or pay-per-use are other types of performance-based offerings. Customers with process-critical assets and with low maintenance knowledge, or no own maintenance function, would benefit from performance-based offerings. Utility-based offerings are linked to the customer's results and the production performance. Focus is on the enterprise level of the user, i.e. a true customer-focused offering by applying an asset management perspective. The utility-based offerings are based on trust between the customer and the service provider. The customer migrates not only a technical risk, but also a business risk, when entering a utility-based contract. This type of offering is therefore suitable for customers with high productivity demands, and where the production is critical, such as power generation or transportation. The customer relationship becomes an important resource – without willingness to share information regarding production and operations, the utility goal might be at risk.

4 Conclusions

Maintenance evidently becomes more important the more complex the production systems and assets becomes. In the smart factories of the future, data-driven and autonomous maintenance will become a key factor for the production, requiring a holistic approach throughout the asset life cycle and on the production, see Fig. 4. Maintenance will be planned, executed, and followed up based on true conditions regarding component and machine health, taking into consideration the whole environment when creating optimal maintenance plans and decision support. Smart systems will conduct simple maintenance by themselves and regulate the performance if needed. Smart factories will rely on smart people also in the future, and digitalization will support the people in daily work, on tactical level as well with making smart strategic decisions. For reaching highest benefit of digitalization projects within maintenance, these should be based on the needs of the organization and run as cross-functional improvement projects rather than as traditional ICT projects. Moreover, maintenance has great potential to become a business opportunity for those who provides maintenance services. Customers could benefit from new service-based business models, as the possibility to acquire integrated support for the asset, or even the whole production, shifts the focus from support activities to the core business activities. The business models of the future therefore require a holistic and integrated approach, putting less focus on maintenance as a support activity, and more on maintenance as a value creating activity.

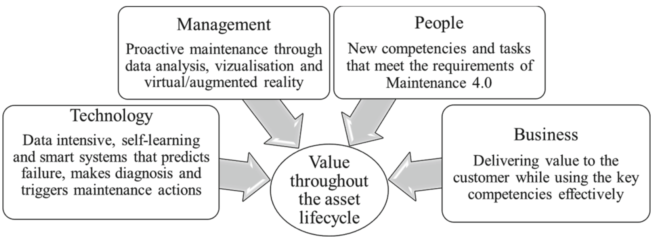


Fig. 4. Value through maintenance 4.0 and service management 4.0

If the author would try to depict the maintenance of the future, following predictions are made. Future research on the technical support for predictive and proactive maintenance is required to realize Maintenance 4.0, as well as on the integration and orchestration of different systems. This includes research on all levels, from smart and self-powered sensors, to decision support, expert support, and visualization. Research is also needed on the managerial and organizational aspects of Maintenance 4.0, for instance factors affecting the efficient implementation of Maintenance 4.0, project management methodologies suitable for maintenance digitalization, as well as information sharing capabilities within the organization and between different actors in the ecosystem. In addition, understanding business logics and creating reliable service-based business models, including risk management, pricing strategies and key performance indicators for managing and follow up, is needed.

References

1. Kans M, Ingwald A (2016) Business model development towards service management 4.0. Procedia CIRP 47:489–494
2. Kans M, Ingwald A (2016) A framework for business model development for reaching service management 4.0. J Maint Eng 1:398–407
3. Braun EL, Pereira GM, Sellitto MA, Borchardt M (2017) Value co-creation in maintenance services: case study in the mechanical industry. Bus Process Manag J 23(5):984–999
4. EN 13306 (2010) Maintenance - Maintenance terminology
5. Kans M (2013) IT practices within maintenance from a systems perspective: study of IT utilisation within firms in Sweden. J Manuf Tech Manage 24:768–791
6. Bokrantz J, Skoogh A, Berlin C, Stahre J (2017) Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. Int J Prod Econ 191:154–169
7. Civerchia F, Bocchino S, Salvadori C, Rossi E, Maggiani L, Petracca M (2017) Industrial Internet of things monitoring solution for advanced predictive maintenance applications. J Ind Inf Integr 7:4–12
8. Kans M, Galar D (2017) The impact of maintenance 4.0 and big data analytics within strategic asset management. In: Maintenance, Performance, Measurement & Management: Conference Proceedings, MPMM 2016, pp 96–103
9. Peng K (2012) Equipment Management in the Post-Maintenance Era. Productivity Press
10. Tien JM (2012) The next industrial revolution: Integrated services and goods. J Syst Sci Syst Eng 21(3):257–296
11. Moubray J (1997) Reliability-centred maintenance. Industrial Press Inc
12. Lasi H, Fettke P, Kemper H-G, Feld T, Hoffmann M (2014) Industry 4.0. Bus Inf Syst Eng 6(4):239–242. https://doi.org/10.1007/s12599-014-0334-4
13. Ashton K (2009) That 'internet of things' thing. RFID J 22(7):97–114
14. Eifert T, Eisen K, Maiwald M, Herwig C (2020) Current and future requirements to industrial analytical infrastructure—part 2: smart sensors. Anal Bioanal Chem 412(9):2037–2045
15. Mollaverdi N, Mirabadi S (2015) Dynamic Maintenance in series – parallel manufacturing systems with consideration of buffer inventory. In: Proceedings of the 2015 International Conference on Industrial Engineering and Operations Management, Dubai, United Arab Emirates (UAE)
16. Osterwalder A, Pigneur Y (2010) Business model generation: a handbook for visionaries, game changers, and challengers. Wiley, Hoboken

17. Kans M, Ingwald A (2019) Modular-based framework of key performance indicators regulating maintenance contracts. In: Joseph Mathew CW, Lim LM, Sands D, Cholette ME, Borghesani P (eds) Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies. LNME. Springer, Cham, pp 301–310. https://doi.org/10.1007/978-3-319-95711-1_30
18. Stenström C, Parida A, Kumar U, Galar D (2013) Performance indicators and terminology for value driven maintenance. J Qual Maint Eng 19(3):222–232
19. Léger JB, Iung B, Morel G (1999) Integrated design of prognosis, diagnosis and monitoring processes for proactive maintenance of manufacturing systems. In: Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics. IEEE SMC, vol. 4, pp 492–498
20. Maletic D, Maletic M, Al-Najjar B, Gomiscek B (2020) An analysis of physical asset management core practices and their influence on operational performance. Sustainability 12(21):1–20
21. Kans M, Galar D, Thaduri A (2016) Maintenance 4.0 in railway transportation industry. In: Koskinen K et al (eds) Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015). Lecture Notes in Mechanical Engineering. Springer, Cham. pp 317–331 https://doi.org/10.1007/978-3-319-27064-7_30
22. Aksa K, Aitouche S, Bentoumi H, Sersa I (2021) Developing a web platform for the management of the predictive maintenance in smart factories. Wireless Pers Commun 119(2):1469–1497. https://doi.org/10.1007/s11277-021-08290-w
23. ISO 13374–2 (2007) Condition monitoring and diagnostics of machines - Data processing, communication and presentation - Part 2: Data processing
24. Rødseth H, Schjølberg P, Marhaug A (2017) Deep digital maintenance. Adv Manuf 5(4):299–310. https://doi.org/10.1007/s40436-017-0202-9
25. Marquez AC, Fernandez, JFG, Fernández, PM-G, Lopez, AG (2020) Maintenance management through intelligent asset management platforms (IAMP). Emerg factors key impact areas data models. Energies 13(15):3762
26. Penna R et al, (2014) Visualization tool for cyber-physical maintenance systems. In: 2014 12th IEEE International Conference on Industrial Informatics (INDIN), pp 566–571
27. Giese H (2017) Formal models and analysis for self-adaptive cyber-physical systems. In: Kouchnarenko O, Khosravi R (eds) Formal Aspects of Component Software. Springer, Cham, pp 3–9. https://doi.org/10.1007/978-3-319-57666-4_1
28. Manca D, Brambilla S, Colombo S (2013) Bridging between virtual reality and accident simulation for training of process-industry operators. Adv Eng Softw 55:1–9
29. Borro D, Suescun Á, Brazález A, González JM, Ortega E, González E (2021) WARM: wearable AR and tablet-based assistant systems for bus maintenance. Appl 11(4):1–20
30. Campos J, Sharma P, Jantunen E, Baglee D, Fumagalli LM (2016) The challenges of cybersecurity frameworks to protect data required for the development of advanced maintenance. Procedia CIRP 47:222–227
31. Kans M, Campos J, Salonen A, Bengtsson M (2017) The thinking industry – an approach for gaining highest advantage of digitalization within maintenance. J Maint Eng 2:147–158
32. Kujala S, Kujala J, Turkulainen V, Artto K, Aaltonen P, Wikström K (2011) Factors influencing the choice of solution-specific business models. Int J Proj Manage 29(8):960–970
33. Visnjic Kastalli I, Van Looy B (2013) Servitization: disentangling the impact of service business model innovation on manufacturing firm performance. Int J Oper Prod Manage 31(4):169–180
34. Kowalkowski C, Witell L, Gustafsson A (2013) Any way goes: Identifying value constellations for service infusion in SMEs. Ind Mark Manage 42(1):18–30
35. Ng ICL, Xin DD, Yip N (2013) Outcome-based contracts as new business model: the role of partnership and value-driven relational assets. Ind Mark Manage 42(5):730–743

36. Famurewa SM, Asplund M, Galar D, Kumar U (2013) Implementation of performance based maintenance contracting in railway industries. Int J Syst Ass Eng Manag 4(3):231–240
37. Moore JF (1993) Predators and prey. Harv Bus Rev 71(3):75–86
38. Ingwald A, Kans M (2016) Service management models for railway infrastructure, an ecosystem perspective. In: Koskinen K et al (eds) Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015). Lecture Notes in Mechanical Engineering. Springer, Cham. pp 289–303. https://doi.org/10.1007/978-3-319-27064-7_28
39. Metso L, Kans M (2017) An ecosystem perspective on asset management information. Manage Syst Prod Eng 25:150–157

Risk, Reliability and Maintenance

Methodology for Optimizing Preventive Maintenance Programs for Equipment on an Electrical Distribution Network

Gabrielle Biard, Karim Brunet-Benkhoucha, Raynald Vaillancourt(✉), and Georges Abdul-Nour

Université du Québec à Trois-Rivières, 150 Rue Sainte-Catherine O, Montréal, QC, Canada
Vaillancourt.raynald@hydro.qc.ca

Abstract. This article outlines the methodology developed by Hydro-Québec Distribution (HQD) for optimizing the maintenance policy related to equipment on an electrical distribution network based on the estimated useful life, reliability and replacement costs. This methodology relies on the failure modes analysis as well as on the statistical analysis of operational data and safety risk. Its application will be demonstrated using MV three-phase gang operated overhead switches, but the methodology is applicable to assets on a distribution network with a preventive maintenance program, including equipment refurbishment. The application of the optimal preventive maintenance policy obtained with the proposed methodology on approximately 5000 equipment generated labour gains of more than 2,000 h per year, a useful life considerably higher than the design life and economic gains of more than $400k per year. The proposed application case contributes to the literature relating MV three-phase overhead switches, which is currently absent, and to the optimization of its maintenance strategy.

1 Introduction

Managing distribution feeder assets presents numerous challenges. The need to achieve a balance between optimizing performance, minimizing risks and optimizing resource utilization are the main challenges [1]. To these must be added ageing equipment which imposes increased maintenance requirements. This balance takes into consideration the desired reliability and the optimal life cycle of assets based on the risk of outages, safety required for the public and workers, and customer satisfaction.

With the aim of improving company performance, while reducing annual spending, optimizing the maintenance strategies of equipment covered by a preventive maintenance program could be a possible approach. This approach would increase equipment and labour availability. However, the availability of data is a challenge for performing this type of analysis.

This article explains the methodology used for assessing the performance, safety and economic benefits of optimized maintenance strategies for distribution network assets with a preventive maintenance program, including equipment refurbishment. This methodology follows the process for estimating the volume and the replacement costs of HQD's assets [2]. A case study on MV three-phase gang operated overhead switches is presented.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 151–162, 2022.
https://doi.org/10.1007/978-3-030-96794-9_14

2 Literature Review

Several authors have studied the optimization of maintenance frequency. However, the proposed methods are based on mathematical models which require high quality historical operational data [3–6]. These methods are also often applied to individual equipment [5, 6], rather than to homogeneous equipment families.

Among the methods proposed, the combination of a mathematical method with the FMECA (Failure Mode and Effect Critical Analysis) method is often used [4, 7]. However, in the case of Fischer, et al. [7], this technique only forms a knowledge base compiling the causes of failure, but does not provide a method for optimizing maintenance periodicity. Also, some authors only combine the step of decomposing the system into components to improve the mathematical model [5].

Methods proposed in the literature are taken from various fields, such as mining [3, 5], infrastructures [6] and wind energy [4, 7]. However, there are few publications in the field of electricity distribution networks. Optimization of the maintenance frequency of MV gang operated overhead switches is also absent from the literature. Hughes [8] discusses a similar topic, but the equipment analyzed is the 11 kV oil filled switchgear. The study carried out by Hughes is based on the principle of failure mode analysis, benchmarking and assessment of the condition of the equipment at the end of the established maintenance interval. However, the methodology does not demonstrate how to establish the critical component, because for this equipment the benchmarking makes it possible to establish it with conviction.

3 Background

The maintenance policy applied to this equipment is specific to HQD. It consists in systematically replacing the network's equipment according to a set frequency in order to refurbish it in the workshop. This periodicity is identified and monitored by the corporate information system, based on the inventory and the type of equipment. This policy aims at ensuring an appropriate level of safety for workers when they operate the equipment as well as the reliability of the device.

HQD owns more than 11,000 MV three-phase gang operated overhead switches on its network. With a scheduled maintenance frequency established at 12 years for all equipment, more than 900 interventions are undertaken annually on the network and in the workshop. Figure 1 shows the equipment breakdown by age group.

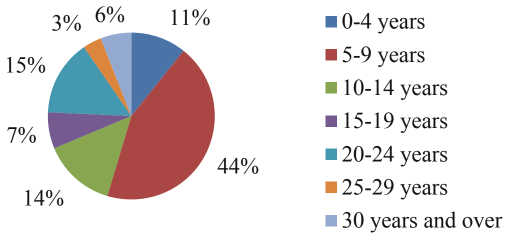


Fig. 1. Equipment in service by age on the network

MV three-phase gang operated overhead switches on the distribution network have undergone technological developments over the past few years. The operating conditions in which this equipment is used also has an impact on its useful life. Considering HQD's context (see [2]), equipment in aggressive environments (5% of medium-voltage three-phase gang operated overhead switches) require a specific maintenance policy appropriate to the operating conditions. Aggressive environments include polluted or saline areas.

The study is limited in scope to overhead switches in operating conditions where most of this asset class is located, i.e., manual switches in non-polluted and non-saline environments. The other categories are excluded from this analysis but could be part of subsequent studies. Table 1 shows the number of equipment, as well as the average age by category.

Table 1. Number and average age of equipment by category

Type of equipment	Total	Polluted/saline environment	Non-polluted/non-saline environment (Compliant)	Non-polluted/non-saline environment (To be upgraded)
Manual	7,964	398	5,095	2,471
Automated	3,124	156	2,968	0
Total	11,088	554	8,063	2,471
Average age		9 years		23 years

Regarding the equipment concerned in this article, the inspection requires the device to be withdrawn from the network for refurbishment in the workshop. Refurbishment makes it possible to replace parts that are showing signs of wear. Design improvements are also made possible by this practice. This policy allows HQD to extend the useful life of this asset, while reducing the frequency and duration of outages.

Equipment lifespan is impacted by the number of operations of the device [9]. At HQD, 99% of equipment will never reach the number of operations dictated by the manufacturer to establish their lifespan. Number of operations is therefore not a factor that affects the lifespan.

4 Methodology

The proposed method is inspired by the method developed by Bertling Tjernberg, et al. [10]. However, it is suitable in a context where the data necessary for the determination of the critical component are limited. The method is therefore combined with the FMECA method to overcome this limitation.The combination of two methods, as prescribed by Bertling Tjernberg, et al. [10], also makes it possible to overcome the limitations of the RCM (Reliability Centered Maintenance) method for the evaluation of cost effectiveness. The use of FMECA, on the other hand, ensures that maintenance efforts are focused on the most critical components and failures [7].

Then, to present the practical and concrete application of the methodology, the case study method is used on equipment in the distribution network.

The methodology suggested in this article consists in applying the process to estimate the volume and replacement costs for HQD's assets, adapted to consider the maintenance analysis process. The process is based on a statistical analysis of the available operational data for equipment with uniform characteristics. Input includes the characterization of the causes of replacement, the evolution of the degradation mechanism, as well as forecasting the number of equipment that will need to be replaced, based on the equipment lifespan.

Figure 2 illustrates the analysis process for developing possible replacement or maintenance strategies. It also presents the steps in assessing gains in performance, risks and costs. The objective of the methodology is to determine the optimal inspection frequency for the equipment. The period for which it is cost-efficient to continue to keep the equipment in operation is also established through this method.

This extended frequency needs to make it possible to maintain the level of safety, both for the public and workers, as well as equipment availability for network operation.

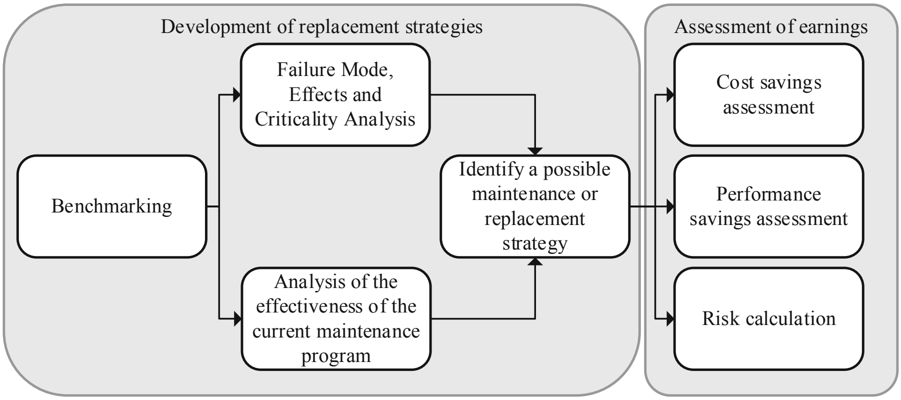


Fig. 2. Process for developing replacement strategies and assessing gains

4.1 Benchmarking

At HQD, the maintenance strategy for this equipment has evolved over the years. Before 2006, full inspections were performed every six years, on the pole. Since 2006, full inspections are carried out every 12 years. Currently, HQD ensures that overhead switches are in good condition by performing two types of inspections:

1. Visual inspection before each operation: to detect all visible defects.
2. Full inspection every 12 years: replacement of the device in order to refurbish it in the workshop. Also aims to ensure that equipment installation and integrity comply with applicable standards.

A survey developed by HQD on the CEATI platform indicated that:

- 27% of distributors are not performing any maintenance.
- 36% are performing inspections on the pole or using infrared.
- 36% use a maintenance frequency higher than HQD, but the type of inspection or where it is performed is not specified.
- HQD is the only company that refurbishes its equipment.

4.2 Effectiveness of the Current Maintenance Program

The steps for analyzing the effectiveness of the current maintenance strategy are:

1. Define the tolerance limits of the asset reliability
2. Assess the causes of withdrawal of equipment in service
3. Assess the equipment's condition at time of inspection
4. Assess the cost of the current maintenance program

4.2.1 Define the Tolerance Limits of the Asset Reliability

Defining the tolerance limits can be impacted by regulatory and environmental requirements. The analysis also takes into consideration the potential health and safety risks for workers and the public, as well as environmental risks.

4.2.2 Assess the Causes of Withdrawal of Equipment in Service

Analyzing causes of network withdrawal makes it possible, among other things, to determine the rate of replacement due to external causes. This rate can be assessed in order to validate if the maintenance program is relevant and if its optimization leads to substantial gains. Any causes that are not related to the degradation mechanism of the equipment itself or to its age are considered external causes. This category comprises accidents caused by the public, fauna and vegetation, meteorological events and corporate decisions. The operating environment, including the climate in Quebec, is responsible for a sizeable proportion of these causes. For equipment where withdrawal due to external causes represents the majority of replacements, implementing or optimizing a maintenance program represents few gains. A technical change in the equipment or in the network configuration would be more appropriate in these situations.

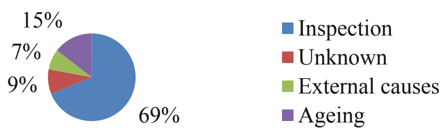


Fig. 3. Causes of replacement of MV three-phase overhead switches on the network

For this study, a sample of 583 jobs performed between 2000 and 2017 were analyzed. The analysis shows that the main cause of replacement was for inspections (Fig. 3).

External causes represented 0.05% of the population by year and ageing-related failures represented 0.1%. According to the company's tolerance limit, it is acceptable to assume that the maintenance frequency is too high.

4.2.3 Assess the Equipment's Condition at the Time of Inspection

Assessing the equipment's condition at the time of inspection calls for an analysis by sampling. For the equipment in the sample, the condition of the components was assessed based on operability, as well as the safety of the public and workers. In this study, 28 three-phase gang operated overhead switches in the workshop for refurbishment were selected randomly. For each of them, the number of components with observed damages affecting the in-service equipment was assessed. For 82% of the equipment observed, all the components were in good condition or were showing signs of wear that wouldn't affect the operability of the equipment. The effects of ageing on the equipment's condition can be observed only after 25 years. Beyond 25 years, more than 30% of switchgear showed components in a condition that prevented its operability. It is thus acceptable to consider lowering maintenance frequency.

4.2.4 Assess the Costs of the Current Maintenance Program

Assessing the costs of the current maintenance program makes it possible to compare the proposed strategy and to calculate the economic gains of optimization. Calculating costs requires a representative sample of maintenance jobs. Jobs with inconsistent or unusual costs are excluded. The costs recorded in the corporate information system for the following categories are identified from the sample:

- Labour costs in the field
- Labour costs in the workshop
- Material costs: validated, if needed, by purchasing data.

Labour costs are validated, if needed, against the estimated time required according to maintenance operators and the people responsible for job planning. In this study, the number of components replaced according to equipment age is estimated using the Delphi method. This method is favoured over the data from the information systems because components are recycled in the workshop according to maintenance practices. Table 2 illustrates the results of the analysis.

Table 2. Proportion of refurbishment costs against purchase costs, according to the components to be replaced based on equipment age

Component	12-year frequency (age)				16-year frequency (age)		
	12	24	36	48	16	32	48
Insulator	0	2	3	6	1	3	3

(*continued*)

Table 2. (*continued*)

Component	12-year frequency (age)				16-year frequency (age)		
	12	24	36	48	16	32	48
Jaw contacts	3	3	3	3	3	3	3
Ice shield	0	3	0	3	1	2	3
Blades assembly	0	0	1	3	0	0	1
Interrupting unit (new)	0	0	3	0	0	0	3
Interrupting unit (refurbished)	1	3	0	1	3	3	0
Compared to purchase price $ (%)	18	57	120	180	32	71	143

Given the current maintenance policy, useful life is established based on an economic boundary. Useful life is reached when maintenance costs are significantly higher than the acquisition costs for a new device (>120%). This economic boundary is considered reached after four inspections, performed every 12 years. The replacement rate for external causes is sufficiently low (0.05% of the equipment each year) to validate the calculation methodology. The maximum lifespan is 48 years.

4.3 Failure Mode, Effects and Criticality Analysis (FMECA)

The failure mode, effects and criticality analysis includes the following four steps:

1. Functional dividing of the equipment
2. Identification of the causes of degradation and failure
3. Establishment of the frequency, severity and detectability scales
4. Assessment of criticality components

4.3.1 Functional Dividing of the Equipment

This is done in collaboration with the experts responsible for the equipment. This step involves dividing the equipment to its smallest maintainable component.

4.3.2 Identification of the Causes of Degradation and Failure

The functional dividing of equipment by component makes it possible to analyze the failure modes specific to each piece of equipment. The main causes of equipment degradation and failure are corrosion, misalignment of blades with fuse clips, failure of the opening and closing mechanism, sealing of the arc-control device and breakage or cracks on the insulator [9]. CEATI [9], the manufacturer and the requirements of standard IEEE C37.30.1–2011 [11] indicate that equipment reliability depends on operating frequency and equipment age. In total, 91% of HQD's equipment is operated five times or less per year Operating frequency is thus not a factor that affects lifespan.

4.3.3 Establishment of the Scales

The frequency, severity and detectability scales are adapted from IIE in order to comply with internal requirements and HQD's values [12]. Considering the types of equipment subject to the methodology, the criticality of components is estimated based on the predetermined scales. Possible effects, as well as mitigation measures are considered. Effects must take into consideration public and worker health and safety, the environment, the network and the integrity of the equipment. For the equipment in question, the severity scale does not consider the environment, because the challenges are non-existent. The residual effect is the possible effect once mitigation measures are implemented.

4.3.4 Assessment of the Criticality of Components

Criticality is calculated by multiplying the frequency, detectability, and severity ratings of the residual effect of the component's failure mode. Frequency is established based on the analysis of causes of equipment withdrawal due to ageing. This analysis focuses on maintenance jobs, including movement of equipment inventory. Refurbished three-phase overhead switches were not subject to inventory tracking before 2015. The results were revised upward (x3) to include, hypothetically and proportionally, equipment replacements without stores movement information from the warehouse. The results are presented in Table 3. The detectability of the breakage as well as its severity are determined using the Delphi method. Table 4 shows the analysis results for one of the components.

Table 3. Frequency of component breakage

Cause specified	Proportion of the population revised upward	1 breakage out of (equipment)	Broken component	1 breakage out of (equipment)
Other	0.017%	5,700	Unknown	1,000
No details	0.087%	1,100		
Burnt	0.028%	3,600	Interrupting unit	2,000
Interrupting unit	0.028%	3,600		
Misaligned	0.010%	9,600	Blades assembly	7,000
Melted blades	0.003%	28,700		
Broken insulator	0.031%	3,200	Insulator	2,000
Salt or pollution	0.021%	4,800		
Inoperative	0.063%	1,600	Jaw contacts	1,600
Total	0.289%	300		300

4.4 Determination of One or Several Replacement Strategies

The various replacement policies are developed with equipment technical experts, based on benchmarking data. They also take into consideration the technical analyses of the

Table 4. FMECA for medium-voltage three-phase overhead switches

Component	Failure mode	Cause	Possible impact without corrective measure			Corrective measures already in place	Possible impact with corrective measures			Frequency	Severity	Detectability	Criticality
			Public safety	Worker safety	Equipment integrity		Public safety	Worker safety	Equipment integrity				
Arc-control device	Loss of performance	Age, number of operations	Not applicable	Projection of melted metal where electrical arcs are generated	Melting of contacts, blades, requiring the replacement of the device	Safe working methods, grounding on the pole, dielectric and seal tests applied when restoring the arc-control device at the time of refurbishment	Not applicable	None	Physical breakage	3	5	5	75

Research Institute (IREQ: Institut de recherche d'Hydro-Québec), as well as the judgment of operating and maintenance staff. The literature review of renowned bodies and international conferences is also considered. Two replacements strategies were analyzed:

1. On-site equipment maintenance and repair
2. In-shop maintenance and repairs with a frequency change

For the first option, it was established that inspections at the pole make it possible to detect and correct most critical parts, except arc-control devices. It is difficult, if not impossible, to detect the start of a failure only by inspecting the arcing plunger or the general condition of the component. As the rest of the arc-control device is sealed, a full inspection of components cannot be performed. On-site inspection would inevitably lead to the failure of some arc-control devices. The second option was therefore selected. The results show that the arc-control device is the most critical component. The optimal maintenance frequency will consequently be dictated by the reliability of this component, based on its age. In collaboration with IREQ, analyses of the effects of ageing were carried out on arc-control devices.

The observations at IREQ made it possible to establish that the seal was the most significant problem with this component. The rapidity of occurrence of this failure mode therefore dictates the new maintenance frequency. Based on this criterion, it would be acceptable to increase the refurbishment periodicity from 12 to 16 years. Considering the component design lifespan and for internal corporate requirements, it was deemed prudent to limit maintenance frequency to 16 years. Additional analyses could be the subject of another study in order to support a longer periodicity.

The two scenarios analyzed to calculate maintenance frequency are as follows:

A. Keeping inspection frequency at 12 years
B. Changing inspection frequency to 16 years

4.5 Calculation of Economic and Performance Gains

The calculation of economic gains resulting from the proposed maintenance policy is based on the total labour and material costs of each inspection. Maintenance frequency and equipment volume make it possible to estimate current annual costs. Economic gains are determined by subtracting the costs of the optimized strategy from the current costs and evaluated in constant dollars. Economic gains also include assessing equipment lifespan against maintenance costs. The current lifespan is compared with the lifespan obtained with the optimized strategy. Equipment is deemed to have reached the end of its useful life when the total maintenance costs are higher than the costs of buying a new piece of equipment. To estimate lifespan, refurbishment costs include only the costs of labour in the workshop and material costs. Labour in the field to replace equipment with new or refurbished equipment is considered equivalent. Performance gains represent gains in labour. These can make it possible to allocate resources to other needs and reduce costs in order to improve the global performance of the company. In this instance, the performance gains and annual costs saved by revising the maintenance frequency to 16 years are shown in Table 1. The 2017–2021 period represents the time frame to upgrade non-compliant equipment. Afterwards, this equipment will benefit from the optimized maintenance periodicity. The lifespan is equivalent to the current situation. The economic criterion is deemed reached after three inspections, performed every 16 years. The maximum lifespan is 48 years (Table 5).

Table 5. Gains in costs and hours of maintenance avoided

Years	2017–2021 (unit/year)					2022+ (unit/year)				
Type	Equipment	Inspections		Savings ($K)	Labour gains (hours)	Equipment	Inspections		Savings ($K)	Labour gains (hours)
		A	B				A	B		
Compliant	5,095	425	318	420	2,354	7,566	631	473	623	3,476
Total	7,566	919	812	420	2,354	7,566	631	473	623	3,476

4.6 Calculation of *R*isk

Risk is estimated by assessing the criticality of the effects of equipment components' failure modes. The impact of the optimized maintenance policy in terms of the frequency, severity or detectability of the effects is estimated and compared to the initial situation. In this study, the risk is considered equivalent for both scenarios. Monitoring indicators are implemented to track the risk.

5 Conclusion

The contribution of this article to the current literature is that it provides a methodology for optimizing the frequency of maintenance of distribution network equipment. The characteristic of the method is that it is applicable when the historical operational data are in quantity or in insufficient quality to allow advanced mathematical analyzes. The method also provides the steps to establish the critical component of an equipment. The proposed application case contributes to the literature relating the optimization of MV three-phase overhead switches maintenance strategy, which is currently absent. The methodology detailed in this article sets itself apart through the development of an FMEA in which the impact is assessed for the following factors in order to determine the optimal maintenance frequency: public safety, worker safety, equipment integrity. The application of an optimal preventive maintenance policy, in combination with access to data on equipment condition at the time of maintenance generated significant economic and labour gains for HQD. It also makes it possible to achieve a lifespan considerably higher than the design life for the case study equipment, while maintaining cost effectiveness. This methodology can be applied to various pieces of equipment on the distribution network subject to the same maintenance policy. An additional study could establish the acceptable levels of risk from the calculation of the criticality of possible failures of equipment components. Furthermore, a complementary study could examine the impact of adapting maintenance frequency based on equipment age. These improvements would lead to an enhanced optimization of the maintenance strategy.

References

1. Brown RE, Humphrey BG (2005) Asset management for transmission and distribution. IEEE Power Energy Mag 3(3):39–45
2. Biard G, Vaillancourt R, Abdulnour G, Langheit C, Gaha M, Houle G (2020) Determining the lifecycle and future replacement cost of distribution network equipment. Int J Mod Eng Res 10(3):51–60
3. Angeles E, Kumral M (2020) Optimal inspection and preventive maintenance scheduling of mining equipment. J Fail Anal Prev 20(4):1408–1416. https://doi.org/10.1007/s11668-020-00949-z
4. Andrawus JA, Watson J, Kishk M, Gordon H (2008) Optimisation of wind turbine inspection intervals. Wind Eng 32(5):477–490
5. Gölbaşı O, Demirel N (2017) Optimisation of dragline inspection intervals with time-counter algorithm. Int J Min Reclam Environ 31(6):412–425. https://doi.org/10.1080/17480930.2017.1339168
6. Mahmoudi M, Elwany A, Shahanaghi K, Gholamian MR (2017) A delay time model with multiple defect types and multiple inspection methods. IEEE Trans Reliab 66(4):1073–1084. https://doi.org/10.1109/TR.2017.2754519
7. Fischer K, Besnard F, Bertling L (2012) Reliability-centered maintenance for wind turbines based on statistical analysis and practical experience. IEEE Trans Energy Convers 27(1):184–195. https://doi.org/10.1109/TEC.2011.2176129
8. Hughes DT (2001) Optimising the maintenance of 11 kV oil filled switchgear-an example of condition based decision making. In: 16th International Conference and Exhibition on Electricity Distribution, 2001. Part 1: Contributions. CIRED. (IEE Conf. Publ No. 482), 18–21 June 2001, vol 1, p 3. https://doi.org/10.1049/cp:20010702
9. CEATI, Distribution System Health Indices, 2016, vol. Report T134700–50/118
10. Bertling L, Allan R, Eriksson R (2005) A reliability-centered asset maintenance method for assessing the impact of maintenance in power distribution systems. IEEE Trans Power Syst 20(1):75–82. https://doi.org/10.1109/TPWRS.2004.840433
11. *Standard Requirements for AC High-Voltage Air Switches Rated Above 1000V*, C37.30.1-2011, IEEE (2011)
12. IIE, FMEA, in *Six sigma black belt*: Institute of Industrial Engineers (2011), ch 3–6

Pragmatic Performance Management
Aligning Objectives Across Different Asset Portfolios

Ype Wijnia(✉)

AssetResolutions B.V., PO box 30113, 8003 CC Zwolle, The Netherlands
ype.wijnia@assetresolutions.nl

Abstract. To manage the realization of value, the core concept of asset management, it is vital to have a good understanding of what that value is, how it is produced and how it can be measured. In this paper, we first make a connection between the concept of value, its production in a value chain and the asset lifecycle and conclude that these concepts are not necessarily aligned. Next we address issues in performance management. Common practice is to use Key Performance Indicators (KPIs) to keep track of results, with multiple KPI's often compiled into a business dashboard or balanced scorecard. To be effective, KPIs need to be valid, functional and legitimate. Unfortunately, a focus on pursuing targets can provoke strategic behavior resulting in destruction of value. Furthermore, targets may be aimed at the wrong lifecycle, be part of an optimization or be blind for optimization across portfolios. To address these challenges, we propose a pragmatic approach based on the cost to benefit ratio of interventions, embedded in a social setting. This model has been applied for an organization with 3 separate but comparable portfolios. The concept still has to be tested in a more diverse setting.

1 Introduction

In the early days of asset management, when it was still branded as terotechnology, assets were mostly regarded as costs. The aim was to realize the lowest costs for a given performance and lifecycle according to Thackara [1].

> Terotechnology is defined here as a combination of management, financial, engineering, building and other practices applied to physical assets in pursuit of economic life cycle costs.

Given its background in the industrial domain, such a focus on economics is a valid interpretation, even in modern times[1]. However, this does not apply to all types of assets. In the infrastructure domain, for example, most assets are not operated, they provide their function by existing. Defining performance for such a passive asset can be stretching the imagination [2]. The awareness that infrastructure assets are different resulted somewhere in the 1990s in an inclusion of risk into the concept, which was documented in various manuals and standards [3–5]. The idea that assets have negative

[1] In his keynote at WCEAM in 2014, John Woodhouse made reference to the British standard on terotechnology and its applicability in asset management.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 163–172, 2022.
https://doi.org/10.1007/978-3-030-96794-9_15

value impacts by means of risk was furthered into the notion that assets produce value, culminating (so far) in the formal definition of asset management in the ISO55000 series [6]:

> Asset management is the coordinated activity of an organization to realize value from assets.

This modern interpretation thus regards the asset base more as a profit center, and management should therefore focus on performance.

The paradigm shift from assets as a cost center to assets as a profit center comes at a cost though. Whereas in the old-fashioned view a focus on costs and technical failures sufficed for most engineering asset managers, the modern approach requires much more awareness of the context in which the assets are operated: what is produced where and when by what assets, who benefits/suffers how, what values are impacted and by how much? The interests of all relevant stakeholders then can be translated into indicators that measure the value production, including the objectives (targets) to be achieved and other constraints that should be adhered to by asset management. Additionally, the system for managing the realization of value has to be established as well. This whole process of creating the right conditions for engineers to deliver value with the asset can be considered asset governance [7, 8][2], as shown in Fig. 1.

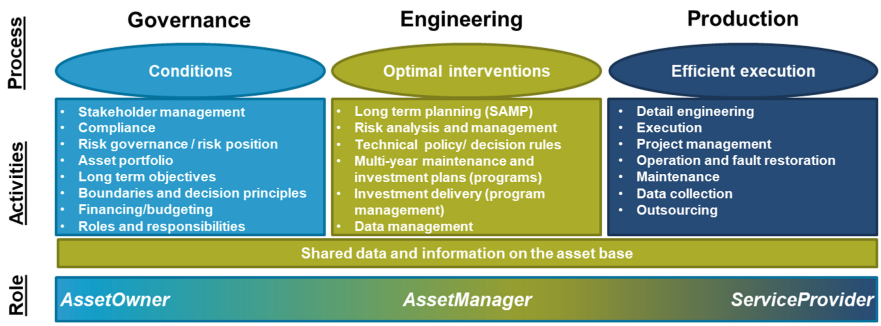


Fig. 1. Processes of asset governance, engineering and production. The split in processes resembles the role division into asset owner, asset manager and service provider, but these roles seem to be implemented with more flexibility in practice.

Given that the combination of assets and context is unique for any asset base, it could be argued on theoretical grounds that the outcome of asset governance (i.e. the management system including indicators) should be unique as well. From a more pragmatic perspective this is undesirable, as unique systems costs more to develop and maintain. Furthermore, uniqueness limits interorganizational learning, e.g. via benchmarks. Despite this clear need for standardization, surprisingly little guidance is available

---

[2] These papers date from before ISO55k and asset governance was framed as additional to asset management. Nowadays, the governance activities are part of asset management as defined by ISO55000 series. Contrast is therefore made with the engineering activities within the standard.

with regard to effective and efficient asset governance and the associated performance management systems.

In this paper we will explore the potential for pragmatic solutions. First we will build a general understanding of value production and its relation to asset management. Next we will look at performance management, potential indicators and their tradeoffs, and define the characteristics of useful indicators. After that we will present a very simple model for selecting and/or constructing pragmatic indicators and apply them to a variety of asset types. We will conclude with an outlook for further developing this idea.

2 The Concept of Value Production

There are several perspectives on value production. Most fundamental perhaps is the 6 capitals model, which can be used to report the impact a company has on the planet as a whole [9]. In this view, there are 6 forms of capital on the world: financial, manufactured, intellectual, human, social and natural. A production process draws from these capitals to make outputs, which may result in outcomes in the form of the 6 capitals. Net value production occurs if the sum over the capitals is increased (Fig. 2).

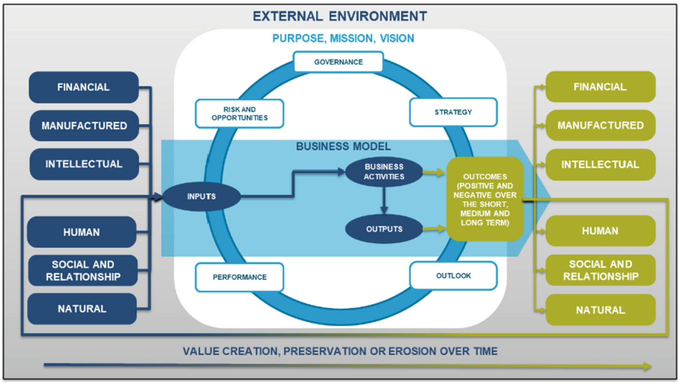


Fig. 2. The 6 capitals value creation process after [9]. Both business activities and outputs generate outcomes.

From a management perspective, such a high level approach to value is not good enough and the business activities need to be considered in more detail. This is the value chain perspective from Porter [10]. A common decomposition is distinguishing primary and supporting processes, and the operational chain is in general divided into the stages incoming logistics, production, outgoing logistics, sales and service. Each of these stages may make use of one or more types of asset (Fig. 3).

Finally, from an asset management perspective the value production process may not be very visible, as the focus is more on guiding the assets through their lifecycle [11]. There are several views on how to divide the lifecycle, but in general[3] phases

[3] This is the technical view. For some asset managers, design may just be the identification of a need and construction the purchase of a turnkey machine. Yet, the machines still have to be designed and constructed.

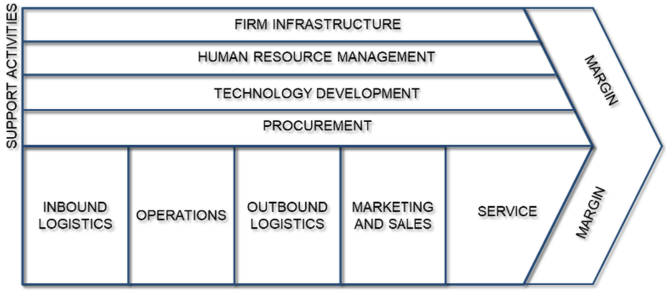


Fig. 3. Value chain model after Porter [10]

like design, construction, operation, maintenance and termination are recognized [12]. It may be tempting to think that assets only generate value in their operational phase, but that would neglect the fact that assets can be products in other value chains, that not all rights to an asset (e.g. ownership and possession or right to use) need to be in one hand, and that these different rights may be part of different value chains simultaneously. For example, take a look at real estate. For the contractor, the building is a product, whereas it is an asset for the investor/owner. The right to use the building is rented out (sometimes via a lease to a broker) to the tenants, for which it becomes an asset if they use it in another value chain (e.g. run a production facility in the building). Over time, the owner may want to change its investment portfolio and sell the property rights to this building (not necessarily affecting the lease and right to use), which becomes a product then once more. The physical object provides value to the user in its operational phase and depends on the condition and thus maintenance, but the value of ownership may depend more on the attractiveness of its location than on the state of the asset. It may even be that activities that create value for one party erodes value for others. In such a case, integrated/holistic performance management would result in radically different outcomes than performance management of the parts. For the purpose of this paper, we will assume the various rights to the asset are more or less aligned and focus on value delivery by the use of the physical object in its operational phase.

3 Measuring Asset Performance

In general, performance management requires performance measurement, following the saying "what gets measures gets done". Common practice is to use Key Performance Indicators (KPIs) to keep track of results, with multiple KPI's often compiled into a business dashboard or balanced scorecard [13, 14]. Performance measurement can serve various purposes, like (sorted by impact) creating transparency, learn and improve, compare, assess and sanctioning [15]. The more impact an indicator has, the higher the demand for quality, as it would be perceived as unfair if a sanction would be based on poor data. However, literature [16] suggests a law of decreasing effectiveness: indicators used for sanctioning undesired performance may incentivize strategic behavior, i.e. doing things that give a good score, not necessarily things that deliver value. Additionally, in order to be effective, indicators should be valid (consistently measure what is intended to be measured), functional (relevant to the objectives) and legitimate (influenceable by the one measured and the target should be achievable).

It is clear that compliance with these criteria is difficult to achieve in a complex relational situation as the aforementioned example for real estate, where the objectives of different stakeholders may not be aligned. However, even in much simpler conditions (asset owner, manager, user and service provider all part of the same organization with aligned objectives) it may not be as straightforward as it seems. This is because the phase where value is delivered (in general the operational phase) is not necessarily the phase where asset management interventions adds most value [17] as shown in Fig. 4 after [2].

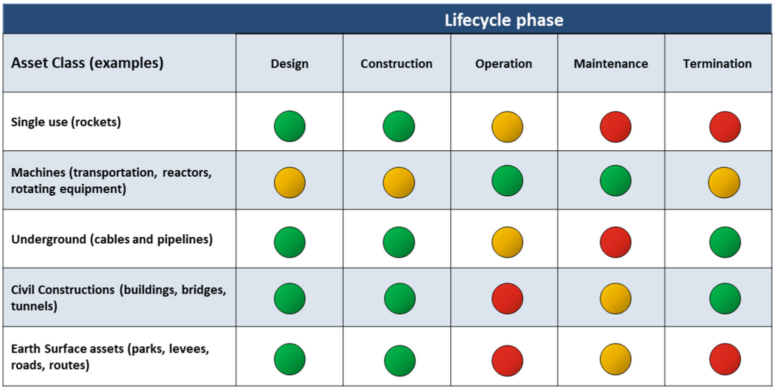


Fig. 4. Impact of interventions on the value delivery over the life cycle of some types of infrastructure assets after Wijnia [2]. The colors indicate the relative importance.

For single use assets (ranging from disposable tools to rockets), virtually all value is determined in design and construction, perhaps a little bit in operation but maintenance is not relevant and termination is designed in. For machines (active/rotating equipment) most of the value is determined in the operations and maintenance phase. Design is not irrelevant, but there are numerous examples of chemical plants operating at double their design capacity. Underground assets like cables and pipelines may be operated but maintenance can be impossible like it is for cables. Civil constructions like buildings and bridges are not operated[4] but they do require maintenance. Finally, earth surface assets like roads and parks may have a very long to even indefinite lifespan as they allow for infinite repairs. The right of use is more important here than the technical asset, and thus termination is not relevant as an intervention.

If performance targets are set in a phase where interventions do no matter much, value may be destroyed. This is for example the case in utilization of power cables. More volume over the same cable sound like a good idea, but the increased energy losses (proportional to the square of the current) may cost more than a new cable with a higher capacity.

It is also important to realize that management only directly controls the deployment of interventions and not performance itself. Deployment of interventions is always an

[4] The movable parts like elevators, gates, bridge decks are regarded as machines.

optimization problem: there will be a moment at which more interventions do not deliver more value or that other interventions provide more value for the same cost [18]. This is demonstrated in Fig. 5 for optimization of the maintenance interval.

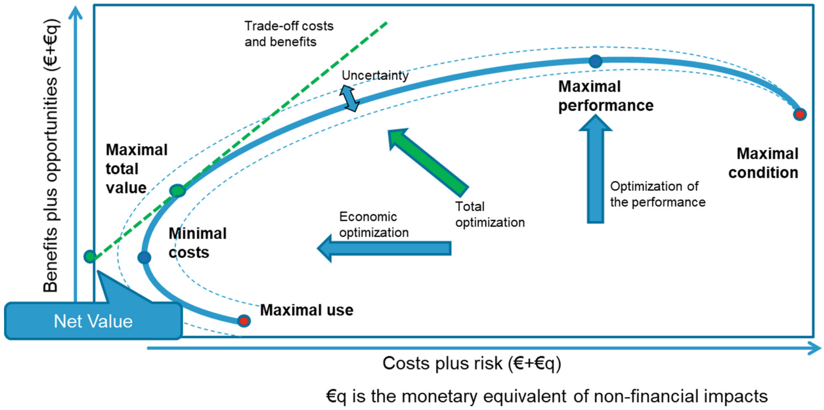


Fig. 5. Optimization of the maintenance interval, running from maximal use (run to failure) to maximal condition after [18].

To keep an asset in the best possible condition requires a significant amount of maintenance to the point where it impacts asset availability[5]. Decreasing the amount of maintenance allows more use and thus more benefits at a lower cost. In the trajectory of the amount of planned maintenance two observable extreme positions will be met. The first is that of maximum Performance, where the decrease in lost performance due to planned maintenance is equal to the increase in lost performance due to unplanned maintenance. The second is that of minimal costs, where the cost reduction of planned maintenance is equal to the cost increase of unplanned maintenance. From that extreme, a further reduction of planned maintenance to zero (maximal use/run to failure) will both increase cost and decrease availability. The optimum is somewhere between the two extreme values. It is the point where the decrease in cost of maintenance is equal to the decrease of benefit of performance. It is the point of lowest cost per unit of performance. This may be observable (e.g. in production facilities with objective production) but for others (e.g. infrastructure or other (semi) public assets) it may require an inherently subjective valuation and thus can in fact be any point in between the observable extremes.

This optimization is complicated by (earlier mentioned) competing interventions on the same asset (like replace by maintenance free items) or even on other assets because the marginal benefit of extra units of interventions is higher. There is little value in optimizing an asset that is near its optimum if there is a very poor performing asset somewhere else in the value chain.

Combining these views, it seems that the whole concept of realizing maximal value/optimal performance by steering towards explicit performance targets may be

[5] The extreme for preserving the condition would be to put the asset in a museum, not having any value in its intended usage.

fatally flawed for many assets, especially those in the public domain like infrastructure assets. Management has no direct control over the production of value, and what is controlled (the interventions) is part of an optimization and thus has no universal gradient for goodness. In the next section we will propose a pragmatic approach to overcome these issues.

4 A Pragmatic Solution

In order to develop a pragmatic solution for the aforementioned obstacles, it is important to step back and reconsider what the purpose of performance management is. Is it really only about delivering an absolute amount, like the produced volume, the availability of an asset or staying within budget limits? Or is it more about assuring that the available resources are used in the most effective and efficient way, like driving towards the best value per unit of cost or the lowest cost per unit of production? This latter perspective seems to be more aligned with the whole philosophy of continual improvement and lean management in which the reduction of waste is essential [19].

Optimal resource allocation is achieved if all portfolios[6] have the same marginal benefit. This is the benefit per unit of cost (yield) for the least yielding/significant intervention within the portfolio. It is highly unlikely that such an alignment of portfolios occurs if targets are set without explicitly considering the yield. This means that if such an optimal resource allocation perspective is adopted, focus should move away from setting the targets directly and shift towards targets derived from the process of prioritizing interventions across portfolios.

For such a process to work, several prerequisites can be formulated. First of all, there needs to be a common understanding of total value and its monetization, so that the yield of a project can be expressed in a uniform ratio [20]. A relatively simple way to achieve this is by implementing a properly designed risk and/or opportunity matrix, in which non-financial values are aligned with the financial value to provide a reasonable estimate for the willingness to pay [21].

The second requirement is that this universal value framework is made specific for each portfolio. This should be preferably as simple as possible, as that facilitates widespread application. For virtually all asset types there is often only a limited number of aspects (2–3) that dominates the performance. For infrastructure assets like roads and cables these typically are risks to safety[7] and (a form of) reliability. But for capital intensive commercial assets like real estate and rotating equipment production facilities, performance is typically more linked to utilization, as that provides the income, either (semi) fixed or variable[8]. If the utilization is already high (say above 80% like in 24/7 production facilities) focus typically shifts to the efficiency (=inputs needed per unit of output) of the production process. This also holds for capital extensive processes which are dominated by the cost of the inputs. For facilities that run at near 100% utilization,

[6] The concept of portfolio is used her as indicating a set of interventions within the same budget.

[7] If safety is heavily regulated, this may be replaced with compliance.

[8] A different pricing strategy is not considered an intervention in this paper. Monopolies (where prices could be set) are typically regulated, and in markets a different price point has impacts on the sales volume.

availability may become dominant. And sometimes these assets lend themselves for an upgrade in the capacity. In all these cases the performance metric is something like the average cost per unit[9]. For assets that can be reused freely like information [22] performance is more likely dominated by the ability to find new applications (the whole idea of big data) as there is little cost and risk in keeping the data once collected. Each asset portfolio thus typically has a most relevant area of attention. These areas are structured in Fig. 6 by polarity and certainty.

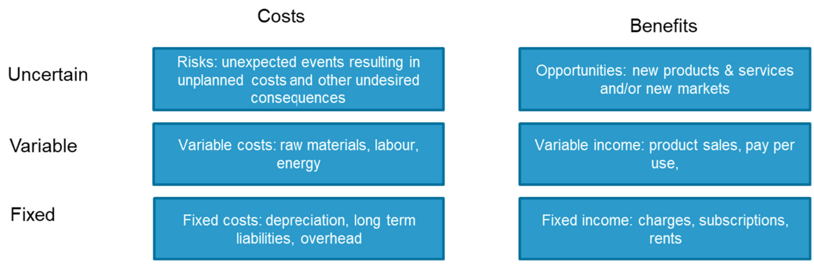


Fig. 6. A division of focus areas by polarity and certainty

That each portfolio has a specific focus area in which improvements should be chased does not mean the others are not relevant. If a good improvement is encountered, it should be implemented, but there is no need to search for those improvements until that sector becomes dominant.

The final requirement is to embed the quantitative prioritization process in a social setting in order to limit strategic behavior. This can be for example a group decision to determine the ranking of the interventions, with options to bypass the calculated priority. This group decision tends to be governed by rules of fairness (e.g. every participant has one joker to bypass the ranking for a specific intervention) and social pressure will limit spreadsheet management to get a good score, as was demonstrated in the combined decision over 3 portfolios (separate but relatively similar) for a distribution network operator [2, 20].

5 Conclusion

Performance management by setting targets is common practice, but proves to be very difficult. Simple direct measures often do not tell the full story and may result in strategic behavior if pursued stringently. They also may be variables for an optimization, meaning that improvements are not always in the same direction. Ratio indicators (like the cost per unit) can perform better, but may cross borders in the span of control and thus violate the criteria for good indicators. Furthermore, ratio's tend to have extremes in what can

[9] This has some overlap with utilization, as that impacts the average cost per unit as well. The distinction is made because for an asset only used 8 h per day for 200 days per year, a huge cost reduction per unit could be made by running double shifts. Such an improvement is virtually impossible by efficiency (inputs per output) or throughput (output per unit of time) measures.

be achieved. Combined together, a more pragmatic approach could be to focus on the process of deploying interventions, requiring that only interventions are implemented that improve the ratio (Business case requirement) or even total value and that a process is in place for identifying the best options for improvement. This may be breaking the current practice of target setting in many organizations, but in terms of value creation it may be the best we have. Elements of this approach have been proven to work, but for a very diverse asset base it has not been tested yet. Further research is needed.

References

1. Thackara AD (1975) Terotechnology - What it is all about. Chart Mech Eng 22(6):88–90
2. Wijnia YC (2016) Processing risk in asset management: exploring the boundaries of risk based optimization under uncertainty for an energy infrastructure asset manager. In: Technology, Policy and Management, Department of Engineering Systems and Services, Section Energy & Industry. Delft University of Technology, Delft
3. IAM (2002) International Infrastructure Managament Manual. V 2.0, UK 2002 ed.: UK Institute of Asset Management
4. BSI (2004) PAS55-1 Asset Management, in Part 1: Specification of the optimal management of physical infrastructure assets: London
5. BSI (2004) PAS 55-2 Asset management in Part 2: Guidelines for the application of PAS 55-1
6. ISO (2014) ISO 55000 in Asset Management-Overview, principles and terminology: Geneva
7. Mardiasmo D, et al (2008) Asset management and governance: analysing vehicle fleets in asset-intensive organisations. In: The Twelfth Annual Conference of the International Research Society for Public Management (IRSPM XII), pp 1–20
8. Wijnia YC, Herder PM (2009) The state of asset management in the Netherlands. In: World Conference on Engineering Asset Management. Springer, Athens
9. IIRC (2021) International Integrated Reporting Framework. 2021 [cited 18 April 2021]. https://integratedreporting.org/wp-content/uploads/2021/01/InternationalIntegratedReportingFramework.pdf
10. Porter M (1985) The value chain and competitive advantage. In: Competitive Advantage: Creating and Sustaining Superior Performance, pp 33–61. Free Press, New York
11. Van der Lei T, Herder P, Wijnia Y (2012) Asset Management The State of the Art in Europe from a Life Cycle Perspective, 1st edn. Topics in Safety, Risk, Reliability and Quality. Springer, Dordrecht, Netherlands
12. IAM (2014) Asset Management - an anatomy Version 2. The institute of asset management
13. Kaplan RS, Norton DP (1996) The Balanced Scorecard: Translating Strategy Into Action. Harvard Business School Press, Boston
14. Arthur D et al (2016) Asset Planning Performance Measurement. Springer International Publishing, Cham
15. Green D, et al (2016) Setting targets in an asset management performance measurement framework. In: 7th International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering. 11th World Congress on Engineering Asset Management
16. Bruijn H (2008) Managing performance in the public sector - by Hans de Bruijn. Public Admin 86:863–865
17. Van der Lei T, Wijnia YC, Herder PM (2010) Towards an integral asset management framework using asset characteristics, asset environment and lifecycle phases as leading principles, In: The 5th World Conference on Engineering Asset Management. Springer, Brisbane, Australia

18. Wijnia Y (2016) Towards quantification of asset management optimality. In: Proceedings of the 10th World Congress on Engineering Asset Management (WCEAM 2015). Springer International Publishing, Cham
19. Gupta S, Sharma M, Vijaya SM (2016) Lean services: a systematic review. Int J Prod Perform Manage. 65:1025–1056
20. Wijnia YC, Warners JP (2006) Prioritizing investment. The value of portfolio decisions in electricity infrastructure management. In: 29th IAEE Annual International Energy Conference 2006: 'Securing Energy in Insecure Times': Potsdam
21. Wijnia YC (2010) Asset risk management: issues in the design and use of the risk matrix. In: The 5th World Conference on Engineering Asset Management Brisbane, Australia
22. Moody D, Walsh P (1999) Measuring the value of information - an asset valuation approach. In: ECIS

Building an Optimal Long-Term Asset Renewals and Modernization Plan Through Quantified Cost/Risk/Performance Value

Saulo Trento, John Woodhouse(✉), and Peter Jay

The Woodhouse Partnership Ltd., 19 Prince Henry House, Kingsclere Business Park, Kingsclere, Hampshire RG20 4SW, UK
john.woodhouse@twpl.com

Abstract. As assets deteriorate and/or new technology becomes available, asset-intensive industries across the world struggle with planning and justifying the necessary reinvestment to renew and modernize their equipment. This paper presents a methodology for rapidly creating an optimized long-term asset renewal plan that targets the maximization of value to the organization. It ensures alignment with top-level strategic objectives, while at the same time is built from the bottom up, based on the assets' condition, system functions and criticalities. It also involves broad participation and buy-in from technical staff, so there is widespread consensus on the emerging priorities.

The methodology is based upon the 6-step SALVO Process for Strategic Asset Lifecycle Value Optimization, the product of a 5-year multi-sector R&D collaboration programme. Benefits of the method include the ability to calculate and demonstrate the monetized value, risks and other business impacts generated by each proposed intervention at different potential timings, and the optimization of combined effects within any overriding constraints (such as budgets, resources or timing commitments). This involves quantifying and modelling the trade-offs between Capex, Opex, risks, performance and sustainability, with mixed quality data and expert/tacit knowledge, using state-of-the-art decision support tools. It also achieves, usually for the first time, true alignment between technical and financial departments, providing a transparent and auditable basis for the interventions and funding requirements. A case study is demonstrated and discussed, with lessons learnt, from the successful creation of a 10-year renewal and modernization plan at a large electricity transmission company (ISA CTEEP) in Brasil. This work formed part of a wider 3-year asset management innovation project under the R&D programme supported by the Brazilian electrical sector regulator, ANEEL.

1 Introduction

Much of the current Brazilian electricity transmission infrastructure was built during the 1970s and 1980s, when state investments in the sector peaked. With the anticipated useful life of the physical assets installed in this period, it is possible to expect a need for major reinvestment and renewal of these assets in the current decade. On the other hand,

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 173–184, 2022.
https://doi.org/10.1007/978-3-030-96794-9_16

the environment of regulatory uncertainty that followed legislation in 2012 (Measure 579/2012) and subsequent further changes, have altered the tariff recognition rules associated with investments in the existing network. This had the practical effect of greatly reducing the investment motivation in many of utility companies. In the case of ISA CTEEP, one of the largest power transmission companies in Brasil, the planning horizon had been reduced to budgeting only for urgent asset reinvestments one year at a time.

In 2017, the company started a substantial effort to adopt best asset management practices. This built a clearer 'line of sight' connectivity between strategic goals and immediate priorities and actions, manifested in a Strategic Asset Management Plan (SAMP). In developing this, it quickly became evident that building a long-term asset renewal plan should be a priority. Many assets are nearing the end of (or already exceeding) their presumed technical life, with unknown risks associated with continued usage. Furthermore, the regulatory situation is now more favorable, encouraging investments in the renewal of assets in many cases.

The construction of an optimized asset 'Renovation Plan' was defined as an asset management objective in the 2018 version of the company's SAMP, and the construction of a 10-year investment plan was completed in 2019, covering the period 2021—2030.

2 Methodology

2.1 The SALVO Process

The methodology adopted for the construction of the Renewal Plan was based on the SALVO Process (Strategic Asset Lifecycle Value Optimization). The SALVO Process was developed by a multi-industry R&D program [1], addressing optimal asset management decision-making in different stages of the life cycle. It consists of six steps that serve as a methodological guide, through a systematic approach to all the factors that must be considered in the decision-making process. The methods have a wide and flexible application for many decision types, considering the interests of all stakeholders and the competing organizational objectives. The processes are people-centric and specifically suited to cases where decisions have to be made with incomplete or uncertain information, which is usually the case in real-world investment planning.

Figure 1 is known as the SALVO "smiley" and illustrates the six generic steps, each of which has process requirements, information and quality attributes, governance and competency needs. The layout guides the *top-down* identification and understanding of the most urgent and important problems to solve, and the identification of potential interventions/options, followed by the *bottom-up* evaluation, justification and timing of such solutions, firstly individually and then in combinations and total delivery programmes.

The adoption of this approach also ensures the alignment of the Renovation Plan with the company's Asset Management Policy and with the requirements of the ISO 55001 standard for good asset management.

2.2 Practical Aspects of Building the Plan

The main work of constructing the Renovation Plan took place over a 10-week period starting in September 2019, and was organized in 5 phases of 2 weeks each. The team

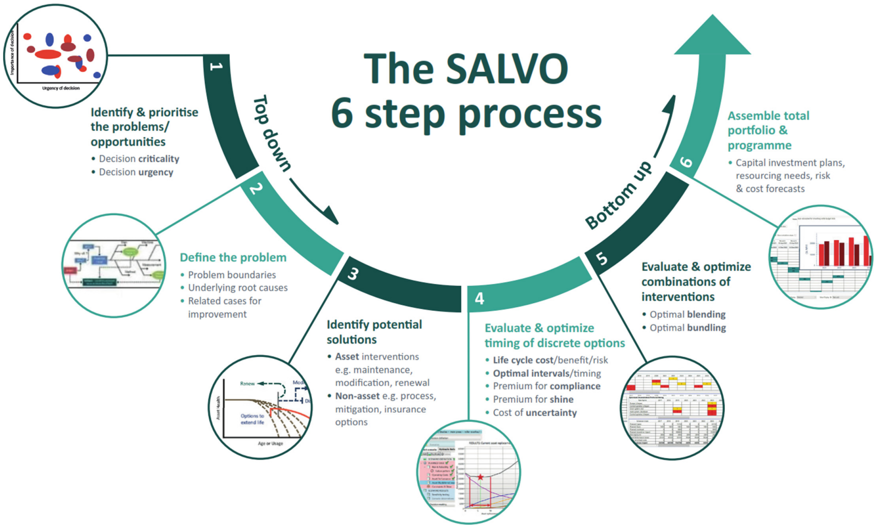


Fig. 1. The SALVO process.

was also able to draw on some previous niche studies, developed as part of the company's learning journey in asset management. In particular this included the identification of health and criticality attributes of assets, and the development of a risk model for the asset-related failures (both useful contributors to SALVO Step 1).

Also, during the period 2018–19, several discrete studies had been carried out for 'bad actor' assets and reliability/risk concerns. These introduced the modelling and 'what if?' evaluation capabilities and DST (Decision Support Tools) [2] developed by the SALVO consortium. These tools provide both the navigational rigour and the calculation of cost/ risk/ performance/ sustainability and other trade-offs in support of steps 4, 5 & 6 of the SALVO Process. The prior experience of these studies provided useful process templates for the wider evaluation of the very large and diverse asset portfolio.

For the systematic study period, a dedicated multidisciplinary team was created, supported by focal points in each discipline department, in each asset class specialism and each business interest group (finance, safety etc.). These stakeholders participated in workshops to capture tacit knowledge, explore options and ensure consensus in the outcomes. This resulted in great interest and participation of people from the technical areas, regional operations areas, engineering standards and work planning groups.

A total of about 70 people thus participated directly in the work, at an average intensity of around 3 days a week, around the core team of 8 people working almost full time throughout the 10-week programme. In total it was estimated that the building of the first comprehensive Renewal Plan incurred around 500 man-days of effort. The work was organized in 5 phases (Table 1):

The objective of this phasing was to start with a vision of needs by asset classes or families, for which some collective strategies could be defined, followed by substations

Table 1. The 5 phases in construction of the renewal plan

Phase 1 Preparation 2 weeks	• Define the core team and ensure its availability • Training of facilitators • Construction of the Investment Evaluation Worksheet, gathering the main technical, financial, regulatory and planning information required
Phase 2 Equipment Class Evaluations - 2 weeks	• Conduct studies for the most critical families / types of equipment • Indications of specific risks and needs by the specialist areas
Phase 3 Substations: Pilot for one region - 2 weeks	• Studies focusing on substations, taking advantage and expanding the work of the previous phase, and identifying synergy opportunities • Intense participation of specialists from the regions, who work in direct contact with the assets
Phase 4 Substations: Other regions - 2 weeks	• Replicate previous phase in the other regions • Core team divided into sub-teams, each working with a regional, in the respective regional headquarters
Phase 5 Consolidation of Optimized Plan - 3 weeks	• Consolidation, incorporating investment plans for lines, telecom system and operation centre • Involvement of project planning and financial planning areas
Lines, Telecom, COT In parallel with phases 2, 3 and 4	• Definition of renovation plans for Lines Transmission, Telecommunications System and Operation Centre • Themes led by the respective specialist areas

(with system contexts and differences) in phases 3 and 4, offering synergy and alignment opportunities for execution of renewal projects involving equipment from different families.

Phase 2 equipment studies were led by engineers from the specialist areas of each asset class, and resulted in the identification and cost/risk justification (i.e. Step 4 of SALVO) of a preferred range for replacement timing (for example "3 to 5 years" or "7 to 10 years"). This information was used in the next stage, led by the regional operations and planning staff, to seek project grouping opportunities ('bundling' in Step 5 of SALVO) and establishing a more specific proposed timing.

In phase 3, the core team moved to the headquarters of one of the regional offices, to establish a standardized work methodology and consensus support for the emerging plan. The hands-on, visual method of post-it's and 'whole picture visualization' was adopted to seek greater engagement in building a histogram of future CAPEX requirements - this proved better than peering at a computer screen.

This approach also enabled parallel working in phase 4, without prejudice to the quality and homogeneity of the criteria used. The combined results were assembled into a total program (SALVO Step 6), with total costs and risk forecasts out to the 2030 horizon.

2.3 Investment Pre-Evaluation

The main tool used in the initial construction of the Renovation Plan was a set of spreadsheets in which a "Pre-Evaluation of Investments" was calculated. Because they included very large amounts of data, and to allow work to be carried out in parallel in the various regional offices, the spreadsheets were sub-structured for regional responsibility, with typically 10 substations per region, each with an average of 3,000 line items. Each line corresponded to a discrete asset or an asset system. And, in each case, the pre-evaluation criteria considered the cost, risk and benefits (e.g. revenue allowance under regulatory tariff rules) represented by the potential replacement of the asset(s).

The calculation of benefits involved a calculation of the various aspects of value to the organization, in a long-term and sustainable viewpoint. Using the 'Shamrock' model (see Fig. 2) from the European MACRO Project (EU1488) [3], we considered all the stakeholder impacts and consequential effects. This included, in addition to CAPEX and revenue effects, effects upon operating costs (such as preventive and corrective maintenance) and changes in reliability, risk and potential penalties due to planned/unplanned downtime. It is the combined effect of these factors ('Total Business Impact' in SALVO terms) that was modelled at the individual asset level. This provided a direct link between the technical view of the assets, risks, obsolescence etc., and the business case for their renewal. The estimated annual cash flows (both planned and probabilistic/risk), considering different possible renewal timings, resulting in spreadsheets with more than 230 columns.

A critical success factor was the alignment of the model with decision-making criteria agreed during the construction of the company's Strategic Asset Management Plan. These criteria were reinforced in the workshop sessions with technical, financial, engineering and management departments. Such validation ensured understanding and acceptance of the model and therefore the credibility of the results.

Phases 3 and 4 involved multidisciplinary work meetings in which the team went through each asset in each substation to verify and agree that the asset should be a candidate for the ten-year plan. If so, the team would then group assets from the same substation into a proposed project bundle, agreeing on the specific year for their execution (Fig. 3).

Fig. 2. (Right) Shamrock model for identifying and quantifying stakeholder impacts

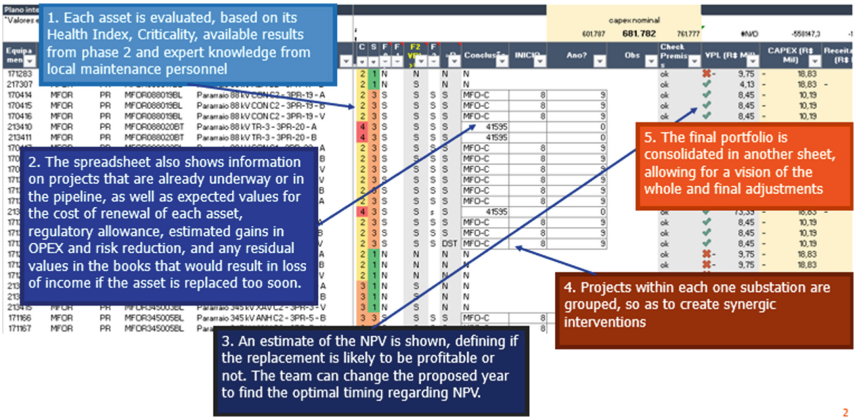


Fig. 3. (Below) screenshot of a "pre-evaluation of investments" worksheet, with explanations about its use.

The four filters used to identify renewal candidates were:

F0: Do you want to propose the asset as a candidate? (if yes, then why?)
F1: *Is an intervention mandatory*? (if Yes, other filters do not matter. This also includes cases of red/'intolerable' risk)
F2: *Is it profitable to replace it*? Field is automatically filled in from an initial NPV calculation. A negative NPV suggests that the asset should not be included in the plan without further studies (which does not prevent the team from requesting such a study anyway)
F3: *Is it the only reasonable alternative*? This filter exists to remind participants that, although the renewal might be profitable, there may be a better alternative - so asset is marked for further analysis (Step 3 of SALVO is designed to stimulate such 'lateral thinking' consideration of alternative options).

When applying these filters and proposing a grouping and timing of projects, the team dig into the health and criticality information about the assets, the calculated NPV, and present value estimates for the impacts of Capex, Opex, risk and revenue of the proposed renewal, as well as information on any projects, overhauls or other tasks already planned for the either the specific asset or nearby assets in the same substation (as they are evident in adjacent lines in the same spreadsheet).

The calculations considered the time value of money, and the cost and risk impacts adjusted according to the year of project completion. This allows the team to investigate timing alternatives quickly, looking for manually-optimized a plan for the substation and avoiding a fragmented plan. Asset renewals are therefore grouped into larger projects, representing efficiency opportunities that need to exceed any cost/risk effects of some assets not therefore being replaced at their personally optimal timings. One of the modules of the DST toolkit (DST Schedule Optimizer™) automates this process, using a genetic algorithm to find optimal bundling of tasks and timings, with least cost/risk/downtime impact.

2.4 The Handling of Risk

The model used has the advantage of being extensible to all asset classes, degradation mechanisms and risk types, allowing a consistent treatment of all vulnerabilities and business impacts.

This incorporated a criticality classification of assets, being a holistic measure of their importance to the company. The classification considered the asset's potential to cause negative impacts such as systemic and consumer impacts, repair cost and damage to the company's image, and is expressed in the form of a criticality score. This has a direct linear relationship with the monetized impact of a risk event, with 1 point assigned to represent an impact equivalent to one thousand Reais (aprox US$ 200). This criticality model was created with cross-functional input, and the general results and methodology validated with CTEEP management.

Our study also incorporated a Health Index of assets, which again had been the focus of previous work within the company, conducted by the maintenance teams from the regions. The Index is intended as a measure of the urgency of attention that an asset has regarding end-of-life interventions, such as replacement or refurbishment. A different methodology was created by the specialist teams per type of asset as appropriate, which may include for example such variables as age, recent test results, obsolescence, failure rate, and others.

A correlation was assumed between the probability of a risked event and the health index of each asset, once beyond a certain threshold. The following formula was developed to represent this relationship:

$$\text{Risk} = \text{k} * \max\{\text{S},\ \text{m}\}^\wedge e * \text{C} \quad (1)$$

Where:

S - Health of the asset (from 0 to 5)
m - Health threshold (actual onset of deterioration) [adjusted value = 1.75]
C - Criticality (0 to 6000)
k - Scale factor [adjusted value = 0.0024 (R$ thousand)]
e - Health exponent [adjusted value = 1.4]

Parameters K and m were adjusted so that:

a) Summation of risks of current assets should be equal to the estimated annual impacts observed in relation to asset failure events (~R $ 100 million);
b) Adding the risks of a *simulated* portfolio of all assets being new (replaced), the total level would drop by half. This was a 'what if?' premise that recognises the fact that some risks are external, and new assets are, themselves, prone to some failure modes (e.g. commissioning risks, design defects);
c) The difference in probability between the worst and best health assets was challenged for reasonableness. The resulting range, which was considered reasonable, represented a factor of 4.3, i.e. an asset in the worst health would represent a risk up to 4.3 times higher than a corresponding new asset.

The correlation of failure probabilities with health indexes is known to be fragile [5]. So the evaluation of intervention decisions and optimal timings can be improved with a multi-factor modelling that can incorporate a mix of degradation mechanisms, economic profiles, risks and uncertainties. This was provided by the DST toolkit, which brings together reliability engineering, degradation, risk and life cycle modelling with financial appraisal techniques and real-time 'what if' analytics. The module that explicitly models remaining asset life and optimal re-investment timing is DST Lifespan Evaluator™.

2.5 DST Lifespan Evaluator™

This tool (see Fig. 4) was developed by the SALVO consortium to provide an analysis of the remaining life of a current asset and the investment, at an optimal time to be determined, to a new item (not necessarily like-for-like) with, in turn, an optimal life cycle (duration and cost) also to be determined.

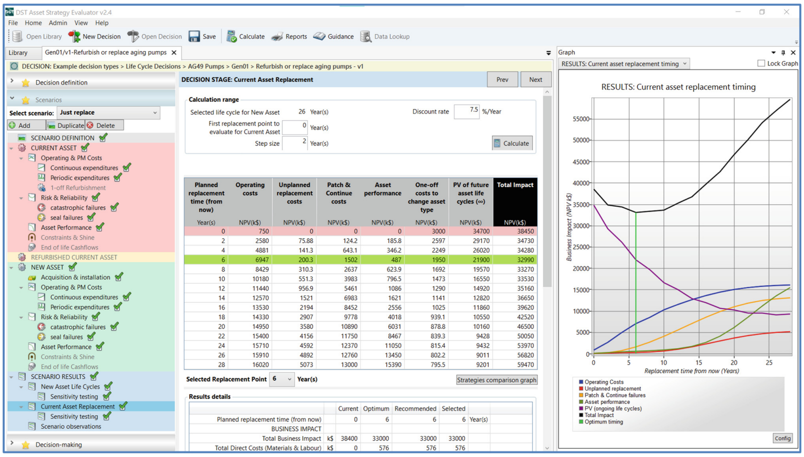


Fig. 4. DST Lifespan Evaluator™ optimisation of asset renewal timing

Treatment of cost, risk and performance elements in this regime are expressed as present values, subject to survival (itself a product of the risk projections), for different replacement timings.

The modelling thus includes:

- Probability of an asset surviving to T, at which point it will be replaced;
- Operating and maintenance costs and how these may change;
- The replacement costs and consequences if the asset fails before T;
- Any additional cash flows occurring at T (e.g. decommissioning costs, or potential resale benefits);
- Procurement and installation costs of the new asset
- Ongoing life cycle attributes (costs, risks and performance patterns) for the new asset (which may not be the same as the current one and based upon calculation of the optimal lifespan within any absolute horizon constraints).

The general form [7] is given by the following expression, to derive the total present value (TVI) of the replacement strategy for the existing asset:

$$\mathrm{TVI} = \mathrm{V}_{\mathrm{PC}} + (\mathrm{V}_{\mathrm{IC}} + C_{OO} + C_{\mathrm{NCEL}}))(R_{NC}(N)D(N) + \int_0^N f_{NC}(t)D(t)dt) \quad (2)$$

- V_{PC} the present value of costs and risks of the pre-cyclic phase, up to either a non-repairable failure of the existing asset or (subject to survival of that asset) time N
- V_{IC} the cost of the planned asset replacement and any initial installation costs
- C_{OO} the operating cost of future life cycles
- $\mathrm{C}_{\mathrm{NCEL}}$ the end-of-life cashflow at end of the pre-cyclic period
- R_{NC} the survival probability during the pre-cyclic period
- f_{NC} the pdf of failure during the pre-cyclic period
- D(t) is the discount factor at time t (and (D(N) is the same at time N)

The diversity of life cycle parameters and flexible modelling options available within DST Lifespan Evaluator™ enable a rich life cycle model [6] to be developed for asset investment strategies. However there are additional complexity issues when rolling up multiple asset replacement decisions to portfolio level, with overlapping system contexts, interacting risks, bundling opportunities, competing for resources and subject to global constraints and performance goals. These complexities are navigated, and the total investment programme optimised, using artificial intelligence algorthms in the DST Schedule Optimizer™ and DST StAMP™ (Strategic Asset Management Planner) modules (Fig. 5).

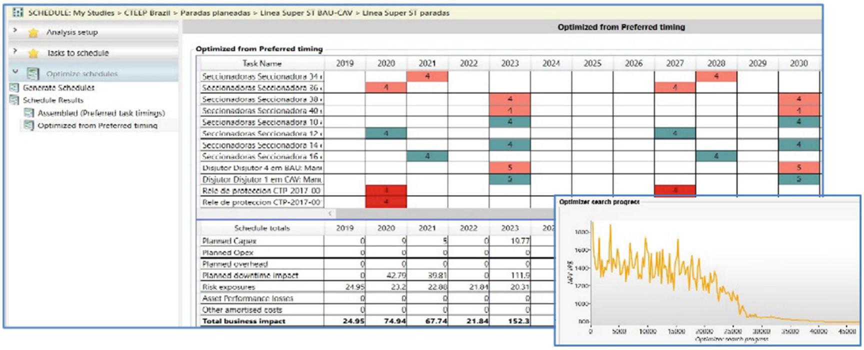


Fig. 5. DST Schedule Optimizer™ seeking best-value bundling opportunities

3 Results

The CTEEP results were compiled in the form of a master table indicating, for each asset, the proposed year for renewal and the proposed project, or an indication that the asset is not expected to be renewed within the plan's 10-year horizon. The results can be filtered and grouped, showing how the investments are distributed by year, location, type of asset, etc. This material was circulated internally to support the formal proposals in a final report explaining the results and basis for the optimal investment programme. The results showed the need for a significant increase in asset replacements and Capex investment (see Fig. 6).

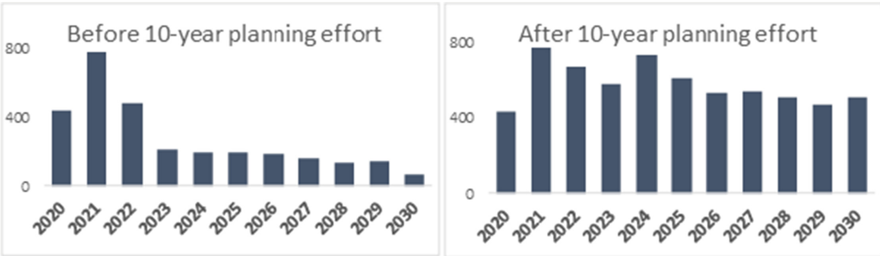


Fig. 6. Capex plan 'before and after'

A total of 408 new asset replacement projects were identified as needed, adding to the 487 projects already planned or underway for the period 2020—2030. The new projects also represent a greater proportion of the total Capex required. This reflects the greater use of bundling opportunities (SALVO step 5), yielding bigger combined project scopes (for example, renovating an entire substation in a single project, instead of several discrete equipment unit replacements).

By the end of 2030, the implementation of the plan will represent the renewal of almost half of the current portfolio of substation assets. Figure 7 shows the expected asset age profile, normalized by their regulation-defined 'useful life'. As mentioned, renewal of assets that have exceeded (>100%) their regulatory useful life enables an increase in revenue through an increased recognized value for the company's infrastructure (the basis for remuneration).

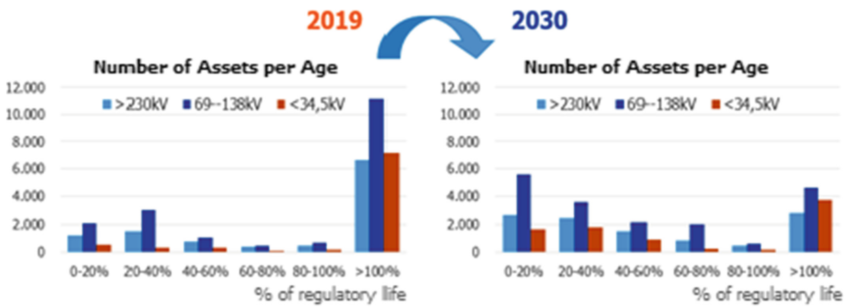


Fig. 7. Asset age profiles, as percentage of their 'regulatory life': current and projected (by 2030).

Not all proposed projects have a positive NPV in all scenarios (see Table 2). Some negative values are found in the pessimistic scenarios explored, and even in some optimistic scenarios (e.g. if a change is mandatory change, or some other overriding stakeholder obligation). The exploration of such ‘overrides’, plus data uncertainties and sensitivities are handled rigorously within the DST tools. These cases are further analysed as the proposed year for the project approaches, to confirm/reject, refine and develop the individual/grouped project justifications and their optimal timings.

Table 2. Distribution of proposed projects according to their current merits

Mandatory	Clearly justified	Probably justified	Not justifiable currently
22%	18%	48%	12%

Subsets of assets that contribute the most to the ‘less profitable’ projects have also been identified, allowing batches of related studies to address families of similar cases. For example, designs that provide for the replacement of electrical disconnectors tend to be less attractive due to the current procurement costs. If better prices can be negotiated (considering the greater volume and more predictable demand that is now known), many of these projects become more economically attractive. It is anticipated that other such cases will emerge as project-specific and programme-level DST modelling is applied or updated, incorporating the rate of change in risks, costs etc.

4 Conclusions

The construction, justification and optimization of the 10-year strategic plan for the renewal of assets was a highly successful study, combining a systematic review of aging asset risks, engagement with all key stakeholders, advanced analytics (of both available hard data and the quantification of expert knowledge) and commercial focus (including business case justification of the results). The outputs achieved several important benefits to the company. These include:

- A credible projection of the future Capex requirements, allowing for better financial planning, with uncertainties recognized and quantified;
- Alignment of technical and financial staff, working together to build the plan;
- Life cycle value demonstration with a ‘TotEx’ vision for the projects, and a clear indication of which projects are marginal and should be further investigated before approval/rejection;
- Demonstration of responsible asset management to shareholders such as government, ANEEL (industry regulator), distributors and customers;
- Visibility of renewal plans, especially important for:
  - procurement planning, being in a better position to negotiate with suppliers for a large and predictable long-term demand;

– Asset managers in regional offices, who can optimize their maintenance strategies and contingency plans.
– Project planners, who can submit their projects for regulatory approval, where necessary, with adequate leadtimes;
– Engineering/design, who now have a greater incentive and opportunity for standardizations and programme efficiencies.

This study showed the value of starting with simple surveys, using extensive consultation and consensus-building, backed by full-rigor modelling of life cycle costs, risks and performance factors, at the levels of both discrete cases (optimal individual project timings) and the total program/portfolio (optimal bundling, scheduling and treatment of global constraints). The methodology has been documented and incorporated into the organisation's annual review and updating of the investment plans as part of the strategic business planning processes.

References

1. http://www.SALVOProject.org
2. http://www.decisionsupporttools.com
3. http://www.MACROproject.org
4. Woodhouse J (2018) Asset management decision-making: The SALVO Process, ISBN 978-0-9563934-5-6. http://www.SALVOproject.org or www.amazon.com
5. Thomas P, Bratvold R, Bickel J (2013) The risk of using risk matrices. SPE Econ Manage 6. https://doi.org/10.2118/166269-MS
6. Calder M et al (2018) Computational modelling for decision-making: where, why, what, who and how. R Soc Open Sci 5:172096. https://doi.org/10.1098/rsos.172096
7. Hunt A, Taylor S, Morgan M, Lange R, (2020) Modelling investment plans at asset portfolio level using optimum plan rationalisation approaches. In: 4th IFAC Workshop on Advanced Maintenance Engineering, Services and Technologies, University of Cambridge, UK

System for Early Detection of Insulation Failures of Electric Machinery

Bruno Renó Gama[1](✉), Wilson Cesar Sant'Ana[2], Luiz Eduardo Borges da Silva[1], Erik Leandro Bonaldi[2], Germano Lambert-Torres[2], Camila Paes Salomon[1], Isac Antonio dos Santos Areias[1], Daniel de Almeida Arantes[1], Fernanda Mitchelly Vilas Boas[1], Fabio Monteiro Steiner[3], and Rafael Bartholomeu Bernardo Carvalho[3]

[1] Universidade Federal de Itajubá (UNIFEI), 1303 BPS Ave., Itajubá, MG 37500903, Brazil
br.gama@unifei.edu.br

[2] Instituto Gnarus, 53 Jose Juliao Simoes St, Itajubá, MG 37502495, Brazil

[3] EDF Norte Fluminense, 330 República do Chile Ave, Rio de Janeiro, RJ 20031170, Brazil

Abstract. This work presents the development of a system for detection of early damage to insulation of electrical machines. This system is composed by a hardware implemented in a FPGA board in conjunction with a software written in C#. The frequency response analysis (FRA) technique is used to infer about the machine insulation condition. The application of FRA consists of obtaining, periodically, the impedance spectra of the device under test (DUT). The obtained spectra are compared with a base spectrum, called baseline. Differences between the baseline and the acquired spectra can indicates a damage or the forming of a failure mechanism in machine insulation. The software implements a sweep frequency algorithm to control the hardware and obtain the machine impedance spectra. This algorithm communicates with the hardware, sending commands for generation and acquisition of signals in a predefined frequency range. With the acquired signals, the software is able to calculates the impedance for each signal and, in the end of the process, the impedance spectrum is obtained. Since the early damage diagnosis is based in comparison between spectra, and a visual analysis requires a well-trained and expert maintenance team, it is proposed in the literature the use of some statistical indexes to compare the data. Those indexes have the advantage of obtaining a more objective diagnosis of the DUT, since visual analysis is subject to subjective interpretations. The software also implements some of the indexes proposed in the literature for a better analysis of the machine insulation. To evaluate the developed system, experimental results are presented using a machine with taps on its windings in order to emulate insulation faults. Since insulation faults represents a considerable percentage of electrical machines failure, the proposed system has great potential in industrial applications, preventing unscheduled stops of the machinery.

1 Introduction

Electrical machines are found in all kind of industry around the world. Some of them are responsible for extremely critical operations, so, an unpredictable failure in one of these

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 185–195, 2022.
https://doi.org/10.1007/978-3-030-96794-9_17

machines can represent huge losses to industry. Among all causes of electrical machine failure, a considerable percentage is related to insulation damage.

A survey performed by CIGRE (Conseil International des Grands Réseaux Électriques, in English: International Council on Large Electric Systems), analysed 1199 electric generators during 10 years in 5 countries, the study showed that a total of 56% of all observed machine failures were caused by insulation damages [1, 2]. Another similar statistical number, presented at [3] and [4], point that for low to medium voltage machines, around 25%, and for high voltage machines, more them 50% of the failures are caused by insulation degradation. With relation to only induction machines, the range of this kind of failure is about 26% to 36% [5]. Therefore, it is incontestable that the research and development of techniques involving the detection and prediction of insulation failures of electrical machines is of huge importance to the academia and to the industry.

Diagnosis techniques are largely found into literature, some of most important are: insulation resistance [6, 7], polarization index [6, 7], ac hipot tests [6–8], dc hipot tests [6–8], power factor (PF) tip-up tests [7], hipot tests for turn insulation [7], partial discharge (PD) tests [7, 9], surge test [7, 8]. The problem with those techniques, according to the patents [10] and [11], is that they are able only to inform the existence or not of the fault, but they are incapable of detecting the developing of the issue, therefore, they cannot detect early insulation failure. Based on this problem, the patents [10] and [11] propose a technique that analyses the machine impedance spectrum. The method basically injects high frequencies signals into machine stator winding and measure the machine impedance for each applied frequency, as a result, an impedance spectrum is obtained. This spectrum is compared with historical data, changes between recent and historical data can indicates damages to machine insulation.

A technique very similar to the one presented by the patents [10] and [11] is called FRA (frequency response analysis), and it is widely used for transformers diagnosis. The procedure is presented at [12]. With relation to rotating machines, its use is still restricted to research due to complex machine stator winding behavior at high frequencies and due to lack of repeatability at measurements caused by the influence of rotor position [14]. The FRA technique, like the technique presented at [10] and [11], compares impedance spectra with a base spectrum, differences between the actual measured spectra with the base spectrum indicates changes into physical characteristics of the machine insulation. The comparison between spectra usually requires a specialist team, which analyses the spectra and provides a verdict about insulation health. This brings two main problems, first, the constant need of specialist team presence. And, second, the conclusion about the real insulation condition is subjected to analyzers opinion, therefore, it is not an objective evaluation method. To solve this problem, literature suggests a wide variety of statistical indexes, that can be used to assist insulation diagnosis. References [15] and [16] present different statistical indexes applied to FRA analysis. Another promissory application of FRA is an on-line analyse, therefore, a system in operation. With this configuration, there is no need to stop the machine to perform the analysis. Like mentioned in [17], the challenge with on-line operation is the connection of the FRA equipment, which operates at a few volts, with the power terminals of an energized machine, which operates at hundreds/thousands of volts. To solve this problem, [13] presents a capacitive based device and [17] presents a C-L-C filter based device for coupling FRA system and machine.

In references [14–17], it was used a PicoScope 5203 based system for FRA application to determine the machine insulation condition. Besides been a relatively expensive system, the PicoScope has no flexibility in terms of hardware configuration, so, in [18] an FPGA approach was proposed to solve this problem. A tutorial regarding the development of the FPGA system for the FRA application is presented in [19]. The system was able to detect early insulation failures, but only one index indicator was used to data analyse, which could limit the diagnostics. Therefore, this work proposes an extension to [18] with a wide range of indexes that can be used to certify a machine fault tendency. Furthermore, the spectra are now subdivided into vary frequency sections (low frequency, medium frequency and high frequency), allowing the tendency classification by spectrum region.

Section 2 reviews FRA analysis and some of statistical indexes proposed in the literature. Section 3 presents the developed system, composed by the hardware and software. Section 4 shows the experimental results of the system applied to two different machines. Finally, Sect. 5 concludes this paper.

2 Frequency Response Analysis Method and Statistical Indexes

2.1 Frequency *R*esponse *A*nalysis (FRA)

The Frequency Response Analysis (FRA) technique, as explained in [20], is an efficient diagnostic method. The idea is to inject a high frequency signal into a machine winding to obtain its impedance. This procedure is repeated for a wide range of frequencies. By the end of the process, an impedance spectrum is obtained. This spectrum is compared with a reference spectrum, called baseline. Differences observed between the baseline and the new measured spectrum suggest some variation in the parameters of windings. [20] also points the existence of two ways of applying the FRA method: with an impulse signal injection into winding, or with a frequency sweep using sinusoidal signals. While impulse application has the advantage of being a faster method, frequency sweep has better signal to noise ratio, has a wider frequency range, and measurement can be executed with less equipment. Since a wider frequency range analysis is preferred compared to a faster analysis, this work is developed with the frequency sweep method.

FRA application must use a circuit to perform its sweep frequency, the one used in this work, and explained in [21], is presented in Fig. 1.

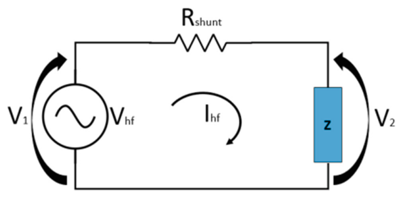


Fig. 1. Measurement circuit.

Since the idea of the technique is to determine the machine impedance spectrum, for each input signal (applied by the source - Vhf), an impedance is determined (Z).

Analyzing the circuit, it can be determined that the impedance of a load, Z, is calculated by the following Eq. 1.

$$Z = R_{shunt} \cdot \frac{V_2}{V_1 - V_2} \tag{1}$$

So, to apply the Eq. 1, it must be measured voltages V_1 and V_2, also, the shunt resistor, R_{shunt}, is known.

2.2 Statistical Indexes

There are a wide variety of statistical indexes presented in FRA literature. Those indexes are useful to allow a less subjective spectrum analysis, since a visual comparison between spectra relies on the experience of the analyst. The indexes compare two sets of data x_i and y_i, each set of data is composed by n values, where n is the number of points presented in the acquired spectrum. By the comparison between the sets of data, the index equation provides a number that represents how different are those compared data. Statistical index information are found at [20, 22–24], and a survey on main statistical indexes is presented in [15]: Correlation Coefficient (CC), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Comparative Standard Deviation (CSD), Sum Squared Ratio Error (SSRE),Sum Squared Max-Min Ratio Error (SSMMRE), Absolute Sum of Logarithmic Error (ASLE), Absolute Difference (DABS), Minimum-Maximum ratio (MM), Hypothesis tests (t-test, f-test), Normalized Capacitance (NC), Sum Squared Measurements Variations (SSMV).

2.3 Frequency Range Analysis

As pointed in [25], the existence of a machine fault does not affect equally the entire spectrum. So, there are some spectrum areas that are more affected by the damage than others. [25] concludes that a better index sensitivity can be achieved with the division of the entire spectrum into smaller sections.

About the spectrum division, [15] affirms that the literature does not have a consensus on the ideal frequency ranges. So, each work applies a division based on personal analysis. This work follows [15], choosing tree regions for range division. The chosen regions are: Low Frequency (LF) from 1000 Hz to 30538 Hz. Medium Frequency (MF), from 32745 Hz to 200923 Hz. High Frequency (HF), from 215443 Hz to 1000000 Hz.

3 Early Insulation Failure Detection System

The proposed system is composed by a hardware and a software. The hardware module is composed by a development board and a developed printed circuit board (PCB). On the development board, there is a FPGA, a dual channel analog-to-digital converter (ADC), and a digital-to-analog converter (DAC), with those components, a sweep frequency system is projected and implemented. The PCB is proposed to implements the measurement circuit, necessary to impedance measurement process. The software, developed using C# language, is responsible for controlling the hardware module and for the calculation of the statistical indexes.

3.1 Developed Hardware

The STEMLab board is used for the sweep frequency system development. This platform is composed by the Xilinx Zynq-7000 family SoC. The SoC comprises an Artix-7 family FPGA and a dual-core ARM Cortex-A9. The board presents a 14-bit ADC and a 14-bit DAC, both of high speed (125 Msps), also, it allows Ethernet connection and has a variety of peripherals [18, 19]. Figure 2 presents a platform overview, with some main components highlighted. The developed system is able to injects and acquires frequencies within the range of 100 Hz to 1 MHz, using this frequency range, the system is able to measure and compare a machine's impedance spectra and infer about its insulation condition.

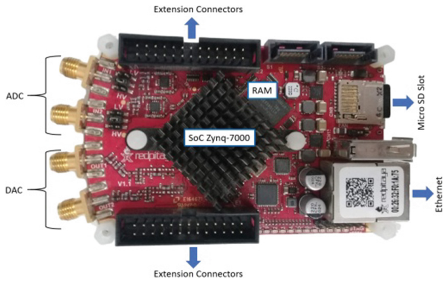


Fig. 2. Development platform. The following components are highlighted: ADC inputs, DAC outputs, SoC Zynq-7000, Extension connectors, RAM, micro-SD slot, Ethernet.

To perform the impedance measurement, and, therefore, obtain the impedance spectrum, of the device under analysis, it is necessary a measurement circuit, like the one presented in Fig. 1. This circuit is implemented in a PCB and is presented in Fig. 3. The

Fig. 3. Measurement circuit. 1: SMA connector for input signal; 2: SMA connector for channel 1 output signal; 3: SMA connector for channel 2 output signal; 4: SMA connector to device under test connection; 5: Shunt resistor; 6: Button on/off; 7: Micro USB power.

PCB couples the development board, this layout prevents problems like bad contacts and external noisy.

Besides the sweep frequency circuit implemented on FPGA, the hardware also implements a TCP/IP based protocol, therefore, a client/server system, where the hardware acts like a server, providing services to a remote client (software) connected to it. Figure 4 presents the server (hardware) basics implementation.

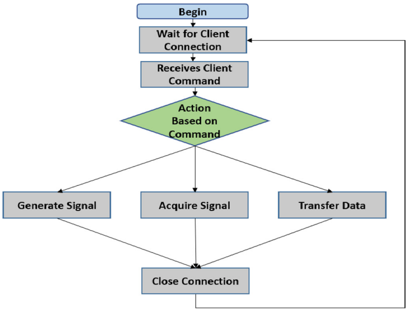


Fig. 4. Server implementation.

3.2 Developed Software

The software communicates with the hardware via Ethernet connection, and it acts like a client, configuring and controlling the remote device. The client has a wide set of configurations available to hardware operation and, also, it implements functions to indexes calculation, allowing a visual analysis of the insulation machine condition. Figure 5 shows the implementation of software sweep frequency algorithm.

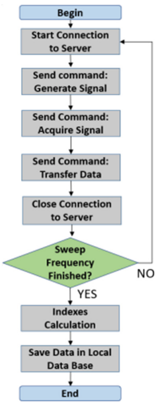


Fig. 5. Software implementation.

4 Experimental Results

The developed system is applied to a 2 kW, 220 V, 2 round-poles synchronous machine (presented in Fig. 6, with developed system coupled). This machine is custom made, therefore, it has access to some winding points (taps). [26] affirms that machine insulation condition is related to machine winding capacitance, therefore, in this work, insulation faults are simulated by introducing different capacitors into taps of machine winding.

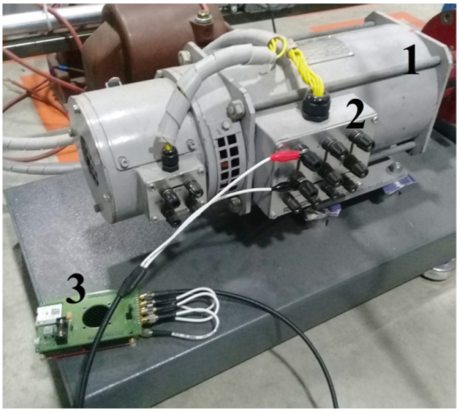


Fig. 6. Developed system applied to synchronous machine. 1: Synchronous machine. 2: Panel to access taps of machine winding. 3: Developed system.

4.1 Results on the Synchronous Machine

For the experiments with the synchronous machine, six different capacitor values were used between two taps of the machine winding. These taps represent 1% of the total machine winding. The values of the capacitors were chosen in order to simulated a progression of the insulation failure. The system was configured to analyze the machine in range of 1 kHz to 1 MHz, for each different condition, twenty impedance spectra are obtained. The spectra are presented in Fig. 7, to avoid visual pollution, it is taken the mean of the twenty spectra obtained for each condition. The capacitances used to fault simulation are: 30 nF, 43 nF, 330 nF, 470 nF, 1 uF, and a short-circuit case.

Four statistical indexes were chosen to analyse the obtained spectra, they are: ASLE, DABS, SSMV, NC. Those statistical indexes were calculated based on the total range (1 kHz–1 MHz), the low frequency (LF) range (1 kHz–30538 Hz), the medium frequency (MF) range (32745 Hz–200923 Hz), and the high frequency (HF) range (215443 Hz–1 MHz). Figures 8, 9, 10 and 11 present statistical curves obtained by the system applied to the synchronous machine, each figure analyse a specific frequency range.

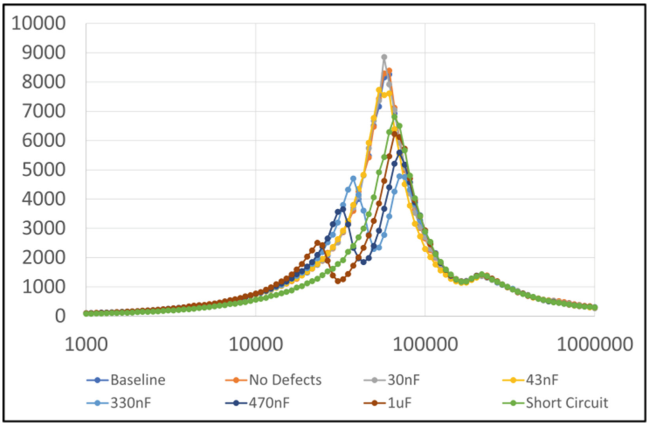


Fig. 7. Synchronous spectra for each insulation condition.

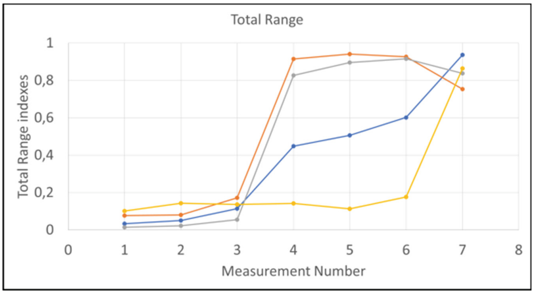


Fig. 8. Synchronous machine statistical indexes curves – Total range. Blue: ASLE; Orange: DABS; Gray: SSMV; Yellow: NC.

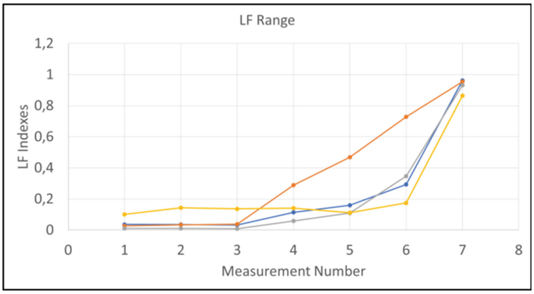


Fig. 9. Synchronous machine statistical indexes curves – Low frequency range. Blue: ASLE; Orange: DABS; Gray: SSMV; Yellow: NC.

In Fig. 8, an overall increasing tendency is observed for all indexes, from condition 1 to 7, however, each index responds differently to the simulated fault. ASLE index has a consistent increasing value, while DABS and SSMV responds aggressively from third

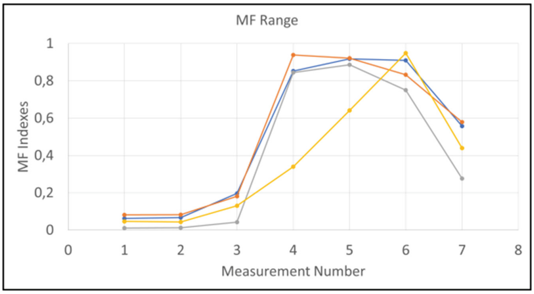


Fig. 10. Synchronous machine statistical indexes curves – Medium frequency range. Blue: ASLE; Orange: DABS; Gray: SSMV; Yellow: NC.

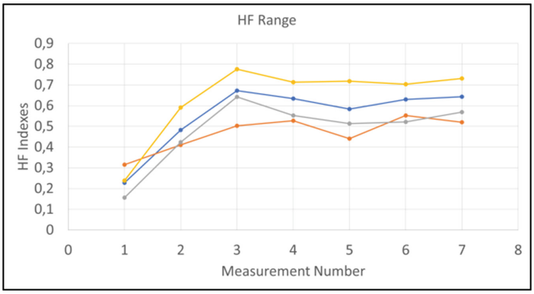


Fig. 11. Synchronous machine statistical indexes curves – High frequency range. Blue: ASLE; Orange: DABS; Gray: SSMV; Yellow: NC.

to fourth condition, and, after that, settle on their maximum value. For NC index, there is a small increase from condition 1 to condition 6, and an intense modification from 6 to 7 condition. For LF range in Fig. 9, the three first conditions all have a similar and neutral response to all index, from condition 3 to condition 4, all indexes present an increasing tendency up to last condition. MF range, presented in Fig. 10, presents a first moment increasing and second moment decreasing movement, for all indexes, during the experiment. Finally, for HF range, presented in Fig. 11, it is observed a higher sensibility for the first three conditions, presenting an increasing tendency for those conditions, and settling on their maximum value for the subsequent conditions.

5 Conclusion

This work presented a FPGA based developed system that is able to early detect insulation faults to electrical machines. The system applies the FRA technique in addition to statistical indexes to detect a fault tendency on a device under test. The faults were simulated with the addition of different capacitors values, simulating a progression of the insulation failure. A synchronous machine was analysed, the system was able to correctly follows the progression of insulation failure. The use of different statistical indexes and different frequency range for the fault analysis allows a more detailed and precise machine diagnosis, therefore, the analyst team is able to better understand and take decisions about the device, preventing unexpected stops due to failure problems with the machine. For future works, it is proposed the use of well-established decision-making process in addition to the presented statistical indexes to better assess the machine condition. Also, the system can be improved to online work, therefore, with a machine in operation, and, finally, statistical techniques can be applied to provide a prognosis of the analysed machine.

Acknowledgments. The authors would like to thank the following Brazilian Research Agencies: CAPES, CNPq, FAPEMIG and ANEEL R&D for the financial support of this project.

References

1. Sumereder C (2008) Statistical lifetime of hydro generators and failure analysis. IEEE Trans Dielectr Electr Insul 15:678–685
2. Brütsch R, Tari M, Fröhlich K, Weiers T, Vogelsang R (2008) Insulation failure mechanisms of power generators [feature article]. IEEE Electr Insul Mag 24:17–25
3. Neti P, Younsi K, Shah MR (2013) A novel high sensitivity differential current transformer for online health monitoring of industrial motor ground-wall insulation. IEEE Energy Convers Congr Expos 2013:2493–2499
4. Jensen WR, Foster SN (2019) Remaining useful life estimation of stator insulation using particle filter. IEEE Energy Convers Congr Expos (ECCE) 2019:7004–7011
5. Drif M, Cardoso AJM (2014) Stator fault diagnostics in squirrel cage three-phase induction motor drives using the instantaneous active and reactive power signature analyses. IEEE Trans Industr Inf 10:1348–1360
6. Klempner G, Kerszenbaum I (2004) Operation and maintenance of large turbo-generators. IEEE press series on power engineering, vol 3. Wiley-IEEE press, pp 1–1032
7. Stone GC (2003) Recent important changes in IEEE motor and generator winding insulation diagnostic testing standards. In: IEEE industry applications society 50th annual petroleum and chemical industry conference, pp 101–110
8. Stone G, Boulter EA, Culbert I, Dhirani H (2004) Electrical insulation for rotating machines: design, evaluation, aging, testing, and repair. IEEE press series on power engineering, vol 2. Wiley-IEEE Press, pp 1–672
9. James RE, Chi S (2008) Condition assessment of high voltage insulation in power system equipment. IET power and energy series. Institution of Engineering and Technology, pp 1–276
10. Dister CJ, DelVecchio PA, Rogovin DN (2000) System to provide low cost excitation to stator winding to generate impedance spectrum for use in stator diagnostics. US Patent 6035265
11. Kendig MW, Rogovin DN (2002) Method of conducting broadband impedance response tests to predict stator winding failure. US Patent 6483319
12. IEEE PC57.149/D9.2 (2012) IEEE draft guide for the application and interpretation of frequency response analysis for oil immersed transformers. In: IEEE PC57.149/D9.2, pp 1–66
13. Gomez-Luna E, Mayor GA, Gonzalez-Garcia C, Guerra JP (2013) Current status and future trends in frequency-response analysis with a transformer in service. IEEE Trans Power Deliv 28:1024–1031
14. Sant'Ana WC et al (2016) Influence of rotor position on the repeatability of frequency response analysis measurements on rotating machines and a statistical approach for more meaningful diagnostics. Electr Power Syst Res 133:71–78

15. Sant'Ana WC et al (2016) A survey on statistical indexes applied on frequency response analysis of electric machinery and a trend based approach for more reliable results. Electr Power Syst Res 137:26–33
16. Sant'Ana WC (2017) On the use of hypothesis tests as statistical indexes for frequency response analysis of electric machinery. Electr Power Syst Res 147:245–253
17. Sant'Ana WC et al (2016) Early detection of insulation failures on electric generators through online Frequency Response Analysis. Electr Power Syst Res 140:337–343
18. Gama BR et al (2019) Desenvolvimento em plataforma FPGA de sistema de detecção precoce de falhas em isolamento de máquinas elétricas. Revista de Engenharia e Tecnologia 11:49–58
19. Gama BR et al (2021) FPGA Prototyping using the STEMlab board with application on Frequency Response Analysis of Electric Machinery. IEEE Access 9:26822–26838
20. Ryder SA (2003) Diagnosing transformer faults using frequency response analysis. IEEE Electr Insul Mag 19:16–22
21. Lamarre L, Picher P (2008) Impedance characterization of hydro generator stator windings and preliminary results of FRA analysis. In: Conference record of the 2008 IEEE international symposium on electrical insulation, pp 227–230
22. Kim JW, Park B, Jeong SC, Kim SW, Park P (2005) Fault diagnosis of a power transformer using an improved frequency-response analysis. IEEE Trans Power Deliv 20:169–178
23. Kennedy GM, McGrail AJ, Lapworth JA (2007) Using cross-correlation coefficients to analyze transformer sweep frequency response analysis (SFRA) traces. In: 2007 IEEE power engineering society conference and exposition in Africa – PowerAfrica, pp 1–6
24. Badgujar KP, Maoyafikuddin M, Kulkarni SV (2012) Alternative statistical techniques for aiding SFRA diagnostics in transformers. IET Gener Trans Distrib 6:189–198
25. Wimmer R, Tenbohlen S, Heindl M, Kraetge A, Kruger M, Christian J (2007) Development of algorithms to assess the FRA. In: Proceedings of the 15th international symposium high voltage engineering
26. Perisse F, Werynski P, Roger D (2007) A new method for AC machine turn insulation diagnostic based on high frequency resonances. IEEE Trans Dielectr Electr Insul 14:1308–1315

Turnaround Maintenance in Process Industry: Challenges and Potential Solutions

Antti Rantala(✉), Helena Kortelainen, and Toni Ahonen

VTT Technical Research Center of Finland, 33101 Tampere, Finland
antti.rantala@vtt.fi

Abstract. Turnaround maintenances (TAMs) are huge projects in terms of manpower and expenditure and therefore they have a direct effect to company's profitability. TAMs include several challenges, such as prioritizing the maintenance tasks, scheduling the project, sharing information among all stakeholders on site and keeping focal company's maintenance data in the IT systems updated. Due to the significance of TAM in economic and safety perspective, solutions for the challenges are needed, and advanced technologies could play a major role in solving these challenges. For example, sensor technology and software could help in evaluating asset condition and prioritizing maintenance tasks. In addition, mobile technology and apps could enable smoother information sharing on site. Moreover, external expertise could be brought into the TAM project by utilizing virtual- and augmented reality.

Keywords: Turnaround maintenance · Advanced technologies · Process industry

1 Introduction

Turnaround maintenance (TAM) is periodic maintenance operation in which plant is shut down to allow for inspections, repairs, replacements and overhauls that can be carried out only when the assets are taken out of service [1]. This category of maintenance is referred to as shutdown, turnaround or outage maintenance [2]. During this period, three types of work are carried out: (1) work on equipment that cannot be performed unless the whole plant is shutdown, (2) work that requires a lengthy maintenance and a large number of maintenance personnel and (3) maintenance on defects that are discovered during the normal operation but cannot be repaired [2].

TAM plays an important role in maintaining consistent means of production delivered by reliable equipment in the process industry. The successful execution of TAM in terms of quality and cost is vital to the profitability of the company and to its competitive advantage [1]. The effective planning and management of TAM is regarded as an extremely critical business process [3]. Because TAM is a significant maintenance and engineering event, a direct connection can be drawn between its successful accomplishment and the profitability of the company [3].

Properly executed TAM minimizes the risk of production losses by ensuring reliable assets [1]. Reliability is essential for any production facility given that an equipment

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 196–207, 2022.
https://doi.org/10.1007/978-3-030-96794-9_18

failure may also result in severe safety consequences (e.g., injuries or loss of life and the company's reputation), financial damages (e.g., production losses and damages to assets) and environmental consequences [2].

Digital transformation in the form of industry 4.0 or industrial internet offers opportunities for asset intensive sectors such as process industry. Advanced technologies are supposed to increase the effectiveness and efficiency of asset management activities. For instance, cloud-based technologies enable on-demand network access to a shared information resources [4]. In addition, a variety of solutions for identification, positioning and tracking and laser scanning, and application of drones and AR/VR solutions have been tested in asset management activities [5] and such solutions could be beneficial also in the turnaround maintenance.

Learning and improvement needs more attention at the practical level, and exploring more effective feedback and learning methods at the research level. At the strategic and long-term planning level, the need for more integration between different functions of the organization utilizing new technologies and methods is quite clear [6]. When taking into consideration the significance of the turnaround maintenance from economic, safety and business continuity perspectives, the research in this field is surprisingly scarce. Previous research [6] highlights issues like optimizing TAM scheduling along the supply chain, developing methods for sharing knowledge, best practices and expertise within the supply chain and learning from previous TAMs as domains requiring future research. The purpose of this study is to contribute to the TAM related research by identifying the main challenges currently encountered in planning, executing and closing down TAM in process industry. The focus is on the management of the supply network and contractors, and information sharing on the operational level. In this paper, we discuss also potential solutions for the identified challenges.

2 Characteristics of TAM

TAMs are huge projects in terms of manpower and financial expenditure. The industrial processes that undergo TAM projects are often of high value and their maintenance operations are intensive, complex and costly [7]. TAM involves individuals from different departments in the own organization and many network partners, suppliers and contractors, and their subcontractors from different companies. Companies strive to complete TAM as quickly as possible to minimize the cost of lost production, resulting in long working weeks and constant overtime work [8].

A major source of complication in TAM planning and execution is the involvement of multiple internal and external stakeholders that significantly influence the success of TAM [6]. Accidents at process plants can occur not only while operating normally but also during shutdown periods. As the number of persons present during a shutdown may be much higher than during the normal plant operation, the consequences of any associated accident might also be more severe. For instance, Malmén et al. [9] raise up the need to address risks due to process chemicals present on site during plant shutdowns and to share this information with all stakeholders involved in TAM activities. Challenges in risk management when several organisations work together on a shared workplace have been highlighted by Heikkilä et al. [10].

TAM includes also specialized working tasks that differs from normal maintenance work [8]. Managing various simultaneously executed tasks and communicating the progress to stakeholders require extensive knowledge from the managers. Involving various stakeholders at an early stage of planning is needed for better coordination during execution, resulting in higher efficiency and utilization of resources [6]. In addition, a proper project management software is essential for TAM execution to help in prioritizing needs and identifying time and budget constraints for each planned activity [11]. IT-tools are needed for information transfer and monitoring the project as well. Delays in schedule and budget overruns are mainly attributed to miscommunication between different units [6].

TAM projects consists of several phases, which can be identified as initiation, preparation, execution and termination [12]. Planning is absolutely critical to the success of TAM [8] and it consists of various factors, such as purchasing spare parts, creating schedules and selecting contractors. TAM planners need to identify objectives and priorities, establish work timetables and organize resources to achieve their goals [13]. Managing TAM projects successfully requires a mixture of skills along with the capability to understand the situation and people and then dynamically integrate appropriate leadership behaviours [14]. Various management skills and knowledge areas are required at different stages of the TAM including planning, cost management, decision-making, quality management, risk management, time management, resource allocation, communication, negotiation and supervision skills [8].

Having timely and accurate feedback of learned lessons helps in continuous improvement in TAM performance. Performance measures play major role in operational improvement. Long-term planning for TAM that covers the time span of the plant is often neglected in planning for maintenance operations in manufacturing systems. The objective of long-term planning is to align maintenance planning with the organization strategic plan. A recent trend in the literature spots the importance of viewing the maintenance function as a strategic activity that extends the life of the asset and adds value to the production system in quality and quantity [6].

In general, the overall objective of TAM is to ensure high plant safety, reliability and production continuity on a long term, and to execute the TAM project within schedule and budget. In more detail, the objectives of TAM project could include e.g. [15]:

- bringing the plant to their original health;
- making the plant machines safe to operate;
- improvement on efficiency and throughput of plant by suitable modification;
- reduction of routine maintenance costs;
- increasing the reliability/availability of equipment during operation; and
- upgrading technology by introducing modern equipment and techniques.

The success of a TAM project is measured often by the time (keeping the schedule) and cost (keeping the budget). Other success elements [15] include safety, environmental impacts, quality and functionality. High quality manifests itself in smooth commissioning and start-up without a need for rework. Functionality means good operational performance and required output volume, and that quality requirements are met.

3 Research Context and Data Collection

In this study we aim at identifying challenges in the planning, executing and closing phases of TAM in process industry. The ongoing co-innovation project "Solid value from digitalisation in forest industry (SEED)" - project [16] offer a context for this research. As stated in the SEED project, the current means to share data and resources, and to support, organize and forecast the progress of the field activities e.g. during a maintenance turnaround are limited. Typically, ERP offers the basic processes for resource management but does not offer an interface to the company network operating on the site during the turnaround.

In this paper, we study the main challenges in TAM processes and propose potential solutions for them. The research applies qualitative research method and the data was collected by expert interviews. The interviews were extended beyond the SEED project to provide a broader insight into the practices of asset-intensive industries. Altogether ten maintenance experts from different companies were interviewed. Individuals with strong experience in maintenance development and especially in the planning of major repair outages or turnaround maintenances were selected for interview. Such interviewees were identified and contacted through the networks of the National Maintenance Association. The information on the interviewed persons is presented in Table 1.

Table 1. Interviewees' industries and roles

Company nr	Role of the interviewee	Industry
1	Senior Manager, Reliability	Steel
2	Maintenance Manager	
3	Maintenance Manager	Forest
4	Maintenance Support Manager	
5	Development Manager, Asset Management	
6	Maintenance Manager	Chemicals
7	Reliability Engineer, Asset Management	Petrochemicals
8	Manager, Production Asset Management	Energy
9	Maintenance Manager together with his experienced staff members	OEM with wide service offering
10	Manager, Maintenance Development	

The interviews were aimed at providing insights into main challenges and practices in the TAM life cycle. Semi-structured interviews were conducted to provide insights into the selected thematic areas, such as TAM planning, TAM execution, collecting feedback and measuring success. Examples of the interview questions included:

- What are the main challenges in turnaround maintenance planning?

- How do you prioritize the maintenance tasks for turnaround maintenance?
- How do you share information with stakeholders during the turnaround maintenance?
- What are the main challenges in executing and tracking turnaround maintenance?

Two of the interviews were held face to face and further interviews as were tele meetings via Microsoft Teams, because COVID-19 precluded face to face interviews. Two or three interviewers participated in the interviews and took notes. The interview data were coded using Nvivo, which is a qualitative data analysis software.

4 Findings from the Interviews

The interview notes were analysed by identifying the main challenges of TAM preparation, execution and termination phases. Also solutions and wishes expressed by the interviewees for improving TAM management were identified from the interview data.

4.1 Challenges in Preparation Phase

Before the TAM execution starts, all pre-work should be precisely done. Planning strives to ensure that the right job is done at the right time by the right people [3]. TAM is a huge event in which any delays and cost overruns may become extremely expensive. Evaluating the condition of each asset and selecting which assets should be repaired or replaced was regarded as the biggest challenge among the interviewees in the preparation phase. Lack of specific knowledge about asset condition creates challenges in selecting and prioritizing the maintenance tasks. This was expressed by an interviewee:

> *"Better knowledge of asset condition would ease the TAM planning, especially in machines, where you can't see the condition from outside."*

Reliable data about asset condition and maintenance history would help in prioritizing maintenance tasks. However, the data is usually stored in various systems that creates challenges in all stages of the TAM project. This challenge was common to all interviewees, described as frustrated as follows:

> *"We should utilize all maintenance data, but everything is in different systems. How could we get all data into a common system? Fragmented data is a big challenge."*

If the recorded maintenance data is insufficient or lacking, expert knowledge plays big role in prioritizing the maintenance tasks. Thus, there will be challenges in TAM planning if key personnel change:

> *"We don't have any tools or models for prioritizing (maintenance tasks). We rely on our experience – probably too much."*

Selecting the contractors may be challenging. One of the interviewees expressed the opinion that maintenance engineers may emphasize familiar and well-known contractors to secure the quality of maintenance work, whereas the purchasing department may prefer cheaper contractors, even though they may not possess the required expertise to execute the challenging maintenance tasks. As a solution, one interviewee proposed that criticality assessments should be done to the most important contractors.

Selecting the TAM starting date and duration can be difficult as well, because market situation is dynamic and companies want to meet the market demand. Best time for TAM would be when demand is low in order to minimize the cost of lost sales. From TAM planning point of view, unexpected schedule changes can compromise a carefully drafted plan. In addition, it is well known that extending the time between successive downtimes increases the risk of machine failure and high consequential costs.

Interviewees emphasized that the preparation phase is the most important phase in TAM projects. Starting the planning early enough was frequently mentioned as a key issue and success factor. In order to secure the success of TAM, work planners should have enough time to do all pre-work, including purchasing the maintenance services, components and machines. The TAM will be executed as well as it is planned. Proper planning reduces the risk of time and budget overruns and enhances outcome quality and operational safety [6].

Further challenges in TAM planning were scheduling in a way that minimizes the contractors' spare time and maximizes the project efficiency. Scheduling of the jobs allocated to contractors and network partners is not always easy and the following comment reflects the challenges ahead:

> *"Biggest target for development is to get the work started effectively. All third parties arrives at the same time – there is a lot of spare time for many workers. Would it be wise to stagger the schedule?"*

4.2 Challenges in TAM Execution

TAM execution is a complex project with a variety of challenges and it is not surprising that information sharing was the most frequently mentioned challenge in the interviews. Managing major turnaround requires continuous change management. There can be thousands of people on the site, various simultaneous working tasks and different kind of unexpected surprises may occur. Information sharing is important also from safety point of view:

> *"During a turnaround maintenance, huge amounts of information is shared and we can't control everything. There can always be surprises such as power outages. If information, for example about power outage, isn't shared, it could cause dangerous situations."*

Flexible and efficient collaboration among maintenance personnel, production personnel, network partners, suppliers and contractors on the site is considered to be of prime importance. The maintenance managers felt that the maintenance personnel carry the main responsibility on the prioritization, and seek for collaboration with production personnel and other stakeholders:

"Collaboration among maintenance personnel, production personnel and subcontractors is a challenge. Total productive maintenance –thinking is still in progress. There is still work to do for doing the right things in a real collaboration."

Quality monitoring of maintenance work was also considered as a big challenge. Poorly done maintenance work may become expensive, and some of the defects can be noticed only when the machines are turned on and poorly done maintenance work may lead to costly rework:

"Monitoring the quality of work is a challenge. Many things can't be measured during a TAM and possible defects can be noticed only when the machines are turned on."

During the TAM, real-time documentation regarding accomplished maintenance tasks is often not possible. The information about executed maintenance tasks is seldom entered into the CMMS or ERP in real time. Therefore, the systems needs to be manually updated after the TAM execution. The interviewees characterized the emerging hardships in the following ways:

"(Big challenge is) real-time documentation. Information is shared via phone calls, WhatsApp, meetings etc. All information is scattered. Utilizing mobile apps could be a possible solution (in information sharing problems)."

"We have a lot of work that is organized by paper notes. We don't have a system for information transfer. There would be a lot of need for that."

4.3 Challenges in Closing Down the TAM Project

Last phase of TAM is named as termination [12] but in a real life, a TAM is a long project which continues after the actual execution even the production is started up. As our interviewees emphasised that the TAM related tasks are carried out weeks or even months we have chosen to call the last TAM life cycle phase "closing down" rather than termination. Updating the CMMS/ERP with the data gathered during the TAM requires a lot of work and may take several months. In addition to entering all the event data generated by the accomplished work to the systems, tracking costs takes time. Some contractors may send invoices even months after the project has ended, which complicates the follow up and reporting. One interviewee described a somewhat surprising problem this way:

"Big challenge is to track the costs. Sometimes it can take even three months to get the invoices from the subcontractors."

Receiving feedback from each stakeholder is an essential part of TAM and supports the objectives to improve the TAM process. However, the challenge is implement the feedback into the next TAM. Feedback from the supply network is also appreciated:

"It is important to get feedback from subcontractors and learn from it. Subcontractors have very good experience and they see constantly all kinds of turnaround

maintenances. Best practices from subcontractors should be implemented in our own actions. In addition, their commitment increase when they are heard"

4.4 Potential Solutions to TAM Challenges

Selecting the TAM starting date and duration was one of many challenges in preparation phase. If two big companies have TAMs at same time, it increases the demand for certain resources not only locally but also on the national level, and it could be difficult to get all the needed contractors on the site. Therefore, overlapping turnaround dates should be avoided. Interviewees wished for a *national TAM calendar* in Finland in order to check which big companies could have TAM at the same time.

In the planning phase, the biggest challenge was evaluating asset condition. Interviewees also wished for better predictability and knowledge about remaining asset life or the expected lifespan for each asset that would facilitate the asset life cycle replacement planning. Sensors and analytics software could be utilized to get better understanding of asset condition. This would help to optimise TAM intervals and to avoid starting maintenance work too early or too late. This wish was expressed by one interviewee:

"Condition based maintenance work should be done more. This is related to integrating and utilizing data. Everything should be recorded in maintenance systems and reporting should be more visual."

Introduction of mobile technologies and apps were regarded as solutions for keeping the IT-systems updated help to monitor the progress of the maintenance works. The suppliers and contractors (and subcontractors) would also have a role in entering maintenance data into the systems, but the issue divides opinions as one interviewee expressed:

"Our goal is that third parties would provide data into our systems, but there is still long way to go"

In addition to the mobile technologies, our interviewees told about COVID 19 induced and successful experiences on using AR (Augmented reality) technology in supervising sensor system installation. The technology provider guided the field technician with AR glasses, instead of traveling from another country to the factory site. Such technologies could be utilized also in some turnaround maintenance tasks to avoid unnecessary traveling and to save costs.

Several interviewees emphasized that it would be important to form a situational picture about the TAM progress. Such a situational picture would ease the information transfer between various stakeholders both during the planning period and during the execution phase. Digital tools could make information transfer smoother and more systematic during TAM. Turnaround dashboard offering a real-time situational picture on the TAM progress was characterised in the following way:

"In the execution phase, it would be important that some kind of dashboard would show the situation picture about the progress. In that case, if some surprises occur, it would be possible to react quickly."

"Real-time turnaround maintenance dashboard would be good, especially if it would understand linkages between tasks. For example, if some work is two hours late, it would shift also other works that are related to the belated work."

5 Summary of the Findings and Discussion

Interviewees expressed various challenges in all life stages of TAM and proposed several solutions for solving their issues. Main challenges and proposed solutions for each TAM phases are summarized in Table 2.

Table 2. Main challenges in current TAM projects and proposed solutions

TAM life cycle stage	Challenge	Solution(s)
Preparation and planning for TAM	- Evaluating asset condition	- Utilizing sensors and software
	- Prioritizing maintenance tasks	- Deeper collaboration between different departments
	- Fragmented data	- Common database for asset data
	- Starting the planning early enough	
	- Changes in key personnel	- Relying more on maintenance data instead of human experience
	- Selecting the starting date and duration	- National TAM calendar for levelling the use of resources - Utilizing data to evaluate maintenance needs
	- Selecting the contractors	- Making criticality assessments for contractors
	- Scheduling the work order	
Execution	- Information transfer	- Utilizing mobile tools on site - Constructing a TAM dashboard
	- Real-time documentation into systems	- Utilizing mobile tools in documentation - Providing third parties access into focal company's IT-systems

(*continued*)

Table 2. (*continued*)

TAM life cycle stage	Challenge	Solution(s)
	- Work quality monitoring	
	- Poor collaboration among different stakeholders	
Closing down the project	- Updating all maintenance data manually into IT-systems	- Enabling third parties to input data into focal company's IT-systems
	- Cost tracking	
	- Learning from feedback	

Interviewees expressed various challenges related to preparation of TAM. Some of the challenges were mentioned also in the previous literature, such as optimizing the scheduling with supply network [6]. Interviewees also mentioned challenges that weren't discussed in the previous TAM literature, such as possible changes in key personnel. One interviewee commented that if maintenance systems would support TAM planning by illustrating different assets' maintenance needs, changes in key personnel wouldn't have so serious consequences.

According to the interviews, the main challenge in execution phase is information transfer on site. It is important to find solutions to this issue, because delays in schedule and budget overruns are mainly attributed to miscommunication between different units [6]. In addition, Al-Turki et al. [6] emphasized that methods for sharing knowledge and expertise within the stakeholders should be developed. As a solution for these information transfer challenges, interviewees proposed that mobile technology could be used for sharing information among the stakeholders. Other idea from interviews was that forming a dashboard of the TAM progress would significantly ease TAM management. In addition, of the interviewees told that AR-technology was successfully utilized in their factory in a sensor system installation. Because TAM includes also specialized working tasks that differs from normal maintenance work [8], we propose that AR-technology could be a solution to obtain external expertise into TAM works.

Interviews also expressed challenges regarding closing down the TAM project. Entering contractors' maintenance data into the focal company's IT-systems after the execution of TAM can take even several months. If third parties could enter their maintenance data straight into focal company's IT-systems, focal company wouldn't need to update the data into systems afterwards.

6 Conclusions and Future Work

The vision of the SEED project states that digitalization offers means for major improvement of the forest industry in the productivity and business renewal. TAM management could profit from cloud solutions, digital planning tools and other novel technologies that have good prospects to solve identified challenges in the TAM life cycle and offer means for efficient information sharing. The value of the shared data is evident.

The SEED research approach is challenge driven and applies scientific methods to find solutions to the practical problems [16]. The findings from the conducted interviews in this research study will be further elaborated towards an industrial use case together with an interested forest company (use case owner). The use case description is a foundation for cooperation between companies and a starting point for one or more solutions that would solve identified problems. In this use case, co-operation between the use case owner, OEMs (Original Equipment Manufacturer) and IT partners is needed. The companies are encouraged to demonstrate the developed solution or solutions in a form of a proof-of-concept.

Further research work is needed in the areas of prioritizing maintenance tasks and asset replacement decisions. Furthermore, research is still needed for developing methods to improve the predictability of the accumulating costs and to create models to increase information sharing throughout the TAM life cycle and among the TAM-project stakeholders. For instance, a research gap can be identified in use case specific data refinement and exchange of the resulted information in a multi-vendor maintenance management context.

Advanced digital solutions could play an essential role in solving the challenges presented in current paper. For example, sensor technology and software could help in evaluating asset condition and its remaining lifetime. Also, these solutions should enable third parties to mark their works done straight to maintenance integrator's or asset owner's IT-systems so they could keep all information updated. In addition, by utilizing virtual- and augmented reality technologies some of the management and guidance could be done remotely. Learnings from the COVID-19 induced experiments could boost this development. Thereby, digital solutions could enable more comprehensive collaboration among the network and contractors on site.

Acknowledgements. The research group have received funding from Business Finland that is the Finnish government organization for innovation funding and trade, travel and investment promotion. VTT Technical Research Center of Finland and the companies participating in the SEED project have been co-funding the study. The authors wish to thank all these organisations for their support.

References

1. Duffuaa S, Ben DM (2004) Turnaround maintenance in petrochemical industry: practices and suggested improvements. J Qual Maint Eng 10(3):184–190
2. Hameed A, Khan H (2013) A framework to estimate the risk-based shutdown interval for a processing plant. J Loss Prev Process Ind 32:18–29
3. Lenahan T (2005) Turnaround, shutdown and outage management: effective planning and step-by-step execution of planned maintenance operations. Elsevier Science & Technology
4. Holgado M, Macchi M, Fumagalli L (2016) Value-in-use of e-maintenance in service provision: survey analysis and future research agenda. IFAC-PapersOnLine 49(28):138–143. https://doi.org/10.1016/j.ifacol.2016.11.024
5. Kortelainen H, Hanski J, Valkokari P (2020) Advanced technologies for effective asset management - two cases in capital intensive branches. IFAC-PapersOnLine 53(3):7–12. https://doi.org/10.1016/j.ifacol.2020.11.002

6. Al-Turki U, Duffuaa S, Bendaya M (2018) Trends in turnaround maintenance planning: literature review. J Qual Maint Eng 25(2):253–271. https://doi.org/10.1108/JQME-10-2017-0074
7. Al-Turki U, Duffuaa S, Bendaya M (2013) A holistic system approach for turnaround performance management. In: Conference of maintenance performance measurement and management, Lappeenranta. https://doi.org/10.13140/2.1.2805.0886
8. Obiajunwa C (2013) Skills for the management of turnaroud maintenance projects. J Qual Maint Eng 19(1):61–73
9. Malmén Y, Nissilä M, Virolainen K, Repola P (2010) Process chemicals: an ever present concern during plant shutdowns. J Loss Prev Process Ind 23(2):249–252. https://doi.org/10.1016/j.jlp.2009.10.002
10. Heikkilä AM, Malmén Y, Nissilä M, Kortelainen H (2010) Challenges in risk management in multi-company industrial parks. Saf Sci 48(4):430–435. https://doi.org/10.1016/j.ssci.2009.12.008
11. Brown MV (2004) Audel managing maintenance planning and scheduling. Wiley, Lavoisier
12. Duffuaa SO, Ben-Daya M (2009) Turnaround maintenance. In: Ben-Daya M, Duffuaa SO, Raouf A, Knezevic J, Ait-Kadi D (eds) Handbook of maintenance management and engineering. Springer, London, pp 223–235. https://doi.org/10.1007/978-1-84882-472-0_10
13. Pettersen N (1991) Selecting project managers: an integrated list of predictors. Proj Manag J 22(2):21–26
14. Pant I, Baroudi B (2007) Project management education: the human skill imperative. Int J Proj Manage 26(2):214–218. https://doi.org/10.1016/j.ijproman.2007.05.010
15. Obiajunwa C (2012) A framework for the evaluation of turnaround maintenance projects. J Qual Maint Eng 18(4):368–383
16. Valkokari K, Valkokari P, Kortelainen H, Nyblom J (2020) Building business impacts of an industry 4.0 ecosystem through collaborative network settings between IT and forest companies. In: Camarinha-Matos LM, Afsarmanesh H, Ortiz A (eds) PRO-VE 2020, vol 598. IAICT. Springer, Cham, pp 463–474. https://doi.org/10.1007/978-3-030-62412-5_38

Modelling the Effect of Maintenance-Induced Failures from Periodic Testing of Safety-Critical Equipment as Part of RCM Analysis in the Oil and Gas Industry

Jon Tømmerås Selvik[1,2](✉) and Hans Petter Lohne[2]

[1] University of Stavanger, Stavanger, Norway
jon.t.selvik@uis.no
[2] NORCE Norwegian Research Centre, Stavanger, Norway

Abstract. Determining appropriate maintenance programmes for technical inventory is recognized as important for quality reliability and safety management in the oil and gas industry. The programme could be achieved through reliability-centred maintenance (RCM) analysis, where safety-critical equipment with potential for hidden failures is given particular attention. Output of the analysis is seen in combination with relevant requirements to perform functional testing of the equipment. The testing involves collecting and analysing data for verification of acceptable reliability and safety levels during the operational phase. This testing is often required in periodic intervals, where shorter intervals might be required initially or after failures for more control. Despite the intention of such activity, it could however influence equipment conditions in a negative way and over time contribute to a reduced reliability performance, i.e., lead to maintenance-induced failures. In this paper, focus is on periodic testing of the component 'downhole safety valve' (DHSV), and mechanisms leading to its failure. We consider the use of an age-adjusting imperfect repair model for analysing the effect of maintenance-induced DHSV failures and discuss the influence of recommended industry guidance. We particularly discuss the benefits of a test strategy having initially one to three months intervals, compared with an alternative strategy with constant six-month or one-year intervals. Based on the analysis, the 12-month interval gives the highest overall probability of failure on demand despite reducing the probability for maintenance-induced failures. There is a marginal difference between the other two alternatives, where then the selected distributions and uncertainties play a larger role. Barrier data collected by the Petroleum Safety Authority Norway (RNNP project data) is used for the analysis.

1 Introduction

Reliability centred-maintenance (RCM) is a widely used technique within reliability engineering practise, with over 40 years of history of successful application in various industries [1], including the oil and gas industry. It provides a way of selecting the appropriate maintenance policy and assessment of periodicity (scheduling), where high

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 208–218, 2022.
https://doi.org/10.1007/978-3-030-96794-9_19

attention is given to safety critical equipment, especially passive equipment with 'hidden failure' potential such as blow-out preventers (BOPs) and downhole safety valves (DHSVs). This paper will focus mainly on the latter.

RCM is a way to identify critical items and to develop appropriate maintenance programs maintaining inherent reliability [2]. Assessment of maintenance intervals (maintenance optimization) being a main step. A description of the steps is given in Rausand et al. [3, p. 392]. The assessment covers attributes such as requirements, consequences, cost, and the probability of failure on demand (PFD).

Functional testing is normally the default for passive safety critical equipment unless there is failure-alerting condition-monitoring. When considering DHSV reliability as part of the RCM analysis, for operations on the Norwegian Continental Shelf, such valves are normally subject to a six-month functional test interval as recommended in NORSOK D-010 [4]. The first year being a bit special, consisting of three one-month and then three three-month intervals, afterwards resuming six-month intervals unless some functional failure occurs. The interval could be adjusted depending on the reliability demonstrated, but also based on planning of other maintenance activity, i.e., grouping of maintenance tasks. Testing the functionality (proof-test) makes it possible to detect DHSV failures before a demand and a potentially dangerous situation occurs. A shorter time between the tests influences the expected time between failure and detection. The idea being that shorter intervals is in favour of reliability and safety. Cost aspects might pull in the opposite direction. One could also have maintenance-induced failures, challenging the actual reliability and safety benefits of frequent testing.

When optimising maintenance, the potential for maintenance-inducing failures could be included as part of an imperfect repair model, although the influence is often ignored in practice [5]. According to Dekker [6] a main reason for why, is the lack of adequate tools and methods for identification of such failures, besides the lack of good data. The lack of sufficient degradation experience leads to extensive use of the exponential failure distribution when estimating the DHSV mean time to failure and PFD. However, for DHSVs such a distribution might not be realistic as the industry recommended test schedule indicate a higher failure rate during the first year of operation. A main objective of this paper is to consider the use of an age-based model, comparing the exponential versus the Weibull distribution, as well as different test strategies for the DHSV maintenance optimisation part of RCM. A Weibull distribution might be more realistic and is often pointed to in qualification, but also way more complex to integrate analytical-wise. This as input for development of an imperfect repair model for scheduling of DHSV proof-tests, where the effect of maintenance-induced failures may be studied.

2 Imperfect Repair Modelling

A typical objective or objective function when optimizing maintenance, is cost; the optimization criteria typically expressed as expected cost over a time period. Reliability might also serve as a criterion, e.g., the number of failures or the PFD. When the prime focus is safety, a key is to ensure acceptable safety integrity performance. In general, the criteria can be expressed through time-based models, which are widely applied in the context of RCM; such as P-F interval models, block-replacement, and age-replacement models (see model descriptions in e.g. Lindqvist [7]; Moubray [1] and, Rausand et al. [3, p. 549]). The influence of maintenance-induced failures for passive safety equipment is discussed in e.g., Hafver et al. [8], where two BOP test schemes are compared, i.e. a scheme with constant testing intervals and a scheme with adaptive scheduling. Adaptive scheduling means that time to next test is adjusted to compensate for changes in failure rate or failure frequency during the operational cycle to maintain the reliability. Such flexibility allows for more efficient test scheduling.

In modelling of repairable systems, such as DHSVs, the impact of imperfect repair is typically reflected by adjustment of the failure frequency *w(t)* as the starting point following (immediately after) a repair action. The basis is then that the failure frequency immediately after the repair should reflect the quality of the maintenance. The *w(t)* is sometimes referred to a *rate of occurrence of failure* (ROCOF).

One of the simplest ways is to assign a repair efficiency function (or index) for the maintenance quality, i.e., the probability of maximum condition improvement, and link this to the *w(t)* or the PFD. This could be modelled binary by assigning ρ(t) as the probability of perfect repair (as-good-as new), with 1−ρ(t) as the probability of the item condition being as immediately before the failure occurs (as-bad-as old). Such a concept could be applied for any maintenance policy. The concept could also be extended to multistate as suggested in e.g., Doyen and Gaudoin [9]. It represents a type of imperfect repair modelling based on *w(t)* adjustment, where each maintenance event could lead to either perfect repair (as the best) or minimal repair (as the worst), or something in between. However, the steepness of the failure rate curve at time *t*, the *w'(t)*, is not influenced by the adjustments. A different way to adjust, is to make time the basis for *w(t)* reduction. This could change the steepness of *w(t)*, where the reduction factor or function moves the failure rate (virtually) back in time (age reduction). However, such an adjustment will not have much effect under assumption of an exponential failure distribution, where the *w(t)* is constant. The reduction factor then gives the percentage of time reduction based on the time elapsed since the previous repair or total elapsed time. The adjustment is then limited to the development recorded for the earlier period.

Assuming that failures only are revealed from proof-testing at intervals with length τ and as-good-as-new after repair, the average PFD can be calculated from:

$$PFD_{AVG} = \frac{1}{\tau}\int_0^{\tau} F(t)dt \tag{1}$$

Where $F(t)$ is the PFD at time t; $P(T < t)$. For each cycle, and immediately after the repair, the item is virtually moved back to $t = 0$. However, if the repair is imperfect, unless the failure rate is decreasing, the $\mathrm{PFD_{AVG}}$ will be increasing for each test cycle of length τ. $F(t)$ can be calculated with reference to the failure rate $z(t)$:

$$\mathrm{F(t)} = \mathrm{PFD} = 1 - \exp\left[-\int_0^{t} z(t)dt\right] \tag{2}$$

In (2) if $z(t)$ is a constant $z(t) = \lambda$ the integral in (2) equals $\lambda \cdot t$, giving the exponential distribution. F(t) for the Weibull distribution is presented in Sect. 4. Following each cycle of tests, the underlying failure density distribution and F(t) may then change depending on the repair quality. Let $F_i(t)$ denote the PFD in the interval from immediately after test number $i-1$ to test i. For constant test intervals with length τ: $F_1(\tau) \rightarrow F_2(2\tau) \rightarrow F_3(3\tau) \rightarrow \ldots \rightarrow F_{n\text{-}1}[(n-1)\tau] \rightarrow F_n(n\tau)$.

A change in F_i would then reflects the change in *z(t)* induced by the maintenance. In addition, the maintenance effect could be modelled by adjusting the virtual start time for cycle *i*. Instead of start at $t = 0$, it could start at a point in time reflecting the current equipment condition at start-up to match the PFD. Such a modelling may allow for adjustment of the initial cycle level and for the development up to next test. The PFD_{AVG} for cycle *n* specifically could be expressed as:

$$PFD_{AVG}(\mathrm{n}) = \frac{1}{\tau}\int_{S(n)}^{S(n)+\tau} F_n(t)dt; \quad \text{for } n > 0 \tag{3}$$

where S(n) is a function assigning virtual start-up time of cycle *n*. This is used as basis for modelling the effects of maintenance-induced DHSV failures in Sect. 4. For more details and presentation of different imperfect repair models, we refer to reviews of imperfect repair models given in e.g., Wang and Pham [10, p. 13], Pham and Wang [11]. See also Rausand et al. [3, p. 455] and Nakagawa [12, p. 171; 13].

3 The Downhole Safety Valve Situation at NCS

DHSV is a main barrier element in offshore wells and plays a key role for safety management for oil and gas facilities. Each year the Petroleum Safety Authorities Norway (PSA) publish a risk level report including a DHSV reliability status. Figure 1 gives an overview of the development in fraction of failed tests from 2002 to 2019 based on data reported from the oil and gas companies to the PSA (see [14]).

In total over the period 89,514 tests are registered, with 2,582 failures; see also Table 1. As indicated by the figure, there is a significant increase in fraction of failures per test, and data show that 35 out of 80 facilities have a fraction above the critical level of 0.02 in 2019. For the full period almost half of the facilities are above this critical level. Assuming a six-month test interval this gives a constant failure rate $\lambda = 6.68 \cdot 10^{-6}$/h, meaning an expected time to failure of 17.3 years. Due to confidentially issues, field-specific data are not presented.

Fig. 1. Fraction of DHSV failures on number of tests

According to the PSA reporting, previous analysis shows that facilities with more than 20 years of operation, is more prone to DHSV failures compared to younger facilities (for the period 2008 to 2017). The analysis also shows that facilities with 6 to 20 years of operation have a significantly lower fraction of failures; supporting a lower failure frequency in the middle of the lifetime.

Table 1. Number of DHSV tests with failures at the Norwegian Continental Shelf

Year	Tests	Failures	Fraction of failed
2002	3 851	31	0.00805
2003	3 098	46	0.01485
2004	3 566	67	0.01879
2005	3 322	80	0.02408
2006	4 787	95	0.01985
2007	5 290	153	0.02892
2008	5 863	130	0.02217
2009	4 993	156	0.03124
2010	4 993	135	0.02704
2011	5 227	149	0.02851
2012	5 624	135	0.02400
2013	5 772	149	0.02581
2014	4 592	169	0.03680
2015	5 016	168	0.03349
2016	5 786	200	0.03457
2017	6 051	252	0.04165
2018	6 032	243	0.04029
2019	5 651	224	0.03964

A main challenge for the failed valves is the failure mode 'leakage in closed position'. A presentation by Molnes [15] shows that this specific failure mode accounts for around 35% of historical failures, being the dominating one. It represents a type of failure that could be traced to the maintenance activity and number of tests performed. The challenge being primarily the robustness of the seal. Number of tests may influence the performance, but also the time between tests if there are long dormant periods (as for having one-year test intervals) with fluids or sand particles eroding the seal, inducing 'sticking' and insufficient closing ability. Several publications, e.g., Vick et al. [16] and Vinzant et al. [17], focus on the sealing technology as a performance limiting factor. We refer to Selvik and Abrahamsen [18] for DHSV reliability review.

4 Modelling the Effect of Maintenance-Induced DHSV Failures

The modelling presented in this section is based on (3), where The PFD_{AVG} for each cycle is interpreted as the mean value for the specific period. The overall PFD_{AVG} can then be calculated from the arithmetic mean over the n periods with length τ:

$$Overall\ PFD_{AVG}(n) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \frac{1}{\tau} \int_{S(i)}^{S(i)+\tau} F_i(t)dt \tag{4}$$

For a test cycle i, the function S(i) is here seen as a function assigning a value in [0, ∞) based on the stresses accumulated by the i tests performed, assuming no use of the valves except the test demands, with $S(i-1) \leq S(i)$; $0 < i \leq n$. S(i) comprise as such the tests' probability of reducing DHSV condition or functionality; with 0 as 'as-good-as-new', τ as 'as-bad-as old', and $t > \tau$ as a decrease in condition and at worst a complete loss of function. It depends on the number of tests, quality of the tests, and the condition at the test. One way is to specify periods of equal length from τ. We select here 45 periods as the number of cycles m assumed before the valve is in a 'as-bad-as-old' condition after testing, giving (in hours) when τ is six months:

$$\mathrm{S(i)} = \frac{\tau}{m}(i-1) = 96.0(i-1) \tag{5}$$

We also assume the F(t) to be independent on the cycle number, i.e., $F_0(t)$ equals $F_i(t)$. As input, we refer to the failure data presented in Sect. 4 and will use these as the field-specific results. Regarding the F(t), we consider two common distributions: the exponential distribution with parameter $\lambda = 6.68 \cdot 10^{-6}[h^{-1}]$; (giving a mean time to failure $MTTF = 17.3$ years), and, the Weibull distribution with rate parameter λ (and scale parameter 1/λ), and shape parameter $k = 1.3$; and with $F(t) = 1 - \exp[-(t \cdot \lambda)^k]$ ($MTTF = 15.8$ years). The Weibull k parameter is derived from available qualification testing results. The comparison is then achieved by considering three distinct testing strategies: the one with one-year shorter initial tests, a constant six-month strategy, and a maximum 12-month strategy.

For the overall PFD_{AVG} formula (4) under assumption of a constant failure rate λ, and fixed test intervals $\tau_i = \tau$, the formula can be expressed as:

$$Overall\ PFD_{AVG}(n) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \frac{1}{\tau_i} \int_{S(i)}^{S(i)+\tau_i} \left(1 - e^{-\lambda \cdot t}\right)dt \tag{6}$$

$$= \frac{1}{n\tau}\sum\nolimits_{i=1}^{n} \left[t + \frac{1}{\lambda}e^{-\lambda t}\right]_{S(i)}^{S(i)+\tau} = 1 + \frac{1}{n\tau}\sum\nolimits_{i=1}^{n} \frac{1}{\lambda}\left(e^{-\lambda \cdot (S(i)+\tau)} - e^{-\lambda \cdot S(i)}\right) \tag{7}$$

The overall $PFD_{AVG}(n)$ when applying the Weibull distribution can then be expressed as:

$$Overall\ PFD_{AVG}(n) = \frac{1}{n}\sum\nolimits_{i=1}^{n} \frac{1}{\tau_i} \int_{S(i)}^{S(i)+\tau_i} \left(1 - e^{-(\lambda_i\, t)^{k_i}}\right)dt \tag{8}$$

In practical applications there will be initially higher failure probability. However, this is not included here. The effect would be a higher probability in the first year, although it is uncertain how large this effect would be.

When assuming an exponential distribution the probability of failure at each test is shown in Fig. 2. Longer test intervals have a higher probability of failure from the beginning, but the difference diminishes over the cycles due to maintenance induced wear. Due to the maintenance induced failures, the PFD at the start of a cycle will be larger for shorter test intervals. The PFD_{AVG} for a short test interval policy will therefore become greater than the PFD_{AVG} for a longer interval policy earlier than Fig. 2 might indicate. The 6-month intervals policy will have a higher PFD_{AVG} than the 12-month interval policy after 21.8 years, and with more frequent initial testing already after 17.8 years.

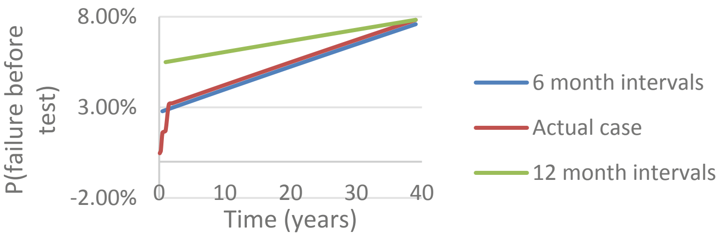


Fig. 2. Probability of the DHSV being in a failed state for a test with exponential distribution, conditioned on that it was functioning at the start of the interval, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

When considering the Weibull distribution, there can be seen an effect of wear in Fig. 3. The more frequent initial test intervals demonstrate a lower probability of being in a failed state for a test during the early years, at the cost of slightly higher probability due to test-induced wear in the later years. There is also a slight curvature to the lines in the plot caused by the increase in failure rate due to the time-adjusted aging. For the Weibull situation, the 6-month intervals policy will have a higher PFD_{AVG} than the 12-month interval policy after 23.0 years, and with more frequent initial testing after 19.2 years; 1.4 years later than for the exponential.

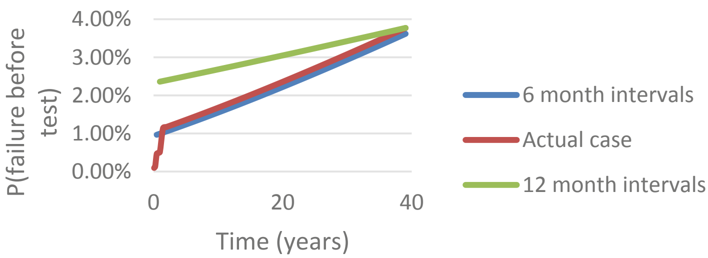


Fig. 3. Probability of the DHSV being in a failed state for a test with Weibull distribution, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

Alternative modelling-approaches are shown below. Figure 4 shows a Weibull distribution where the scale parameter is reduced by a fixed amount per interval. Even though the four additional tests performed in the first year only changes the probability by a small amount initially, the effect grows larger later in life. If the scale is reduced by a percentwise reduction per test a similar difference increase is seen, as shown in Fig. 5. Finally, a scenario where the maintenance activity does not modify the distribution, but rather an independent increasing probability of failure occurring due to the maintenance itself that is then added to the failures that can occur within the interval is shown in Fig. 6. In this case each maintenance activity causes a fixed increase that remains constant over time.

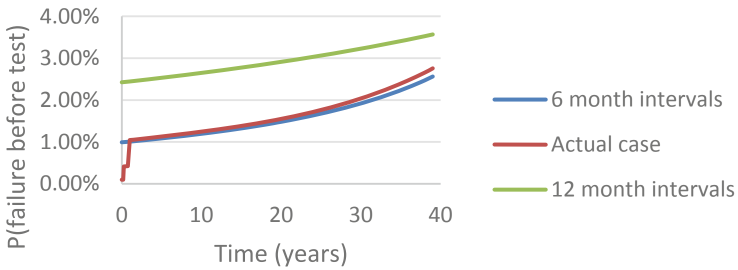


Fig. 4. Probability of DHSV being in a failed state for a test using Weibull distribution, where the scale parameter is reduced linearly by each test, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

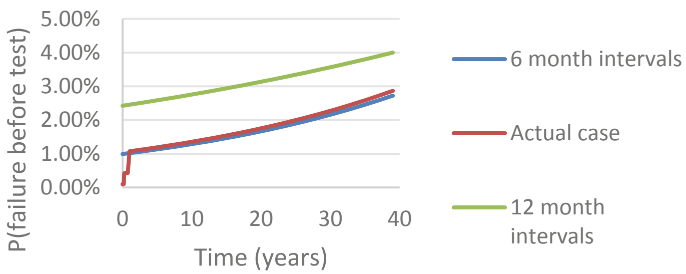


Fig. 5. Probability of DHSV being in a failed state for a test using Weibull distribution, where the scale parameter is reduced by a percent for each test, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

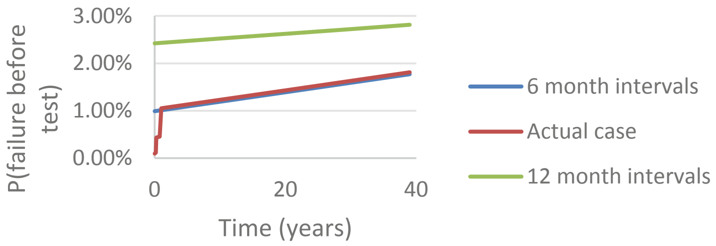


Fig. 6. Probability of DHSV being in a failed state for a test using Weibull distribution, where the maintenance activity can cause failures independently of wear state, when the test is performed either at 6 or 12-month intervals or 6-month intervals with more frequent tests initially.

Overall PFD_{AVG} for the different cases are shown in Table 2 below.

Table 2. PFD_{AVG} for different maintenance schemes and different maintenance-induced failure models.

Test intervals	Weibull			Exponential		
	6 months	Initially more frequent	12 months	6 months	Initially more frequent	12 months
Time adjustment	1.54%	1.59%	1.64%	3.85%	3.94%	4.04%
Linear scale reduction	0.72%	0.75%	1.36%	2.01%	2.07%	3.31%
Percentage scale reduction	0.79%	0.83%	1.46%	2.16%	2.22%	3.50%
Independent failure increase	0.85%	0.88%	1.31%	4.24%	4.51%	3.06%

Table 2 presents then the average probability of failure over the 39 years calculated for the different models. The maintenance scheme with 6-month intervals results gives the lowest values in most cases, although for failure independent of running time longer intervals are better with the exponential distribution.

5 Discussion

5.1 The Effect of Maintenance-Induced DHSV Failures

A key is the difference in effect between the test policies. As seen in the above calculations, frequent testing in the initial phase will reduce the maximum probability reached of being in a failed state at a test, at the cost of slightly higher probabilities in later years.

From the overall average probability of failure on demand, it is seen that 12-month intervals are worse off in most cases. However, this depends on the severity of damage caused by the testing. It is nevertheless obvious that maintenance is a trade-off between reducing the highest probability of being in a failed state and the overall probability of being in a failed state (such as during the initial phase). This depends on the failure distribution as well.

How maintenance-induced failures occur matters. In most of the examples where the increased testing shifts towards a state of increased wear, either through time-adjustment or through parameter adjustment, the changes seen are not drastic but will accumulate to a greater overall probability of failure over the lifetime. In the examples where failures are actually induced by maintenance, rather than just causing an increase in failure rate during the next cycle, the failure probability will here shift the curves upwards, which illustrates how frequent initial testing can deteriorate the valves quickly. Similar behaviour is difficult to achieve with models that simply "age" the valves when wear-out failures occur relatively late in the example distribution used. Distributions with a finite support (i.e., components with limited lifespan) would be able to develop much stronger differences.

5.2 RCM Value

The modelling of maintenance-induced failures is important to understand if frequent early initial testing is worth the wear to reduce the probability of being in a failed state within a test interval. If there is a fixed cost (in terms of failures) to wear, such as if there is a constant probability of failure due to the maintenance, then the frequency of maintenance can be done by focusing on keeping the maximum probability of failure in an interval as low as possible without considering additional long-term effects. In other cases, the long-term effect of maintenance should be considered, as the total probability of failure during the time the valve is installed can be larger.

When discussing maintenance, other aspects can also be considered, such as the associated costs as well as the needs. If during the early stages there is greater uncertainty in subsurface conditions there is also a greater need for a functioning DHSV, while if the uncertainty is low and therefore low probability of a demand for the DHSV the condition of the valve is less precarious. In such situations one could discuss having more frequent tests when a demand is more likely, and rather increase the test intervals when a demand is less likely to avoid excessive wear.

While several deterministic models are shown here to illustrate some different alternatives to modelling maintenance induced failures, when making decision related to maintenance based on reliability modelling the uncertainty of these models and the underlying failure causes they attempt to represent must be considered.

6 Conclusions

In this paper, DHSV failures and imperfect repair modelling has been presented, including a dataset from the NCS. A Weibull distribution and exponential distribution based on this dataset was then used to illustrate the effect of different models for maintenance-induced DHSV failures. The results illustrate the cost frequent testing can have on later reliability. To make correct decisions regarding testing, it is important that also the uncertainty around maintenance-induced failures, wear during operation and their interactions are included.

References

1. Moubray J (2012) Reliability centered maintenance, 3rd edn. Industrial Press, Incorporated, New York
2. Selvik JT, Aven T (2011) A framework for reliability and risk centered maintenance. Reliab Eng Syst Saf 96(2):324–331
3. Rausand M, Barros A, Høyland A (2021) System reliability theory – models, statistical methods, and applications, 3rd edn. Wiley, Hoboken
4. NORSOK D-010 (2021) Well integrity in drilling and well operations, 5th edn. Standards Norway, Oslo
5. Jackson C, Mailler B (2013) Post-servicing failure rates: optimizing preventive maintenance interval and quantifying maintenance induced failure in repairable systems. In: Proceedings annual reliability and maintainability symposium (RAMS), Orlando, FL, USA
6. Dekker R (1996) Applications of maintenance optimization: a review and analysis. Reliab Eng Syst Saf 51:229–240
7. Lindqvist B (2006) On the statistical modelling and analysis of repairable systems. Stat Sci 21(4):532–551
8. Hafver A, Oliveira LF, Pedersen FB (2019) Optimal scheduling of tests of safety systems, considering test-induced degradation. In: Proceedings of the 29th European safety and reliability conference, Hannover, Germany
9. Doyen L, Gaudoin O (2011) Modeling and assessment of aging and efficiency of corrective and planned preventive maintenance. IEEE Trans Reliab 60(4):759–769
10. Wang H, Pham H (2006) Reliability and optimal maintenance. Springer, London. https://doi.org/10.1007/b138077
11. Pham H, Wang H (1996) Imperfect maintenance. Eur J Oper Res 94(3):425–438
12. Nakagawa T (2005) Maintenance theory of reliability. Springer, London. https://doi.org/10.1007/1-84628-221-7
13. Nakagawa T (2008) Replacement and preventive maintenance models. In: Misra KB (ed.) Handbook of performability engineering. Springer, London. https://doi.org/10.1007/978-1-84800-131-2_49
14. PSA (2020) Summary report – Norwegian Continental Shelf 2019 Trends in risk level in the Norwegian Petroleum industry. Petroleum Safety Authority Norway (PSA). https://www.ptil.no/en/technical-competence/rnnp/rnnp-2019/summary-report/
15. Molnes E (2012) Leveraging data for high value decisions on well OpEx optimization. Presentation at the US-Norway technology partnership conference, Houston, Texas, USA, 29 March
16. Vick J, et al (2005) Unique safety valve design using a floating magnetic coupler and metal-to-metal sealing capability to provide step-change improvements in reliability, safety, and cost. OTC paper 17455. http://www.onepetro.org
17. Vinzant M, et al (2004) Enhanced reliability with unique tubing-retrievable safety valve design. SPE paper 88627. http://www.onepetro.org
18. Selvik JT, Abrahamsen EB (2013) How to describe well barrier reliability: a SCSSV review study. In: COMADEM 2013: proceedings of the 26th international congress on condition monitoring and diagnostics engineering management, Helsinki, Finland

Asset Information Systems

Facilitating Change Towards Predictive Maintenance

Martine van den Boomen(✉), Marc Botermans, Thijs de Weerd, and Andreas Burzel

Rotterdam University of Applied Sciences, Heijplaatstraat 23,
3089 Rotterdam, JB, The Netherlands
m.van.den.boomen@hr.nl

Abstract. Predictive maintenance is the promise of the future in infrastructure asset management. Predictive maintenance more and more uses sensor data. Sensors are relatively cheap, and their data mostly comes in huge quantities. Trends or flags may be observed in the data, sometimes with traditional statistical analyses and more often with advanced analyses such as machine learning techniques. These trends and flags may indicate a developing problem, allowing maintenance professionals to act before a failure occurs. However, as of today predictive maintenance is far from being common practice in infrastructure asset management. Failure mechanisms are often extremely complex. Besides knowing how to search for flags and trends, one should first know what to search for. Also, on the organisational side barriers are found. Often, big data is available in infrastructure organisations but underutilised for various reasons such as inaccessibility of data, client unfriendly user interfaces and a lack of tools to analyse data effectively by maintenance engineers. The current research investigates the potential for more predictive maintenance in current professional practices, not by adding new sensors but through exploitation of existing data (data mining) and removal of barriers which professionals experience in using such data.

1 Introduction

Predictive maintenance is the latest trend in maintenance and seen as the way to go forward [1]. To understand its added value first the traditional view on maintenance is elaborated on. Maintenance in infrastructure asset management is generally categorised in corrective and preventive maintenance [2] as depicted in Fig. 1. Corrective maintenance comprises of both run-to-failure maintenance and repair of critical failures. Run-to-failure maintenance is a dedicated strategy in which non-critical assets are allowed to fail and repaired upon failure. Such strategy is deployed for non-critical assets when it is cheaper to let them fail, and no harm is done. In contrast, repair of critical failures occurs when preventive maintenance has failed. As such repair of critical failures is not a dedicated maintenance strategy but a failed maintenance strategy. Each maintenance organisation aims to reduce the number critical failures and their repairs. This is done by either reducing the criticality of an asset, for example by a modification or redesign (e.g., to build in redundancy) or by deploying more preventive maintenance.

Preventive maintenance is generally classified in time-based maintenance and predictive maintenance [2, 3]. Time-based maintenance performs maintenance activities at

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 221–230, 2022.
https://doi.org/10.1007/978-3-030-96794-9_20

fixed intervals in calendar time or run-time hours. This strategy is effective in preventing failures if the mean time between failures is predictable and fairly constant. The other preventive strategy is predictive maintenance which started-off as condition-based maintenance. Here the condition of an asset is periodically assessed (measured) and follow-up maintenance is scheduled accordingly. This strategy is effective in the presence of a so-called pf-interval which stands for a measurable point of condition deterioration after which a mean time to failure is fairly predictable. An associated maintenance strategy is called detective maintenance. It is debatable whether this strategy belongs to corrective or preventive maintenance. Detective maintenance is periodic testing for hidden failures. Protective devices such as non-return valves, smoke alarms or stand-by equipment are subject to this type of maintenance. It is sometimes seen as preventive maintenance because failure of such devices does not necessarily lead to a critical failure in the primary process. Adequate detective maintenance can therefore prevent critical failures. It can also be seen as a type of corrective maintenance because such devices are tested for still being functional and repaired or replaced when found in a failed state.

Condition-based maintenance has advanced to structural health monitoring (SHM) of assets. SHM is sometimes equated with condition-based maintenance, however, SHM measures much more frequently - directly or indirectly - the health of assets by for example the application of sensors. This generates big data and analysis of this data with advanced statistics or machine learning techniques allows for spotting deviations, establishing trends and forecasting of condition deterioration. Condition-based maintenance could therefore be seen as a precursor of SHM and both intend to predict maintenance needs.

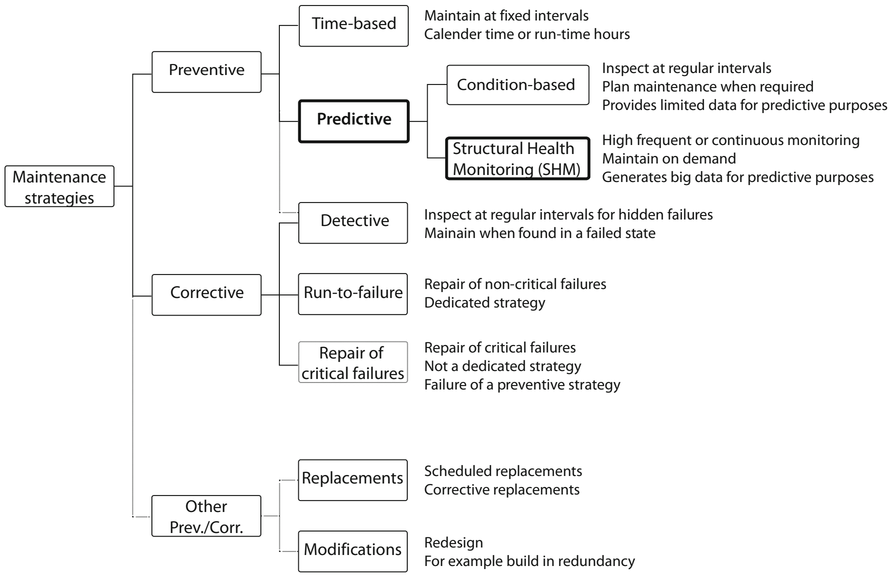


Fig. 1. Taxonomy of maintenance strategies and the position of predictive maintenance and structural health monitoring

Several decision criteria for selecting appropriate maintenance strategies for assets exist which are well summarised by one of the founders of modern reliability centred maintenance [2]: "an effective maintenance strategy should reduce the risk to an acceptable level, is technically feasible and economically worthwhile". In this context risk is defined by the multiplication of a failure probability, which has a relation with the type and intensity of maintenance, with a ranking for the impact upon failure, resulting in a so-called risk priority number.

2 The Promise of Predictive Maintenance

Predictive maintenance is promising because it transforms current standardised block maintenance activities into tailor-made maintenance activities for individual assets depending on their performances [4]. In current practice similar maintenance activities are scheduled for similar assets but in real life assets are hardly ever similar. Performances depend on the quality of instalment, usage, quality of maintenance and factors such as settlements, vibrations, temperature, humidity, etc. Even two identical assets in similar circumstances will seldom obey their estimated mean time to failure or behave identically. Therefore, predictive maintenance reduces failures because it spots deviations in time and a maintenance engineer can act accordingly. It also increases the effectiveness because maintenance is done when required, not too late and not too early.

Another promise of predictive maintenance is the ability to exploit data for mid and long-term predictions by using advanced statistics and machine learning. Mid- and long-term predictions support availability calculations and long-term asset planning. A conventional long-term asset plan (LTAP) visualises and budgets maintenance activities over a period of 50 years. Maintenance activities, their intervals which are derived from estimates of the mean time between failures, and costs are projected on a timeline [5]. Although analyses of past failure data support LTAP's to some extent, the estimates of maintenance intervals are still mainly based on expert judgement for several reasons. First is that hardly any maintenance organisation has sufficient failure data on individual assets to support the derivation of failure probability distributions because maintenance organisations aim to prevent failures. Just a small proportion of maintenance activities is run-to-failure maintenance. A second reason is that maintenance influences the failure probability of individual assets. In general (not always), when maintenance is intensified, the failure probability of assets will be reduced and vice versa. It is very difficult to establish generic mathematical relationships between the intensity of maintenance activities and the failure probability of assets based on past failure data and maintenance records. Some statistical approaches exist for estimating the relationship between maintenance frequency and failure probability [6–8]. However, in practice these are difficult to apply because of the absence of sufficient failure data. Even if such relationships are found, past performances do not guarantee future performances.

Moreover, condition deterioration and failure behaviour in practice are extremely complex and many maintenance engineers question whether common reliability distributions sufficiently grasp this complexity. This argument holds for many infrastructure organisations but is especially valid for organisations that manage critical protective infrastructures like storm surge barriers which are often unique and seldom operated.

Standby equipment which is tested periodically behaves differently than similar equipment that is put into full use and functional most of the time. Failure distributions provided by suppliers of equipment therefore have limited value in these circumstances.

This also explains why predictive maintenance is embraced as a way to move forward. Predictive maintenance builds on high-frequent monitoring and big data analysis. The big data are often provided by sensors. Sensor data are generally derived quantities such as temperature, vibration, tension, position, or wattage. These are relatively cheap to obtain, and many infrastructure organisations already have sensors installed and big data collected. Although these derived quantities do not directly measure condition or failures, they function as signal indicators for the health of assets. Advanced statistics and machine learning techniques can be used to find patterns in huge quantities of data. Also, data series can be combined. Even the full absence of failure data does not have to be a show-stopper because unsupervised learning techniques can handle the absence of result variables and still provide meaningful information about the assets' health [9–11].

3 Research Question and Methods

Despite the many advantages of predictive maintenance and the availability of underutilised sensor data in many organisations that manage infrastructures, predictive maintenance is far from being common practice. The current research asked the question why and investigated the barriers for the transition towards more predictive maintenance. This research was done within an organisation that manages critical infrastructures in the Netherlands and which possesses huge quantities of sensor data. The research question is explored by guidance of the Theory of Change (ToC) [12]. This theory emphasises the role of conscious and unconscious assumptions in transition processes. Assumptions are often taken for granted and can both hamper and accelerate change. It is important to make unconscious assumptions known and to share assumptions to facilitate change. The current research followed the first 5 steps of the ToC [12]:

1. Identification of the desired change, why and for whom?
2. Analysis of the (change) system and the current situation;
3. Mapping the pathways for change;
4. The assumptions and barriers underlying the change;
5. Strategic options.

A focus group discussion, individual discussions and data analyses were conducted for addressing these steps of the ToC. In the following sections, the findings of this research are elaborated on by following the steps of the ToC. As part of the ToC approach, a case was built for demonstrating the potential of underutilised data for predictive maintenance. The results were shared with a team of maintenance and reliability engineers responsible for different critical infrastructures at different geographical locations in the Netherlands. Further, the barriers for a transition towards more predictive maintenance were discussed. This resulted in strategic options to facilitate such change. The key findings are summarised in Fig. 3. The findings may also benefit other infrastructure organisations to discover the potential of underutilised data and to make a shift towards more predictive maintenance.

4 The Case Study

A case study was provided by Rijkswaterstaat, a governmental organisation in the Netherlands and responsible for managing national infrastructures like highways, bridges, primary waterways, locks and storm surge barriers. The current research is geared at the storm surge barriers and especially the Maeslant barrier. Storm surge barriers are critical infrastructures and of national importance. Another characteristic is their uniqueness: each barrier is one of a kind. Moreover, storm surge barriers are rarely functional. They are tested yearly on closure and additionally put into operation in extreme circumstances, when a critical threshold of the sea water level or river water level is reached.

The Maeslant barrier is a massive storm surge barrier and was put into service in 1997. It is designed for a lifetime of 100 years with a closing frequency of 1:10 years in the first half of its life and 1:5 years in the second half of its life. The Maeslant barrier will close the Nieuwe Waterweg at a sea water level of 3 m above a normalised level at Rotterdam, the second city of the Netherlands. The Maeslant barrier protects the densely populated urban delta of Rotterdam and its hinterland. Since it was put into service, it has closed twice, in 2007 and in 2018, due to extreme sea water levels. The Maeslant barrier has two massive doors of each 210 m in length and 22 m in height. Under normal conditions these doors rest in their docking stations at the shores at each side of the waterway. To close the waterway, the rolling gates of the docking stations are opened, and the doors are navigated on the waterway until they reach each other. When in position, the doors are filled with water, sink to the bottom and close the Nieuwe Waterweg. In reverse, water is pumped out of the doors, the doors rise to the water surface and are navigated back to their docking stations (Fig. 2).

Fig. 2. Maeslant storm surge barrier [13]

Although the Maeslant barrier is seldom used, it is a heavenly maintained infrastructure because of its importance for the South-Western Netherlands. Each year preceding the storm season, the Maeslant barrier is tested on closure and reopening. Potential issues emerging from the test closure are resolved. During the storm season maintenance activities are minimised as the barrier should be prepared to be called upon at any time. Most maintenance is planned outside the storm season and follows conservative time-based principles such as time-based replacements of components and service maintenance. In addition, regular tests of crucial systems are carried out and detective maintenance (inspection on standby equipment) is planned with regular intervals. As such, standby equipment is periodically tested on being functional and repaired or replaced when found in a failed state.

Step 1. Identification of the desired change, why and for whom?
Predictive maintenance as described in the introduction is not yet part of the maintenance approach. Here the maintenance and reliability engineers see opportunities because storm surge barriers and also the Maeslant barrier are probably the best monitored infrastructures in the Netherlands. A huge quantity of sensor data is available and monitored but underutilised for predictive purposes. Analyses of past sensor data may assist in establishing sharper inspection and maintenance intervals based on individual components' performances. Analyses of sensor data may also transform detective maintenance into predictive maintenance. Instead of testing whether a component is still functional (not in a failed state), analyses of sensor data during tests and comparing these with the sensor data from previous tests may indicate a piece of equipment is about to fail. And as such preventive maintenance can be deployed to prevent failures of standby equipment which makes it even more interesting for storm surge barriers which are seldom called upon but should work if so.

Summarising: the maintenance engineers observed two potential advantages of predictive maintenance. First, making preventive maintenance intervals tailor-made, designed on the specific needs of individual assets and second, reducing corrective repairs and replacements following from detective maintenance on standby equipment. In the current case study safety is not the prime issue for a transition towards predictive maintenance as equipment is tested before the storm season. However, a transition towards more predictive maintenance could make the current time-based planning more effective and efficient. Moreover, the components of the Maeslant barrier are unique in how they are operated. Therefore, generic supplier information on lifetime distributions cannot be used to support the maintenance intervals and remaining life estimations. On the plus side, the Maeslant barrier is intensively monitored, and this generates a huge amount of data during tests. The question is how to exploit this data for predictive purposes and how to take the maintenance organisation along in a transition towards more predictive maintenance.

Step 2. Analysis of the (change) system and the current situation
Storm surge barriers are unique assets and vital for the protection of urban deltas and their hinterlands. Maintenance and reliability engineers working on these assets are highly specialised and staff turnover is deliberately low. Highly specialised engineers need arguments for change. They want 'proof'. Interesting is that the research question did not emerge from external researchers but from the maintenance and reliability engineers themselves. Based on their long-term experience in working with these assets and the data collection, they sense that more value can be obtained from the data but they lack the data analysis skills and time to demonstrate the added value. Furthermore, the

maintenance and reliability engineers experience that current workload and day-to-day matters leave little time for experimenting with new ideas and approaches required for a transition to more predictive maintenance. On the other side, the maintenance and reliability engineers express that after obtaining new skills, predictive maintenance may ease their current workload.

Therefore, facilitating change starts with facilitating the maintenance and reliability engineers in the current situation. This means that the added value of predictive maintenance first needs to be demonstrated in current practice. Hereafter the maintenance and reliability engineers need enablers to facilitate change. Moreover, the maintenance and reliability engineers working on the storm surge barriers are embedded in a large organisation. For effective change, the maintenance and reliability engineers need managerial support from top level as well.

Step 3. Mapping the pathways for change

The mapping of pathways for change has three milestones: (1) building a case and providing evidence for the added value of predictive maintenance; (2) discussing this evidence with the maintenance and reliability engineers and (3) discussing the barriers for change and how to overcome these.

The case addressed a hypothesis from one of the maintenance engineers. The opening and closure of the rolling gates of the docking stations which contain the doors are monthly tested. The rolling gates are opened and closed by electric winches. These winches are on standby most of the time. During the test, which takes about 20 min, the wattage is measured by sensors. The hypothesis of the maintenance engineer was that analysis of past trends of this wattage could be useful for future predictions. Instead of waiting for a winch to find it in a failed state when tested, the observation of the wattage and its trend analysis may indicate future problems. A winch could then preventively be revised or replaced instead of correctively, which would put less pressure on the maintenance organisation and also would be less expensive. However, the sensor data of the wattage of the winches come in huge quantities, are stored in a database and are not easily accessible. Also, the knowledge on how to analyse big data for predictive purposes was lacking as maintenance and reliability engineers are not trained as data scientists. Together with the maintenance and reliability engineers, the researchers subtracted the data, cleaned the data and performed fundamental descriptive statistics. Indeed, expected trends could be observed. Maintenance and reliability engineers were able to explain anomalies based on their in-depth asset knowledge. Visualising the trends and discussing them with the maintenance and reliability engineers already provided sufficient evidence for them. Maintenance engineers could recall higher or lower water levels explaining deviations in trends and also the impact of a revision was clearly visible in the data.

Results were plenary discussed in a focus group with maintenance and reliability engineers from all storm surge barriers in the Netherlands. There was broad consensus on the added value for more predictive maintenance followed by a constructive discussion on the barriers which needs to be removed to facilitate such change.

Step 4. The assumptions and barriers underlying the change

The prime barriers for change mentioned by the maintenance and reliability engineers are a (perceived) lack of knowledge in data science and a lack of time to specialise oneself in a new knowledge domain. Maintenance and reliability engineers have high workloads and time to experiment with data analysis is scarce. During the focus group discussion, several participants indicated that their current workload does not enable them to take on extra tasks. As a result, the ambition to put time and energy into the transition to predictive maintenance remains idle and did not receive further actions yet.

Moreover, some underlying assumptions were observed such as: predictive maintenance is perceived as extremely difficult; without a thorough knowledge of advanced statistics and machine learning, big data cannot be analysed. Another barrier mentioned was that the current database lacks a user-friendly interface. It does not invite for 'playing with data'. In addition, maintenance and reliability engineers indicated that it takes them at least a full day to access and analyse certain data and to derive useful information from it.

It was also observed that the implementation of more predictive maintenance lacks a sense of urgency and can even have a negative connotation of optimising maintenance intervals on the expense of safety and reliability. Underlying fears or assumptions can be that predictive maintenance motivates management to cut budgets or to dispose of specialised personnel. However, this focus group discussion was held with highly specialised maintenance and reliability engineers who are primarily driven by a professional pride and an ambition for personal development and growth. The primary barrier for change is the perceived gap between sound current maintenance practice and what the specialists see as an ideal situation. A lack of knowledge and (time-saving) tooling are considered the dominant barriers for change towards more predictive maintenance.

Step 5. Strategic Options

There is energy for change towards more predictive maintenance at the storm surge barriers. The following strategic options are suggested for mobilising this energy. Adding data scientists to the maintenance and reliability teams may boost the transition. Data scientists need object knowledge which can only be provided by maintenance and reliability engineers. A data scientist or algorithm can never replace maintenance and reliability engineers but can enhance their effectiveness.

Another suggestion is progressing in small steps. The promise of predictive maintenance is so huge that it may paralyse maintenance and reliability teams. Fundamental descriptive statistics on available data are the first step forwards: learning how to subtract and clean data, how to visualise trends and using expert knowledge to explain these trends. These fundamental statistics are easily learned in trainings.

User-friendly interfaces to perform data analysis tasks are also considered as great facilitators of change. Otherwise, big data analyses can be very time consuming and frustrating. Also important is to facilitate platforms where experiences with predictive maintenance can be shared. Storm surge barriers are unique assets, however, not unique in being unique. Much can be learned from practices in other asset management domains like bridges and process installations.

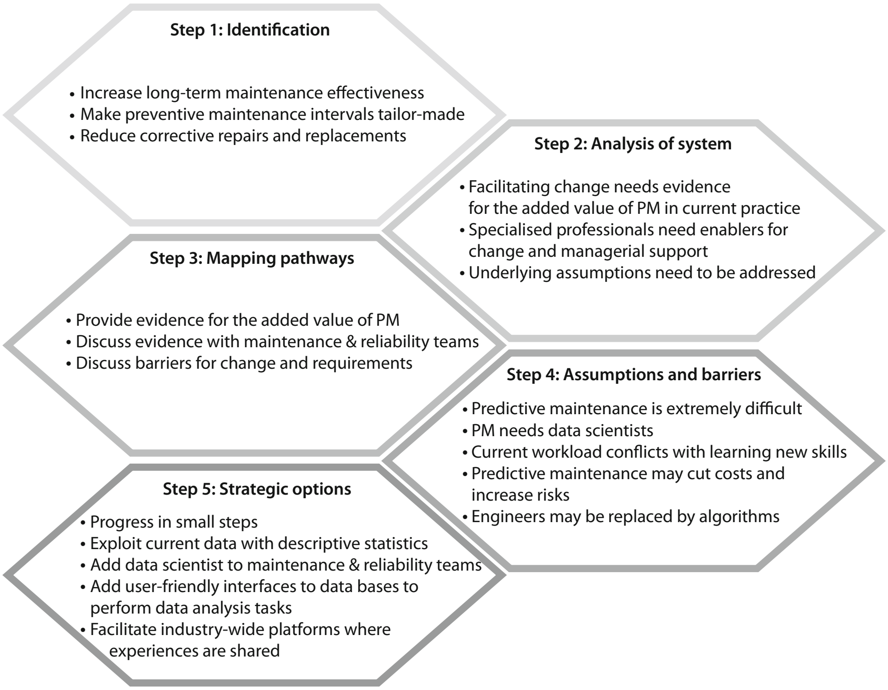


Fig. 3. Summary of the first steps of the Theory of Change applied on the case study

5 Conclusions

Predictive maintenance is promising. However, transition from current maintenance practise towards more predictive maintenance does not occur by itself. The following barriers were found in the case study of this research:

- The potential of predictive maintenance is not yet a proven concept for maintenance and reliability engineers.
- Sensor data are still locked-up in systems and not easily accessible for further analysis; user friendly interfaces for advanced data analysis are not yet available.
- Maintenance and reliability engineers are unfamiliar with statistical techniques for big data analysis.
- Data scientists are not yet part of a maintenance organisation.
- Maintenance organisations are under pressure. Finding time for learning new skills and techniques is challenging.
- A shift towards more predictive maintenance requires management of change which also needs to address unconscious assumptions; maintenance engineers with in-depth asset knowledge are indispensable but may fear that they will be replaced by data scientists and algorithms.

To overcome these barriers collecting good practices and evidence on using sensor data in order to better predict maintenance intervals will be a first step forward. Also, the development of user-friendly interfaces for big data analysis will greatly facilitate change. Maintenance and reliability engineers will benefit from training in fundamental data analysis, which will enhance their current maintenance decision making. Data scientists can be added to maintenance organisations, not to replace but to support maintenance and reliability engineers with advanced data analysis. Finally, for a successful transition towards predictive maintenance, addressing unconscious assumptions and possible fears is also essential. Predictive maintenance should not be seen as a strategy to cut budgets but embraced as a process for professionalising current maintenance practice.

Acknowledgements. We are extremely grateful for the expert knowledge, data, time and fruitful discussions provided by the maintenance and reliability engineers, and the managers of the Dutch storm surge barriers; especially those of the Maeslant barrier: Pieter Jacobs, Peter Oskam, Linneke van der Veeken and Jaco van Voorst.

References

1. Márquez AC, González-Prida V, Fernández JFG (2018) Advanced maintenance modelling for asset management: techniques and methods for complex industrial systems. Springer, Cham
2. Moubray J (1997) Reliability centered maintenance. Industrial Press, New York
3. Bradley E (2017). Reliability engineering: a life cycle approach. Taylor & Francis, CRC Press
4. Selcuk S (2017) Predictive maintenance, its implementation and latest trends. Proc Inst Mech Eng Part B: J Eng Manuf 231(9):1670–1679. https://doi.org/10.1177/0954405415601640
5. Van de Ven-Glastra MJ, Verberk JQJC, Dominicus J, Bindt A (2017) Strategic scenario analysis of long-term asset management planning. Water Pract Technol 12(1):48–54. doi:https://doi.org/10.2166/wpt.2017.009
6. Jardine AKS, Tsang AHC (2017) Maintenance replacement and reliability. CRC Press, Boca Raton
7. Mondro MJ (2002) Approximation of mean time between failure when a system has periodic maintenance. IEEE Trans Reliab 51(2):166–167. https://doi.org/10.1109/TR.2002.1011521
8. Wang W, Dell'Anno M, Zeh C (2007) Approximation of mean time between failures with periodic maintenance. Int J Perform Eng 3(3):382–384
9. Sajid S, Abid H, Shashi B, Mohd J, Tarum G, Manoj M (2021) Data science applications for predictive maintenance and materials science in context to Industry 4.0. In: Materials today: proceedings 2021. https://doi.org/10.1016/j.matpr.2021.01.357
10. Lughofer E, Sayed-Mouchaweh M (eds) (2019) Predictive maintenance in dynamic systems. Springer, Cham. doi:https://doi.org/10.1007/978-3-030-05645-2
11. Carvalho TP, Soares FA, Vita R, Francisco R, Basto JP, Alcalá SGS (2019) A systematic literature review of machine learning methods applied to predictive maintenance. Comput Ind Eng 2019:137. https://doi.org/10.1016/j.cie.2019.106024
12. Van Es M, Guijt I, Vogel I (2015) Theory of change thinking in practice: a stepwise approach. In: Hivos ToC guidelines. Hivos, The Hague
13. Rijkswaterstaat (2005) Maeslant storm surge barrier. Rijkswaterstaat. Retrieved from https://beeldbank.rws.nl/MediaObject/Details/313508

An Intangible Asset Management Proposal Based on ISO 55001 and ISO 30401 for Knowledge Management

Vicente González-Prida[1,2](✉), Antonio Guillén[1], Carlos Parra[1], Eduardo Candón[1], and Pablo Martínez-Galán[1]

[1] University of Seville, 41092 Seville, Spain
{vgonzalezprida,ajguillen}@us.es, parrac@ingecon.net.in
[2] UNED, 28040 Madrid, Spain

Abstract. This contribution is intended to provide a view on the standards ISO 55001 (requirements for an asset management system) and the ISO 30401 (requirements for a knowledge management system), in order to consider knowledge and human asset management, as a relevant dimension for all engineering and industrial sectors. An intangible asset management framework is proposed in this paper, considering the principles and requirements of the above-mentioned standards, together with methodologies already developed for physical asset management, in order to coordinate and realize value (in this case) from the industrial knowledge. This proposal is intended to be a helpful decision support tool in order to align the different knowledge areas to the industry strategy and, in particular, to the business drivers of the company. Such a proposal will require first the identification of the key company knowledge areas, which are needed to sustain and grow the business, supporting strategic decision-making. After that priorization, a gap analysis shall be performed in order to reckon if the core knowledge and the key industrial capabilities match with the current company resources and where they lie (people expertise, document repositories, etc.). This analysis will help not only to detect core capabilities to be developed and/or acquired by the organization, but also to reassign efficiently the current company resources to more critical activities with more added value. Finally, connections to risk and uncertainty references, the digitalization industry process, as well as to possible future research lines are commented as a conclusion.

1 Introduction

According to ISO 55000 [1], an asset is something that has potential or actual value to an organization. That means in other words, an item that has the capability to impact in the company throughout its full lifecycle. Every company will decide what an asset is and will set priorities accordingly. This is important to start mentioning since, according to this definition, an asset may be a tangible element (like manufacturing equipment, testing tools, etc.), but it may be as well an intangible element (like knowledge, human resources, competences, etc.). Together with the concept of asset, it is commonly define the term of asset management. This term will refer to all those coordinated activities of

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 231–240, 2022.
https://doi.org/10.1007/978-3-030-96794-9_21

an organization with the aim of realizing value from its assets. According to the kind of asset (tangible or intangible), the corresponding management will not be necessary the same. However, there will be main lines that can be similar.

Nowadays, there are standards that help organizations to manage their different systems. In particular, the ISO 55001 [1] helps to manage assets. That standard does not specify if such assets have to be tangible (physical) elements or not. Nevertheless, organizations usually apply this standard to manage their physical assets. On the other hand, It has been recently published the ISO 30401 [2] that refers to the establishment of a knowledge management system. As any other management system defined by the ISO, it follows too the well-known PDCA (Plan-Do-Check-Act) cycle, also named Deming Cycle. In general, terms, Asset Management based on ISO 55000 [1] helps to respond to specific questions and supports the decision-making according to gathered data. In this sense, these questions are like the following ones:

- What is the Asset portfolio?
- How important are these assets for the company?
- Are the resources efficiently assigned for each asset?
- How much will the assets cost during their whole life cycle?
- How are the assets linked to the business goals?

In order to respond to these questions, organizations apply a set of methodologies like Criticality Analysis (CA); Reliability Centered Maintenance (RCM); Reliability, Availability, Maintainability and Safety Analysis (RAMS), Life Cycle Cost Analysis (LCCA), among many others. As commented, organizations usually apply these methodologies in a cyclical way following a kind of PDCA or Deming cycle, in order to manage their physical assets with a continuous improvement background. In others words, asset management demand the accurate definition of asset concept and the use of analysis tools in order to extract information. Such information is that one required for the decision making and for the provision of support to the management function. In terms of decision making model, an asset is a parametric representation of an element that provides value to a system or organization. This kind of management can be applied for intangible asset in a similar way as physical asset. In this sense, considering intangible asset as knowledge, through the definition of management system, it will respond to questions like:

- What knowledge is required? Why?
- How can knowledge be measured and parameterized? How can a knowledge asset portfolio be define?
- Who has/where is this knowledge available?
- How important is that knowledge?
- How efficiently is the organization using its knowledge?
- How is this knowledge protected and developed?
- What is the added value of this knowledge to the business strategy?

According to the ISO 30401 [2], knowledge management will refer to the combination of processes, actions methodologies and solutions that enable the creation, maintenance, distribution and access to organizational knowledge. Consequently, the corresponding management system will be that set of interrelated or interacting elements of an organization that establish, embed and enable the knowledge management policies and objectives, as well as the processes to achieve those objectives.

In addition to this, that standard compiles a set of other definitions. Among them, there is the concept of Competence, which is defined as the ability to apply knowledge and skills in order to achieve intended results for the organization. Of course, such intended result has to be aligned with the organization's strategy. Other definition to competence can be found in the DoD Instruction 1400.25 [3], where it is depicted as an observable, measurable pattern of knowledge, skills, abilities, behaviors, and other characteristics needed to perform work roles or occupational functions successfully. In any case, added to this concept, the standard specifies which may be the core competences, as well as the model and framework where these competences have to be described and structured. As a remark to the standard, the ISO 30401 [2] is probably quite focused on competences related to human resources, where it is supposed that the company knowledge stays. However, it is important to underline as well that knowledge may rely in documental repositories too, and must be taken into account that in the next generation (or even today), much of these competences will belong to AI.Knowledge and competence definition, together with management are experimenting a revolution as a consequence of digital transformation, especially, the AI development.

2 Intangible AM (IAM) from a Physical AM (PAM) Model

Back to the general asset management system, any asset management framework will have to consider the business context. That refers to those company's strategy, policies, objectives, plans, but also the activities to do (Fig. 1).

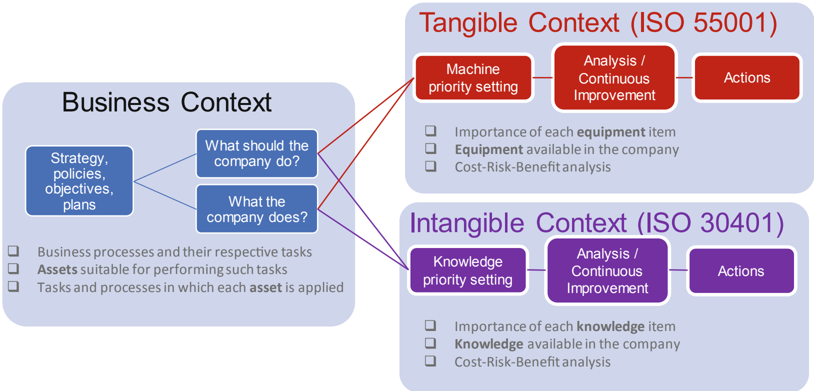


Fig. 1. Business context and asset management

Regarding such activities, it is important to identify what the company should do, and what the company really does. In other words, for the definition of the business context, it will be needed:

- Business processes and their respective tasks
- Assets suitable for performing such tasks
- Tasks and processes in which each asset is applied

Since asset may be tangible (machines) or intangible (knowledge), the framework should follow and consider both possibilities. Together with the framework, it is important to take into account the possible already existing procedures in the company. In this sense and regarding the business context, the company probably has already established different manuals related to its goals, risk & opportunities, strategic planning, audits etc. Regarding the tangible context, procedures that may exist already in the company may refer to manufacturing, non-conformity management system, investments procedures, etc. On the other hand, regarding intangible context, the organization probably has already HR plans, control of documented information, public affairs or talent management system among other procedures. Considering physical asset management, there are many publications dealing with that topic, as for example the 8-phase management framework defined by Crespo (2007) [4].

This paper on the contrary, is interested to focus on intangible asset management, considering knowledge as such an intangible asset. Consequently, adapting the above-mentioned 8-phase management framework (which is devoted to physical assets) to a knowledge management framework (which is devoted to intangible assets), the phases can be for instance the following ones (Fig. 2):

- Phase 1: Definition of Competences/Knowledge Areas According to Business Strategy
- Phase 2: Competences/knowledge areas priority setting
- Phase 3: Intervention on high impact competence-resources problems
- Phase 4: Design of competence plans and resources requirements
- Phase 5: Competence plan, schedule and resources optimization
- Phase 6: Resources assessment and control
- Phase 7: Competences and resources life cycle analysis
- Phase 8: Continuous improvement and new techniques utilization

Similarly to the physical asset management, in this case it will be needed to determine a specific intervention level. That means that the organization should have to consider as much as possible the following aspects before defining a knowledge management framework:

- Business processes and their respective tasks
- Knowledge suitable for performing such tasks
- Importance of each knowledge item
- Knowledge available in the company (documents)
- Knowledge nature: tacit or explicit
- Task and processes in which each knowledge item is used

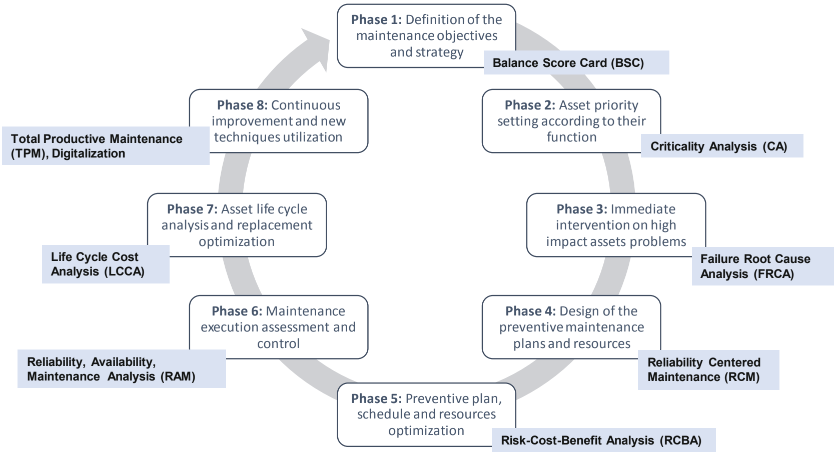


Fig. 2. Asset management framework proposed by Crespo (2007) [4]

3 An Example of IAM: An Asset Knowledge Management Framework

According to the ISO 30401:2018 [2], knowledge is defined as the human or organizational asset enabling effective decisions and action in context. On the other hand, the knowhow is defined by the Merriam Webster dictionary as the knowledge of how to get things done, or the knowledge gained by actually doing or living through something. Knowhow is usually described with characteristics like a personal and reusable asset that may promote knowledge. In the introductory section, some questions were stablished in regard to physical asset management, In agreement to such questions, one of the first steps to stablish a knowledge management framework will be to identify the critical knowledge or how to prioritize it. In that sense and with that purpose, it is possible to consider questions like:

- Is the knowledge aligned to the business mission/vision?
- Is it specific or general? Is it secret or open?
- Does it provide any competitive advantage?
- Is it clearly linked to competences and experts?
- …?

3.1 Phase 1: Definition of Knowledge According to Business Strategy

According to the above-mentioned questions, it will be crucial to identify competences in order to determine role characteristics (Fig. 3).

In addition to this, it is important to identify functions and structure in the organization, in order to determine HR as well as document characteristics. With all this, it will be obtained as a result a competence matrix, together with a kind of repository where the different experts and documents are identified for each specific competence.

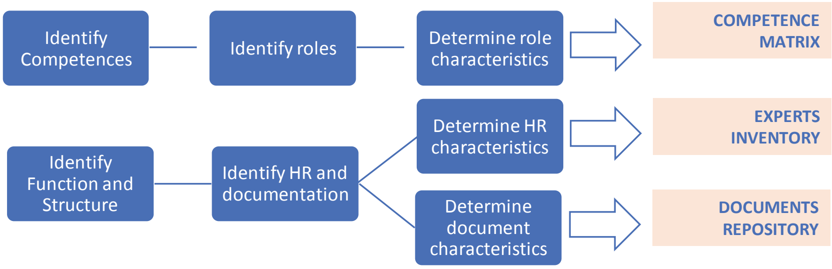


Fig. 3. Identification of what the company should have, and what the company really has

3.2 Phase 2: Knowledge Priority Setting

Before the contrast between desired competences with existing resources, it is important first to prioritize such competences or required knowledge areas. Since the matter under study is intangible, this activity will be performed by using a qualitative assessment extracted from experts' judgement, evaluating how align is each competence with the business strategy. With this, the result should be the identification of core competences for the business. As an example, Fig. 4 illustrates the methodology followed by the European Defence Technological and Industrial Base (EDTIB), exploring key skills and competences for defence in order to be sustained into the future [5].

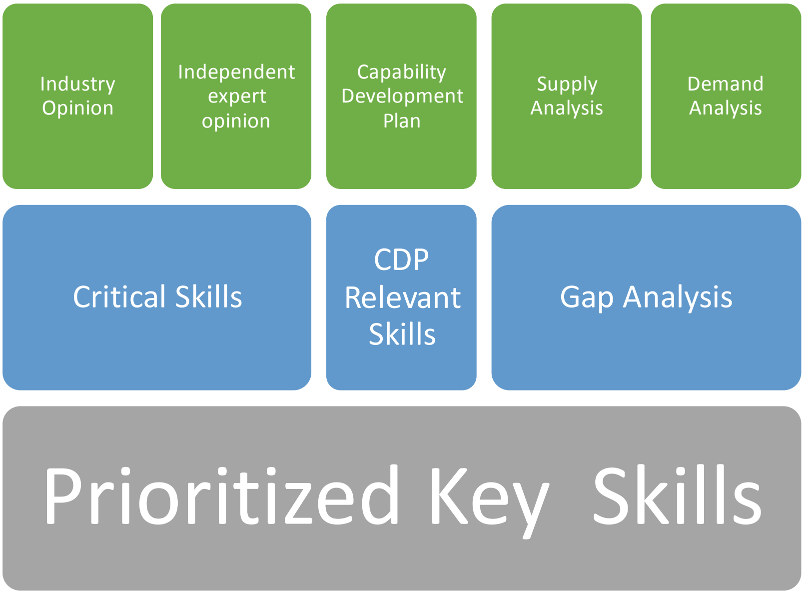


Fig. 4. Methodology for prioritisation of key skills and competences for the EDTIB (adapted from [5]).

3.3 Phase 3: Intervention on High Impact Competence-Resources Problems

Once identified the core competences for the business, and the existing resources, not only in the company, but also in the labor market. In that sense, Defence Growth Partnership (DGP) presents a detailed map of skill's criticality and availability as identified by the Skills Survey [5]. As depicted in such reference, it represents the scale of the 'critical skills issue' as a proportion of the overall response. In that survey, criticality is understood in terms of the ability of companies' to meet their business needs within the labor market. Additionally, it is important to determine if the company resources are well assigned to the core competences, and if there are gaps. These gaps can be in terms of competences without any resources assigned to them, or, on the other hand, gaps in terms of existing resources that are not devoted to defined competences (if there are people devoted to these competences or not, and vice versa). With this, it is important to study the severity of the deficiencies as well as the risk of knowledge loss.

3.4 Phase 4: Design of Competence Plans and Resources Requirements Phase 5: Competence Plan, Schedule and Resources Optimization

The above analysis will be useful in order to align competences and resources. Such alignment will be obtained applying a kind of action plan where competences and resources requirements have to be depicted. In other words, the company will find core competences with a gap in resources, and existing resources that do not match with the desired competences. The matching between competence plan and resource requirements will need the application in the Human Resources (HR) area of change management technics, in terms of training and/or rotating the staff, or acquiring the knowledge from the labor market. It will suppose to the employees (technicians, managers, etc.) an incremental and transformational change. A possible process for non-core competences is suggested in Fig. 5.

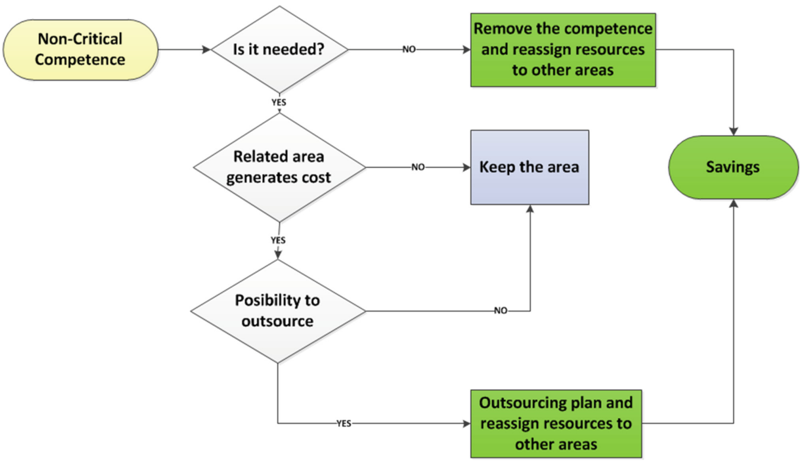


Fig. 5. Possible process for non-core competences

3.5 Phase 6: Resources assessment and Knowledge control

Phase 7: Competencies and resources life cycle analysis

In order to control and assess company resources linked to specific core knowledge, there are today many software and applications that helps the management of such success factors. The usual process for that control and assessment can be described in the following steps:

- Manager sets goals
- Manager guides employee through the goals
- Employee works towards goals
- Manager reviews employee performance

In order to review the performance, it is important to define properly specific measures and indicators. The variation of such indicators with the course of the time will help to observe the progress degree, mainly when specific measures have been applied. These aspects will help the company to detect necessities so, it will require the implementation of a change plan that must be communicated to the organization. Such a change plan will need of course an implementation process and, after that, and assessment. At the end of the day, the needed changes are expected to be absorbed by the business culture.

3.6 Phase 8: Continuous Improvement and New Techniques Utilization

The above-mentioned phases will require of course to be repeated continuously following a Deming cycle as a usual continuous improvement process.

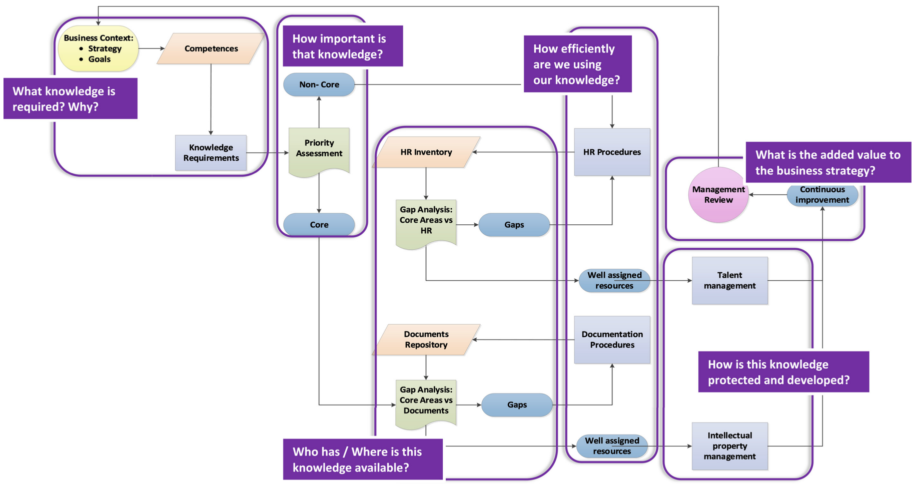


Fig. 6. Proposed knowledge management process

Figure 6 is intended to summarize the knowledge management framework, showing a cyclical process where the above-mentioned phases are represented, together with the initial questions stablished for an appropriate intangible asset management system. Therefore, this is just an example of the process related to the proposed knowledge management framework. Nevertheless, there are of course others found in the literature (for example: [6, 7]), together with interesting analysis and study cases that link asset management (in the traditional way) with aspects related to knowledge and other intangible elements ([8–10] or [11] among others). All the commented phases will require to gather data and to analyze them in order to get conclusions that helps to improvement de knowledge management system itself. Nowadays, there are many easy-to-use tools that can aim and support all this process, since new business intelligence platforms may provide significant and disruptive tools helpful to the proposed framework.

4 Summary and Conclusion

"Knowledge is power" has become a common but true cliché in all areas of industry. In the knowledge economy, there is a consensus that this intangible asset is one of the main resources for wealth creation in companies. Therefore, it seems to be crucial that companies have to manage their knowledge in order to obtain the maximum possible added value. Of course, all organizations manage their knowledge, although not all of them always managed it optimally. Due to that reason, this document has presented a management framework based in already well-known standards like the ISO 55001 for asset management and the ISO 30401 for knowledge management. Although the concept of competence has been depicted in relation to ISO 30401 and asset knowledge management framework, some other recent work has been done in relation to competences and asset management [12]. In addition to that, this contribution has adapted the methodology depicted for physical assets, in order to be useful to prioritize competences, check if they are aligned with the existing resources and to fulfil the gaps by training, relocating or acquiring new resources. Basically, those companies that are aware on how to manage efficiently their knowledge are in a better competitive situation. Future discussions can be focused on the interpretation of these findings, particularly in light of the continuous improvement over this topic, as well as their limitations.

References

1. ISO 55001:2015. Asset management—Management systems—Requirements. International Organization for Standardization
2. ISO 30401:2018. Knowledge management systems—Requirements. International Organization for Standardization
3. DoD Instruction 1400.25, Volume 250, Civilian Strategic Human Capital Planning, November 18, 2008
4. Crespo A (2007) The maintenance management framework: models and methods for complex systems maintenance. Springer, London. https://doi.org/10.1007/978-1-84628-821-0. ISBN 9781846288210
5. Retter L, Taggart L, Freeman J (2015) Executive summary. Key skills and competences for defence. EDA contract reference: 14.CPS.OP.030. RAND Europe RR-1226-EDA
6. Kriege LK, Jooste JL, Vlok PJ (2016) A framework for establishing a human asset register for the improved management of people in physical asset management. S Afr J Ind Eng 27(4):77–89. https://doi.org/10.7166/27-4-1549

7. Colakoglu S, Hong Y, Lepak DP (2010) Models of strategic human resource management. In: Wilkinson A, Bacon N, Redman T, Snell S (eds) The SAGE handbook of human resource management. SAGE Publications, London
8. Kriege L, Vlok PJ (2015) Human Resources Within ISO 55000—The Hidden Backbone to the Asset Management System. In: Amadi-Echendu J, Hoohlo C, Mathew J (eds) 9th WCEAM Research Papers. LNME. Springer, Cham, pp 435–446. https://doi.org/10.1007/978-3-319-15536-4_34
9. Blair MM (2011) An economic perspective on the notion of human capital. In: Burton-Jones A, Spender J-C (eds) The Oxford handbook of human capital. Oxford University Press, Oxford, pp 49–70
10. Amadi-Echendu J (2010) Behavioural preferences for engineering asset management. Def Concepts Scope Eng Asset Manage 1(1):347–355
11. Maheshwarkar M, Sohani N (2016) Evaluation of knowledge management levels based on multi criteria analysis. Manage Sci Eng 10(2): 5–23. https://doi.org/10.3968/8561. Canadian Research & Development Center of Sciences and Cultures. ISSN 1913–0341
12. Maletič D, Pačaiová H, Nagyová A, Gomišček B, Maletič M (2021) Framework development of an asset manager selection based on risk management and performance improvement competences. Safety 7(1):10

The Potential for Digital Twin Applications in Railway Infrastructure Management

Gerhardus Christiaan Doubell[1], Karel Kruger[1](✉), Anton Herman Basson[1], and Pieter Conradie[2]

[1] Department of Mechanical and Mechatronic Engineering, Stellenbosch University, Stellenbosch 7600, South Africa
kkruger@sun.ac.za

[2] Department of Industrial Engineering, Stellenbosch University, Stellenbosch 7600, South Africa

Abstract. The potential of digital twin technology has become apparent in recent literature, occurring evermore frequently in literature as the world moves on to the fourth industrial revolution. The use of digital twins in industries such as manufacturing, aerospace and aviation, and healthcare, have illustrated its value in lifecycle data management, control, monitoring, and more. This paper presents a review of digital twin applications in railway infrastructure. Considering digital twin adoption for public infrastructure, the rail industry is still at an early stage with few recorded implementations. However, digital twins present the possibility of addressing the emerging needs of infrastructure data management in the rail sector. Identified needs include the integration of data from various sources, validation of management paradigms, and the processing of large volumes of data.

1 Introduction

Digital twin is a concept that is at the forefront of development with the fourth industrial revolution [1, 2]. From inception, the concept of a digital twin has been utilized, among others, in the context of complex systems. Application domains such as smart manufacturing, precision medicine and aerospace have sparked great interest. However, the application of digital twins in complex railway systems and, in particular, railway infrastructure management, is yet to be sufficiently explored.

From the onset, it should be noted that a universally agreed-upon definition of the concept of a digital twin is yet to be established [3]. However, in the context of railway infrastructure (which can be seen as part of the built environment), a digital twin can generally be understood to be a digital representation of a physical entity that integrates different static and dynamic data, such as that acquired from building information modelling (BIM) and condition monitoring systems [4].

While the use of digital twin technology has been rapidly expanding in industries such as manufacturing [5], the use of digital twin technology in the railway industry is still in its infancy [4]. As such, the contextual needs that can be addressed, and the value to be gained from digital twins, might not be clear yet. This paper aims to provide

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 241–249, 2022.
https://doi.org/10.1007/978-3-030-96794-9_22

insight into the potential for digital twin applications in the management of railway infrastructure assets.

The paper will provide a brief introduction to the concept and potential benefits of digital twins in Sect. 2. The use of digital twins in rail infrastructure, as reported in literature to date, is then explored in Sect. 3. Thereafter, the current needs of asset management in the context of railway infrastructure is discussed in Sect. 4. Finally, the paper concludes in Sect. 5 with ideas for future research.

2 Digital Twin Overview

The concept of the digital twin was first introduced in the context of product lifecycle management (PLM) [6]. The proposed model presented most of the major, universally agreed-upon elements that would come to be associated with a digital twin and is illustrated in Fig. 1, below.

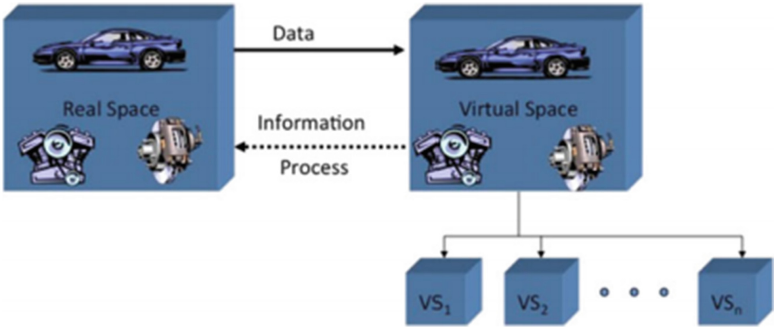


Fig. 1. The conceptual ideal for PLM - a front-runner for the digital twin concept [6]

The concept was introduced as having a "real space", representing the physical entity or "twin"; a "virtual space", containing a virtual (digital) representation of the physical entity in the "real space", as well as data and information flows between the real and virtual spaces. The model also made provision for information flow from the virtual space to "virtual sub-spaces".

In 2003, Främling et al. [7] built on this idea, presenting "an agent-based architecture where each product item has a corresponding virtual counterpart or agent associated with it". The authors argued that this approach presented a means to ease product lifecycle data management, providing a more efficient means to accommodate product-specific information changes over the entire lifecycle.

NASA researchers, in their roadmap [8], presented the concept as a digital twin – applying it to astronautics and aerospace to provide a means of remote inspection and scenario investigation of launched satellites. Digital twins have since been widely researched, with particular emphasis on manufacturing [9], aviation [10–12], and healthcare [13, 14].

After NASA's roadmap publications on digital twins in industrial operations reached more than 500 publications by 2019 [1]. It is thus clear that the concept has captured the attention of researchers, with "digital twin" service offerings also emerging in commercial solutions from large companies, such as General Electric, Siemens, Ansys, and AspenTech.

The possibilities and advantages offered by digital twin use are broad – with different use cases and application scenarios utilising different characteristics of digital twins to their advantage. The example of NASA leveraging digital twins to perform remote inspection and scenario investigation, contrasted to using them for effective PLM, attests to this. Other advantages include the use of digital twins to enable individualized structural management plans [10, 11], prognostics and health monitoring (PHM) [15], as well as using digital twins as virtual sensors, to analyse and predict the structural integrity of aircraft [10–12].

3 Digital Twins in Railway Infrastructure

This section reviews the published literature on the utilization of digital twin technology in the management of railway infrastructure. Railway infrastructure can be subdivided into five sectors: track (also referred to as the "perway" or "permanent way"); civil structures (such as bridges, culverts and tunnels); electrical; telecommunications; and signalling systems (or "train authorisation systems"). Research on the application of digital twin technology to one or more of these sectors existing in railway infrastructure are considered in this section.

3.1 Research Towards Implementation

The application of digital twin technology in rail infrastructure is still in its infancy [4], with only a few publications on the subject to date. In these publications, the research on digital twin implementation has focused on how to efficiently produce a digital twin of the rail infrastructure geometry [17–20]. The largest body of work outside of this has been in the civil structures infrastructure sector, with research focusing on digital twins for structural health monitoring (SHM) of railway bridges [16, 21].

Other publications include a description of a conceptual model for the digital modelling of infrastructure facilities and rolling stock [22], a discussion on infrastructure maintenance from the viewpoint of life-cycle management through cyber-physical systems (which form the necessary foundation for implementing digital twins [23]), and an approach to implement a digital twin for the control and monitoring of Electric Railway Power Systems (ERPSs) [24]. Further arguments have been made for the value of implementing digital twins in rail infrastructure by looking at its value in performing PHM, the context of smart cities, and the potential services to be offered [15, 25–27]. Finally, the testing of signaling systems prior to commissioning and integration into the existing network is also proposed [28].

Geometrical Digital Twins

A general proposition in digital twin development is to establish the digital twin of the physical component and perform most of its development in the design phase of the project lifecycle [2, 15, 29]. However, since the overwhelming majority of rail infrastructure is already established this would, in most cases, not be possible.

As a proposed first step in generating digital twins for established rail infrastructure, Ariyachandra, *et al.* [17–20] investigated the challenges to create digital twins from point cloud data of existing infrastructure. However, in this case the digital twin only presents the geometric model of the existing railway elements and should thus rather be seen as a building block to digital twin implementations (as defined in Sect. 1). Ariyachandra, *et al.* initially found that the average time required to generate a geometric digital twin from point cloud data with software such as Autodesk Revit 2016, to be about 10 times more than the time required to obtain the point could data itself [17]. After a series of research projects, a methodology was presented that utilized the fairly unique (and predictable) characteristics of railway topology. By utilizing their methodology, the time taken to generate geometry-only digital twins of railway elements with no prior information, should now be 98.6% faster than manually generating the model [19].

Railway Bridge Structural Health Monitoring
Digital twin implementations for railway bridges have taken the form a unified data structure encapsulating data from multiple data sources, and integrating that with multiple simulation models to provide more confident predictions of the general state of the bridge [16, 21]. Both publications on the topic have used arrays of fiber optic strain sensors to identify anomalies in the data (which might result from for example faulty sensors, or anomalous behavior), as well as to analyze the strain distribution of the bridge over time.

In structural health monitoring, there are generally two approaches for processing and interpreting bridge monitoring data: a physics-based approach and a data-driven approach. [16]. These two approaches can be combined in what is called a Data-Centric Engineering (DCE) approach, where the advantages and capabilities of both are exploited – thereby adding greater value to bridge SHM data [30]. Ye *et al.* considers this integration to be key in the successful development of digital twins for bridges [16].

In the research done by Ye *et al.* the benefits of using a digital twin for a railway bridge was concluded to be the efficient query of relevant data, integrated capabilities of data processing and interpretation, and the provision of a collaborative environment for various stages of a bridge project.

Febrianto, Butler, Girolami, and Cirak implemented such a digital twin, using an approach based on the Statistical Finite Element Method, which "allows one to make predictions about the true system behaviour in light of measurement data" [16]. This approach leads to improved long-term SHM. It provides reasonable strain distribution predictions for locations with no measurement data, enabling more reliable "what-if" analyses for loading scenarios at these locations. Having a digital twin of the bridge in this form also enables optimization of sensor placement, thereby reducing costs.

Other Literature
It is interesting to note that by focusing more on the services that can be gained from digital twin technology, Boschert, Heinrich and Rosen [26] use the example of a point machine to illustrate the benefits of using a digital twin. Though still conceptual, they present fairly specific implementation ideas, such as the use of similarity search algorithms to filter the data contained in each digital twin. This idea enables different services to be performed with the same (fixed) digital twin configuration.

3.2 Digital Twin Implementations for Rail Infrastructure by Industry

The term digital twin appears on the websites of many major companies involved with digitisation. The ambiguity of the exact definition of a digital twin has led to companies advertising different custom service deliveries that vary slightly in their implementation and value propositions.

One of the more recent applications in the rail industry is Siemens Mobility's project to establish a "Signalling simulation centre" for the Singapore Downtown line. The signalling centre will provide a digital twin of the entire Downtown line's signalling system. A core value proposition of the applications is that it will enable testing and integration of new features, signalling system functionalities, and key hardware interfaces before deployment – digitally simulating these operations to avoid service disruptions. This means that system enhancements, troubleshooting, vulnerability checks, system patching, as well as the testing of new software releases can be done remotely and digitally, without impacting operations or passenger service [31]. It is interesting to note that this idea strongly correlates with the recent publication on signalling systems [28] mentioned in the previous section.

This simulation functionality of the digital twin will also increase management teams' capability for training their technical staff. Furthermore, having the signalling system represented as a digital twin presents the added advantage of enabling rapid, in-depth, and technical analysis on signalling-related incidents [31].

Another example of digital twin implementation in rail industry is AlmavivA's massive digital twin project in Italy. The project entails the creation of a digital twin for the infrastructure of the entire Italian rail network, which includes more than 16,000 km of track. The digital twin of the network will contain component information obtainable through GPS and camera sensors. The digital twin will be able to provide "back-and-forth feedback between the twin and its real-world counterpart" [32].

From these two industrial implementations, it is clear how different the implementations of a "digital twin" can be. The AlmavivA implementation presents a greater focus on the visual representation, which is in contrast to the Siemens implementation that focus more on simulation.

4 Analysis of the Needs of Railway Infrastructure Data Management

From previous sections it is clear that digital twins have great potential to add value to the railway infrastructure management application domain. This section presents three of the major needs currently encountered in the railway infrastructure domain that can be addressed through the implementation of digital twins. These needs have been identified through a review of literature and discussions with engineers in management positions at the Passenger Railway Agency of South Africa (PRASA).

4.1 Integration of Information

With the world moving into the fourth industrial revolution, industries often face the same challenge of legacy information systems having been custom-made to meet specific user and technical requirements. This has led to information now existing in silos – often within specific software applications, enterprise divisions or engineering disciplines – leaving information systems functionally isolated from each another [33]. These systems often develop organically over several years – with custom applications being developed to meet changing business requirements and priorities.

Managing these isolated and disparate information systems generate "a suboptimal foundation needed to support a modern digital environment" [4]. It is much preferred to have data shared and integrated so as to support optimal data-based decision making, with as little effort as possible. Having most data in information silos greatly increase the effort of data integration.

Other than having information shared, it is also important for the data being used for decision making to be dynamic and accurate. This ensures that the decisions remain relevant, useful and informed by the most recent information.

Instead of grouping information according to application, division or discipline, digital twins can integrate data according to the physical entity represented [2]. Ye *et al.* actually specifies a common data environment to be a key feature of a digital twin [16]. This allows data to be used in a much more flexible way, with use cases that might not have been envisioned at the time of development still being feasible as a result of the ease of data access and integration being offered by digital twins.

4.2 Validation of Existing Maintenance Paradigms

Optimal maintenance paradigms can have a dramatic effect on organizational expenditure. A good example of the difference that appropriate maintenance strategy can have on costs, is Van der Westhuizen and Gräbe's research that conclude that organizations can save up to 35% in maintenance costs, and increase their capacity by approximately the same percentage, if condition-based and condition performance-based maintenance is applied to railway infrastructure [34].

Successful validation of existing maintenance paradigms require data from the infrastructure and thus ties into the first need discussed. Digital twins can thus also assist with the validation of maintenance paradigms for the same reasons that it is appropriate for the integration of data. However, digital twins also often incorporate a simulation component [16], which enables the digital twin of a certain part of the railway infrastructure to predict its own infrastructural degradation patterns based on historic and real-time data. This paves the way for predictive maintenance and other exciting data analytics that could be used to verify or improve existing maintenance paradigms.

Prognostics and health management (PHM), considered by some to be the "optimal maintenance plan" [15], would benefit greatly from digital twin implementation. Xiaodong, *et al.* [15] considers digital twins to be an essential enabling technology for PHM, fitting perfectly within the complex nature of railway systems.

The behavior of each individual asset for which a digital twin exists should also be able to be separately monitored (assuming the asset provides sensor data, or the digital twin at least possesses an accurate model to provide it with information of that asset). This detailed and customized view of the infrastructure can lead to greater insight and inform decisions regarding maintenance that might otherwise only be possible through laborious manual data processing.

4.3 Processing Large Volumes of Data

Railway infrastructure, when equipped with sensors, may easily generate vast amounts of data. Processing this data can be an extremely laborious and time-consuming process that is not only inefficient, but also hinders the utilization of the data. The automation of data transfer, processing and storage from the physical asset to the digital storage space is recommended and can be effectively supported by digital twin implementations.

5 Conclusion and Future Work

This paper presented an overview of the potential of digital twin technology for railway infrastructure data management. The still slow uptake of digital twin technology in the railway infrastructure domain was discussed. Finally, a brief perspective is given on what the current needs of railway infrastructure management are, together with how digital twins might have a role to play in meeting them. Future research work will delve into an actual implementation of digital twin technology for railway infrastructure – gaining more insight into the opportunities and challenges presented by a case study implementation.

Acknowledgments. This paper was developed with support from the MAD Research Group, as well as the PRASA Research Chair at Stellenbosch University. The authors gratefully acknowledge the contribution and expertise of the members involved in both research groups.

References

1. Melesse TY, Di Pasquale V, Riemma S (2020) Digital twin models in industrial operations: A systematic literature review. Procedia Manuf. 42(2019):267–272. https://doi.org/10.1016/j.promfg.2020.02.084
2. Fuller A, Fan Z, Day C, Barlow C (2020) Digital twin: enabling technologies, challenges and open research. IEEE Access 8:108952–108971. https://doi.org/10.1109/ACCESS.2020.2998358
3. van der Valk H, Haße H, Möller F, Arbter M, Henning JL, Otto B (2020) A taxonomy of digital twins. In: 26th Americas Conference on Information Systems (AMCIS), August 2020
4. Johnson A et al (2021) Informing the information requirements of a digital twin: a rail industry case study. In: Proceedings of the Institution of Civil Engineers - Smart Infrastructure and Construction, pp 1–13. https://doi.org/10.1680/jsmic.20.00017
5. Overton J, Brigham JC (2017) The Digital Twin. Data-driven simulations innovate the manufacturing process, White Paper. https://assets1.dxc.technology/analytics/downloads/DXC-Analytics-Digital-Twin.pdf
6. Grieves M, Vickers J (2017) Transdisciplinary Perspectives on Complex Systems. Springer, Heidelberg
7. Främling K, Holmström J, Ala-Risku T, Kärkkäinen M (2003) Product agents for handling information about physical objects. Rep. Lab. Inf. Process. Sci. Ser. B 153(3):1–20
8. Shafto M et al (2010) Modeling, simulation, information technology & processing roadmap. Natl Aeronaut Space Adm 11:27

9. Assad Neto A, Ribeiro Da Silva E, Deschamps F, Pinheiro De Lima E (2021) Digital twins in manufacturing: an assessment of key features. Procedia CIRP 97:178–183. https://doi.org/10.1016/J.PROCIR.2020.05.222
10. Tuegel EJ (2012) The airframe digital twin: Some challenges to realization. In: Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. https://doi.org/10.2514/6.2012-1812
11. Gockel BT, Tudor AW, Brandyberry MD, Penmetsa RC, Tuegel EJ (2012) Challenges with structural life forecasting using realistic mission profiles. In: Collection of Technical Papers - AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference. https://doi.org/10.2514/6.2012-1813
12. Tuegel EJ, Ingraffea AR, Eason TG, Spottswood SM (2011) Reengineering aircraft structural life prediction using a digital twin. Int J Aerosp Eng 2011:1–14. https://doi.org/10.1155/2011/154798
13. Chen Y, Yang O, Sampat C, Bhalode P, Ramachandran R, Ierapetritou M (2020) Digital twins in pharmaceutical and biopharmaceutical manufacturing: a literature review. Process 8(9):1088. https://doi.org/10.3390/PR8091088
14. Tadayoni R, Maier-Rabler U, Angus D Digital Twins in Healthcare
15. Xiaodong W, Feng L, Junhua R, Rongyu L (2020) A survey of digital twin technology for PHM. Adv Intell Syst Comput 1031:397–403. https://doi.org/10.1007/978-981-13-9406-5_48
16. Ye C et al (2019) A digital twin of bridges for structural health monitoring for proceedings of the 12. In: 12th International Workshop on Structural Health Monitoring
17. Ariyachandra MRMF, Brilakis I (2019) Understanding the challenge of digitally twinning the geometry of existing rail infrastructure. In: 12th International Research Conference of Faculty of Architecture Research Unit (FARU), no. December, pp 25–32. https://doi.org/10.17863/CAM.47494
18. Ariyachandra MRMF, Brilakis I (2020) Detection of railway masts in airborne LiDAR data. J Constr Eng Manage 146(9):04020105. https://doi.org/10.1061/(asce)co.1943-7862.0001894
19. Ariyachandra MRMF, Brilakis I (2021) Application of railway topology for the automated generation of geometric digital twins of railway masts. In: Proceedings of the 13th European Conference on Product & Process Modelling, vol 2040
20. Ariyachandra MRMF, Samarakkody A, Perera BAKS (2019) Real-virtual synchronisation: a review on the state-of-the-art geometric digital twinning of infrastructure, no. November, pp 502–512. https://doi.org/10.31705/wcs.2019.50
21. Febrianto E, Butler L, Girolami M, Cirak F (2021) A Self-Sensing Digital Twin of a Railway Bridge using the Statistical Finite Element Method, pp 1–15
22. Efanov D, Shilenko AS, Khoroshev VV (2020) Digital modeling in railway infrastructure and rolling stock objects at all stages life cycle: features. In: International Russian Automation Conference (RusAutoCon), pp 29–34, Sep 2020. https://doi.org/10.1109/RUSAUTOCON49822.2020.9208088
23. Redelinghuys AJH, Kruger K, Basson AH (2020) An Architecture for the Digital Twin of a Manufacturing Cell, no. March, p 210
24. Ahmadi M, Kaleybar HJ, Brenna M, Castelli-Dezza F, Carmeli MS (2021) Adapting digital twin technology in electric railway power systems. In: 12th Power Electronics, Drive Systems, and Technologies Conference PEDSTC 2021. https://doi.org/10.1109/PEDSTC52094.2021.9405876
25. Tchana Y, Ducellier G, Remy S (2019) Designing a unique digital twin for linear infrastructures lifecycle management. Procedia CIRP 84:545–549. https://doi.org/10.1016/j.procir.2019.04.176
26. Boschert S, Heinrich C, Rosen R (May 2018) Next generation digital twin. In: Implementation, application and utilization of smart systems

27. Zheng H (2021) Research and Analysis on the Application of Digital Twin Technology in Urban Rail Transit, pp. 1067–1070. https://doi.org/10.1109/ipec51340.2021.9421186
28. Kochan A (2020) Digital twin of the ECTS application. WUT J Trans Eng 131:87–98. https://doi.org/10.5604/01.3001.0014.8199
29. Barricelli BR, Casiraghi E, Fogli D (2019) A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access 7(Ml):167653–167671. https://doi.org/10.1109/ACCESS.2019.2953499
30. Lau FDH, Adams NM, Girolami MA, Butler LJ, Elshafie MZEB (2018) The role of statistics in data-centric engineering. Stat Probab Lett 136:58–62. https://doi.org/10.1016/J.SPL.2018.02.035
31. Siemens Mobility GmbH (2019) Siemens Mobility to set up Singapore's Downtown Line signalling simulation center. Press Company, Siemens, 14 Oct 2019. https://press.siemens.com/global/en/pressrelease/siemens-mobility-set-singapores-downtown-line-signalling-simulation-center. Accessed 16 Feb 2021
32. Mangan T (2021) How Digital Twin Technology is Helping Build a Smart Railway System in Italy, 20 May 2020. https://www.nutanix.com/theforecastbynutanix/industry/how-digital-twin-technology-is-helping-build-a-smart-railway-system-in-italy. Accessed 16 Feb 2021
33. Moore WJ, Starr AG (2006) An intelligent maintenance system for continuous cost-based prioritisation of maintenance activities. Comput Ind 57(6):595–606. https://doi.org/10.1016/J.COMPIND.2006.02.008
34. Van Der Westhuizen NJ, Gräbe PJ (2013) The integration of railway asset management information to ensure maintenance effectiveness. J South Afr Inst Civ Eng 55(3):18–29

Digital Twins in Asset Management: Potential Application Use Cases in Rail and Road Infrastructures

João Vieira[1,2(✉)], João Clara[1], Hugo Patrício[2], Nuno Almeida[1], and João Poças Martins[3]

[1] Instituto Superior Técnico, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
{joao.cardoso.vieira,joaoclara, nunomarquesalmeida}@tecnico.ulisboa.pt
[2] Infraestruturas de Portugal, S.A., Praça da Portagem, 2809-013 Almada, Portugal
hugo.patricio@infraestruturasdeportugal.pt
[3] CONSTRUCT-GEQUALTEC – Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, 4200-465 Porto, Portugal
jppm@gcloud.fe.up.pt

Abstract. Asset management is data-intensive and new tools and processes are often necessary to collect, manage, analyse, and use asset data. The use of these tools can improve organisational knowledge and decision-making. Industry 4.0 tools are prompting the digital transformation of organisations and emerging innovation opportunities. Among these tools are those supporting the concept of Digital Twin (DT). The concept started being mentioned a few decades ago, but the discussion around its definition and potential applications still continues. This paper explores some of the interpretations of the concept of DT and its interrelation with some Industry 4.0 tools. Besides presenting some of the known benefits and opportunities related to DT applications in specific industries such as aerospace and manufacturing, namely in the early stage of asset lifecycle, the paper seeks to emphasise the vast exploratory potential of DT use in infrastructure asset management, especially in the operation and maintenance (O&M) phases. This presentation includes the description of an exploratory project for DT implementation in rail and road networks by the largest infrastructure management body in Portugal.

1 Introduction

Asset management has been the focus of studies in both the scientific and industrial environments [1]. The research community has been aiming at establishing and advancing its body of knowledge and enhance levels of service, improve lifecycle cost modelling and risk management, develop investment decision strategies, among others [2].

It is worth mentioning some important challenges related with significant backlogs of insufficient expenditure leading to under-performing infrastructures [3]. Additionally, the financial resources of organisations are limited and are seldomly readily available, which leads to constant competition and the need to prioritise asset-related investments [2]. Thus, asset managers must continuously analyse these restrictions and take informed decisions for each asset or asset system to extract the best global value [4]. To support their decisions, asset managers need appropriate and timely asset information [5].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 250–260, 2022.
https://doi.org/10.1007/978-3-030-96794-9_23

The emergence of new digital technologies that are components of Industry 4.0, enables new approaches to deal with these complex challenges [6] and is promoting industrial digital transformation [7].

This paper aims to explore the potential of application of one of these trending approaches, namely Digital Twin (DT). It includes the description of the early stages of an exploratory project for DT implementation in road and railway networks managed by *Infraestruturas de Portugal, S.A.* (IP). IP is the largest public infrastructure management body in Portugal and a member of the Shift2Rail project In2Track-3, which further explores DT applications in asset classes such as tunnels, bridges, switches and crossings and rail track. In the road sector. the World Road Association (PIARC) undertook a global research project [8] about the different approaches used world-wide to collect, process and use road related data, focusing on the state-of-the-art methods which can deliver efficiencies, facilitate the operation and management of the road network, and improve the provision of services to road users. The development and implementation of data related initiatives such as DT was recommended to be further explored.

2 Background Knowledge on Digital Twins

Over the past few years, the DT trend has been gaining momentum [9] as a trend amongst other Industry 4.0 tools [10], that before were not available for feasible use [11]. This has led to an increasing number of DT applications since then [12].

According to Deloitte [14], the deployment of DT capabilities has accelerated due to a list of factors:

- Simulation: tools for building DT are getting more powerful and sophisticated, and machine learning (ML) is enhancing the value of insights;
- New data sources: data from real-time asset monitoring technologies (such as LIDAR) are incorporated into simulations. IoT sensors embedded in assets or processes can provide operational data to simulations, enabling real-time monitoring;
- Interoperability: improved industry standards for communications between IoT sensors, operational hardware, and efforts to integrate multiple platforms;
- Visualisation: Advanced data visualisation allows to overcome analysis problems associated with the sheer volume of data, by filtering and extracting information in real-time. The most recent tools go beyond basic dashboards and standard visualisation capabilities to include 3D or Virtual and Augmented Reality;
- Instrumentation: IoT sensors, both embedded and external, are becoming smaller, more accurate, cheaper, and more powerful;
- Platform: Increased access to powerful and less expensive computing power, network, and storage (cloud-based platforms, IoT, and analytics capabilities).

Nevertheless, DT has been interpreted in many different ways. Based on the state-of-the-art research work developed by some authors [10, 13, 15] and some additional DT definition proposals [1, 10, 16–25] it is possible to conclude that there is still no harmonised definition. DT has been attracting increased interest in multiple sectors of activity, leading to distinct definitions depending on the context in which the term is used [1, 10, 15, 26].

Kritzinger et al. [26] propose the distinction of three different types of DT, depending on the level of data integration (Fig. 1):

- Digital Model, a digital representation of a physical object that does not use automated data exchange between the physical and the digital objects. An instant change in the state of the physical object does not impact the digital object and vice-versa;
- Digital Shadow, a digital model with one-way automated data flow between the physical object and the digital object. A change in the object state leads to a change in the state of the digital object, but not vice-versa;
- Digital Twin, a model where data flows between the physical and the virtual objects are fully automated, even allowing the digital object to control the operation of the physical object. A change in the state of any of the objects impacts the state of the other.

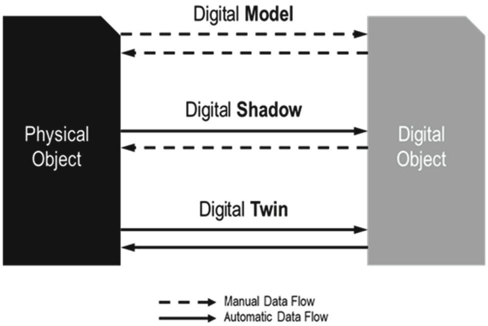


Fig. 1. Data flows in digital models, shadows and twins (adapted from [26])

According to Kritzinger et al. [26], most of the existing literature around the subject of DT actually refers to Digital Models and Shadows, whilst true DT literature is still scarce, which reveals that DT development is still in an early stage. The literature review developed by Liu et al. [27] concludes that over half of the studies described digital models or digital shadows, although the authors claimed to have studied DT's. Negri et al. [10] state that the polarisation around DT concepts suggests that DT scientific literature is still in its infancy.

Tchana et al. [7] suggest that, although DT and Digital Model seem to have the same definition, they are distinct concepts. Woods and Freas [11] inform that a DT is often seen as a synonym of a 3D or 4D representation of an asset or system. Throughout the literature, it is also frequent to find close connections between DT and BIM. Kaewunruen and Lian [28] state that BIM is even sometimes described as a DT. For example, SNCF [29] defines DT as a 3D model or a BIM system.

Although BIM can manage digital information in 3D [30], and be considered an asset information model and a digital model of a physical asset or system, it does not fulfil all the requirements to be considered a true DT as defined in the literature. To be considered so, according to the most consensual definitions, BIM not only needs to include all the necessary data about the real asset or asset system but also the connection to the physical twin, improving decision-making by analysing data from the physical asset and providing feedback to it [7, 23].

3 DT and Decision-Making in Asset Management

Asset management focuses on balancing costs, opportunities, and risks against the desired performance of assets to achieve organisational objectives. Asset management does not focus on the asset itself, but on the value (tangible or intangible, financial or non-financial) that the asset can provide to the organisation according to the stakeholders and the organisational objectives [31].

Asset management emphasises the importance of risk-based decisions taken over the asset lifecycle and the critical role of information in supporting those decisions [32]. Because asset management is data-intensive and good decisions require good quality information, asset management needs adequate tools and processes to collect, assemble, manage, analyse, and use asset data. The use of these tools can improve organisational knowledge and decision-making [31].

DT is expected to enhance asset information systems, contributing to unlock value in asset management decision-making [1, 23, 27]. This can be achieved through the data coming from the physical asset, creating an opportunity for feedback into the physical twin [23]. There is a wide variety of decisions supported by DT's in different asset lifecycle phases and at different asset levels [1]. Liu et al. [27] summarise DT industrial applications in each lifecycle stage (Fig. 2).

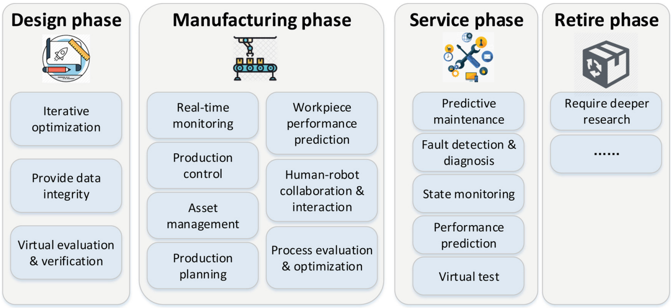


Fig. 2. Industrial DT applications according to each lifecycle stage [27]

According to Macchi et al. [1], the benefits of DT are consistent with the ones discussed by the literature: i) system lifecycle mirror, by predicting performances and long-term behaviour of assets, and by ensuring data digital continuity along the lifecycle phases of the system; ii) improved O&M decision-making [1].

As previously discussed, asset management faces the challenge of managing asset information throughout the lifecycle, in a systematic and structured way. However, Heaton et al. [33] state that most organisations still have manual and *ad-hoc* approaches to this activity, and there is still an insufficient definition of asset management requirements. On the other hand, BIM has been widely recognised as a key enabler for the development of information management processes [33], which could provide a structured and normative framework (ISO 19650) to the development of DT, while there is an absence of official standards.

4 Potential Applications of DT in Road and Rail Infrastructures

Road and rail networks have proved to be vital to the development of societies and they continue to have a crucial role on society and economy. It is estimated that, within the EU, road and rail transportation account for more than 63% of goods transport and almost 90% of passenger transport [34]. These infrastructures ensure the daily mobility of people and goods while seeking to provide an efficient and safe service [35]. Therefore, it is essential that road and rail infrastructure managers have access to the best possible data regarding those assets, so they can make the best decisions during the lifecycle.

The development of Industry 4.0 tools and the DT approach uncover new opportunities for improving data-driven decision-making. Although most DT literature focuses on other sectors and frameworks for DT application, a few use cases are being explored for applying DT in road and rail infrastructures (Table 1).

Most DT applications relate to the O&M phases, and the opportunities of application (e.g., asset condition monitoring, maintenance planning, fault prediction, and scenario simulation) are consistent with those discussed previously. Hence, it is possible to conclude that the application of DT in road and rail infrastructures is still in a very early stage and that more application use cases will emerge in the future as the technology matures and diffusion of innovation occurs [36, 37].

The use cases of DT in road and rail infrastructures are at an early stage of exploitation in *Infraestruturas de Portugal, S.A.* (IP). IP is the biggest transport infrastructure manager in Portugal and has a certified ISO 55001 Asset Management System. IP is a public body and is responsible for the construction, operation, and maintenance of road and railway networks. This organisation manages more than 15.000 km of national roads and 3.000 km of railway networks [38], distributed by multiple asset groups (road pavements, rail track, switches and crossings, power equipment, engineering structures, etc.). These circumstances mean that IP has a vast network of stakeholders, with competing interests [39], while seeking to provide a multimodal, efficient, safe, and sustainable transportation service [38].

IP identifies the power of information and technological innovation as two of the five major trends in the transportation sector. Within these subjects, IP highlights subtopics such as Big Data, AI, ML, automation, connectivity, and new infrastructure monitoring processes. The use of sensors in vehicles and infrastructures allows real-time monitoring of assets, namely local weather conditions, asset physical condition, and traffic. The large-scale management of asset information will need the support of complementary techniques, such as ML and AI, to enhance the efficiency of infrastructure management [38, 40].

The simultaneous management of both rail and road networks generates opportunities but also unique challenges. For example, while in road networks the demand is ruled by the intentions of each vehicle user, in rail networks the demand follows operational contracts and route schedules, established with each rail transportation company [39]. On the other hand, the condition of rail and road infrastructures also has distinct impacts on the operations of each network. Thus, the coordination of operations is particularly relevant in asset management for achieving the best organisational outcome. This coordination may be enhanced with DT applications. Certain DT applications can use data coming from operations and process it to support asset-management decision-making (e.g., train sensors providing feedback on rail track condition) or vice-versa (e.g., track sensors providing condition-related data to adjust travel speed).

Table 1. Overview of DT application examples in rail and road infrastructures

Network	Asset group	Challenges	Opportunities for DT	Reference
Rail	Switches and Crossings	Faults cause of delays; safety issues; significant proportion of maintenance and renewal	Failure prediction	[40–44]
	Track	Defects and welding points are discontinuities in railway track	Detect and monitor the structural health of rail track discontinuities	[45, 46]
		Differential stiffness in track foundation and transition zones increase track deformations	Monitoring track loads, long term settlements and geometric deformations according to safety limits	[40]
	Railway Stations	Operation efficiency and level of service	Operational risks and emergency response, customer service (comfort, train information, etc.)	[47]
	Power systems	High complexity	Energy management, power flow analysis, power quality monitoring, fault diagnostic and maintenance, condition monitoring, operating profile optimisation	[48]
Road	Pavement	Condition monitoring, planning of maintenance, and renewal	Pavement performance prediction	[49, 50]
Rail/Road	Bridges	Condition monitoring, planning of maintenance and renewal; inspections are resource and time-consuming; safety	Structural health monitoring and prediction, enhance predictive maintenance, manage inspection planning	[40, 51]

(*continued*)

Table 1. (*continued*)

Network	Asset group	Challenges	Opportunities for DT	Reference
	Tunnels	Condition monitoring, planning of maintenance and renewal; energy use and safety	Monitoring of energy use and system health; maintenance scheduling and event simulation	[52]

IP identifies the power of information and technological innovation as two of the five major trends in the transportation sector. Within these subjects, IP highlights subtopics such as Big Data, AI, ML, automation, connectivity, and new infrastructure monitoring processes. The use of sensors in vehicles and infrastructures allows real-time monitoring of assets, namely local weather conditions, asset physical condition, and traffic. The large-scale management of asset information will need the support of complementary techniques, such as ML and AI, to enhance the efficiency of infrastructure management [38, 53].

The simultaneous management of both rail and road networks generates opportunities but also unique challenges. For example, while in road networks the demand is ruled by the intentions of each vehicle user, in rail networks the demand follows operational contracts and route schedules, established with each rail transportation company [39]. On the other hand, the condition of rail and road infrastructures also has distinct impacts on the operations of each network. Thus, the coordination of operations is particularly relevant in asset management for achieving the best organisational outcome. This coordination may be enhanced with DT applications. Certain DT applications can use data coming from operations and process it to support asset-management decision-making (e.g., train sensors providing feedback on rail track condition) or vice-versa (e.g., track sensors providing condition-related data to adjust travel speed).

IP manages a significant variety of asset groups, especially in the case of rail infrastructures. The rail network comprises linear (e.g., rail track, catenary) and vertical assets (e.g., substations), which require a wide range of expertise areas. Moreover, IP dedicates significant effort to inspection activities. In the road network, annual routine inspections correspond to about 50% of the total network length [38]. This suggests that there is potential for inspection digitisation, allowing IP to increase its efficiency and knowledge, by inspecting a higher number of assets per year. The application of DT could also help inspection teams to schedule and prioritise their activities according to real-time data coming from the assets. For example, monitoring the physical condition of a bridge with a DT could help to identify the most critical points (e.g., condition, evolution over time, etc.) and suggest changes in inspection time planning.

DT application can ultimately improve maintenance and renewal decisions [1], as more data about the asset performance and condition is available. On a broader scale, this could enhance investment and resource planning, by predicting the optimal time for interventions in advance and with more accuracy.

Despite most of the discussed benefits of DT use are associated with the O&M phases, DT can also be potentially valuable during the initial lifecycle stages. For instance, during planning or construction of a rail or road project, a DT can assist in project visualisation, prediction of changes arising from project execution, and risk prediction brought by non-conformant execution. During project commissioning DT can also optimise site testing and accelerate project validation and certification, reducing "time to market" and the impact on operations [53].

However, investing in infrastructure digitisation and connectivity also brings relevant cybersecurity challenges to infrastructure organisations, who often deal with sensitive information. Therefore, an organisation such as IP must guarantee system redundancy, vigilance, and backup measures [54].

IP has a vast potential for DT application, not only at the asset level but also at higher integration levels (asset group, system, or network). The ongoing R&D projects at IP are exploring the DT application use cases that derive most benefits for infrastructure asset management.

5 Conclusions

The emergence of new digital technologies that are components of Industry 4.0 enables new approaches to deal with complex asset management challenges. These tools support the industrial digital transformation. This paper presents a discussion on the definition of DT and an overview of potential applications for road and rail infrastructures. Although the literature shows that DT enhances asset information systems, contributing to unlock value in asset management decision-making, the most relevant experience comes mainly from specific industries such as aerospace and manufacturing, with an emphasis at the early stages of assets lifecycle. Some studies discuss the potential application of DT to the operation, maintenance, and decommissioning phases, but experience with regards to this is still at an early stage. Regarding the use of DT in rail and road infrastructures, the authors foresee a vast application potential. Few use cases were found concerning DT application on the main asset groups of road and rail networks but ongoing R&D projects aim to fill that knowledge gap and enhance asset management capabilities.

Acknowledgments. This work was supported by *Infraestruturas de Portugal, S.A.* and the Shift2Rail Joint Undertaking funded by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101012456 (In2track3).

References

1. Macchi M, Roda I, Negri E, Fumagalli L (2018) Exploring the role of digital twin for asset lifecycle management. IFAC-PapersOnLine 51:790–795. https://doi.org/10.1016/j.ifacol.2018.08.415
2. de Almeida NM, Vieira J, Silva JG, e Castro C (2021) The impact of asset management development programs in infrastructure organizations. In: Rodrigues H, Gaspar F, Fernandes P, Mateus A (eds) Sustainability and Automation in Smart Constructions. ASTI. Springer, Cham, pp 247–258. https://doi.org/10.1007/978-3-030-35533-3_29

3. ASCE (2009) Report card for America's infrastructure. American Society of Civil Engineers, Reston, VA
4. Trindade M, Almeida N, Finger M, Ferreira D (2019) Design and development of a value-based decision making process for asset intensive organizations. In: Mathew J, Lim CW, Ma L, Sands D, Cholette ME, Borghesani P (eds) Asset Intelligence through Integration and Interoperability and Contemporary Vibration Engineering Technologies. LNME. Springer, Cham, pp 605–623. https://doi.org/10.1007/978-3-319-95711-1_60
5. ISO (2018) ISO 19650-1:2018 - Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM) - Information management using building information modelling
6. Ozturk GB (2021) Digital Twin Research in the AECO-FM Industry. J Build Eng 40(2021):102730. https://doi.org/10.1016/j.jobe.2021.102730
7. Tchana Y, Ducellier G, Remy S (2019) Designing a unique digital twin for linear infrastructures lifecycle management. Procedia CIRP 84:545–549. ISSN 2212-8271. https://doi.org/10.1016/j.procir.2019.04.176
8. PIARC (2020) Road related data and how to use it. World Road Association. ISBN: 978-2-84060-627-7
9. Deloitte (2021) Tech Trends 2021. Deloitte Insights
10. Negri E, Fumagalli L, Macchi M (2017) A Review of the roles of digital twin in CPS-based production systems. Procedia Manuf 11:939–948. https://doi.org/10.1016/j.promfg.2017.07.198
11. Woods E, Freas B (2019) Creating Zero Carbon Communities: The Role of Digital Twins. Navigant Research, Navigant Consulting, Inc.
12. Tao F, Zhang H, Liu A, Nee YC (2019) Digital twin in industry: state-of-the-art. IEEE Trans Ind Inf 15(4):2405–2415. https://doi.org/10.1109/TII.2018.2873186
13. Tao F, Cheng J, Qi Q, Zhang M, Zhang H, Sui F (2017) Digital twin-driven product design, manufacturing and service with big data. Int J Adv Manuf Technol 94(9–12):3563–3576. https://doi.org/10.1007/s00170-017-0233-1
14. Deloitte (2020) Tech Trends 2020. Deloitte Insights
15. Lamb K (2019) Principle-based digital twins: a scoping review. https://doi.org/10.17863/CAM.47094
16. Shafto M, Conroy M, Doyle R, Glaessgen E, Kemp C, LeMoigne J, Wang L (2010) Technology Area 11: Modeling, Simulation, Information Technology and Processing Roadmap. NASA Office of Chief Technologist
17. Tuegel EJ (2012) The airframe digital twin: some challenges to realization. In: Proceedings 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference. https://doi.org/10.2514/6.2012-1812
18. Lee J, Lapira E, Bagheri B, Kao HA (2013) Recent advances and trends in predictive manufacturing systems in big data environment. Manuf Lett 1:38–41. https://doi.org/10.1016/j.mfglet.2013.09.005
19. Reifsnider K, Majumdar P (2013) Multiphysics stimulated simulation digital twin methods for fleet management. In: 54th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conferencem vol 1578. https://doi.org/10.2514/6.2013-1578
20. Hochhalter J, Leser WP, Newman JA et al. (2014) Coupling damage-sensing particles to the digital twin concept. NASA Center for AeroSpace Information
21. Rosen R, Wichert G, Lo G, Bettenhausen KD (2015) About the importance of autonomy and digital twins for the future of manufacturing. IFAC-PapersOnLine 48:567–572. https://doi.org/10.1016/j.ifacol.2015.06.141
22. Gabor T, Belzner L, Kiermeier M, Beck MT, Neitz A (2016) A simulation-based architecture for smart cyber-physical systems. In: 2016 IEEE international conference on autonomic computing (ICAC). IEEE, pp 374–379. https://doi.org/10.1109/ICAC.2016.29

23. Bolton A, Enzer M, Schooling J et al. (2018) The Gemini Principles: Guiding values for the national digital twin and information management framework. Centre for Digital Built Britain and Digital Framework Task Group. https://doi.org/10.17863/CAM.32260
24. Deloite (2019) Digital capital projects: the capital project of the future. Deloitte
25. Vivi QL, Parlikad AK, Woodall P, Ranasinghe GD, Heaton J (2019) Developing a dynamic digital twin at a building Level: using Cambridge campus as case study. In: International conference on smart infrastructure and construction 2019 (ICSIC), pp 67–75. https://doi.org/10.1680/icsic.64669.067
26. Kritzinger W, Karner M, Traar G, Henjes J, Sihn W (2018) Digital twin in manufacturing: a categorical literature review and classification. IFAC-PapersOnLine 51(11):1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474
27. Liu M, Fang S, Dong H, Xu C (2021) Review of digital twin about concepts, technologies, and industrial applications. J Manuf Syst 58(Part B):346–361. https://doi.org/10.1016/j.jmsy.2020.06.017
28. Kaewunruen S, Lian Q (2019) Digital twin aided sustainability-based lifecycle management for railway turnout systems. J Cleaner Prod 228:1537–1551. ISSN 0959–6526, https://doi.org/10.1016/j.jclepro.2019.04.156
29. SNCF (2019) The digital twin. Société Nationale des Chemins de fer Français. Accessed 07 July 2021, https://www.sncf-reseau.com/en/entreprise/newsroom/sujet/the-digital-twin
30. EFCA (2019) BIM and ISO 19650 from a project management perspective - Booklet on ISO Standard 19650. European Federation of Engineering Consultancy Associations. ISBN: 9789075085075
31. ISO (2014) ISO 55000:2014 - Asset management — Overview, principles and terminology
32. UIC (2016) UIC Rail Application Guide – Practical implementation of Asset Management through ISO 55001. International Union of Rails, Asset Management Working Group. ISBN: 978-2-7461-2521-6
33. Heaton J, Parkilad AK, Schooling J (2019) Design and development of BIM models to support operations and maintenance. Comput Ind 111:172–186. https://doi.org/10.1016/j.compind.2019.08.001
34. EU (2020) Statistical Pocketbook 2020 – EU transport in figures. European Union. ISBN 978-92-76-17565-0
35. OECD (2020) Transport Bridging Divides. OECD Urban Studies, OECD Publishing, Paris. https://doi.org/10.1787/55ae1fd8-en
36. Schwabe O, et al (2020) A maturity model for rapid diffusion of innovation in high value manufacturing. Procedia CIRP 96:195–200. https://doi.org/10.1016/j.procir.2021.01.074
37. Schwabe O, Schneider L, de Almeida NM, Salvado AF (2021) A framework for accelerating innovation through innovation webs. In: Rodrigues H, Gaspar F, Fernandes P, Mateus A (eds) Sustainability and Automation in Smart Constructions. ASTI. Springer, Cham, pp 205–210. https://doi.org/10.1007/978-3-030-35533-3_23
38. IP (2020) Relatório e Contas 2020. Infraestruturas de Portugal
39. Patrício H, Almeia N (2021) A common risk framework for road and rail infrastructures. In: Crespo Márquez A, Komljenovic D, Amadi-Echendu J (eds) WCEAM 2019. LNME. Springer, Cham, pp 113–125. https://doi.org/10.1007/978-3-030-64228-0_11
40. EC (2020a) Grant Agreement number 101012456 - IN2TRACK3. European Comission
41. Pillai N, Shih J-Y, Roberts C (2021) Evaluation of numerical simulation approaches for simulating train-track interactions and predicting rail damage in railway switches and crossings (S&Cs). Infrastructures 6:63. https://doi.org/10.3390/infrastructures6050063
42. Vatn J (2018) Industry 4.0 and real-time synchronization of operation and maintenance. In: Haugen, et al (eds) Safety and Reliability – Safe Societies in a Changing World. Taylor & Francis Group, London. ISBN 978-0-8153-8682-7

43. Boschert S, Rosen R (2018) Digital Twin: a second life for engineering models. ERCIM News 115:8–9. ISSN 0926-4981
44. S-CODE (2019) D3.1 Report – Next generation control: monitoring and sensor systems. Switch and Crossing Optimal Design and Evaluation
45. Ramatlo D, Long C, Loveday P, Wilke D (2020) A modelling framework for simulation of ultrasonic guided wave-based inspection of welded rail tracks. Ultrasonics 108:106215. ISSN 0041–624X, https://doi.org/10.1016/j.ultras.2020.106215
46. EC (2018) Grant Agreement number 826255 - IN2TRACK2. European Comission
47. Li D, Yang X, Xu X (2020) A framework of smart railway passenger station based on digital twin. In: CICTP 2020, pp 2623–2634. https://doi.org/10.1061/9780784482933.226
48. Ahmadi M, Kaleybar HJ, Brenna M, Castelli-Dezza F, Carmeli MS (2021) Adapting digital twin technology in electric railway power systems. In: 12th power electronics, drive systems, and technologies conference (PEDSTC), pp 1–6. https://doi.org/10.1109/PEDSTC52094.2021.9405876
49. Yu G, Zhang S, Hu M, Wang YK (2020) Prediction of highway tunnel pavement performance based on digital twin and multiple time series stacking. Hindawi Adv Civil Eng 2020:1–21. https://doi.org/10.1155/2020/8824135
50. Meža S, Pranjić AM, Vezočnik R, Osmokrović I, Lenart S (2021) Digital twins and road construction using secondary raw materials. Hindawi J Adv Transp 2021:1–12. https://doi.org/10.1155/2021/8833058
51. Wenner M, Meyer-Westphal M, Herbrand M, Ullerich C (2021) The concept of digital twin to revolutionise infrastructure maintenance: the pilot project smartBRIDGE Hamburg. In: 27th ITS world congress, Hamburg, Germany, 11–15 October 2021
52. Tijs K (2020) Digital Tunnel Twin – Enriching the Maintenance & Operation of Dutch Tunnels. MSc. Thesis, TU Delft
53. EC (2020b) Draft proposal for a European Partnership under Horizon Europe - Transforming Europe's Rail System. European Commission
54. IP (2019) 50 Desafios de Investigação, Desenvolvimento e Inovação. Infraestruturas de Portugal

Asset Management Decisions

Proposition of a Generic Decision Framework for Prescriptive Maintenance

Pedro Dias Longhitano[1(✉)], Khaoula Tidriri[2], Christophe Bérenguer[2], and Benjamin Echard[1]

[1] Volvo Group Trucks Technology, 1 Henri German, Saint Priest, France
pedro.dias.longhitano@volvo.com
[2] GIPSA-lab, Univ. Grenoble Alpes CNRS Grenoble INP, 11 Rue des Mathématiques, Saint-Martin-d'Héres, France

Abstract. The digitalization of the economy in the past decades has made data availability grow and become more important. From the maintenance point of view, clients are more demanding, wanting systems that will not have breakdowns while reducing exploitation costs. This challenging scenario has pushed companies in the direction of more intelligent maintenance solutions that involve choosing the best course of action in terms of system availability. Nowadays, these solutions are usually called prescriptive maintenance. This term is vaguely defined and its use is often unjustified. In this article we will discuss what really characterizes prescriptive maintenance, review some of the work published with this term and propose a generic framework to guide the development of such solutions. In the end, we will illustrate the use of the generic framework in a practical case.

1 Introduction

In the competitive and technological scenario of contemporaneous industry, the importance of efficient maintenance solutions has become a key factor of success. [1] has estimated that, in industrial firms, maintenance cost varies from 15% to 40% and even in simpler systems, such as industrial vehicles, this source of expense is far from neglectable. The National Road Committee (CNR) of France estimated that the maintenance costs of a long-haul truck accounted on average for 8.2% of the total expenses. For trucks in particular, not only this cost cannot be overseen but the importance of effective maintenance is crucial in the transportation business. Internal reports conducted with Volvo trucks clients suggest that when a truck undergoes a breakdown, all the annual revenue generated by this vehicle is compromised.

Since maintenance is so important in all sort of different domains, it is only natural that it has been the center of interest of several researchers. In recent years, academic works on new maintenance solutions have been published and terms such as condition-based maintenance (CM) [2], predictive maintenance (PdM) [3] and more recently prescriptive maintenance (PsM) [4] have gained some popularity. One of the main issues of employing the term prescriptive maintenance is that there is no clear definition for it. The boundaries between predictive and prescriptive maintenance are not well defined, and we have to be careful when employing those terms to avoid confusion. A rigorous

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 263–273, 2022.
https://doi.org/10.1007/978-3-030-96794-9_24

definition of prescriptive maintenance is important to guide and frame future work on the area and to develop the necessary tools to design and implement solutions that really achieve reliability maximization and cost minimization in industrial applications. In this paper, after a discussion on the use of the term prescriptive maintenance, a modeling framework that highlights the differences between predictive and prescriptive maintenance will be presented. This framework will help to guide the development of generic decision-making algorithms for up-time maximization and avoid the unjustified use of jargon.

Therefore, this document is organized according to the following structure: Sect. 2 will highlight the vagueness of the definition of prescriptive maintenance, briefly reviewing some of the work published using this term. A generic framework of PsM will be detailed in Sect. 3 - hopefully its use will avoid ambiguity and guide future work in the area. Finally, through Sect. 4, a practical example of PsM applied to the automotive domain will be given.

2 Prescriptive Maintenance

2.1 Prescriptive Maintenance in the Literature

In recent years, a few authors have employed the term prescriptive maintenance in their works [4–6]. It appears that, even if these works focus on discussing conceptually PsM, they do not always present a formal definition of the term, but base their definition on the broad idea of choosing the correct course of action for a system. This general understanding of PsM seems to be an extension of the concept of prescriptive analytics, which focus on prescribing the best decisions in order to take advantage of the predicted future utilizing large amounts of data [7].

Hence, one can think of prescriptive maintenance as the use of prescriptive analytics to maintenance. According to this definition, a PsM solution should use failure predictions or, data-driven degradation models, to quantitatively give the best course of action in terms of up-time maximization. Some of the works published on PsM do not fit this definition perfectly.

For example, the authors in [5] developed a prescriptive maintenance solution built on a threshold-based rule, meaning that an action will be taken on the system only when a quality indicator overpasses a threshold. The second issue is that PsM should quantitatively assesses what is the best action to take. To that end, it is crucial to have an objective measure of the impact of different actions. However, in [5], actions are chosen based on previous engineering knowledge, working as thumb rules. The main problem when using this rudimentary notion of PsM as the simple application of prescriptive analytics to maintenance is that it does not help to distinguish between PsM, PdM and CM. For example, [4] uses the term prescriptive maintenance to provide a solution that chooses the best maintenance and inspection schedule for a system subjected to degradation. Similar problems were addressed by different authors before [8], without using the term prescriptive maintenance. In fact, a big part of research conducted on PdM could be classified as decision-making and therefore as an application of prescriptive analytics, which makes this notion insufficient to make the distinction between PdM and PsM.

2.2 Defining Prescriptive Maintenance

An attempt of differentiating PsM from PdM is given in [9]. According to the author, a crucial difference lies in the fact that PsM takes into account all the functionalities of the system, by extending the notion of maintenance to the one supported by the Prognostics and Health Management (PHM) community. Thus, the objective of PsM is to provide the best course of action to minimize the overall cost of systems exploitation. These actions may include use mode recommendations, tasks management, parameters reconfiguration, etc.

Many research works are increasingly interested in developing one or many of these actions in addition to maintenance scheduling. In [10] for example, the authors focused on finding the best moment to perform maintenance while managing spare parts. Another example is given in [11] where a dynamic method was developed to jointly schedule missions and maintenance operations while taking into account the system's deterioration, in order to minimize the maintenance costs. It is worth noticing that most of the previous actions will have an impact on the system's usage and on the degradation process, which should in turn affect the choice of the next actions to be applied. Therefore, PsM should take this notion into account by following a closed-loop structure, meaning that algorithms should be robust, deal with uncertainty and assess the effect of the decisions chosen on the system to adapt them as often as required.

In the next section, a PsM framework that relies on the concept of closed-loop decision process will be proposed.

3 A Generic Framework for Prescriptive Maintenance

One key aspect of PsM is the notion of closed loop, as represented in Fig. 1. Due to the randomness of degradation processes, the outcome of previously chosen actions must always be monitored. As the system evolves through time, actions may be chosen considering previous unexpected behavior, as well as new inputs.

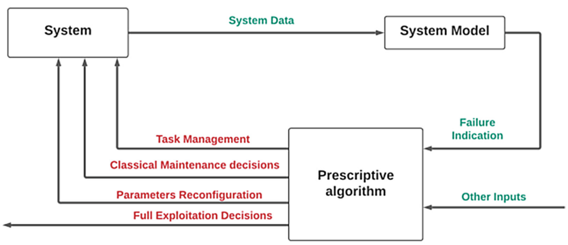


Fig. 1. Example of PsM algorithm

To arrive at such structure, which characterizes PsM, a generic framework is proposed. It is composed of three main steps: system modeling, action modeling and optimization.

3.1 System Modeling

To quantitatively decide between different actions, it is necessary to understand how the system behaves and how it fails. Actions may affect the remaining useful life (RUL) of a system and quantifying this effect is crucial for choosing how to exploit it. Since the PsM solution may prescribe changes in usage conditions and environment, it is crucial to model the degradation process of the system in a way that all different operational conditions are considered. To this aim, data availability is not enough. It is also important to ensure that the variables which have the biggest impact on the degradation process were identified and monitored, and that the degradation model took them into account.

In the following sections, three classes of techniques, which are the most common in the literature for modeling the behavior of the system in terms of RUL, are presented.

3.1.1 Degradation Modeling

Degradation models are usually developed based on degradation data combined with the understanding of the physics of the process. These models can be deterministic or stochastic, with the latter usually presenting more flexibility and robustness.

In the literature, several stochastic degradation models have been used. For PsM, models with covariates may be used to handle applications where several different factors affect the degradation. As an example, one can cite the variance gamma process combined with Markov chains [12]. It is important to highlight that in some simple cases, where the possible actions to be applied affect only one variable, or where the degradation process is well defined by only one stress factor, classical models such as the gamma [13] and the wiener processes [14] can be used.

3.1.2 Reliability Distributions

In several real-world applications, the degradation process may be too complex to model, or there may not be enough data available to infer the parameters of degradation models. In these cases, other approaches, such as reliability distributions, may be used to directly estimate a failure probability in given usage conditions. In this case, the failure time of a system is modeled as a random variable with known distribution. It is worth noting that most of the classical distributions, such as Weibull, Gamma and Exponential, do not account for covariables and hence they do not account for the possible effect of the chosen actions and for different usage conditions. Therefore, they may not be suitable for PsM applications. Some examples of reliability distributions that are relevant for PsM applications can be found in [15].

Moreover, it is important to highlight that, although reliability distributions can be useful for system modeling, they provide less insight on the system than degradation models. Indeed, the latter can always be used to derive distributions of the time to failure, but the opposite is not always true. Therefore, degradation models should be used whenever it is possible because they contain more information on the system.

3.1.3 Black Box Data Driven Approaches

Data driven approaches are an alternative to the model-based approaches discussed before. Black Box Data-driven approaches can infer patterns from data without prior hypothesis on the degradation process nature or on the failure time distribution and incorporate the effect of several different covariables.

Data-driven algorithms can be used to predict the failure time, estimate a failure probability, or classify systems in different categories according to the severity of their usage. Examples of data-driven approaches for prognostics can be found in [16]. However, it is important to highlight, that the accuracy of such models depend on data availability and quality, and an understanding of different failure modes and stress factors can be necessary to employ them satisfactorily.

3.2 Action Modeling

Once the system is modeled, the following step is to list and model all the actions that can be applied to it. It is important to keep in mind that each system is unique and that different classes of actions may apply in each case. In the following, different examples of actions are presented.

3.2.1 Classical Maintenance Decisions

PsM solutions must account for the classical maintenance decisions that are usually found in PdM and CM literature. The vast majority of literature on decision-making for maintenance focuses on choosing the maintenance date, assessing how to make this decision under different circumstances, i.e. perfect and imperfect maintenance operations [17], perfect and imperfect information [18], etc. Alongside maintenance date decision, one can find several articles interested in defining inspection intervals, such as [8].

Therefore, it appears that classical maintenance decision is the core of every intelligent maintenance policy. A PsM must take it into account and go beyond, exploiting other dimensions of the decision-making process.

3.2.2 Task Management

Systems can be composed of more than one subsystem (e.g. a fleet of vehicles). Some subsystems are more prone to failure than others and, therefore, the decision regarding which subsystem to use to perform a task must be made taking the different degradation levels into account. Similarly, different tasks may present different levels of severity, which makes the order in which they are performed impact the evolution of the RUL as well.

One example of task management considering degradation information can be found in [11], where the mission plan of a fleet of trucks is decided based on the severity of each displacement and the current health state of each vehicle.

3.2.3 Parameters Reconfiguration

PsM relies on the fact that the actions applied on a system can affect its degradation process. The same actions could be used to, at some extent, control the degradation process directly. One example is the action related to parameter reconfiguration. Indeed, controlling the RUL of a system could be achieved by modifying, in a suitable way, the parameters of the system.

Two aspects should be considered when modeling this type of action: (i) the impact of the parameter reconfiguration on the degradation model (e.g., limiting the power of a machine via software change can postpone maintenance operations) and (ii) the impact of the parameter reconfiguration on the system performance (e.g., reducing the power of the machine will reduce productivity and the software change has a financial cost). On this topic, one can cite the article [19] where a wind turbine RUL is controlled based on its torque.

3.2.4 Full Exploitation Decisions

As PsM considers a holistic vision of the system, it also addresses actions that do not impact the degradation process or the maintenance strategy but are rather affected by it. Indeed, not all decisions will result in an acceleration or a deceleration of the degradation process. One example is given in [20] where deterioration information is used to decide the inventory level, providing insight on the best spare part management strategy.

3.3 Optimization

Once every relevant aspect of the system is modeled, an optimization layer is necessary to prescribe the best course of action.

3.3.1 Cost Function

To choose the best actions to apply, a metric has to be established so that actions and their expected outcomes can be quantitatively compared. The chosen actions are the ones that minimize this metric, generally referred to as the cost function.

The cost function has to capture the systems exploitation trade-offs. The example given previously in Sect. 3.2.3 illustrates this trade-off. If reducing machine power can postpone maintenance operations but will reduce productivity, the cost function has to be defined such that it is possible to compare those outcomes. The solution will then decide when and how to reduce machine power to minimize exploitation cost. The cost function must also take into consideration other inputs such as operational constraints, e.g., availability of spare parts, deadlines, workload, etc. The PsM solution should be built on a realistic cost function, that captures all the reality of system exploitation, considering trade-offs and constraints, in order to account for all the functionalities of the system.

3.3.2 Optimization Technique

Once the optimization criteria are chosen, different techniques can be employed in order to find the best course of action. The choice must be made based on the needed reaction time of the PsM solution and the complexity of the space of actions.

In cases where the choice of possible actions is limited, optimization methods that are guaranteed to reach the cost function minimum, such as dynamic programming, can be employed [20]. Whenever the space of actions becomes bigger or the dynamic of the system requires fast adaptability, meta-heuristic methods should be used [11].

4 PsM Application

In this section, a closed-loop PsM solution for the joint maintenance and missions assignment of industrial vehicles is presented. The critical component chosen to illustrate the proposed framework is the brake-pad.

The main hypothesis are:

- Missions are defined as the deliveries that a vehicle has to make from one point to another. All the distances and durations from point to point are considered to be known, and are stored respectively in matrices D and T.
- A decision epoch P_k arises when a new set of missions has to be accomplished. For example, every day a fleet owner has to make deliveries and therefore, decisions will happen at the beginning of each day, before sending vehicles from the headquarter.
- Only breakdowns related to the brake-pad are considered.

4.1 System Modeling

Brake-pads failure is commonly caused by wear and tear with usage. The brake-pad must be replaced whenever its thickness falls under a critical threshold. Figure 2 shows the thickness evolution of a real brake-pad.

The brake-pad thickness evolution exhibits a mean trend close to be directly proportional to the traveled distance. The thickness degradation phenomena can be modeled with a Wiener process with a linear drift, as shown in the following equation:

$$Y(x) = Y_0 + \lambda x + \sigma_B B(x) + \varepsilon \tag{1}$$

where Y is the brake-pad thickness, Y_0 is the initial thickness, λ is the negative drift, x is the traveled distance, $B(x)$ is the standard Brownian motion, σ_B is its variance, and ε is the measurement noise considered to be white and Gaussian. Details on how to estimate those parameters from data can be found in [16].

4.2 Actions Modeling

For this practical example, two types of actions are considered: classical maintenance and tasks management through missions scheduling.

4.2.1 Maintenance Date Decision

Vehicles come regularly to workshops to perform maintenance operations, such as oil change. To avoid extra workshop visits, brake-pad should only be replaced at those visits. The two next scheduled dates of workshop visit t_{wkp_1}, and t_{wkp_2} are known for each vehicle i. The maintenance decision comes down to choosing to replace the brake-pads at t_{wkp_1} or to postpone the change at least until t_{wkp_2}.

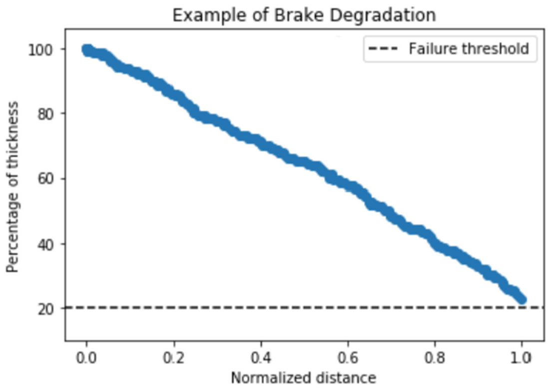


Fig. 2. Thickness evolution

To decide the maintenance date, Monte-Carlo simulations are made based on Eq. 1 and the distribution of traveled distances, in order to compare the cost related to the possibility of a failure between t_{wkp_1} and t_{wkp_2} and the cost related to the expected amount of thickness wasted if a replacement occurs at t_{wkp_1}. If the cost associated to the failure probability is bigger than the one associated to the wasted thickness, the next workshop visit (t_{wkp_1}) is chosen to replace the brake-pad.

4.2.2 Missions Scheduling

The scheduler will receive the list of missions to be accomplished, alongside with the last available degradation measurement. It will then proceed to find the best schedule possible.

It is important to highlight that both maintenance date decision and mission scheduling affect each other. Indeed, the mission schedule will be followed by the fleet and new degradation measures will be collected. These measurements will be used as an input by the maintenance date decision maker which will decide if a vehicle will replace its brake-pads in the next workshop visit. Replacement dates are then updated and used to change the scheduler cost function.

4.3 Optimization

The cost function used by the scheduler, which is responsible for finding the optimal mission schedule, is presented below:

$$C = C_{dist} + C_{delay} + C_{deg} + C_{waste} \quad (2)$$

It integrates four costs. C_{dist} and C_{delay} capture the operational costs and are proportional respectively to the total distance of a schedule and to the delays. C_{deg} and C_{waste} are costs related to the brake-pad thickness. The first accounts for the expected amount of thickness that will be consumed when following a mission schedule. It only considers vehicles that have not established a maintenance date for the brakes. On the other hand, C_{waste} considers only vehicles that will replace brakes in the next workshop visit and accounts for the cost related to the surplus of thickness that may be wasted in that replacement.

This cost function was chosen so that the resulting schedule will reduce operational costs, minimize the brakes degradation and, at the same time, avoid wasting thickness.

4.4 Results

To validate the proposed solution, a comparison between the proposed PsM algorithm and a real client exploitation strategy, which focuses only on distance and delay costs, is made. The simulations emulate the exploitation of a fleet of vehicles that have to perform the same missions every week, scheduling missions and brake-pad replacements in different ways. The results showed that operational costs were identical for both strategies, meaning that the PsM scheduler was able to reduce distance and delay costs as well. On the other hand, maintenance costs were different, as can be shown in Fig. 3.

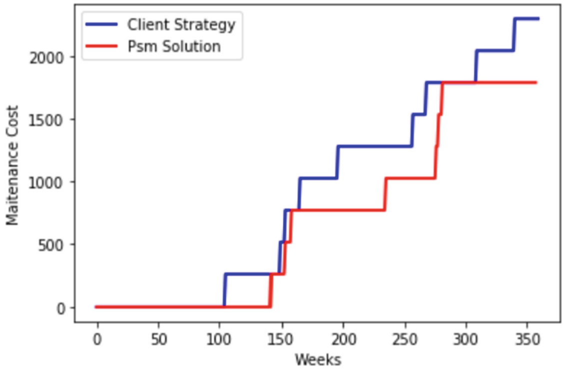


Fig. 3. Total cost comparison

The PsM algorithm performed better, postponing maintenance operations and avoiding waste while respecting operational constraints. This is due to the use of dynamic scheduling that accounts for maintenance dates and often uses more degraded vehicles to perform the least demanding missions.

5 Conclusion and Future Work

Maintenance is a key factor to ensure competitiveness and to minimize costs in many industrial systems. In the literature, maintenance has gathered a lot of attention from researchers who worked on maintenance solutions using concepts such as CM, PdM and more recently PsM. The term “prescriptive maintenance” has not been well defined in the literature. This lack of a robust definition represents an obstacle in the development of PsM solutions.

In this article, a discussion on the use of the term PsM in the literature and its differences regarding PdM was presented and key elements of a PsM structure were highlighted. A framework with three steps was proposed to guide future PsM solutions and avoid the unjustified use of the term prescriptive maintenance.

To illustrate the proposed approach, a practical example of PsM applied to the automotive domain was given, in which maintenance dates and missions are planned for a fleet of vehicles, minimizing the overall exploitation cost.

In future works, a deeper inquiry on the use of PsM in the literature can be conducted to improve the framework and applications involving more complex systems and components will be developed.

References

1. Löfsten H (2000) Measuring maintenance performance – in search for a maintenance productivity index. Int J Prod Econ 63:47–58. ISSN 0925–5273
2. Huynh KT, Barros A, Berenguer C (2012) Maintenance decision-making for systems operating under indirect condition monitoring: value of online information and impact of measurement uncertainty. IEEE Trans Reliab 61(2):410–425. https://doi.org/10.1109/TR.2012.2194174
3. Kimera D, Nangolo FN (2020) Predictive maintenance for ballast pumps on ship repair yards via machine learning. Transp Eng 2:100020. ISSN 2666–691X
4. Liu B, Lin J, Zhang L, Kumar U (2019) A dynamic prescriptive maintenance model considering system aging and degradation. IEEE Access 7:94931–94943. https://doi.org/10.1109/ACCESS.2019.2928587
5. Matyas K, Nemeth T, Kovacs K, Glawar R (2017) 'A procedural approach for realizing prescriptive maintenance planning in manufacturing industries.' CIRP Ann 66(1):461–464
6. Ansari F, Glawar R, Nemeth T (2019) PriMa: a prescriptive maintenance model for cyber-physical production systems. Int J Comput Integr Manuf 32(4–5):482–503. https://doi.org/10.1080/0951192X.2019.1571236
7. Lepenioti K, Bousdekis A, Apostolou D, Mentzas G (2020) Prescriptive analytics: literature review and research challenges. Int J Inf Manag 50: 57–70. ISSN 0268–4012
8. Huynh KT, Barros A, Bérenguer C, Castro IT (2011) A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. Reliab Eng Syst Saf 96(4):497–508
9. Iung B (2019) De la maintenance prédictive à la maintenance prescriptive: Une évolution nécessaire pour l'industrie du futur. In: Conference on complexity analysis of industrial systems and advanced modeling, CAISAM 2019, Ben Guerir, Maroc. ffhal-02126720f
10. Wang L, Chu J, Mao W (2009) A condition-based replacement and spare provisioning policy for deteriorating systems with uncertain deterioration to failure. Eur J Oper Res 194(1):184–205. ISSN 0377–2217
11. Robert E, Bérenguer C, Bouvard K, Tedie H, Lesobre R (2018) Joint dynamic scheduling of missions and maintenance for a commercial heavy vehicle: value of on-line information. In: 10th IFAC symposium on fault detection, supervision and safety for technical processes (SAFEPROCESS'18), IFAC, Aug 2018, Warsaw, Poland, pp 837–842. https://doi.org/10.1016/j.ifacol.2018.09.672, hal-01902212
12. Belhaj Salem M, Fouladirad M, Deloux E (2021) Prognostic and classification of dynamic degradation in a mechanical system using variance gamma process. Mathematics 9:254. https://doi.org/10.3390/math9030254

13. Park SH, Kim JH (2017) Application of gamma process model to estimate the lifetime of photovoltaic modules. Solar Energy 147:390–398. ISSN 0038–092X
14. Tang S, Yu C, Wang X, Guo X, Si X (2014) Remaining useful life prediction of lithium-ion batteries based on the wiener process with measurement error. Energies 7:520–547
15. Kumar D, Klefsjö B (1994) Proportional hazards model: a review. Reliabil Eng Syst Saf 44(2): 177–188. ISSN 0951–8320
16. Heimes FO (2008) Recurrent neural networks for remaining useful life estimation. In: 2008 international conference on prognostics and health management, Denver, CO, USA, pp 1–6. https://doi.org/10.1109/PHM.2008.4711422
17. Do Van P, Voisin A, Levrat E, Iung B (2013) Remaining useful life based maintenance decision making for deteriorating systems with both perfect and imperfect maintenance actions. In: 2013 IEEE Conference on Prognostics and Health Management (PHM), Gaithersburg, MD, USA, pp 1–9. https://doi.org/10.1109/ICPHM.2013.6621442
18. Ghasemi A, Yacout S, Ouali MS (2007) Optimal condition based maintenance with imperfect information and the proportional hazards model. Int J Prod Res 45(4):989–1012. https://doi.org/10.1080/00207540600596882
19. Obando DR, Molina JJM, Bérenguer C (2018) An architecture for controlling the remaining useful lifetime of a friction drivesystem. In: 10th IFAC symposium on faultdetection, supervision and safety for technical processes (SAFEPROCESS 2018), IFAC, pp 861–866
20. Wang J, Zhu X (2021) Joint optimization of condition-based maintenance and inventory control for a k-out-of-n:f system of multi-state degrading components. Eur J Oper Res 290(2):514–529

Asset Information Management Systems: Critical Success Factors in the Brazilian Electricity Sector

Alexandro Teixeira Gomes and Fernando Silva Parreiras(✉)

Laboratory for Advanced Information Systems – LAIS, FUMEC University, Rua Cobre 200, Belo Horizonte, MG 30330-190, Brazil
alexsandro.teixeira@cemig.com.br, fernando.parreiras@fumec.br

Abstract. As the Electricity Sector has a strategic role in the socioeconomic development of a country, shortages have a damaging effect for all consumers and for the company itself. Therefore, companies in the Electricity Sector is considered asset-intensive, as their performance depends on the performance of their assets. This work has as objective to evaluate the critical success factors, promoting good performance on the Asset Information Management Systems in the Brazilian Electricity Sector. To achieve this objective, we defined a model that analyzes the impact of implementing Asset Information Management Systems into organizational performance indicators of companies in the Electric Sector. To validate this model, we simulate connections that were performed through structural equation modeling, based on data obtained directly from professionals from different sectors of Generation, Transmission and Distribution companies in Brazil, in a previously defined form for this research. The relevance of the research is the analysis of the relation between impact and correlation in the Asset Information Management Systems, so that companies in the Electric Sector can prioritize efforts in factors of greater impact. The research is limited to evaluating only the Information within the Management Systems of Assets, with suggestions for future research to evaluate the other elements.

1 Introduction

The relevance of Electric Energy is related to everything that makes it viable and promotes, being an indispensable resource to our daily life, with the Electric Sector having a strategic role from the social and economic point of view in order to supply [1–3].

In this context, Asset Management can be seen as an essential tool for the Electric Sector, considering that it is strongly dependent on the performance of its assets, with a focus on the balance between Cost, Risk and Performance [4] and [5], these being the performance constructs of the Electric Sector used in this work. The Management Systems of Asset consists in the elements necessary to achieve the objectives of Asset Management (policies, plans, business processes and information systems) [5], with emphasis on the Asset Information Management Systems (information systems) for study in this work. Based on the theoretical framework, the Asset Management Maturity Model was developed, developed by IAM - UK Asset Management Institute [6], and

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 274–286, 2022.
https://doi.org/10.1007/978-3-030-96794-9_25

the themes (constructs) related to the Management System of Asset Information, for application in this research: IAM-22 - Asset Information Strategy, IAM-23 - Asset Information Standards/Specifications, IAM-24 - Asset Information Systems and IAM-25 - Data and Information Management.

To achieve this objective, a model was defined that relates the performance factors of companies in the Electric Sector, with the factors that involves implementing a Management System of Asset Information. The validation of the connections establish for this model were carried out through the modeling of structural equations, from data obtained directly from people from different sectors of Generation, Transmission and Distribution companies in Brazil.

2 The Electric Sector, Asset Management and Asset Information Management Systems

It is important to highlight that the Electricity Sector Companies are considered "active-intensive", whose performance depends on the performance of their assets [7]. In this line, there is an alignment with the objectives and benefits of Asset Management, which can be defined as the coordinated activity of an organization to obtain value from the assets, which involves a balance between costs, risks, opportunities and benefits of performance [5].

Understanding that a Management System of Asset Information is a set of interrelated and interacting elements in a organization, whose function is to establish the Asset Management policy, objectives and necessary processes to achieve these objectives [5], it is clear that importance of these elements to the Management System of Asset Information. The relationship between the Electric Sector, Asset Management and Management System of Asset Information situates an issue in this Research: "What are the relationships (impacts and correlations) between the factors that react to the implementation of a Management System of Asset Information in the Brazilian Electricity Sector within the scope of Asset Management?".

3 Methods

To evaluate the relationships between the Performance Constructs and the Constructs of the Asset Information Management Systems presented, a model of structural equations was developed using the PLS (Partial Least Square) approach and to verify the correlation between the indicators, the Spearman Correlation coefficient was used.

The PLS approach is based on the covariance matrix (CBSEM - Covariance Based Structural Equation Modeling), being a technique that offers greater flexibility in data modeling, without requiring multivariate normality of data, independence between observations and large sample size [14]. For this research, the PLS approach was necessary in view of the sample size and because the data set does not present a normal distribution.

Considering the division of the structural equation modeling process into two parts: Measurement Model and Structural Model [15] and [16], the flow of structural equation modeling applied in this work was followed.

In measuring reliability, the indicators Cronbach's alpha (AC) and composite reliability (CC) [22] were used. From the point of view of validation, the AC or CC indicators must be greater than 0.70 for an indication of the construct's reliability [23], which is the case of this work. In exploratory research, values above 0.60 are also accepted for validation [17].

The discriminant validation was verified by the Barclay criterion, where a construct reaches discriminant validity when the factorial load of all its items is greater than the respective crossed factorial loads [24]. That is, such items have a greater relationship with the construct indicated in the theoretical model. Regarding the validation of the structural model, R2 (Pearson's Coefficients of Determination) was used to assess the quality of the adjustments, and for future comparisons of adherence of different samples to the model, the GoF (Goodness of Fit) was also used [23]. The GoF is a geometric average of the AVEs of the constructs and the R^2 of the model and varies from 0% to 100%. The GoF in PLS does not have the ability to discriminate valid models from invalid ones, in addition to not being applied to models with formative constructs [20], it only allows a synthesis of the AVEs and the R^2 of the model in a single statistic, and can be useful for future comparisons of adherence of different samples to the model. To check the correlation between the indicators, Spearman's correlation coefficient was used, which is a limited measure between –1 and 1, and the closer the coefficient is to –1 the greater the negative correlation and the closer the coefficient is to 1 greater the positive correlation [28]. According to Bauer (2007), the correlation coefficient of Spearman's correlation coefficient is used to estimate linear correlations when the data are not normal, which is the case of this work. The survey questionnaire was applied to professionals from the three basic segments of the Electric Sector Generation, Transmission and Distribution, based specifically on the author's network of contacts, using only the means of communication: e-mail, WhatsApp, and LinkedIn.

According to Hair et al. (2009) the sample size required for structural equation models must be between 5 and 10 individuals for each item present in the research instrument (variable). Considering that this survey totals 48 observable variables in the model, the expected number of respondents would be between 240 (5×48) and 480 (10×48) respondents. In this sense, it was established that a survey would need to contain at least 240 respondents, considering that the questionnaire is comprised of 48 questions, 36 of which are related to the Asset Information Management System, and 12 are related to Asset Management Performance.

4 Research Instrument

The Performance Constructs of this research were defined based on the dimensions or benefits expected from Asset Management: Costs, Risks and Performance, whose issues are presented in Table 1, applied in survey as presented in the previous item.

Table 1. Issues related to performance constructs.

Constructs	Questions		References
Cost	D01	The financial result of your company is positive, and your company is well positioned in relation to the players in the sector	[5, 33, 34]
	D02	The decision-making process in your company about investing in assets is based on evidence, data, and facts	[5, 33]
	D03	There are initiatives in your company to reduce capital and operating costs	[13]
	D04	Your company analyzes and considers the life cycle costs of the assets, and not just the initial costs, mana. the end of life of the assets	[35]
Risk	D05	Your company performs a good management of the risks and opportunities of the assets	[5, 33]
	D06	There are controls and initiatives to reduce the exposure to Safety, Environmental and Technical risks	[13]
	D07	Your company monitors and mitigates risks by measuring and valuing these risks	[35]
	D08	There is good management on regulatory compliance and the legislation in force in your company	[36, 37]
Performance	D09	The results of the service quality indicators are good, with improvement of products and services	[5]
	D10	Notifications and penalties for your company are minimal or within an acceptable standard	[36, 37]
	D11	Asset reliability indicators are good and are improving	[34]
	D12	Asset availability indicators are good and are improving	[13, 34, 35]

In Fig. 1 the simplified theoretical structural model that was simulated, whose hypotheses are:

- H01 - There is a positive impact of the Asset Information Strategy (IAM-22) on Costs.
- H02 - There is a positive impact of the Asset Information Strategy (IAM-22) on Risks.

- H03 - There is a positive impact of the Asset Information Strategy (IAM-22) on Performance.
- H04 - There is a positive impact of the Standards/Specifications of asset information (IAM-23) on Costs.
- H05 - There is a positive impact of the Standards/Specifications for asset information (IAM-23) on Risks.
- H06 - There is a positive impact of the Standards/Specifications of asset information (IAM-23) on Performance.
- H07 - There is a positive impact of Asset Information Systems (IAM-24) on Costs.
- H08 - There is a positive impact of Asset Information Systems (IAM-24) on Risks.
- H09 - There is a positive impact of Asset Information Systems (IAM-24) on Performance.
- H10 - There is a positive impact of Data and Information Management (IAM-25) on Costs.
- H11 - There is a positive impact of Data and Information Management (IAM-25) on Risks.
- H12 - There is a positive impact of Data and Information Management (IAM-25) on Performance.
- H13 - There is a positive correlation between Costs and Risks.
- H14 - There is a positive correlation between Costs and Performance.
- H15 - There is a positive correlation between Risks and Performance.

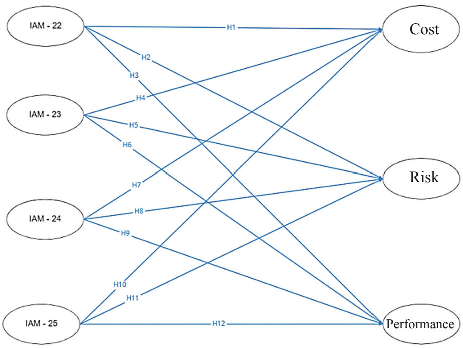


Fig. 1. Simplified theoretical structural model that was simulated.

Based on the simplified theoretical model from Figure 4, hypotheses 1 to 12 were analyzed by modeling structural equations and hypotheses 13, 14 and 15 via correlation.

5 Results

The database of the research questionnaire is composed of 53 variables, 5 of which characterize the individual and 48 corresponding to the items of a questionnaire with a Likert scale of agreement ranging from 1 to 5. The questionnaire was answered by 280 individuals, thus adding 14,840 observations, of these, no missing value was found.

5.1 Measurement Model (Outer Model)

In the analysis of the measurement model, the convergent validity, the discriminant validity, and the reliability of the constructs are verified. Convergent validity ensures that the indicators of a construct are correlated enough to measure the latent concept. The discriminant validity verifies whether the constructs effectively measure different aspects of the phenomenon of interest. Reliability reveals the consistency of the measures in measuring the concept that is intended to be measured.

The Table 2 shows the measurement model of the constructs, where it is noted that any had a factor loading less than 0.50, it is not necessary to remove any item. In addition, there was discriminant validation in all constructs since the maximum value of the crossed factorial loads of each item was lower than the respective factor load.

Table 2. Construct measurement model

Construct	Items	CF[a]	CFC[b]	Com.[c]	Weight
Costs (CT)	D01	0.66	0.45	0.44	0.28
	D02	0.81	0.56	0.65	0.41
	D03	0.66	0.49	0.43	0.26
	D04	0.79	0.57	0.62	0.40
Risks (RC)	D05	0.80	0.64	0.64	0.32
	D06	0.83	0.54	0.69	0.30
	D07	0.86	0.64	0.74	0.34
	D08	0.73	0.52	0.53	0.28
Performance (DP)	D09	0.74	0.47	0.54	0.33
	D10	0.64	0.37	0.41	0.26
	D11	0.85	0.53	0.72	0.37
	D12	0.84	0.50	0.71	0.33
Asset information strategy (IAM-22)	S01	0.82	0.69	0.67	0.17
	S02	0.82	0.71	0.68	0.15
	S03	0.84	0.73	0.70	0.17
	S04	0.87	0.73	0.75	0.17
	S05	0.73	0.60	0.53	0.14
	S06	0.66	0.63	0.44	0.13
	S07	0.75	0.69	0.57	0.14
	S08	0.60	0.57	0.35	0.10
	S09	0.67	0.66	0.45	0.14

(*continued*)

Table 2. (*continued*)

Construct	Items	CF[a]	CFC[b]	Com.[c]	Weight
Asset information standards/specifications (IAM-23)	S10	0.81	0.73	0.65	0.15
	S11	0.82	0.70	0.67	0.14
	S12	0.81	0.74	0.65	0.14
	S13	0.83	0.73	0.69	0.15
	S14	0.85	0.76	0.72	0.15
	S15	0.83	0.76	0.69	0.15
	S16	0.70	0.63	0.49	0.11
	S17	0.76	0.68	0.58	0.13
	S18	0.81	0.72	0.66	0.14
Asset information systems (IAM-24)	S19	0.72	0.64	0.53	0.13
	S20	0.78	0.67	0.60	0.15
	S21	0.77	0.64	0.59	0.15
	S22	0.79	0.66	0.62	0.13
	S23	0.81	0.69	0.65	0.14
	S24	0.83	0.73	0.69	0.16
	S25	0.83	0.73	0.69	0.14
	S26	0.75	0.68	0.56	0.13
	S27	0.80	0.77	0.64	0.13
Data and information management (IAM-25)	S28	0.76	0.64	0.58	0.14
	S29	0.78	0.65	0.62	0.15
	S30	0.81	0.67	0.65	0.14
	S31	0.83	0.69	0.68	0.13
	S32	0.86	0.74	0.74	0.15
	S33	0.79	0.68	0.63	0.14
	S34	0.80	0.76	0.63	0.15
	S35	0.78	0.71	0.60	0.14
	S36	0.65	0.65	0.42	0.13

[a]Factorial Load; [b]Maximum of the crossed factorial loads; [c]Communality

The Table 3 shows the analysis of convergent validity, dimensionality, and reliability for validation of the constructs of the measurement model, considering that:

- There was convergent validation in all constructs, since the AVE's were greater than 0.40.
- In all constructs, the AC and/or CC reliability indexes were higher than 0.60, thus showing their reliability.

- According to Kaiser's criterion [25], all constructs were one-dimensional.

Table 3. Validation of the model for measuring the constructs.

Construct	Items	AVE[a]	A.C[b]	C.C[c]	Dim.[d]
Custos	4	0,54	0,71	0,82	1
Risks	4	0,65	0,82	0,88	1
Performance	4	0,60	0,77	0,85	1
IAM-22	9	0,57	0,90	0,92	1
IAM-23	9	0,64	0,93	0,94	1
IAM-24	9	0,62	0,92	0,94	1
IAM-25	9	0,62	0,92	0,94	1

[a]Average Variance extracted; [b]Cronbach's alpha; [c]Composed Reliability; [d]Dimensionality

5.2 Structural Model (Inner Model)

The Table 4 shows the structural model and Fig. 2 illustrates this model.

Table 4. Structural model

Endogenous	Exogenous	β	EP[a]	CI (95%)[b]	P-value	R[b]
Costs	IAM-22	0.58	0.10	[0.33; 0.82]	**<0.001**	41.78%
	IAM-23	–0.03	0.12	[–0.29; 0.24]	0.819	
	IAM-24	0.01	0.10	[–0.23; 0.22]	0.915	
	IAM-25	0.10	0.10	[–0.09; 0.32]	0.320	
Risks	IAM-22	0.50	0.09	[0.32; 0.72]	**<0.001**	53.45%
	IAM-23	0.17	0.10	[–0.12; 0.36]	0.099	
	IAM-24	–0.15	0.09	[–0.32; 0.04]	0.105	
	IAM-25	0.23	0.09	[0.08; 0.40]	**0.011**	
Performance	IAM-22	0.45	0.11	[0.24; 0.69]	**<0.001**	34.82%
	IAM-23	–0.05	0.12	[–0.28; 0.18]	0.691	
	IAM-24	0.23	0.11	[0.01; 0.45]	**0.035**	
	IAM-25	–0.02	0.11	[–0.22; 0.26]	0.847	

[a]Standard error; [b]Bootstrap interval (95%); Gof = 51.34%

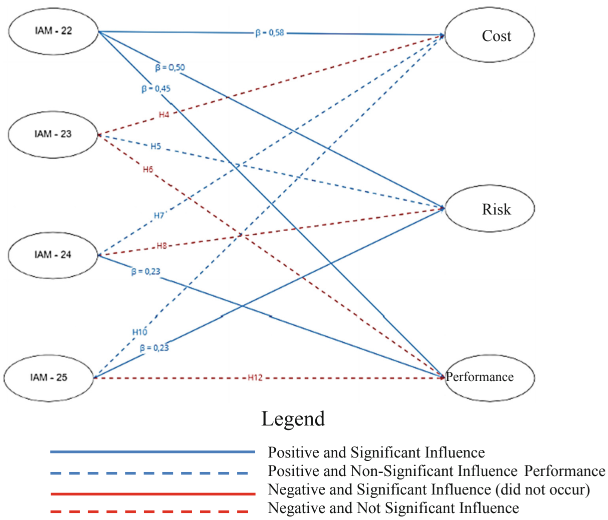


Fig. 2. Illustration of the structural model

From the information on the structural model (Table 8), it is highlighted that:

Cost

- There was a significant (p-value < 0.001) and positive ($\beta = 0.58$) influence of the Asset Information Strategy (IAM-22) on Costs, therefore, the higher the IAM-22 score, the higher the tendency to be. average cost score.
- There was no significant influence (p-value > 0.50) of IAM-23, IAM-24, and IAM-25 on costs.
- IAM-22, IAM-23, IAM-24, and IAM-25 were able to explain 41.78% of the cost variability, that is, R2 represents moderate explanatory capacity (Hair et al. 2009).

Risk

- There was a significant (p-value < 0.001) and positive ($\beta = 0.50$) influence of the Asset Information Strategy (IAM-22) on Risks, so the higher the IAM-22 score, the higher the tendency to be average risk score.
- There was a significant (p-value $= 0.011$) and positive ($\beta = 0.23$) influence of Data and Information Management (IAM-25) on Risks, so the higher the IAM-25 score, the higher the tendency to be. average risk score.
- There was no significant influence (p-value > 0.50) of IAM-23 and IAM-24 on Risks.
- IAM-22, IAM-23, IAM-24, and IAM-25 were able to explain 53.45% of the variability of Risks, that is, R2 represents substantial explanatory capacity (Hair et al. 2009).

Performance

- There was a significant (p-value < 0.001) and positive ($\beta = 0.45$) influence of the Asset Information Strategy (IAM-22) on Performance, so the higher the IAM-22 score, the higher the tendency to be average performance score.
- There was a significant (p-value $= 0.035$) and positive ($\beta = 0.23$) influence of the Asset Information Systems (IAM-24) on Performance, so the higher the IAM-24 score, the higher the tendency to be average performance score.
- There was no significant influence (p-value > 0.50) of IAM-23 and IAM-25 on Performance.
- IAM-22, IAM-23, IAM-24, and IAM-25 were able to explain 34.82% of the performance variability, that is, R2 represents moderate explanatory capacity (Hair et al. 2009).

In addition, the model had a GoF of 51.34%.

5.3 Checking the Hypotheses

Considering the results of the structure model and the correlation analysis, for the hypotheses presented in item 4, it is observed that:

- By modeling structural equations, hypotheses 1, 2, 3, 9 and 11 were confirmed since there was a positive ($B > 0$) and significant (p-value < 0.050) influence between the constructs.
- By modeling structural equations, hypotheses 4, 6, 8 and 12 were not confirmed since the relationship between the constructs was negative ($B < 0$) and not significant (p-value > 0.050).
- By modeling structural equations, hypotheses 5, 7, and 10 were partially confirmed since the relationship between the constructs was positive ($B > 0$) but was not significant (p-value > 0.050).
- By correlation, hypotheses 13, 14 and 15 were confirmed since there was a positive ($r > 0$) and significant (p-value < 0.050) correlation between the constructs.

6 Final Considerations

The relevance of Asset Management for the electricity sector is clear, in view of the total dependence on its performance on the assets themselves, which, if managed, favor the performance of companies by maximizing availability and reliability.

Trying to contribute to this process, this research aimed to evaluate the factors that promote performance in the implementation of the Management System of Asset Information in the Brazilian Electricity Sector within the scope of Asset Management, analyzing the relationships (impacts and correlations).

With regard to the theoretical model, the definition of Performance Constructs (Cost, Risk and Performance), whose balance enables the benefits of Asset Management; as well as the IAM Asset Management Maturity Model from the perspective of Group 4 -

Asset Information and its four themes: IAM-22 - Asset Information Strategy, IAM-23 - Asset Information Standards/Specifications, IAM -24 - Asset Information Systems and IAM-25 - Data and information management; were validated in this research, since:

- The linearity of the items (questions) of each construct was attested through the Bartlett test, proving that they are in the same direction and correlated with each other, thus validating the theoretical model.
- All items (questions) in the measurement model had a factor load greater than the minimum of 0.5, proving that all of them contribute significantly to the formation of the related construct.
- There was discriminant validation in all constructs since the maximum value of the cross-factorial loads of each item was lower than the respective factor loads minimum factor load for validating the measurement model, that is, each item is properly related to its construct.
- There was convergent validation in all constructs, ensuring that the items in the constructs are correlated enough to measure the latent concept.
- The reliability of the constructs was evidenced through the reliability indexes AC – Cronbach's Alpha and/or CC - Composite Reliability.

In this context, individuals tended to agree with the constructs, with greater agreement with the constructs related to Asset Management Performance, indicating that companies in the electricity sector need to pay greater attention to the Asset Information Management Systems within the scope of asset Management.

As the result of the modeling of structural equations, the explanatory capacity of the constructs related to the Asset Information Management Systems was verified, with the constructs related to the Performance in Asset Management, concluding that:

- AMI-22, AMI-23, AMI-24, and AMI-25 were able to explain 41.78% of the cost variability (moderate).
- IAM-22, IAM-23, IAM-24, and IAM-25 were able to explain 53.45% of the variability of the Risks (substantial).
- AMI-22, AMI-23, AMI-24, and AMI-25 were able to explain 34.82% of the performance variability (moderate).

Despite the good explanatory capacity of the constructs related to the Asset Information Management Systems, using the structural equation model, hypotheses H1 (IAM-22 → Costs), H2 (IAM-22 → Risks), H3 (IAM-22 → Performance), H9 (IAM-24 → Performance) and H11 (IAM-25 → Risks), hypotheses 5 (IAM-23 → Risks), 7 (IAM-24 → Costs), and 10 (IAM-25 → Costs), and hypotheses H4 (IAM-23 → Costs), 6 (IAM-23 → Performance), 8 (IAM-24 → Risks) and 12 (IAM-22 → Performance) were not confirmed.

Using the Spearman correlation coefficient, it was found that there was a significant and positive correlation between all the constructs of the model), thus, the higher the value of one indicator, the higher the value of other indicators tends to be, also confirming Hypotheses H13, H14 and H15, that is, the correlation between Cost, Risk and Performance, as foreseen in the theoretical model.

References

1. FGV (2018) Análise do ambiente concorrencial do setor elétrico no Brasil, p 60
2. Furtado AS, Fodra M (2020) Estudo comparativo da rentabilidade em empresas do setor elétrico brasileiro no período pós-privatizações. Desenvolve Revista de Gestão do Unilasalle 9(3), Art. no 3. https://doi.org/10.18316/desenv.v9i3.7070
3. ANEEL (2020) Saiba mais sobre o setor elétrico brasileiro - Agência Nacional de Energia Elétrica - ANEEL. https://www.aneel.gov.br. acessado out. 05, 2020
4. Campbell JD, Jardine AKS, McGlynn J (2016) Asset Management Excellence Optimizing Equipment Life-Cycle Decisions
5. ABNT (31 Jan 2014) ABNT NBR ISO 55000 - Gestão de ativos — Visão geral, princípios e terminologia
6. IAM (2015) IAM - Subjects 22, 23 and 25: Asset Information, Strategy, Standards and Data Management. https://theiam.org. acessado set. 29, 2020
7. ICA e Procobre (2020) Gestão de Ativos: oportunidades e desafios para o setor elétrico
8. Kardec A, Esmeraldo J, Lafraia JR, Nascif J (2014) Gestão de Ativos. Qualitymark
9. Tancredo FVP (2018) Modelo para determinação da maturidade de uma organização na gestão de ativos físicos. masterThesis, Instituto Superior de Engenharia de Lisboa, 2018. Acessado: abr. 02, 2021. Disponível em: https://repositorio.ipl.pt/handle/10400.21/8841
10. Arese MC (2018) maturidade da gestão de ativos físicos de processo: uma perspectiva sustentável baseada na pas-55, utilizando lógica nebulosa | Doutorado em Sistemas de Gestão Sustentáveis, dez. 18
11. IAM – BSI (2019) IAM - BSI PAS 55 - Asset Management. https://theiam.org/knowledge/bsi-pas-55/. acessado set. 29, 2020
12. Arese MC, França SLB, Bonina N, Denizot de AM, Domingos LC (2017) Gestão de ativos, modelo de maturidade e sustentabilidade: Uma análise bibliométrica, pp 22
13. IAM (2015) Asset Management - an anatomy. Disponível em: https://theiam.org
14. Vinzi VE, Chin WW, Henseler J, Wang H (2010) Handbook of Partial Least Squares. Springer, Chem
15. Ringle CM, da Silva D, de S. Bido D (2014) Modelagem de Equações Estruturais com Utilização do Smartpls. Revista Brasileira de Marketing 13(2), Art. no 2. https://doi.org/10.5585/remark.v13i2.2717
16. Pilati R, Laros JA (2007) Modelos de equações estruturais em psicologia: conceitos e aplicações. Psicologia: Teoria e Pesquisa 23(2):205–216. https://doi.org/10.1590/S0102-37722007000200011
17. Hair JF, Black WC, Babin BJ, Anderson RE, Tatham RL (2009) Análise multivariada de dados - 6ed. Bookman Editora
18. Leimeister S (2010) IT Outsourcing Governance: Client Types and Their Management Strategies. Springer Science & Business Media, Berlin
19. Fornell C, Larcker DF (1981) Evaluating structural equation models with unobservable variables and measurement error. J Market Res 18(1):39–50. https://doi.org/10.2307/3151312
20. Henseler J, Ringle C, Sinkovics R (2009) The use of partial least squares path modeling in international marketing. Adv Int Market 20:277–319. https://doi.org/10.1108/S1474-7979(2009)0000020014
21. Nunnally JC, Bernstein IH (1994) Psychometric Theory. McGraw-Hill, New York
22. Chin W, Marcoulides G (1998) The partial least squares approach to structural equation modeling. Mod Methods Bus Res 8:295–336
23. Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C (2005) PLS path modeling. Comput Stat Data Anal 48(1):159–205

24. Barclay D, Thompson R, Higgins C (1995) The partial least squares (PLS) approach to causal modeling: personal computer use as an illustration. Technol Stud 2:285–309
25. Kaiser HF (1958) The varimax criterion for analytic rotation in factor analysis. Psychometrika 23(3):187–200. https://doi.org/10.1007/BF02289233
26. Baumgartner H, Homburg C (1996) Applications of structural equation modeling in marketing and consumer research: a review. Int J Res Market 13(2):139–161. https://doi.org/10.1016/0167-8116(95)00038-0
27. Efron B, Tibshirani R (1993) An Introduction to The Bootstrap. Chapman & Hall, New York
28. Hollander M, Wolfe DA, Chicken E (1999) Nonparametric Statistical Methods, 3rd edn. Wiley, New York
29. Creswell J, Creswell J (2018) Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 5a edição. Sage Publications, Los Angeles
30. Rocha JA (2017) Livro completo sobre métodos de pesquisa, fev. 31
31. Pinsonneault A, Kraemer K (1993) Research methodology in management information systems. J Manage Inf Syst Special Sect Strateg Compet Inf Syst Arch 10:75–105
32. ICA e Procobre (2018) Gestão de Ativos - Guia para aplicação da norma ABNT NBR ISO 55001 - Considerando as diretrizes da ISO 55002:2018
33. de Figueirêdo AG (2019) A importância da Gestão dos Ativos Técnicos para o Negócio de Distribuição de Energia Elétrica no Brasil, vol 01, pp 18
34. Duarte LHS et al (2019) Uma reflexão sobre a Transmissão no Brasil sob o prisma das melhores práticas em gestão de ativos
35. ABNT, "ABNT NBR ISO 55001 - Gestão de ativos - Sistemas de gestão — Requisitos". jan. 31, 2014
36. ABNT (2020) ABNT NBR ISO 55002 - Gestão de ativos — Sistemas de gestão — Diretrizes para a aplicação da ABNT NBR ISO 55001. abr. 30, 2020
37. LGAM (2020) Asset Management Checklist - LGAM Knowledge Base. http://www.lgam.info/asset-management-checklist. acessado out. 01, 2020
38. Vaziri e M. Mohsenzadeh R (2012) A Questionnaire-Based Data Quality Methodology. ResearchGate 2012. https://www.researchgate.net/publication/276198070_A_Questionnaire-Based_Data_Quality_Methodology. acessado out. 01, 2020
39. Hawkins DM (1980) Identification of Outliers. Chapman and Hall, London
40. Ross SM (2010) Introduction to Probability Models, 10th edn. Academic Press, Amsterdam, Boston
41. Mingoti SA (2005) Análise de Dados Através de Métodos de Estatística Multivariada: Um Enfoque Aplicado. UFMG

Benchmarking Asset Information Quality of a Utility Company in Brazil

Bruno Pinto Vieira, José Ricardo Gonçalves, Bárbara Guimarães Penna, Marco Antônio Calijorne Soares, and Fernando Silva Parreiras(✉)

LIAISE Pesquisa Desenvolvimento Inovação, Av. Afonso Pena, 4000 – 30.130-009, Belo Horizonte, MG, Brazil
{ma.calijorne,fparreiras}@liaise.tech

Abstract. Context: While asset information management and asset information monitoring play an important role in asset management, low levels of asset information quality can have consequences, such as missed business opportunities, inadequate decisions and flawed risk analysis. **Objective:** To perform a gap analysis on the perception of employees of a Utility Company in Brazil, about the asset information quality in their respective areas, roles, segments and hierarchical levels. **Method**: We applied a questionnaire based on the AIMQ model [20], using five-point Likert items. We collected 70 answered questionnaires. The AIMQ model assesses asset information quality on the following dimensions: Accessibility, Appropriate, Amount, Believability, Completeness, Conciseness, Representation, Consistency, Representation, Ease of Operation, Free-of-Error, Interpretability, Objectivity, Relevance, Reputation, Security, Timeliness, and Understandability. **Results:** The results present a small satisfaction rate, considering hierarchical levels and professional work experience. Only the "Relevance" dimension reached an average higher than 3, which illustrates the low perception of respondents about the quality of information. The analysis of the results (using quality standards, verification metrics, and the created rules) shows that only two samples reached a percentage above 60% accuracy, reaching an overall average of only 25.95%. Regarding completeness, the global average was slightly higher, reaching 35.23%. Some resulting correlations from this study are: Free-of-Error x Objectivity; Accessibility x Ease of Operation; Understandability x Interpretability; Believability x Free-of-Error. **Conclusions:** Benchmarks developed from the AIMQ model help comparing asset information quality across organizations, and provide a baseline for assessing IQ improvements. This study presents important correlations, choosing further automatic techniques for evaluating data quality (DQ) on asset management databases.

1 Introduction

Nowadays, a large amount of information has been produced by various systems, transactional or not, including mobile devices [15]. In 2012, about 2.5 exabytes of data were created per day, and that number doubles every 40 months. Today,

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 287–299, 2022.
https://doi.org/10.1007/978-3-030-96794-9_26

more data travels over the internet every second than all the storage carried out in the last 20 years for this same source. This allows companies to work with a large amount of data [17].

In this context, a concern arises to establish and maintain the Quality of Information (IQ), which has come to be treated as a product that needs to be defined, measured, analyzed and constantly improved to meet the needs of consumers, meaning that the processes use these information in decision making [1,2]. Low levels of IQ can have long and short-term consequences, such as missed business opportunities, inadequate decisions and issues in risk analysis [2,8,21]. In addition to the importance of IQ for the business, its management and monitoring also plays a fundamental role, as the company systems, users and goals are constantly evolving [24].

In companies around the world, Asset Information Management areas deal with huge amount of data and can suffer from low IQ. This area is responsible for monitor, control and deal with all companies assets, especially the ones that are related to their services or products [6,12,22]

Considering this scenario, Asset Information Management represents huge importance for the business health and growth, and the data which is observed in their process could impact on decision-making. The present study aims to assess the information quality of an electricity generation and distribution company in the state of Minas Gerais - CEMIG to answer the following research question: **What is the perception of employees about the information quality processed in Asset Information Management area?**

The perception of employees in the Asset Management sector of the company CEMIG was evaluated. In order to carry out this evaluation, a questionnaire was used, built according to the Methodology for Information Quality Assessment (AIMQ) model [13]. The AIMQ model presents several dimensions to assess the quality of information, such as accuracy, integrity, consistency and validity [13]. Dimensions are not only used to measure the level of DQ, but also to identify gaps and opportunities for further improvement. Therefore, understanding and identifying the dimensions is necessary to provide meaningful assistance to measure, manage and improve the quality of data regarding the process improvement goal [18].

In this study, we were motivated to identify systems and databases gaps regarding their information quality, aiming to build an Asset Register to support a three area (financial, operational and regulatory) asset management decision making structure.

Our study paper is organized as Sect. 2, where we provide the related work; Sect. 3, which addresses the background for concept understanding; Sect. 4 we present the gap analysis, it's methods and execution, Sect. 5, details our findings, and finally Sect. 6, which consists in our conclusion.

2 Related Work

Data quality assessment can be defined by identifying erroneous data, missing data, and measuring the impact of various data-driven business processes. The

DQ management is an important process for assessing data quality, as understanding several different steps involves several groups of people within the organization.

The studies [13,16,19], the AIMQ methodology forms a basis for assessing the Quality of Information and benchmarking. The authors report the use of a questionnaire to measure IQ, and analysis techniques to interpret IQ measures. These are three components of the methodology for structuring what IQ means to consumers and information managers. The first one is the 2×2 model, which has four quadrants to assess whether the information is a product or a service, and positions whether improvements can be assessed in relation to a formal specification or customer expectation. The second component is the questionnaire that measures IQ according to IQ dimensions. Several of these dimensions together measure the IQ for each quadrant of the 2×2 model. The third component is the analysis techniques to interpret the assessments captured by the questionnaire.

Given the diversity of techniques presented to evaluate and improve the quality of the data, the work [3] focused on the definition of methodologies that assist in the selection, customization and application of evaluation and improvement techniques in order to provide a systematic comparison of such methodologies. A methodology is considered to be included in a phase or step if it provides at least a discussion of the corresponding phase or step and possibly original methodological and technical guidelines. In this study, thirteen methodologies used for measuring and improving DQ were compared. Different approaches were used and the measurement of quality steps were carried out with questionnaires in the AIMQ, combining subjective and objective metrics in the Data Quality Assessment (DQA) [4,20], or with statistical analysis in the methodology for the Quality Assessment on Financial Data (QAFD) [4]. QAFD recommends collecting target quality levels including business experts and financial operators, but does not assist in reconciling incompatible DQ levels. In contrast, the management of the improvement solution step is explicitly performed only by Total Data Quality Management (TDQM) [7,9].

The study [11,14,24] reports the importance of IQ for a dataset to achieve a specific objective (scientific or practical), using a certain method of empirical analysis. It also reports the difference in IQ in relation to QD and the Quality of Analysis (QA), however, it depends on these components and the relationship among them. Our study uses statistical methods to increase IQ in the study design and post-data collection phases. Eight dimensions were used for this evaluation: data resolution, data structure, data integration, temporal relevance, generalization, data and objective timeline, operationalization of the construct, and communication.

Literature reports how professionals and researchers recognize the importance of achieving and maintaining DQ. In this context, this study [5], comparatively analysis the DQ structures focused on methodologies that apply in a wide range of business environments with a decision guide for different structures.

3 Dimensions of Information Quality

Dimensions are attributes that, when measured correctly, indicate the overall level of IQ. Depending on the focus, different structures recognize different attributes for IQ in their methodology. The dimensions are highly context-dependent and their relevance and importance can vary from one organization to another and according to data nature [5].

Dimensions are not only used to measure the level of IQ, but also to identify gaps and opportunities for further improvement. Therefore, understanding and identifying dimensions is necessary to provide meaningful assistance to measure, manage and improve IQ against the process enhancement objective [18].

The dimensions of IQ can be grouped into four categories: intrinsic, contextual, representational and accessibility. The intrinsic quality of data implies that the information has quality in itself. The contextual highlights that quality of data must consider in the context of the task in question must be relevant, timely, complete and appropriate, in order to add value. The representational and accessibility categories emphasize the importance of computer systems that store and provide access to information. In other words, the information must be interpretable, easy to understand and manipulate and represented in a concise and consistent way. In addition, the system must be accessible but secure [13]. AIMQ model recognizes DQ dimensions and classifications, however, the categorization differs from those mentioned above and considers four types: Sound, dependable, useful and usable information [5,10]. These categories are classified into four quadrants [5]. Further details of the quadrant assignments are as follows:

- **Sound Information:** The dimensions in this quadrant are tangible, measurable and the solidity of the information is generally independent of the task and decision. The information must be error free and well represented, and consistent representation ensures a minimum level of interpretability and understanding [10].
- **Dependable Information:** The dimensions in this quadrant generally can not be evaluated from the characteristics of the data in a database. Like any service, the delivery of information can only be evaluated after the occurrence. Reliable information is current, secure and provided in a timely manner to support the task at hand [10].
- **Useful Information:** The dimensions in this quadrant are task dependent. For the consumer's task, the information is relevant and sufficient to support decision making [10].
- **Usable Information:** The dimensions in this quadrant distinguish one service from the other. This can only be assessed from the consumers' point of view, based on the task or decision. These dimensions depend on the computer systems, existing between the consumer and the stored data [10].

The classification in four quadrants is called the **PSP/IQ model (Product and Service Performance Model for Information Quality)**, developed from the perspective of information of the consumers, in order to organize the dimensions so that significant decisions can be made about IQ improvement. This PSP/IQ model is the basis for the AIMQ methodology, which has a set of dimensions that cover important aspects of IQ for information consumers [13]. The dimensions are described in Table 1.

The accessibility dimension is considered as a quality of service and it can be assessed by consumer expectations, while the completeness dimension is a quality of the product that can be assessed by specifications. This classification is useful in the process of deciding which aspects of IQ require improvement [5]. Completeness is also defined in terms of missing values, but only for values used or needed by information consumers [13].

To measure the relevant dimensions using a questionnaire-based methodology is a method that can be applied in several purposes. First it provides a

Table 1. Data quality dimensions.

Dimensions	Definitions
Accessibility	The extent to which information is available, or easily and quickly retrievable
Appropriate Amount of Information	The extent to which the volume of information is appropriate for the task at hand
Believability	The extent to which information is regarded as true and credible
Completeness	The extent to which information is not missing and is of sufficient breadth and depth for the task at hand
Concise representation	The extent to which information is compactly represented
Consistent representation	The extent to which information is presented in the same format
Ease of manipulation	The extent to which information is easy to manipulate and apply to different tasks
Free-of-error	The extent to which information is correct and reliable
Interpretability	The extent to which information is inappropriate languages, symbols, and units and the definitions are clear
Objectivity	The extent to which information is unbiased, unprejudiced and impartial
Relevancy	The extent to which information is applicable and helpful for the task at hand
Reputation	The extent to which information is highly regarded in terms of its source or content
Security	The extent to which access to information is restricted appropriately to maintain its security
Timeliness	The extent to which the information is sufficiently up-to-date for the task at hand
Understandability	The extent to which information is easily comprehended
Value-added	The extent to which information is beneficial and provides advantages from its use

general measure of the current quality of information in the organization; second it identifies the dimensions that require urgent and complete attention in the organization, and third it provides values for evaluation purposes. That is, the evaluation occurs when we compare data to reference values, which in this case, can be an ideal organization with high quality data [23].

4 Methods

To perform the GAP analysis proposed in this study, we developed a questionnaire according to the AIMQ method, listing dimensions for data quality analysis. Respondents initially provided information regarding their level of education, time of professional experience, business segment in which they operate and current position. The answers were given anonymously, preventing respondents of their unique identification.

Our analysis considered 70 participants, employees of an electricity generation, transmission and distribution company in the state of Minas Gerais in the area of Asset Management.

Each question has a set of answer options related to a numerical value. Table 2 shows the answer options for each dimension of the IQ.

Table 2. Set of answer options for questions.

Dimensions	Answer options
Frequency	Never: 0
	Rarely: 1
	Sometimes: 2
	Oftentimes: 3
	Ever: 4
Accessibility	Strongly disagree: 0
Completeness	Partially disagree: 1
Understandability	I don't agree nor disagree: 2
Believability	Partially agree: 3
Ease of operation	Totally agree: 4
Interpretability	
Free-of-error	
Objectivity	
Timeliness	
Appropriate amount	
Relevancy	
Concise representation	
Consistent representation	
Reputation	
Security	

Some questions have a reverse answer (R), so the numerical values assigned to each answer have been inverted.

5 Results

When we analyzed the questionnaire responses, we were able to identify GAPs in the quality of asset management data, such as:

- Existence of 11 dimensions with very low score;
- Few areas influenced in the increase of information quality perception.

Analyzing the average score obtained in the questionnaires, considering the interviewee's experience and the dimensions of the IQ, we can conclude that the employees with more experience tend to have a more positive perception of the data quality. Professionals with time experience between 8 to 13 years and more than 21 years remained constant in all analyzed dimensions, but relevancy dimension, which determines the appropriate and applicable use of the information in the work performed presented a considerable increase among the respondents. The results can be seen in Fig. 1.

When the analysis is carried out considering the average score obtained in the questionnaires, and the respondents total experience time and dimensions, there is a scenario change, where more time of total experience does not translate into a better perception of IQ, as shown in Fig. 2.

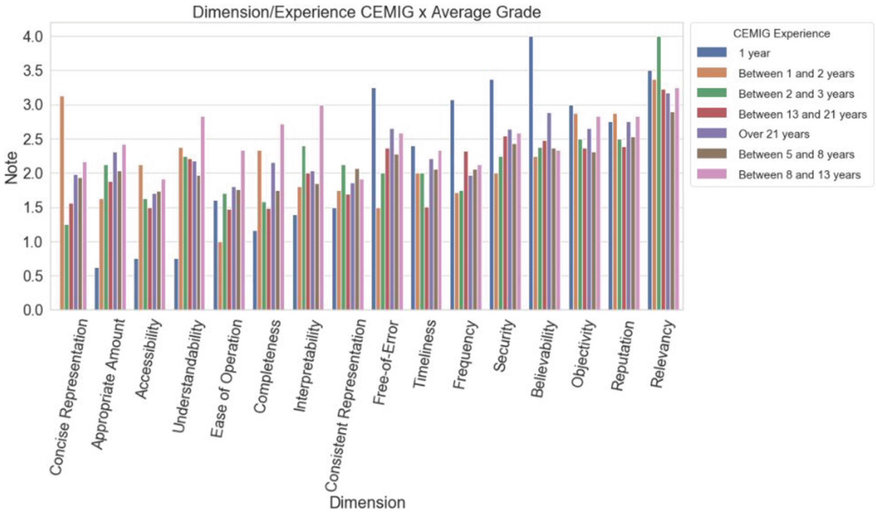


Fig. 1. Average score according to the experience of respondent's at CEMIG and the dimensions of analysis.

Respondents with total professional experience between 2 to 3 years demonstrate constancy in almost all dimensions evaluated, with emphasis on the dimension of accessibility, which means information that is easy to obtain, accessible and retrievable, and the dimension of relevance, which determines the proper use, and information applicable to the work performed.

Analyzing the average score obtained in the questionnaires in relation to the respondents' asset management experience time and dimensions, it is possible to see a dispersion in the perception of IQ, as shown in Fig. 3.

Analyzing the simple average of the scores by dimension, we can see that the respondent users had more difficulty with data/information access **(average grade 1.66)**. With the exception of the relevancy dimension, all the others did not reach an average score higher than 3, indicating that the perception of the respondents' information quality is not satisfactory. With the exception of the relevancy dimension, all the others did not reach an average score higher than 3, indicating that the perception of the respondents about information quality is not satisfactory.

According to the PSP/IQ model, the relevance dimension predominated in the respondents' analysis and, according to the quadrant models, it represents that the organizations information meets or exceeds the consumers expectations and it is useful information.

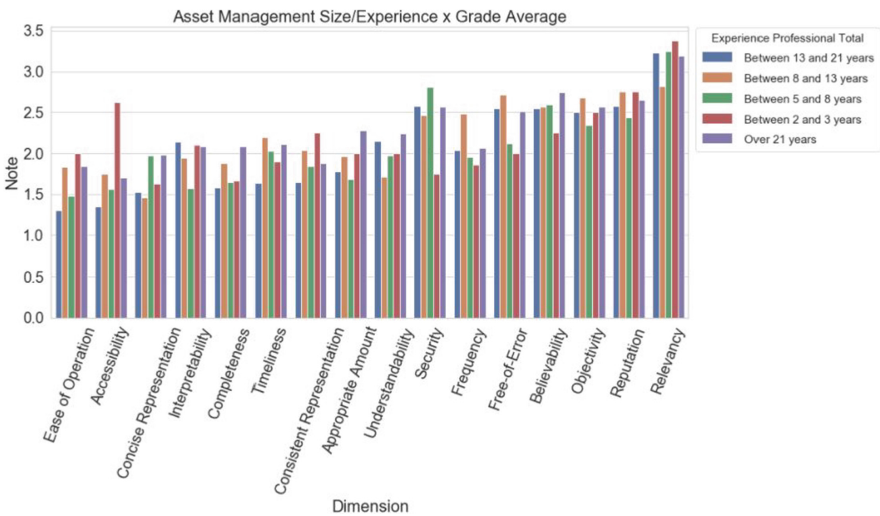


Fig. 2. Average score according to the respondent's total experience time and the dimensions of analysis

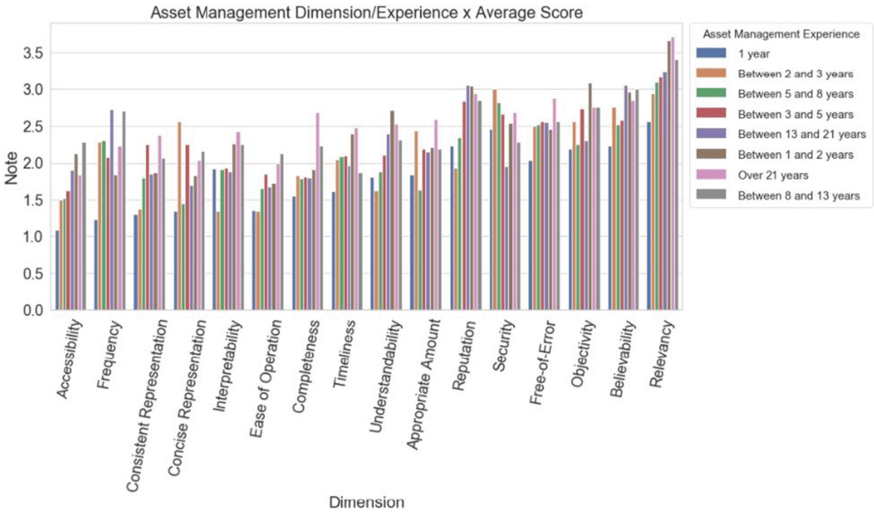


Fig. 3. Average score according to the respondent's total experience in asset management and the dimensions of analysis

5.1 Correlation

We performed a correlation analysis to verify the relationships among the dimensions. The value of the correlations can vary from −1 to 1. Values close to −1 indicate a strong negative correlation. Values close to 0 do not indicate the presence of correlation. Values close to 1 indicate strong positive correlation. The result of this analysis is shown in Fig. 4.

Checking the results of the correlation among the dimensions, where strong correlations are shown in a darker red shade and weaker correlations are shown in a lighter red shade, we can identify the following stronger correlations, meaning the ones that showed higher absolute values:

- Free-of-Error X Objectivity;
- Accessibility X Ease of Operation;
- Understandability X Interpretability;
- Believability X Free-of-Error;
- Believability X Objectivity;
- Objectivity X Relevancy;
- Completeness X Concise Representation;
- Completeness X Consistent Representation;
- Completeness X Appropriate Amount;
- Completeness X Timeliness;
- Free-of-Error X Reputation.

A simple analysis shows that Free-of-Error and Completeness dimension have a greater correlation in relation to the others, showing that they are the most important dimensions to be analyzed.

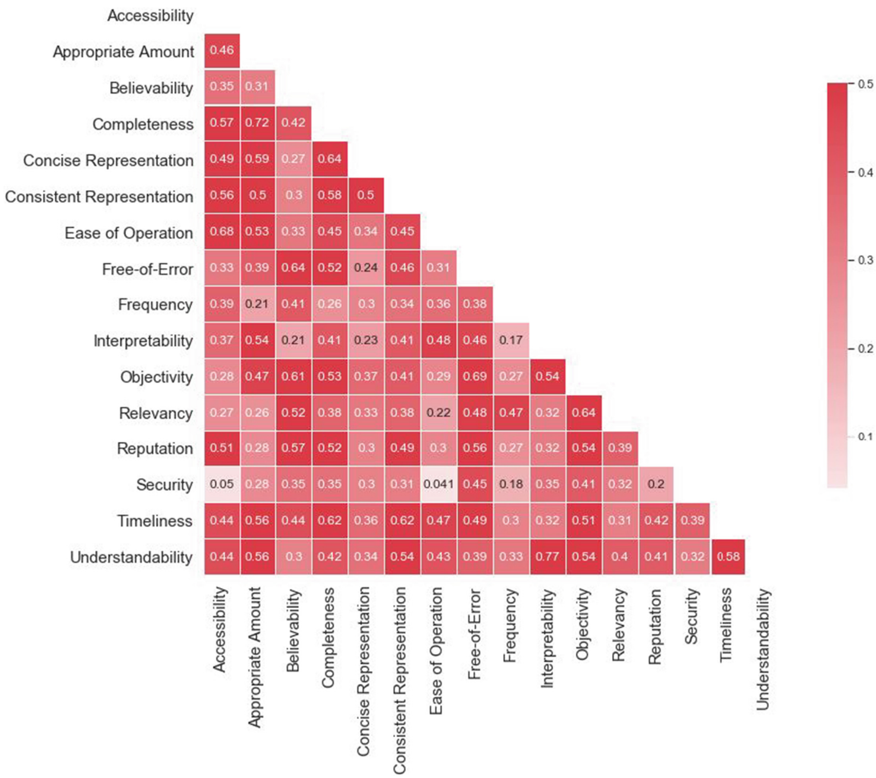


Fig. 4. Correlation of dimensions

5.2 Questionnaire Reliability

The assessment of items to measure each construct is calculated using Cronbach's alphas. This analysis measures the internal consistency of the questionnaire, that is, the extent to which the items that are part of this questionnaire measure the same concept. The alpha result is analyzed according to the values presented in Table 3, indicating the degree of consistency of the questionnaire [13].

Table 3. Degree of questionnaire consistency.

Alfa Value	Internal consistency
Bigger than 0,80	Almost perfect
From 0,80 to 0,61	Substantial
From 0,60 to 0,41	Moderate
From 0,40 to 0,21	Reasonable
Smaller than 0,21	Small

In the present study, the calculation of Cronbach's Alpha was performed using python language and reached a **0.9576** value, demonstrating that internal consistency is suitable.

6 Conclusion

AIMQ methodology and the applied questionnaire enabled us to observe that employees of Asset Management area discovered their data as low quality type.

The needs of the various areas related to asset management were raised, and gathered here, in order to identify the information gaps that must be addressed. As detailed in Sect. 4, the gap identification process started from the asset information management current scenario *("As is")*, carried out with a survey through the applied questionnaire, and the deepening performed in this work to identify the existing gaps. The transformation from *"As is"* to *"To be"* begins when we compare the reported results with the survey of the state of the art and the desired objectives.

The analyzes for the identification and classification of gaps were structured in three principles:

1. Thematic gaps, in terms of the asset management maturity scale.
2. Diagnosis of the data quality level that exists in the concessionaire.
3. Gaps related to the concessionaire's current systems and database, especially regarding to data consumption for the indicators and KPIs generation.

The gaps related to the diagnosis of maturity in asset management were classified into two fronts, Processes and Tools. The analyzes, based on the results, show that some processes relevant to asset management have not yet been defined and established. For instance, the asset criticality matrix, the process of deactivation and disposal of assets, the process of distinguishing between life assets and mature life, and a Strategic Asset Management Plan. Among the tools to support decisions, there is the absence of elements of analysis that consider: balance between corrective and preventive maintenance, the costs of risks (monetized risk), LCC techniques for prioritizing investments, business penalties, trade off between possessions and costs of possessions as to parts inventories.

With regard to gaps in the perception of IQ, it is concluded that the professionals view varies according to the correspondent work experience, since the longer the professional works in the organization, a better perception of IQ was reported in the study. Another conclusion is regarding how employees sees quality on their that when they have access to it, not having permission to see information is, for an expressive amount of interviewed people a label of low information quality.

Future works can be addressed integrating emerging technologies, such as, artificial intelligence, seeking to automate data quality analysis, lowing operational costs and increasing results.

References

1. Araújo CAÁ (2014) Fundamentos da ciência da informação: correntes teóricas e o conceito de informação. Perspectivas em Gestão & Conhecimento 4(1):57–79
2. Barua A, Mani D, Mukherjee R (2012) Measuring the business impacts of effective data. http://www.sybase.com/files/White_Papers. Accessed 15 Sep 2012
3. Batini C, Cappiello C, Francalanci C, Maurino A (2009) Methodologies for data quality assessment and improvement. ACM Comput Surv (CSUR) 41(3):1–52
4. Blomqvist N (2019) Data quality methodologies and improvement in a data warehousing environment with financial data. Ph.D. thesis, Lappeenranta University of Technology
5. Cichy C, Rass S (2019) An overview of data quality frameworks. IEEE Access 7:24634–24648
6. Faiz R, Edirisinghe EA (2009) Decision making for predictive maintenance in asset information management. Interdiscip J Inf Knowl Manag 4(1):23–36
7. Francisco MM, Alves-Souza SN, Campos EG, De Souza LS (2017) Total data quality management and total information quality management applied to costumer relationship management. In: Proceedings of the 9th international conference on information management and engineering, pp 40–45
8. Friedman T, Smith M (2011) Measuring the business value of data quality. Gartner, Stamford 464
9. Ibrahim A, Mohamed I, Satar NSM (2021) Factors influencing master data quality: a systematic review. Int J Adv Comput Sci Appl (IJACSA)
10. Kahn BK, Strong DM, Wang RY (2002) Information quality benchmarks: product and service performance. Commun ACM 45(4):184–192
11. Kenett RS, Shmueli G (2014) On information quality. J R Stat Soc A Stat Soc 177(1):3–38
12. Kyle BR, Vanier DJ, Kosovac B, Froese TM (2000) Information needs towards service life asset management. In: Proceedings of the 17th international conference of the committee on data for science and technology, CODATA, Baverno, Italy, pp 15–19
13. Lee YW, Strong DM, Kahn BK, Wang RY (2002) AIMQ: a methodology for information quality assessment. Inf Manage 40(2):133–146
14. Liaw ST et al (2021) Quality assessment of real-world data repositories across the data life cycle: a literature review. J Am Med Inf Assoc 28:1591–1599
15. Lóscio BF, Burle C, Calegari N (2016) Data on the web best practices: challenges and benefits. In: ODRS 2016: open data research symposium
16. Magdalena R (2018) AIMQ: methodology for accounting information system technology assessment. Int J Curr Res
17. McAfee A, Brynjolfsson E, Davenport TH, Patil D, Barton D (2012) Big data: the management revolution. Harv Bus Rev 90(10):60–68
18. Panahy PHS, Sidi F, Affendey LS, Jabar MA (2014) The impact of data quality dimensions on business process improvement. In: WICT 2014: 2014 4th world congress on information and communication technologies. IEEE, pp 70–73
19. Pecina, R.: Information quality research on close range spatial location providing services using AIMQ from the user perspective. Ph.D. thesis, University of Nebraska at Omaha (2017)
20. Pipino LL, Lee YW, Wang RY (2002) Data quality assessment. Commun ACM 45(4):211–218

21. Silva Neto EC (2016) Um perfil de qualidade para fontes de dados dinâmicas. Master's thesis, Universidade Federal de Pernambuco
22. Vanier DD (2001) Why industry needs asset management tools. J Comput Civ Eng 15(1):35–43
23. Vaziri R, Mohsenzadeh M (2012) A questionnaire-based data quality methodology. Int J Database Manage Syst 4(2):55
24. Zadeh PA, Wang G, Cavka HB, Staub-French S, Pottinger R (2017) Information quality assessment for facility management. Adv Eng Inform 33:181–205

The Value of Business Process Management to Understand Complex Asset Management Processes

Kanika Goel[1](✉), Michael E. Cholette[1], Moe T. Wynn[1], Lutfiye Manli[2], and Lara Meyers[2]

[1] 2 George Street, Brisbane, QLD 4001, Australia
goelk@qut.edu.au
[2] 33 Harold Street, Virginia, QLD 4014, Australia

Abstract. Asset Management (AM) processes play a significant role in organisations' profitability. Clearly documented and managed AM processes improve the delivery potential of assets and minimise the costs and risks involved. Business Process Management (BPM) is a discipline that uses various methods, tools, and techniques to discover, model, analyse, measure, improve, optimise, and automate business processes. Despite the prevalence and proven effectiveness of BPM in a wide variety of domains, there has been little research investigating its potential for describing AM processes. This paper presents a case study that explores the application of BPM to power transmission assets. BPM principles were applied for decision modelling and to capture the lifecycle of power transmission assets. The case study demonstrates how BPM application to AM processes can result in greater clarity of processes, increased collaboration, a better understanding of data, external rules, and regulations, and serve as an internal point of audit.

1 Introduction

Organisations worldwide spend trillions of dollars managing their portfolio of assets [1], and improvements in the management of physical assets remain one of the largest business improvement opportunities of the 21st century [2]. An organisation often conducts a wide variety of interlinked Engineering Asset Management (EAM) activities (e.g., data collection, condition assessment, and refurbishment), resulting in complex processes that involve a diverse team of personnel from different business units. Consequently, it is often quite challenging to obtain a comprehensive view of the tasks, data, supporting standards, etc., involved in key asset decisions. However, a comprehensive view of EAM processes is essential to evaluate their quality, communicate EAM processes clearly across the business to enable a common understanding, and understand where process improvements will have the most significant impact [3]. Managing these complex EAM processes improves business performance [4] and integrating core EAM practices in business is evidenced to improve operational performance [5].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 300–311, 2022.
https://doi.org/10.1007/978-3-030-96794-9_27

Business Process Management (BPM) is a discipline that focuses on enhancing ways organisations identify, describe, analyse, support, and monitor their business processes in order to increase their operational performance [6]. Several industries, e.g., healthcare, finance, and retail, have used BPM to improve their business processes. However, its use in asset management remains relatively inconspicuous. Techniques such as process architecture and process modelling can be used for EAM processes by organisations to identify their capabilities and visualise, describe, and understand their business processes. Additionally, BPM can help develop data requirements, coordinate different EAM activities with the personnel, generate workflow to develop EAM information systems and assist in integrating the current EAM information system with other systems [3].

Given the apparent advantages of applying BPM to EAM processes, this paper presents the application of BPM techniques, particularly process and decision modelling, at Powerlink Queensland—a State Government Owned Corporation that owns, develops, operates, and maintains 15,339 circuit kilometres of high voltage electricity transmission lines and 147 substations transmission network. Powerlink's asset management system seeks to ensure that assets are managed in consistence with asset management policy and overall corporate objectives to deliver cost-effective and efficient services. Powerlink was interested in identifying, visualising, and describing key decision-making processes during the asset's lifecycle. BPM techniques were applied at Powerlink over six months to develop process models elucidating key decisions taken during reinvestment of transmission asset lifecycle.

The rest of the paper is structures as follows. Section 2 presents the methodology adopted in this study. Section 3 presents the findings, and a discussion in followed in Sect. 4. The paper concludes with an overview of contributions, limitations, and future work.

2 Methodology

This section provides an overview of the key steps undertaken and the BPM techniques applied to identify, visualise, and describe key decision-making processes at Powerlink during the reinvestment of Transmission Line Assets (TLA).

2.1 Interviews

Interviews were the primary mode of data collection in this study. Interviews are an important data collection method in qualitative research [7] and one of the most popular methods for qualitative organisational research [8]. They assist in obtaining the personal perspective of the interviewee and create an interactional situation in a face-to-face encounter between researchers and participants. In this method, the interviewer acts as an instrument and, with the help of carefully designed questions, attempts to elicit the other person's opinions, attitudes or knowledge about a given topic. Semi-structured interviews were used in this project. According to Mabry [9], semi-structured interviews allow for "*probative follow-up questions and exploration of topics unanticipated by the interviewer, facilitate development of subtle understanding of what happens and why*".

Since the intention was to gather insights from Powerlink's stakeholders regarding the processes and key decisions involved during the reinvestment of TLAs, semi-structured interviews were considered suitable. Interviews were conducted with subject matter experts (SMEs) of various teams involved in the reinvestment of TLAs. Nine SMEs were interviewed in total, which included two SMEs from bigger departments and one SME from smaller departments involved in reinvestment of TLAs. Table 1 provides a profile of respondents involved in this study.

The interview started with questions regarding the role and previous decisions made. Following this, the questions centred around the key activities undertaken in their teams, interaction with other teams, external documents, regulations used, and the role of IT systems. The interviews concluded by asking SMEs about room for improvement at Powerlink. Each interview lasted for around 60 min and was audio recorded with the permission of the respondent. The interviews were then transcribed, and the information was used to model the initial decision-making process. Once the entire process model was derived, follow-up interviews were held with the SMEs to verify and validate the process model. Follow-up interviews lasted under 30 min. The process model was explained to the SME, and they commented on the appropriateness of the process model. The process model was updated based on feedback from follow-up interviews. This process continued until all the SMEs were satisfied with the output.

Table 1. Overview of interview respondents.

Department	# of SMEs	Role
Line Strategies	2	Provide options to replace/refurbish based on corrosion data
Project Portfolio and Operations (PPO) and Network and Alternative Solutions	2	Decide to replace/refurbish TLA
Maintenance facilitator	1	Facilitate communication among maintenance planner, line strategies, and projects
Projects	2	Costing and planning resources for reinvestment options
Maintenance planner	1	Programming SAP for maintenance actions
Works Control Manager	1	Responsible for routine and corrective inspection management

2.2 Document Analysis

While interviews were the primary form of data collection in this project, document analysis was also conducted. Document analysis involves using both printed and electronic documents by the research team to give voice and meaning to the topic. Like other qualitative methods, document analysis requires that the data be interpreted to obtain meaning, gain understanding, and develop empirical observations [10]. Therefore, the research team collected and analysed documents such as internal policies, external and internal standards, and organisation's operating procedures before, during, and after interviews to augment the contextual information gained during interviews. The documents also assisted in filling in the gaps when building the initial process model, which were verified through follow-up interviews.

2.3 Process and Decision Modelling

Process modelling, a BPM technique, was used to identify, visualise, and describe the process during the reinvestment of TLAs. Organisations are increasingly structuring themselves around their business processes to improve responsiveness to business opportunities and threats, and also for adopting integrated software solutions that support the organisation's core business process needs [11, 12]. To support these aims, a better understanding of business processes is required. "*Process modelling is an approach for visually depicting how businesses conduct their operations; defining and depicting business processes, including entities, activities, enablers and the relationships between them*" [13]. Process models enable capturing systems, data, people, and control flow into a logical framework, which can thereafter be used for analysis and improvement purposes. Process modelling for EAM can assist in visualising EAM processes, demystifying interaction among various personnel, and generating workflow, which can help improve EAM processes. EAM processes have been widely used to guide EAM practices; however, these processes are usually modelled using a flowchart. Flowcharts lack comprehensiveness, which is essential to capture the complexity involved in EAM processes [3]. To capture the complexities of EAM processes in business process models, we used Business Process Model and Notation (BPMN), version 2.0. BPMN provides businesses with the capability to express their business processes using a graphical notation, which enables enterprises to communicate these processes in a standard manner. BPMN is maintained by Object Management Group (OMG) since 2005. Complete details related to BPMN2.0 notation are available at OMG [14]).

Additionally, we used decision modelling to model the key decisions undertaken by the stakeholders during the reinvestment of TLAs. In most process models, decisions are embedded within the models and scattered over constructs of process models, resulting in difficulty in maintainability [15]. Decision modelling assists in modelling decisions in a more precise and transparent manner, separately from the process [16]. According to recent research, decision modelling alongside processes enables better management of complexity and supports the flexibility of processes [17]. To model decisions, we used Decision Model and Notation (DMN), a recent standard of the OMG. It allows efficient detailing and modelling of repeatable decisions in an organisation [16]. DMN provides a notation that is readable by businesses as well as IT users. DMN has been designed to work with BPMN providing a separation of concerns between the decisions and the process [18].

3 Findings

Based on the interviews with SMEs, we built a comprehensive process model encapsulating TLA lifecycle decisions taken during reinvestment. One over-arching process model and seven linked DMN sub-models were developed. Figure 1 shows this process model, which was built using BPMN and DMN on Signavio[1].

Each lane refers to a team at Powerlink. The rectangular boxes with rounded ends indicate an activity conducted by a resource in a team. A rounded rectangular box with a table at the top left shows a decision taken, and when the table is clicked, the DMN sub-model is displayed, which provides more detail on the decision-making process. A document attached to an activity shows an external or internal regulation used for the activity. Similarly, an IT system attached to an activity displays the IT system used by the activity. A diamond symbol refers to a gateway. A diamond with an 'X sign' is an XOR gateway, which communicates the process instance's optional paths from the previous activity. A diamond with a '+ sign' in an AND gateway indicates parallelism, i.e., both paths are undertaken in any order from the previous activity to proceed to the next activity.

3.1 Description of Process Model

The Works Control Manager (WCM) prioritises routine and corrective maintenance tasks. Figure 2 shows the data, which is used as input by the WCM.

In DMN notation, the rectangular box (with a table icon at the top left) shows the decision that is to be taken. The rectangle with a rounded left and right side displays the input required to make the decision. The document-shaped box displays the special knowledge necessary to execute the decision.

Prioritisation of maintenance activities is a decision taken by the WCM as they are responsible for the dissemination of routine and corrective maintenance work orders. To prioritise these activities, the key input used are budget, KPIs, and data from SAP. Additionally, expert advice from the lines maintenance facilitator is also used. A priority risk system is the IT system, which is used to prioritise the tasks.

Once the tasks are prioritised, the WCM assesses the restrictions in discussion with network planning/operations. The planned outages document is used as an input here. Based on these details, the corrective and routine maintenance tasks are planned with the assistance of MS project. This data is fed into the works manager system. Then, work orders are assigned to Maintenance Service Providers (MSPs). On receiving the work order from the WCM, the MSPs conduct inspection using works manager or Line Asset Measuring Points (LAMP). The data to be collected is documented in detail in Appendix E of the maintenance specification document of Powerlink. The data is directly updated from the works manager in SAP, which finishes the work order. The data collected using LAMP is checked for quality by the maintenance planner. The maintenance planner then performs basic consistency and quality checks on inspection data, which is then

[1] https://www.signavio.com/.

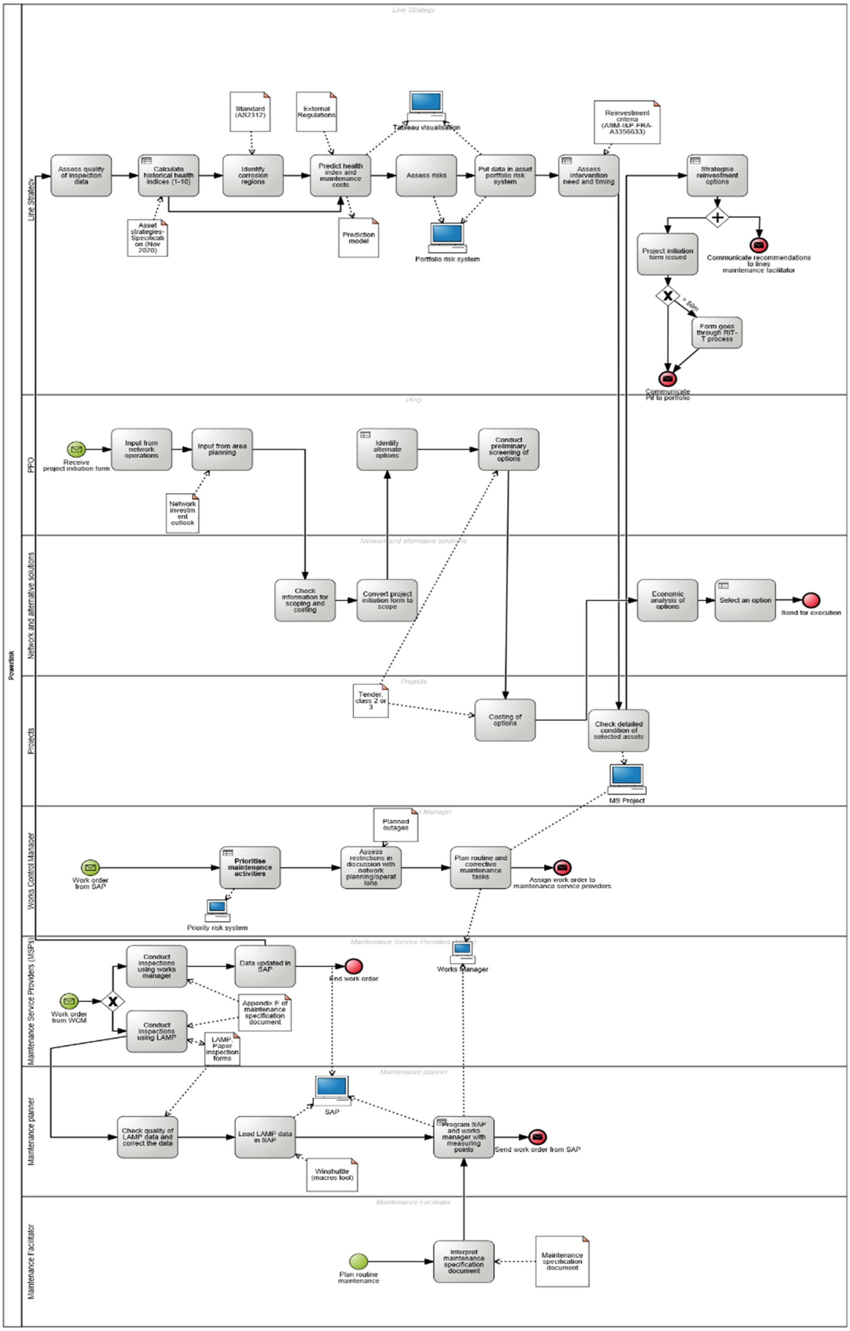


Fig. 1. A Process model capturing key decisions made during the refurbishment and end-of-life phase of TLA.

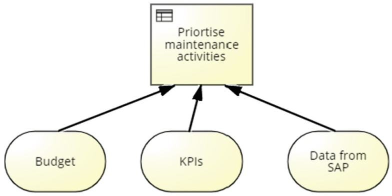


Fig. 2. A DMN model to prioritise maintenance activities.

loaded into SAP. The winshuttle macros tool is used to load the data. Additionally, the maintenance planner programs SAP and works manager with measuring points used by MSPs to enter inspection data. Figure 3 displays the input data that is used to program SAP. This programming is done in consultation with the maintenance facilitator, who interprets the maintenance specification document and provides expert input.

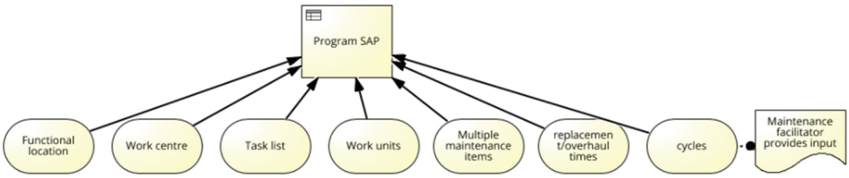


Fig. 3. A DMN model depicting data required to program SAP.

The Line Strategies team uses the data updated in SAP regarding the inspection of TLA. They first assess the quality of data. When the data is ascertained to be of adequate quality, the health index of built sections of TLAs is calculated. The input required for the calculation of the health index is captured in Fig. 4.

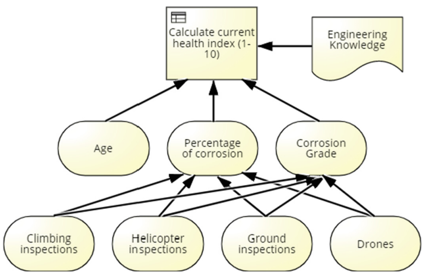


Fig. 4. A DMN model to calculate the current health index.

The age of the asset, percentage of corrosion, and corrosion grade are the three main inputs that determine the health index. Percentage of corrosion and corrosion grade is calculated using the data recorded in SAP during climbing, helicopter, ground, and drone inspections. Input from SAP and engineering knowledge that comes from experience helps in deriving the health index of the built section of TLAs. The ‘asset strategies specification (Nov 2020)' document provides details regarding the calculation of the health index. Then the corrosion regions are identified using the standard AS2312. Next, the health index and maintenance costs are predicted. External regulations, such as the Regulatory Investment Test for Transmission (RIT-T), play an important role. Predictions are made using a prediction model, visualised on Tableau. Following this, the risks are assessed using the portfolio risk system, which is also visualised on Tableau. Based on all the data calculated so far, the timing for intervention is assessed. This decision is captured in Fig. 5.

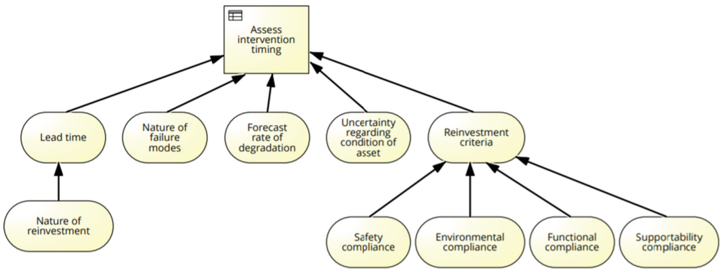


Fig. 5. A model to decide intervention timing.

Figure 5 displays the key input required to make the decision. Nature of failure modes, rate of degradation forecasted using health index and visualised on Tableau, and uncertainty regarding the condition of an asset directly impact assessment of when should intervention be made. Further, the lead time, which is dependent on the nature of the investment, is another input to this decision. Finally, the reinvestment criteria, which is an internal document at Powerlink, also influence intervention timing. The reinvestment criteria outline four areas that need to be considered when deciding on intervention – safety, environmental, functional, and supportability compliance. To further strategise reinvestment options and their timing, the Line Strategies team sends the identified need and options to projects. The Projects team does a detailed assessment of the asset's condition and sends that information back to the Line Strategies team. This data is then used to strategise reinvestment options. The details related to this decision are captured in Fig. 6.

Strategising reinvestment options is dependent on multiple decisions, as shown in Fig. 6, e.g., insulator replacement, damper retrofitting and replacement, etc. For each of these and the main decision, the primary input required is the condition data from SAP, details of risk exposure, and the associated refurbishment costs they receive from the projects team. The strategised options are communicated to the lines maintenance facilitator. At the same time, a project initiation form is issued. If the amount is above $6m, the form goes through the RIT-T process. Then the form is sent to the Project Portfolio and Operations (PPO) team.

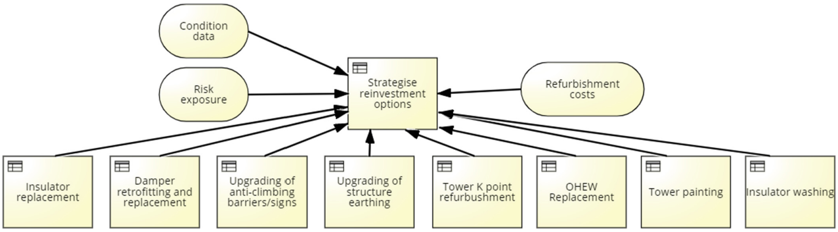


Fig. 6. A DMN model to strategise reinvestment options.

The project initiation form is received by the PPO team of the network portfolio. The team seeks input from network operations and then area planning. The network investment outlook (an internal document) is a crucial input here. Next, the project initiation form options are sent to the Network and Alternative Solutions (NAS) team of the network portfolio, who check the information for scoping and costing. The project initiation form is then converted to scope, which is sent back to the PPO team. The team reviews the scope and identifies alternate options. The input data required to identify alternate options is presented in Fig. 7. There are two high-level options, network configuration or non-network options, to consider. If network configuration needs to be done, the options considered are targeted refurbishment, extensive refurbishment, rebuilding of transmission line, or decommissioning a line. For non-network configuration, options considered include the use of batteries, generation support, and demand management investment, among others.

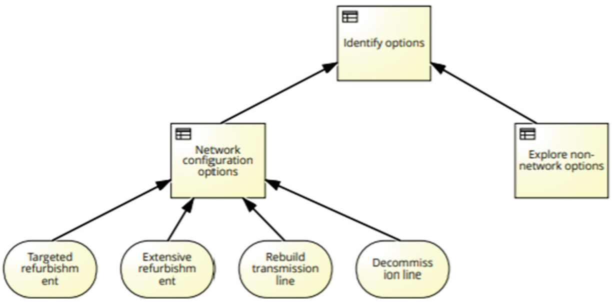


Fig. 7. A DMN model to identify alternative options for reinvestment.

Following this, a preliminary screening of options is done. For this screening, the options are sent to the Projects team, which calculates the costing of options using tender, class 2, or class 3. The costing information is sent to the NAS team, which conducts an economic analysis of options. Finally, an option is selected (Fig. 8), and the decided option is sent for execution. The costing and scoping of the project and the capital expenditure budget are the three main inputs required to decide an option for reinvestment.

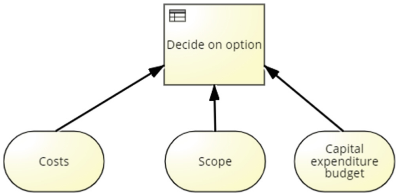


Fig. 8. A DMN model depicting the input required to decide on an option.

4 Discussion

The comprehensive process model developed in this study explains the critical asset management activities and decisions undertaken by the key stakeholders when managing the reinvestment of TLAs. In particular, the process model: (a) allows an *understanding of data/activities/systems/resources* that are not documented but are used in the process, (b) enables people to *understand the key roles and responsibilities* of the teams in an organisation, and (c) assist the *identification of high impact processes* that would benefit from centralisation or IT infrastructure. For example, we found that the health index is a crucial indicator for many decisions made in the organisation, which was not explicit in the documentation. Similarly, we evidenced that Tableau visualisation used by Line Strategies is central to the reinvestment decisions and developing a mature and widely accessible system could aid in communication and assessment across multiple teams. The process model also brought forth the *significance of data quality checks* as it is that data used to make crucial decisions involving a significant financial impact later. The DMN models embedded in the process model were also found *beneficial to capture the key requirements* for decision-making while ensuring that the overall process model is manageable.

Numerous other advantages of the comprehensive process model were noted. The process model can help in *knowledge transfer* and assist a new person entering the organisation to visualise and understand the process. It can also be used as a tool to explain internal processes to auditors/regulators. The model assists in greater *transparency* around the execution of processes as well as their upstream and downstream impact on the organisation. Moreover, the process model acts as a *tactical blueprint*, enabling cross-process comparison (e.g., comparing TLA asset management processes with substations) and identifying areas of high impact.

Additionally, a process model like the one developed in this project can *complement existing documentation* and *check alignment* with ISO55000 or other standards/practices. For example, ISO55000 outlines the significance of having a clearly defined method and criteria for decision-making. While elements of this were present in tactical documents of Powerlink, the detailed process model made the underlying process explicit. As was seen in the case of TLAs, many of the lower-level processes in ISO55000 will likely not be documented at the strategic level or even in more tactical documents—they often lie in the tacit knowledge of the teams and individuals working together. The process model can help make this knowledge explicit. For instance, a requirement by ISO55000 is the clear identification of roles and responsibilities. We found that some roles have not been elucidated in the tactical documents of Powerlink, but the teams have well-defined roles and responsibilities, most being tacit knowledge. Moreover, ISO55000 indicates the significance of taking actions to address risks when managing assets. While Powerlink provided the document, Overview of Asset Risk Methodology, it did not contain some elements that the process modelling exercise revealed were important for risk assessment, e.g., health index calculation. The process model provides more detail on the assessment of risks via health indices, who is involved in it (line strategies, projects, and network portfolio), and what data supports it.

Finally, a clear, detailed process model can be used for further enhancement activities in the organisation. For example, it can be used to automate mature processes and divert attention to more resource-intensive processes. It can also be used as input for automated process improvement techniques such as process mining.

5 Conclusion

This paper presents the application of Business Process Management, in particular process and decision modelling, to model complex EAM processes for TLAs at Powerlink. The project lasted for a period of six months. Semi-structured interviews were conducted with SMEs, documents were analysed, and BPMN2.0 and DMN standards were used to model the EAM processes. The overall process model representing the key activities, decisions, personnel involved, and the system as well as data used was presented. The process model captures the complexities involved in making reinvestment decisions for TLAs and provides a detailed understanding of the underlying process. Overall, the process model was considered beneficial by Powerlink providing the stakeholders a detailed understanding of – the process itself, the interaction among various teams, the key data used for making decisions, the key decisions made by stakeholders, and identify areas of possible improvement. The project clearly demonstrates the significance of application of process and decision modelling, for asset management processes. We acknowledge the limitation of applying process modelling to only reinvestment decisions of TLAs and only one organisation. However, Powerlink is a big organisation with a diverse set of stakeholders, and all stakeholders well received the process models. For future research, process modelling can be applied to other assets and through other phases of an asset lifecycle. Additionally, other BPM techniques such as process automation and process mining can also be applied to improve EAM processes further.

Acknowledgments. The authors gratefully acknowledge the funding provided by Powerlink Queensland for this project and the input from Powerlink SMEs. The QUT research ethics committee approved the data collection methods employed in this project.

References

1. Frolov V, Mengel D, Bandara W, Sun Y, Ma L (2010) Building an ontology and process architecture for engineering asset management. In: Kiritsis D, Emmanouilidis C, Koronios A, Mathew J (eds) Engineering asset lifecycle management. Springer, London, pp 86–97. https://doi.org/10.1007/978-0-85729-320-6_11

2. Greyling BT, Jooste W (2017) The application of business process mining to improving a physical asset management process: a case study. S Afr J Ind Eng 28(2):120–132
3. Sun Y, Ma L, Mathew J (2012) Improving asset management process modelling and integration. In: Amadi-Echendu JE, Willett R, Brown K, Mathew J (eds) Asset Condition, Information Systems and Decision Models. EAMR. Springer, London, pp 71–87. https://doi.org/10.1007/978-1-4471-2924-0_3
4. Lima ES, McMahon P, Costa APCS (2021) Establishing the relationship between asset management and business performance. Int J Prod Econ 232:107937
5. Maletič D, Maletič M, Al-Najjar B, Gomišček B (2020) An analysis of physical asset management core practices and their influence on operational performance. Sustainability 12(21):9097. https://doi.org/10.3390/su12219097
6. Weske M (2007) Business process management–concepts, languages, architectures. Springer, Heidelberg. https://doi.org/10.1007/978-3-540-73522-9
7. Myers MD (1997) Qualitative research in information systems. MIS Q 21(2):241–242
8. Kwon S-W (2008) Does the standardization process matter? A study of cost effectiveness in hospital drug formularies. Manag Sci 54(6):1065–1079
9. Mabry L (2008) Case study in social research. In: The Sage handbook of social research methods, pp 214–227
10. Bowen AG (2009) Document analysis as a qualitative research method. Qual Res J 9(2):27–40
11. Sedera W, Rosemann M, Gable G (2002) Measuring process modelling success. In: ECIS, pp 331–341
12. Hammer M, Champy J (2006) Reengineering the Corporation: A Manifesto for Business Revolution Collins Business Essentials
13. Gill P (1999) Application Development: Business Snapshot-Business Modeling Tools Help Companies Align Their Business and Technology Goals. Information week, April 1999
14. OMG (2011) Business Process Model and Notation (BPMN) Version 2.0. Object Management Group. https://www.omg.org/spec/BPMN/2.0
15. Janssens L, Bazhenova E, De Smedt J, Vanthienen J, Denecker M (2016) Consistent integration of decision (DMN) and process (BPMN) models. In: CAiSE forum, pp 121–128
16. Figl K, Mendling J, Tokdemir G, Vanthienen J (2018) What we know and what we do not know about DMN. Enterp Model Inf Syst Archit (EMISAJ) 13:1–16
17. Taylor J, Fish A, Vincent P (2013) Emerging Standards in Decision Modeling-An Introduction to Decision Model & Notation in iBPMS Intelligent BPM Systems: Impact and Opportunity. Future Strategies Inc., Brampton
18. OMG (2020) Decision Model and Notation Version 1.3. Object Management Group. https://www.omg.org/spec/DMN/1.3

Towards Evidence-Based Decision Making in Asset Management

Helena Kortelainen(✉), Antti Rantala, Toni Ahonen, and Jesse Tervo

VTT Technical Research Centre of Finland, 33101 Tampere, Finland
Helena.kortelainen@vtt.fi

Abstract. Evidence-based asset management aims at making right decisions and optimizing asset management processes with best available information. Asset information systems are widely applied in industrial companies to collect and store asset related data. However, competence and experience of people, - i.e. tacit knowledge - has a crucial role in the decision-making. In this paper, we discuss information transfer, usability of current IT-systems and data utilization in daily tasks of different user groups. In addition, we outline a solution that supports the way towards evidence-based approach in process industry.

Keywords: Evidence-based asset management · Maintenance · Asset information systems · CMMS · Tacit knowledge · Data collection and retrieving

1 Introduction

Industrial companies use a variety of information systems to generate, store, retrieve and process data. Such asset information systems include Enterprise Resource Management (ERP), Enterprise Asset Management (EAM) and Maintenance Management (CMMS) systems. Asset information systems contain information on what work was done to an asset, when, where and how, and, in a course of time, this information builds up a valuable asset history data base. Unfortunately this data base is often underused [1, 2]. Data related to maintenance and failure events is typically recorded manually by the maintenance personnel. Further sources of failure data, like control room diaries that operators use to share and transfer acute issues in their everyday work, provide the maintenance organisation further source of disturbance and failure data [3]. The manually recorded data is often scarce and the data quality leaves space for improvement [4].

The data collected in everyday operations may also be scarce simply because the asset item is new, or its inherent reliability is high, and therefore the number of recorded maintenance or failure events is small for obvious reasons. However, this does not mean that there is no information available. In addition to the asset information systems, experience of people and tacit knowledge [5, 6] are crucial in any asset related decision making. The understanding of the production system, its functions and interrelationships strongly relies on the competence of experts. One crucial question is then, how to extract and exploit the tacit knowledge that the experts do have [7]. Using the expert knowledge often also involves the idea that expert knowledge should be used as a basic starting

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 312–323, 2022.
https://doi.org/10.1007/978-3-030-96794-9_28

model which will be improved and updated in the process of real data arriving (e.g. [8]). Continuous development of the maintenance program also involves both expert and event data [25]. Ahonen et al. [17] proposes a process that could tie together expert knowledge and the different pieces of information refined from data, and use that according to the asset hierarchy.

In the model presented by Ackoff [9] and Rowley [10], data refers to the symbols that represent the properties of objects and events and they are the products of observation. Information consists of processed data (e.g. charts, descriptions, drawings) and this processing aims at increasing its usefulness. Knowledge can be obtained either by transmission from another who has it, by instruction, or by extracting it from experience. Each step up the knowledge pyramid adds value to the data as the data is enriched with meaning and context. In this process, the data is refined to knowledge and insights that allow making interpretations and applying the derived knowledge beyond the original data collection scheme [9].

Data, information and knowledge, and the capability of the organisation to make evidence-based decisions on proposed changes are in the core of asset management activities [10]. For organizations to realize the best value from their assets, decisions should be assessed with analysis and evaluation of opportunities and outcomes from the perspectives of performance, cost and risk [e.g. 11]. Information systems are needed to support competent people in making decisions in a timely manner, through the provision of good quality data and information. In real decision making situations data in different formats and from different sources need to be integrated. In addition, digital technologies and IoT (Internet of Things) solutions entering to the market will further shape the asset management operations [12, 13] and offer novel possibilities to share and retrieve data from various information systems [14]. Evidence-Based Asset Management (EBAM) approach presented by [16] aims at making right decisions and optimizing asset management processes with best available information. In this process, recording maintenance activities in a CMMS is of great importance. EBAM practices could help maintenance managers in improving maintenance practices. The concept for hierarchical replacement and maintenance investment elaboration proposed by [15] is an example of making most of the CMMS data by combining the data with formalized expert knowledge.

Expert elicitation as a source of tacit knowledge is an important constituent of the evidence-based decision making process [8, 16]. Structured elicitation processes and visual templates have proved to be effective in collecting tacit knowledge [17]. Typical methodologies for expert elicitation processes include methods familiar from reliability engineering like failure trees (FTA), reliability block diagrams (RBD), Failure modes and effects analysis (FMEA and FMECA), reliability analyses (e.g. RCM) and Hazard and Operability study (HAZOP).

Evidence-based approach emphasize fact-based analysis rather than decisions based on rules of thumb, intuition and mere experience. However, several investigations indicate that - in addition to the numerous barriers to record data in the first place - the decision analysis capability is often missing in existing CMMSs [2] and the user interfaces (UI) should be simpler and more intuitive. Mahlamäki and Nieminen [4] propose role-based UIs (User Interfaces) to overcome the challenges arising from the local contexts and working procedures. The interesting questions are then "*How help an organisation to* *make evidence-based decisions on all levels?"* and *"What kind of UIs would provide necessary evidence for people in different roles and real life decision making situations?"* In this paper, we discuss on the information needs of different user groups and their needs concerning tacit knowledge integration, analyses for data refinement and visualization. Based on extensive interviews we outline a solution that support evidence-based approach for asset management activities in process industry.

2 Research Context and Approach

Creation of knowledge and understanding of the problem domain, to understand the real challenges that the practitioners face, is in the core of the design-science paradigm [18]. In our study, we apply design science research (DSR) process. An essential part of DSR process is the conduct of research to understand the problem. DSR is then applied to construct and evaluate artifacts (e.g. models and methods) that enable the transformation to a desired direction [19].

DSR aims at developing means-end propositions that solve real problems that typically are *ill-structured* [20]. "Ill-structured" describes decision situations where decision makers may not know or agree on the goals of the decision, and even if the goals are known, the means by which these goals are achieved are not known and requisite solution designs (e.g., technologies) to solve the problem may not even exist. DSR also calls for developing a synthesis from multiple disciplines or, at the very least, applying the expertise and insight from one knowledge domain to another. The process applied in this study (Table 1) follow the process presented by Peffers et al. [21].

Even though the activities in the DSR process seem to follow each other, the process incorporates learning, feedback and iteration. Evaluation could include such items as a comparison of the artifact's functionality with the stated objectives and client feedback. The evaluation may reveal the need to improve the effectiveness of the artifact or to continue on to communication and leave further improvement to subsequent projects [21].

Table 1. DSR research process [21] and application of in this study

DSR process activity		Application is this study
1	Problem identification and motivation	Outline of the specific research problem and justification the value of a solution (Sect. 2.1)
2	Objectives of a solution	The objectives of a solution derived from the problem definition (Sect. 2.1)
3	Design and development	Insight study phase including interviews form the basis for the development (Table 2)
4	Demonstration	A case study was chosen to develop and illustrate the solution (Sects. 2.1, 4 and 5)
5	Evaluation	Comparing the characteristics of the developed artefact with the defined objectives (Sect. 6)
6	Communication	Communication is being done through this paper

2.1 Problem Identification and Motivation

The research activities in the ongoing co-innovation project "Solid value from digitalisation in forest industry (SEED)" - project [22] offer complex and multidimensional practical problems to solve. In the forest companies of the SEED project, the personnel involved in the maintenance activities have already a variety of information systems and data sources available. Supporting decision making with evidence in diverse and varying asset management tasks can surely be characterized as an ill-structured problem.

The motivation to the study arises from the challenges to find relevant data and to access to the necessary information in real life decision situations. In addition, employees in different roles need different information: a field technician needs different information during the implementation of a specific maintenance task than a maintenance work planner who allocates resources, or an asset manager pondering replacement decisions. They all benefit from data integration and visualization, and the companies would benefit from evidence-based decision making in the asset management activities.

2.2 Case Study and Context

The forest company use cases in the SEED project start by an insight study phase that targeted into identifying user needs and requirements and building shared insight on possible solutions. In our use case, employee interviews within three production sites (Table 2) provided insight on the roles and responsibilities of various actors, their knowledge (data) needs and on their expectations for the system to be developed.

In the case study presented in this paper, our objective is to outline elements of a user interface that provides knowledge from different data sources to a field technician when responding the urgent need for maintenance. On a general level, access to all available information including tacit knowledge would be indispensable when aiming towards evidence-based asset management.

2.3 Research Data Collection (Insight Study)

Employees from three different plants representing different positions and roles were interviewed to gain a comprehensive view of all aspects of the study. Semi-structured interviews were carried out to provide insights into the thematic areas that were derived from the problem statement (Sect. 2.1). The thematic areas included challenges in information transfer, usability of current IT-systems and level of data utilization in daily tasks. Examples of the specific interview questions include:

- What kind of challenges do you have in collecting and reporting information?
- How often do you need different kind of documents in your working tasks?
- In which form would you like to collect and report information?
- What areas of development are there in utilizing data in different tasks, especially in production, maintenance and planning?
- What are the main challenges in usability of your current IT-systems?
- Who produces the information that you need in your work?

Table 2. Research data collection by interviews in the SEED-project

Interviewees by organizational group	Number of interviewed persons
Mill management (plant director, production management, maintenance management)	16
Middle management, supervisors and specialists (e.g. maintenance engineers, maintenance development engineers, supervisors, work planners)	52
Operators (production department workforce)	13
Field maintenance technicians (maintenance department workforce)	13

The interviewees represented a cross-section of the entire mill staff from the top management to shop floor operators and maintenance technicians. The interviewed employees are grouped according to their job duties and titles in Table 2. As the duties and titles were somewhat different at different plants the group "Middle management and expert" is the most heterogeneous one. The persons in this group have either supervisory tasks or they are experts in their specific field.

The interviews were arranged by the company contact person who was also responsible for choosing the interviewees. Each remote meetings that took 1–1,5 h. Two interviewers participated in the meeting and both took notes from the interviews. In addition, the interviews were recorded. The interviewees participated the remote meeting as small groups (1–5 persons). The interview data was coded using a qualitative data analysis software Nvivo.

3 Findings from the Interviews

The interview results were analysed from the perspective of data collection, data sharing and transfer, interaction and how to get help in difficult situations. In addition, issues dealing with tacit knowledge were identified. The findings are presented within the context of the roles shown in Table 2 and presented in Table 3.

The interviews confirmed the initial problem statement and clearly highlighted four key challenges. Firstly, the employees on all levels use several IT-systems when carrying out their daily tasks. The systems include not only specific systems like ERP, CMMS and Control room diary, and IT systems used in plant documentation, but office systems and many others. Secondly, the information transfer between different IT-systems is not fluent but information has to be searched in several systems, and a lot of manual work is required. Thirdly, the data quality is regarded as poor and there are several barriers - both technical and operational - that hamper data collection. However, all interviewees shared the opinion that good data quality is beneficial in every job. And finally, there is a variety of technical, operational and motivational issues that make using the systems difficult or even reluctant. Novel, unprecedented faults and problems tend to be a big challenge to everyone in the maintenance department. One interviewee described the

Table 3. Summary of the insight study

Organization group	Key challenges encountered and needs identified
Mill management	• Data is spread across many systems that are difficult to use. The users' skills may also be inadequate • Data entry is crucial, as is recording feedback • Data must move from one system to another, now same data is recorded in many places • People lack motivation to share their knowledge and the ability to write long texts such as accurate descriptions of problems • Technical documentation contain lot of useful information stored, but it is difficult to access and search
Middle management and specialists	• Too many systems on different platforms. Search for information is tedious. All information should be in one place or more easily linked • Event data is not recorded at all or the entry is incomplete. Event descriptions should be more comprehensive • Lack of feedback and responses to event data entries • Different departments, units and shifts use the IT-systems in a different way. Policies should be harmonized • Retrieving information is difficult, data has to be retrieved from several systems. Easier-to-use systems lower threshold for using them • ERP is cumbersome and clumsy to use. Often deployment is not optimal and systems remain underutilised. Immature systems extinguish people's enthusiasm. Much more should be invested in establishing, training and familiarizing • When unprecedented problems arise, technical documents can be useful but it may be out of date or not available • Help sought from third parties if in-house competence insufficient
Field maintenance technicians	• Too many systems on different platforms. Same things need to be recorded in several systems. Programs should be combined • Incomplete or missing event descriptions hamper troubleshooting • Retrieving information is difficult, sometimes paper archives needed • Documents are not always up to date, especially small changes remain often unrecorded • In the field oral communication in a central role. Operators inform on urgent jobs by phone, rarely through the ERP

(*continued*)

Table 3. (*continued*)

Organization group	Key challenges encountered and needs identified
Operators in the production department	• Diary entries are usually made once a shift. Entries are short and often superficial, sometimes left undone • ERP is difficult to use, as is retrieving information from it. Information is shared orally, in meetings, on paper notes • All means of communication are applied: e-mail, telephone, diary, radio telephone, facial expressions, gestures, hand signals • A lot of obstacles to using the systems incl. lack of support during deployment, problems when logging in, computer slowness

situation:"Problem is if you have to repair an equipment that is new to everybody". If the technicians don't have experience of the problem, they often need to browse through documentation or contact other technicians, supervisors or third parties. Crucial knowledge on how to deal with the issue might be tacit and not easily accessible, because the knowledge has not been recorded anywhere or because the records are unavailable. This causes trouble and delays in fixing the faults.

As highlighted also in Table 3 the interviewees also put forward a number of suggested solutions to address the problems identified. The need to collect various information systems in such a way that they would be accessible through a single UI and preferably also by one user ID code was described by one interviewed as *"Stockbroker's screen with all the necessary information displayed"*.

4 Developing and Illustrating the Solution

Evidence-based decision making [16] aims at making right decisions with best available information. The insight study (Table 2) suggest that the major obstacle on the way towards EBAM in everyday decision situations deals with the availability of the information. In an industrial maintenance setting, information must be easily accessible, the data must be comprehensive and, it must be possible to retrieve the data using various search functions. A systematic collection and integration of the tacit knowledge with digital information sources is then a key to transfer the accumulated knowledge to different user groups and to support decision making at all levels.

DSR process encourages to construct artefacts that help to perceive solutions [19]. In our study, we described the employee's activities as a process of collecting and retrieving data, and through this process, we constructed a description of the data that needs to be displayed, and of required role-based UI. The very first idea of such solution is presented in Fig. 1. The illustrative name "stockbroker's screen" was picked up from a comment from an interviewee as this concept highlights a situation in which quick decisions are made and information needs to be acquired at a glance.

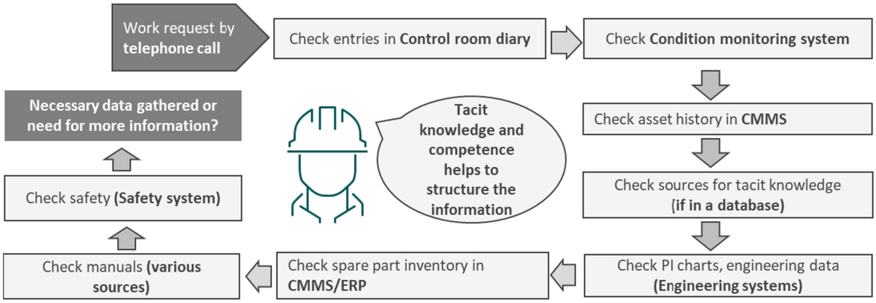


Fig. 1. The field technician responding the work request by retrieving necessary information through a “Stockbroker” screen” from several information systems.

The information compiled in Fig. 1 is contained in several information systems including CMMS, Control room diaries and ERP. The Stockbroker’s screen - solution could collect information from these sources and make it available without logging in and out to several systems. We could expect that advanced cloud-based industrial internet systems make this kind of information retrieving and processing possible.

5 Iterating the Solution and Future Work

The next step to take in applying an evidence-based approach in maintenance and improving utilization of data sources is to map specific needs of different user groups based on the interviews. In order to understand better the decision situations and contexts user groups will be transformed into user personas. The personas will have use cases for an UI which combines data sources and possibly displays essential descriptive graphics as a “Stockbroker screen” or an analytics dashboard. Use cases will be depicted with BPMN (Business Process Model and Notation) models. BPMN is a standard commonly used in process documentation, but it is also a popular tool in software requirements engineering because of its usefulness in addressing and communicating requirements with stakeholders [23]. One such use case is the situation (Fig. 2) in which a field technician responds to a fault notification and retrieves necessary information for fault diagnostics and analyzes it before starting the actual maintenance work. In this case, an integrative interface could streamline the process so that the technician does not need to visit all the involved information systems one by one.

As we also want to show that such a system can be helpful in practice, we will demonstrate the use cases with role-based proof of concept user interfaces which display the data sources in a comprehensive and useful way for the specific roles they are designed for. This means that field technicians will be able see task-relevant Control room diary and CMMS information via a simple search function and have easy access to tacit knowledge in the form of asset-specific analyses and documentation containing tacit knowledge [18] stored in a document database. Other user groups will be connected to the same external systems but their information needs and decision making situations

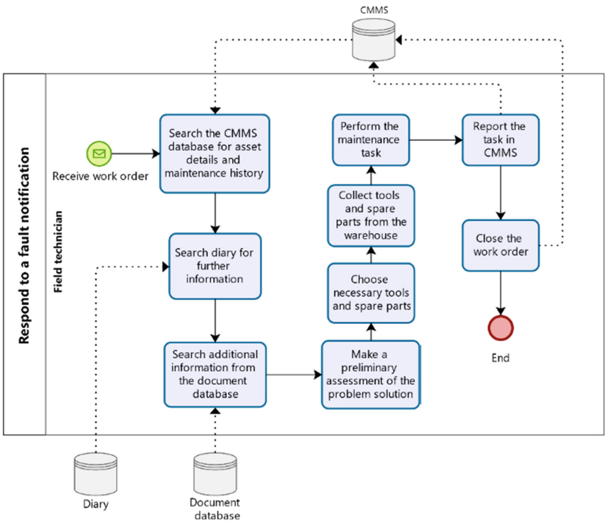


Fig. 2. BPMN model for the technician responding to the urgent fault notification

differ from those of the field technician. For example, a maintenance manager who is making equipment replacement decisions should see some of the same asset history data and documentation but also additional graphs and financial information to get a holistic view of the factors affecting the decision.

6 Conclusions and Discussion

The vision of the SEED project states that digitalization offers the means for major improvement of the forest industry in the productivity and business renewal. In this paper we address the use case offered by a forest company and look for solutions to the challenges to retrieve, search and display in real life decision situations. The original research question "*How help an organisation to make evidence-based decisions on all levels?*" shifted the focus to a more central problem, namely access to information. It is clear that access to all available information including tacit knowledge would be indispensable when aiming towards evidence-based asset management. In this paper, we suggested an approach that would answer on the second research question *"What kind of UIs would provide necessary evidence for people in different roles and real life decision making situations?"* Recent literature has demonstrated the limitations of manually gathered data [4] and acknowledged the importance of people's expertise and tacit knowledge alongside the advanced analytics [7]. Also our comprehensive and extensive interview material highlighted the fact that maintenance and asset management activities suffer from scattered data and lack of models that would bring together relevant pieces of information, display and refine the information in such a way that available information could respond to the various needs of the specific purposes of different user groups. The insight study suggested that current asset information systems do not respond the identified needs and tacit knowledge is difficult to use because the knowledge has not been recorded or because the records are unavailable. There is a clear need to bring the evidence on the display when making decisions. This "evidence" may be in different formats: as recorded data in CMMS or in control room diaries, as information in the documents or as tacit knowledge in human minds. Based on extensive interviews we have outlined a solution that could support evidence-based approach for asset management activities in process industry by removing obstacles that are posed by current silo-like data management systems. The proposed solution - "Stockbroker's screen" - applies the IoT key principles that include transparency, increasing information sharing from various IT-systems and between various stakeholders [14]. In this solution, the data and information stored in different IT-system siloes will be picked up onto display by using identified user profiles and role-based UIs. In the core of the solution is the understanding of the processes - what data is needed, by whom and in which decision making situation. The paper presents the first ideas of approach. In the future research, the solution and its functionalities will be demonstrated in a proof-of-concept setting. The demonstration will entail data from CMMS, Control room diary and tacit knowledge. In the demonstration phase the solution will also be evaluated in a real industrial setting. Holmström, Ketokivi and Hameri [21] state that the strength of the design science approach is its explicit focus on improving practice. The other side of the coin is then that the process does not lead to novel theoretical insight and no new knowledge is created. In our ill-structured problem of supporting decision making with evidence in diverse and varying asset management tasks, DSR process proved to be a useful approach. The process led us to outline solutions that would help to collect the necessary information on a "stockbroker screen". In this study, looking for a solution to a practical problem contributes also to the interdisciplinary research of evidence-based asset management. While data and information are currently scattered in various systems and finding relevant information for the context is difficult, a framework is needed to support the information content, automated refinement of the data and easy access to the databases. The "stockbroker's screen" framework presented in this paper provides a solution where different use cases of decision-making situations are identified according to the requirements of the user groups. In order to make most of the data, data refinement should be interlinked with the maintenance processes. For instance, there is a need to use the gathered data better in enriching recent failure and root cause analysis and also to make these enriched analyses better available for the field purposes. Furthermore, there is a need for automation in such reverse analytics, e.g. to identify potential root causes for a failure based on the coarse description of the symptoms or to identify candidate solutions for a bottleneck. Integration of these approaches, adjusting them with different maintenance processes and development of the specific analytics frameworks to support this are suggested as a focal area of topics for further research.

Acknowledgements. The research group have received funding from Business Finland that is the Finnish government organization for innovation funding and trade, travel and investment promotion. VTT Technical Research Center of Finland and the companies participating in the SEED project have been co-funding the study. The authors wish to thank all these organisations for their support.

References

1. Hodkiewicz M, Ho MTW (2016) Cleaning historical maintenance work order data for reliability analysis. J Qual Maint Eng 22(2):146–163
2. Galán MH (2019) Evidence-based asset management applied to maintenance function control. J Qual Maint Eng 25(4):635–644
3. Kortelainen H, Komonen K, Laitinen J (2021) Knowledge-based life cycle management. Kunnossapitoyhdistys Promaintry, Helsinki, Finland. (in Finnish)
4. Mahlamäki K, Nieminen M (2019) Analysis of manual data collection in maintenance context. J Qual Maint Eng 26(1):104–119
5. Nonaka I, Takeuchi H (1995) The knowledge-creating company. Oxford University Press, Oxford
6. Redding LE, Hockley CJ, Menhen J, Roy R (2015) The role of maintenance, repair, and overhaul (MRO) knowledge in facilitating service led design: a nozzle guide vane case study. In: Amadi-Echendu J, Hoohlo C, Mathew J (eds) 9th WCEAM research papers. LNME. Springer, Cham, pp 379–396. https://doi.org/10.1007/978-3-319-15536-4_31
7. Amadi-Echendu J, de Smid M (2015) Integrating tacit knowledge for condition assessment of continuous mining machines. In: Amadi-Echendu J, Hoohlo C, Mathew J (eds) 9th WCEAM research papers. LNME. Springer, Cham, pp 119–130. https://doi.org/10.1007/978-3-319-15536-4_10
8. Zuashkiani A, Banjevic D, Jardine AKS (2006) Incorporating expert knowledge when estimating parameters of the proportional hazards model. In: Annual reliability and maintainability symposium - 2006 proceedings. IEEE Computer Society, pp 402–408
9. Kortelainen H, Hanski J, Kunttu S, Kinnunen S-K, Marttonen-Arola S (2017) Fleet service creation in business ecosystems – from data to decisions, Espoo, Finland
10. ISO 55002 (2018) ISO 55002- Asset management – Management systems – Guidelines for the application of ISO 55001, 2nd edn. International Organization for Standardization
11. Trindade M, Almeida N, Finger M, Ferreira D (2019) Design and development of a value-based decision making process for asset intensive organizations. In: Mathew J, Lim CV, Ma L, Sands D, Cholette M, Borghesani P (eds) Asset intelligence through integration and interoperability and contemporary vibration engineering technologies. LNME. Springer, Cham, pp 605–623. https://doi.org/10.1007/978-3-319-95711-1_60
12. Kortelainen H, Happonen A, Hanski J (2019) From asset provider to knowledge company—transformation in the digital era. In: Mathew J, Lim CV, Ma L, Sands D, Cholette M, Borghesani P (eds) Asset intelligence through integration and interoperability and contemporary vibration engineering technologies. LNME. Springer, Cham, pp 333–341. https://doi.org/10.1007/978-3-319-95711-1_33
13. Kortelainen H, Hanski J, Valkokari P (2020) Advanced technologies for effective asset management - two cases in capital intensive branches. IFAC-PapersOnLine 53(3):7–12
14. Juhanko J et al (2015) Industrial internet transforms Finland's challenges into opportunities: background synthesis, Helsinki
15. Ahonen T, Rantala A, Kortelainen H, Valkokari P (2021) Long-term history data and integrated expert knowledge in asset lifecycle management. In: Machado J, Soares F, Trojanowska J, Ivanov V (eds) Innovations in industrial engineering. icieng 2021. LNME. Springer, Cham, pp 1–11. https://doi.org/10.1007/978-3-030-78170-5_1
16. Zuashkiani A (2016) Evidence-based asset management. In: Campbell J, Reyes-Picknell J (eds.) Uptime: strategies for excellence in maintenance management, 3rd edn. CRC Press, Taylor & Francis Group, New York, p 489
17. Molarius R (2016) Foreseeing risks associated with new technologies: Co-operation of the authorities to prevent technical risks due to new technologies. Dissertation, VTT Technical Research Centre of Finland. (in Finnish)

18. Bichler M (2006) Design science in information systems research. Wirtschaftsinformatik 48(2):133–135. https://doi.org/10.1007/s11576-006-0028-8
19. March ST, Storey VC (2008) Design science in the information systems discipline: an introduction to the special issue on design science research. MIS Q Exec 32(4):725–730
20. Holmström J, Ketokivi M, Hameri AP (2009) Bridging practice and theory: a design science approach. Decis Sci 40(1):65–87
21. Peffers K et al (2006) Design science research process: a model for producing and presenting information systems research. In: First international conference on design science research in information systems and technology, pp 83–106
22. Valkokari K, Valkokari P, Kortelainen H, Nyblom J (2020) Building business impacts of an industry 4.0 ecosystem through collaborative network settings between IT and forest companies. In: Camarinha-Matos LM, Afsarmanesh H, Ortiz A (eds) PRO-VE 2020, vol 598. IAICT. Springer, Cham, pp 463–474. https://doi.org/10.1007/978-3-030-62412-5_38
23. Bakar NWA, Musa S, Mohamad AH (2020) A mini comparative study of requirements modelling diagrams towards swimlane: evidence of enterprise resource planning system. J Phys Conf Ser 1529(5):052054

A Case Study on Probabilistic Technico-Economic Analysis Including Maintenance Cost for Hydroelectric Turbine Fatigue Risk

Michel Blain(✉), Denis Thibault, Ariane Immarigeon, Rafaël Guay, Jérôme Lonchampt, and Pascal Hernu

Hydro-Quebec, Québec, Canada
blain@ireq.ca

Abstract. This paper presents a novel approach for planning replacement projects using prognostic developed for turbines and generators to try to defer unit replacements by maximizing economic gains while minimizing risk. In the same analysis in which we optimize replacement dates, we would like to also optimize operation and maintenance plans since each turbine design is affected differently by these plans. The current maintenance plans scheduled at fixed intervals and the unit commitment strategy solely based on hydraulic efficiency are therefore also questioned. The goal of this analysis is to demonstrate that maintenance and operating plans need to be adjusted to minimize the probability of regret when optimizing NPV (Net Present Value) by changing replacement dates while considering equipment prognostic. For this study, the impact of inspections, frequency of unit start-up and deferred equipment replacement have been evaluated by comparing them to a reference scenario where the equipment would have been operated under traditional frameworks. The analysis demonstrated the following points:

- It is possible to observe the impacts of the improvements (frequency of inspection, start-up, replacement date, etc.) separately in the technico-economic analyses;
- Unexpected breakdowns can be taken into account for informed risk management;
- Investment deferrals can be analyzed by taking into account the risk;
- Maintenance and repair can be rigorously taken into account by the model.

In fact, the proposed model allows to define the required maintenance and operation plans according to the replacement strategy that will be chosen in a global optimization of the unit fleet. The sensitivity analysis presented shows the importance of adjusting operation and maintenance according to the specific design of each runner and the importance of the chosen deferral date. With this kind of analysis, the decision-maker could possibly (while controlling the risk):

- Distribute the number of start-ups or other damaging conditions differently in order to protect some units while maximizing the use and wear of others;

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 324–332, 2022.
https://doi.org/10.1007/978-3-030-96794-9_29

- Postpone as much as possible the replacement of those equipments that show very little risk of failure.

The paper will be structured as follows. First, an overview of the modelling strategy will be presented. Then, the parameters of the case study and the methodology will be detailed. A sensitivity analysis will show the impacts of the different assumptions on the mean net present value and the probability of regret. Finally, we will discuss the results for several units in terms of applicability for decision making and asset management.

Keywords: Hydroelectric turbines · Asset management · Reliability · Monte-Carlo simulation · Net Present Value · Optimal operation and maintenance · Anticipated turbine replacement

1 Introduction

Large utilities need to optimize the investments made to maintain their assets. For a utility like Hydro-Québec (37 GW), an important part of these investments is intended for the replacement and rehabilitation of turbines and generators. The current strategy for planning replacement projects is based partly on the diagnosis of the equipment but mainly on the accounting life of these assets. Since several power plants were built at the same time, Hydro-Québec is facing a major challenge; the need to replace or rehabilitate many turbines and generators at the same time puts pressure on the company and on its supply chain (Fig. 1).

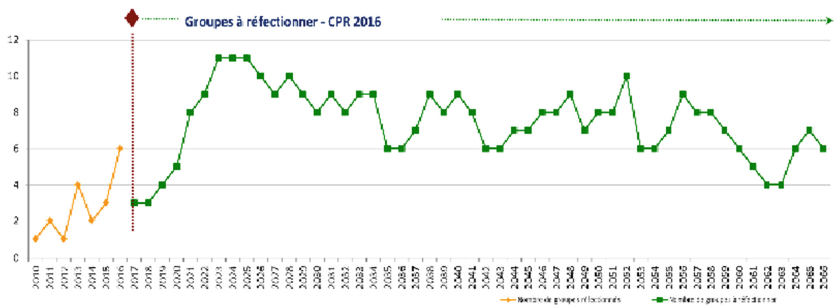


Fig. 1. Number of units to be refurbished (green) compared to number of units that the suppliers have managed to do in recent years (yellow)

The idea of this paper is to improve the current strategy for planning replacement projects by using the prognosis of turbines and generators to try to defer groups by maximizing economic gains while minimizing risk. In the same analysis in which we optimize replacement dates, we strive to also optimize the operation and maintenance plans since each turbine design is affected differently by these plans. The current maintenance frequency of once in every six years, and the systematic start-up of the most efficient units first are therefore also questioned.

The technical-economic analysis on the replacement of the turbines and generators of the plant presented is performed using the VME tool developed by EDF R&D [1]. The event-driven model created allows one to evaluate, in a probabilistic way (using several thousands of scenarios), the economic value and the risk associated with a maintenance plan, an operating scenario or even the different runner designs.

For the hydroelectric turbine runner, fatigue and cavitation are the two main degradation mechanisms. However, of these two mechanisms, fatigue is often considered difficult to manage because damage cannot be monitored until significant cracks appear, which are then automatically considered critical.

For the generator, the prognosis will eventually come from a complex model including all the failure mechanisms known to experts [2]. For the present paper, we consider it as a variable of the sensitivity study to demonstrate at first the relevance of this kind of study.

In short, the goal of this analysis is to demonstrate that maintenance and operating plans need to be adjusted to minimize the probability of regret when optimizing NPV by playing on replacement dates while considering equipment prognoses.

The paper will be structured as follows. First, an overview of the modeling strategy will be presented. Then, the parameters of the case study and the methodology used will be detailed. A sensitivity analysis will show the impact of the different assumptions on the mean net present value and the probability of regret. Finally, we will discuss the results for several units in terms of applicability for decision making and asset management.

2 Modeling

For this study, the impacts of inspections, start-up frequency, and early turbine runner replacement were evaluated by comparing them to a baseline scenario where the turbine runner would have been operated following traditional framing. It is assumed here that the replacement turbine runner has a new design that minimizes the risk of cracking as specified in the new Hydro-Quebec specifications.

The net present value (NPV) is a function of the probability of cracking, failure and the impact of these events. This probability is determined by several factors depending on design, operation and maintenance. The impact of failure is calculated through maintenance costs, downtime and required investments. We will use VME to perform a Monte Carlo simulation that calculates an average NPV and a probability of regret from thousands of simulations.

3 Basic Assumptions

Two failure modes of the turbine runner are present in the model:

- "pre-existing defect", i.e. presence of an early crack requiring repair at the first inspection;
- "cracking as a function of time" that evaluates the risk that a defect of a given initial size becomes sensitive to operation. The associated risk increases with time depending on the number of start-ups experienced and the magnitude of steady state stress.

More than 20 scenarios of cracking probabilities were calculated according to the PreDDIT methodology [3–6] by varying the following 3 parameters:

- The dynamic stress amplitude indicated in MPa (directly proportional link with the fatigue damage of the runners, i.e. their risk of cracking):
  - current design: steady-state stress amplitude = ~3a
  - improved design: steady-state stress amplitude = ~2a
  - design in the new specifications: steady-state stress amplitude = a
- The diameter of the initial defect (3 mm, 1.4 mm and 0.1 mm)
- The frequency of start-ups (1 per day, 1 per week, 1 in 70 years)

The method for calculating the prognosis from the turbine runner stress measurement is presented in the referenced papers [7–9]. Some important impacts used in the analysis are as follows (costs are confidential):

- Non-destructive inspections (NDT) valued at ~3x k$ including the assembly and disassembly of the platform and 2–3 types of inspections (visual, liquid penetrating and magnetic)
- Complete replacement of a unit (including temporary repairs, disassembly, supply and testing over 3 years) of $ yM
- Repairs of ~x k$ (crack of 5” and less) to ~7x k$. Repairs are usually made during inspections, so it is assumed that the platform is already assembled.

In general, the downtime costs associated with inspections and repairs are considered to be virtually negligible, as this work is primarily performed from March to November, when there are no maintenance spills and production is resumed elsewhere.

4 Impacts of Start-Ups

The work carried out has made it possible to study the impact of the frequency of start-ups. It was concluded that with the current estimated stress amplitude (current design), the unit remains at risk of cracking even if the frequency of start-ups is reduced. The analysis performed shows that with the current estimated stress amplitude, the economic value of decreasing the start-ups is low since it will not prevent failures as shown in Fig. 2. These scenarios are mainly affected by the supplier and downtime costs associated with an unplanned turbine replacement. The choice of an early (planned) replacement then becomes an economically enviable alternative.

However, for units with improved design, the impact of start-ups becomes significant. It is then possible to significantly reduce the risks and inspection requirements by reducing the number of start-ups, which leads to a reduction in operating costs.

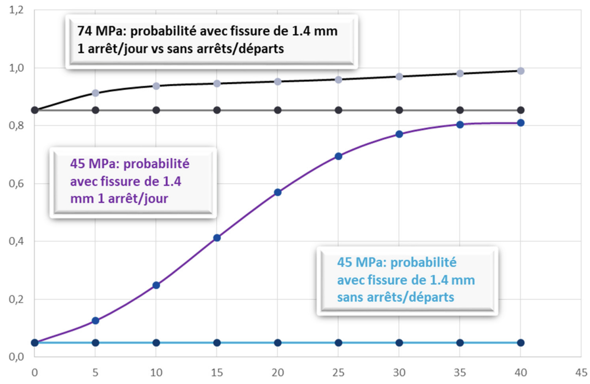


Fig. 2. Probabilities of cracking according to start-up frequency for the 'current design' scenario and for the 'improved **design' scenario.**

5 Impacts of Inspection Frequency

The study shows that the optimal inspection frequency should be customized to the design (Table 1). For the current design scenario, inspecting units more frequently is cost-effective despite the additional inspection cost. By changing the frequency from 6 to 3 years, a monetary value of v k$ per unit is realized. The probability of regret for this decision is 42%. The approach used also shows that it is possible to optimize maintenance costs, because by inspecting these units every 5 years instead of every 6 years, a positive value of 4v k$ per unit is generated with a probability of regret of only 20%.

Table 1. Economic value for different inspection scenarios by design.

Current design	NPV	Pregret
Inspection 5 years	4v k$	20%
Inspection 3 years	v k$	42%
Improved design	NPV	Pregret
Inspection 5 years	−w k$	55%
Inspection 3 years	−10w k$	99%

However, the story is different for units with improved designs. Increasing the inspection frequency to every 3 years would result in an anticipated loss ten times greater than at 5 years and a 99% probability of regret. In practical terms, this means that the reduced risk of cracking does not compensate for the additional inspection costs.

6 Failures Associated with Unscheduled Replacements

The probability of a failure leading to an unscheduled replacement is estimated at 1% to 3% depending on the inspection frequency, respectively 1 per year and 1 in 6 years. This corresponds to a major blade failure leading to a premature replacement of the runner with important economic consequences in the order of several tens of M$ for the replacement of a runner. An additional amount is also anticipated for the downtime caused by the failure. Since it is impossible to predict when this failure will occur, an average cost over the year was used. This cost, provided by the energetic transactions department, cannot be disclosed here. Inspections are treated as preventive actions, as they reduce the probability of operating the unit in the presence of an undetected failure and the presence (or not) of cracks. When considering failures, the frequency of inspection now becomes important even for units with improved design. Risk management becomes a more complex trade-off because it is impossible to simultaneously obtain the maximum NPV and the minimum probability of regret (e.g., the scenario of a unit with no start-up, no deferral of replacement, and inspections every 5). Since the stochastic analysis allows for risk management, the replacement date is moved to 2100 in the evaluated strategy from the baseline of 2070. Table 2 shows the NPVs for the 5-year inspection frequencies estimated above when failures and deferrals are added into the model.

Table 2. Economic value for different inspection scenarios by design.

Current design	NPV	Pregret
Inspection 5 years	~z M$	11%
Improved design	NPV	Pregret
Inspection 5 years	~2z M$	0%

7 Replacement Date Set by the Condition of the Generator

In the case where the condition of the generator does not allow the replacement to be postponed to 2100, the maintenance and operation plans must again be adjusted. Let's take the case for example, where following the prognosis of the turbine-connected generator in the 'improved design' scenario, the experts decide to replace the unit in 2080. In the case where a faster degradation of the generator is observed, the inspection frequency may be lower for the turbine and higher for the generator when updating the technico-economic analysis after the observation. Furthermore, if the condition of the generator is not influenced by the start-up/shut-down frequency, start-up/shut-downs will be inflicted on the unit with the least reliable generator to extend the life of the other units that will be replaced only in 2100. Remember that not all start-ups/shutdowns can be avoided since there is a peak in demand every morning and evening.

8 Maintenance and Operation Adapted to the Design to Promote Investment Deferrals

In general, the study shows that it is possible to manage risk and optimize maintenance, operation and turbine replacement plans based on their design. Thus, by comparing the reference strategy with the improved strategy (presented in Table 3), it is shown that it is possible to generate considerable value in excess of several tens of millions of dollars for the eight units corresponding to the 'current design' and 'improved design' scenarios. The key assumptions used for the simulation of 100,000 scenarios are presented in Table 3.

Table 3. Technical and economic analysis for several units.

	Ref strategy	Improved strategy	
	All units	Current design	Improved design
Inspection frequency	6 years	5 years	5 years
Start-up/shut-down frequency	1X day	1X day	minimal
Replacement date	2070	2100	2100
NPV		~z M$/unit	~2z M$/unit

The constraints for this analysis are global:

- The number of start-ups of all the plant's units must be respected in order to follow the energy demand (e.g.: morning/evening peaks);
- Not enough resources and operating budgets to inspect all units each year;
- Hiring more resources is not a viable solution because downtime can increase rapidly if several groups are shut-down.

The improved strategy chosen for the simulation is hypothetical and the objective here is only to demonstrate the relevance of a technico-economic analysis. Before identifying this improved strategy, it is necessary to better understand how the scheduling of the units is performed at the plant. Often, it is the most efficient turbine that starts first and is shut down last, since the Hydro-Quebec wishes to use this turbine as long as possible. On the other hand, the last turbine started (much less often used) and all the others are therefore subject to more shutdowns/start-ups. If prognosis is considered rather than performance for scheduling start-ups, the intuitive solution would probably be to spare the turbines that have the current design by applying the least amount of shut-downs/start-ups (first started). However, the five turbines with the current design will be used in the averaged-sollicited start-up frequency range even though they may wear out a little more because the effect of shut-downs/start-ups is negligible compared to the effect of their steady state stress magnitude. They will probably have to be replaced sooner than expected anyway. These turbines will be inspected every five years, but the analysis needs to be updated because this frequency should change as soon as a crack appears, for example. The three other turbines (improved design) could, if necessary, be inspected less frequently, especially in the context of limited inspection resources and early cracking not being expected if we limit start-ups/shut-downs for these turbines. For now, the inspection of these turbines is performed every 5 years in the simulation to optimize the NPV. In this first analysis, which demonstrates the relevance of the exercise, the generators are all considered as reliable as the turbines. Since this is only an improved scenario in the sensitivity analysis, a deeper research, including an optimization tool, should bring additional gains.

9 Conclusion

It was difficult to define accurately the hypotheses that had to be considered in the model for the eight units in the present study, but this work has allowed the most important economic levers to be exposed. Therefore, identifying the right simulations, measurements, inspections/maintenance and other R&D work to be done can help improve the accuracy of the hypotheses considered for the inputs to the calculations/simulations and can also help optimize the various impacts (solutions and R&D projects) resulting from these studies.

The present analysis demonstrates that maintenance and operating plans need to be adjusted to minimize the probability of regret when optimizing NPV by playing on replacement dates while considering equipment prognoses. For this study, the impact of inspection frequency, start-up frequency and deferred equipment replacement have been evaluated by comparing them to a reference scenario where the equipment would have been operated under traditional frameworks. The analysis demonstrates the following points:

- It is possible to observe the impacts of the improvements (frequency of inspection, of shut-downs/start-ups, replacement date, etc.) separately in the technico-economic analyses;
- Unexpected breakdowns can be taken into account for informed risk management;
- Investment deferrals can be analyzed taking into account the risk;
- Risk management proposes the updating of models (e.g. when a crack appears).

In fact, the model presented allows to define the required maintenance and operation plans according to the postponement that will be chosen in a global optimization of the power station park. The sensitivity analysis therefore shows the importance that operation and maintenance forecast be carried out according to the specific design of each runner and the deferral date chosen in the context of this global optimization. With this kind of analysis, the decision-maker could possibly (while controlling the risk):

- Distribute the fatigue load over the different turbine runners by increasing the wear on some of them to protect others in order to delay their replacement;
- Postpone as much as possible the replacement of those runners that will not be subjected to fatigue loading.

The present analysis is based only on the actual condition of the two main components: generator and turbine runner. We could consider all the components that are involved in the dismantling of the unit. Elsewhere, the benefits of power increases will eventually have to be considered to provide full value to the proposed study. Finally, this kind of study could allow one to see the impact of an increase in start-up/shutdown frequency when adding 2 more solar energy sources in the network for example.

Acknowledgments. This report would not be possible without contribution from the EDF R&D team and the utilization of VME they provide us for our modeling and case study. Financial support by the Hydro-Québec research institute in Varennes (Québec) is also gratefully acknowledged.

References

1. Lonchampt J (2017) VME a tool for risk informed engineering asset man-agement
2. Blancke O, Tahana A, Komljenovic D, Amyot N, Lévesque M, Hudon C (2018) A holistic multi-failure mode prognosis approach for complex equipment
3. Immarigeon A, Internal note, EP-16208-GR-GÉNÉRAL-GTAH-TURBINE # HydroDoc Prod 116754927
4. Immarigeon A, Internal note EP-17202-GR-GÉNÉRAL-GTAH-TURBINE # HydroDoc Prod 120350748
5. Immarigeon A, Internal note EP-17202-GR-GÉNÉRAL-GTAH-TURBINE # HydroDoc Prod 120503675
6. Immarigeon A, Internal note, EP-17202-GR-GÉNÉRAL-GTAH-TURBINE # HydroDoc Prod 120621811
7. Gagnon M, Tahan A, Bocher P, Thibault D (2012) A probabilistic model for the onset of High Cycle Fatigue (HCF) crack porpagation: application to hydroelectric turbine runner. Int J Fatigue 47:300–307
8. Kitagawa H, Takahashi S (1976) Applicability of fracture mechanics to very small crack. In: ASM proceedings of 2nd international conference. mechanical behaviour of materials, Metalspark, Ohio, pp 627–631
9. Gagnon M, Tahan A, Bocher P, Thibault D (2013) On the fatigue reliability of hydroelectric Francis runners présentée à 5th Fatigue Conference, Fatigue Design

Condition Monitoring and Assessment

Condition Assessment of Engineered Assets in the Era of Society 5.0

Joe Amadi-Echendu(✉), George Botlholo, and Keaton Raman

Department of Engineering and Technology Management,
University of Pretoria, Pretoria, South Africa
joe.amadi-echendu@up.ac.za

Abstract. In the era of *Society 5.0* powered by fourth industrial revolution technologies, the pervading cliché of "information about everything" is aggressively transforming how we monitor and assess the reliability, resilience and vulnerability of engineered assets such as personal gadgets, equipment, machinery, interconnected and interdependent facilities and infrastructure that constitute modern day cyber physical systems. Two case studies of conventional approaches to condition assessments are briefly discussed in the paper. Given that the era of *Society 5.0* proffers huge technology-driven paradigm shifts, the contention is that the sustainability imperative demands a wider and more holistic approach to condition and performance assessments of engineered discrete assets and asset systems.

1 Introduction

The era of *Society 5.0* and fourth industrial revolution (4IR) is characterised by the technology-driven fusion of cyber physical systems with socio-economic and socio-political endeavour inextricably interwoven within the natural environment and ecology. The era also features very rapid evolutionary and transformative change due to exponential interconnectivity and interdependence between the physical and social systems and their respective and combined interaction with the environment and ecology. *Society 5.0* is a "society of intelligence" founded upon the complex consequence of rapid and high degree of convergence between the virtual (cyber) and real (physical) realms [1]. The era is also regarded in terms of the trendy acronym: "VUCA i.e., *volatility, uncertainty, complexity,* and *ambiguity"* [2] and characterised by (i) unexpected or unstable change, (ii) cause and effect uncertainty, (iii) vagarious hyper- connectivity and communicability, and (iv) "unknown unknowns".

In otherwords, VUCA manifests the technology-driven cybernetic fusing and inextricable interaction between physical and social systems and the natural world. Notwithstanding the VUCA challenge, it is generally acknowledged that 4IR technologies such as artificial intelligence (AI), augmented reality, distributed ledger (block chain), and internet-of-things (IoT) provide unprecedented possibilities and opportunities for smart and intelligent management of engineered discrete assets and asset systems [3].

In this era, the hyper interconnectivity and inter-communicability capabilities provided by 4IR technology platforms transforms our traditional view from discrete assets

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 335–343, 2022.
https://doi.org/10.1007/978-3-030-96794-9_30

(i.e., equipment, gadgets, machines, infrastructure, etc.) into asset systems (conglomeration of interdependent discrete assets). This notion of cyber physical systems (CPS) is widely acknowledged – that is, interconnecting, inter-communicating and interdependent systems that are built from, and depend upon, the seamless integration of computational algorithms embedded within man-made things. The embedding of computing and communication capabilities in man-made things[1] in the physical environment enables the monitoring, control, and coordination of the operations of such IoT-enabled[2] and interdependent gadgets, equipment, machinery, facilities and infrastructure. It is remarkable that human-machine interfaces and humanoid robots are becoming increasingly uncanny; such robotic assets often make sophisticated decisions in a way that tantalizes human intelligence.

IoT-enabled things generate tremendous amounts of data and information; and the profusion has been captured in slogans like "data is the new gold" and "information about everything" [4]. The exponentially increasing amounts and variety of data demand the application AI algorithms to provide factual insights necessary for informed decision-making.

The era *Society 5.0* also magnifies the sustainability paradigm in the sense that CPSs are inextricably interwoven with human social endeavour within the natural world. Consequently, human activity in conjunction with natural world effects represents sources of risk to engineered asset systems [5]. Hence the monitoring and assessment of risk (i.e., reliability, criticality and integrity), resilience and vulnerability have become indispensable for sustainable management of CPS. The proposition in this paper is that risk, resilience and vulnerability assessments should be integrated to indicate an overall condition that informs holistic decisions regarding management of 4IR interdependent discrete and asset systems.

The discourse in this paper examines two case studies that highlight the need to integrate risk, resilience and vulnerability assessments into an overall indication of asset condition for sustainable management of cyber physical systems. A theoretical review on risk, resilience, and vulnerability is briefly presented in Sect. 2. This is followed by a concise summary of two case studies in Sect. 3. Section 4 includes some discussions and concluding emphasis on the importance of making higher value asset management decisions based on holistic condition assessment of discrete assets and asset systems.

2 Condition, Risk, Resilience and Vulnerability

A primary reason for monitoring the technical condition of a discrete asset, and for assessing the risk, resilience, and vulnerability of an asset system is to obtain data and information for making relevant decisions. Monitoring involves observation and tracking of progress at regular intervals. There is extensive literature on conventional or technical condition monitoring (see, for example, [6–9]) which generally deals with the technicalities of sensing particular physical parameters, and tracking trends or patterns of changes manifesting in the signals obtained from the sensors. Sophisticated computational techniques are applied to manipulate and analyse data derived from the sensor signals. The

[1] …and now also in human beings!

[2] containing advanced sensing, communication, and connectivity capabilities.

analyses results provide further data and information for the identification, diagnosis, and prognosis of deviations from pre-defined normalcy. The popular tendency is to focus on the technical characteristics of the composition of a discrete asset or system. Further analyses may be done to diagnose the causes of the technical deviations, and/or to forecast/predict the likelihood of deviations. A crucial point here is that monitoring may occur continuously or at regular time intervals. This strict requirement of regularity means that the time intervals must be pre-determined *ab initio*, otherwise the validity of monitoring becomes questionable.

An assessment invariably includes monitoring but more importantly involves appraisal, evaluation or judgment against an established benchmark or desired performance. With regard to a discrete asset or a system of interconnected assets, various parameters can be monitored prelude to assessment, ranging from basic physical and purely quantitative parameters like pressure, temperature and cost to more qualitative parameters like safety, utility, environmental impact and overall value. Monitoring may indicate the incipience or incidence of change or deviation; an assessment may then be done to ascertain the magnitude and impact of the deviation beyond an established or desired norm. An assessment may be conducted at any point in time as situation demands (e.g., as part of an investment appraisal exercise, or for insurance purpose), so the requirement of time interval regularity is not strict. From the foregoing discourse, it is important to reiterate that assessment encompasses monitoring, albeit that monitoring provides input to, and forms part and parcel of assessment. The paramountcy of the value ethos in asset management [10] means that condition assessment inherently combines risk, resilience and vulnerability through the conflation of quantitative and qualitative parameters, even though people with technical persuasions zealously prefer quantitative descriptions. For brevity, the concepts of risk, resilience, vulnerability, and condition assessments are summarized as follows.

2.1 Risk Assessment

ISO [11] defines risk as the negative 'effect of uncertainty on objectives'; this is essentially a reduction of uncertainty into threats. By this convention, risk management involves identification, evaluation and prioritization of threats (i.e., sources of risks) followed by coordinated application of resources to monitor the probabilities, and to mitigate, and/or to control the consequences of the threats. Considering that a threat to an asset represents a source of risk, it follows that managing risk is an implicit task of managing assets [12, 13]. After all, any source of risk represents a threat to something of value.

2.2 Resilience Assessment

Resilience is generally regarded in terms of the disruptive effects of *vis major* or *casus fortuitus* natural phenomena like earthquakes, floods, lightning, etc. The stresses arising from such supervening events not only tend to degrade performance but also cause damage to discrete assets and CPS. The motivation for managing resilience is to provide assurance that a discrete asset or a system of engineered assets will not only absorb the shocks and recover from disturbing stressors but also, adapt and transform in consonance with intransigent, evolutionary, and vagarious nature of supervening stressors. Preliminary assessment of resilience involves identification of robustness, i.e., in terms of the strengths or inherent redundancies of an asset [14]. Detailed assessment of resilience involves (i) retrospective identification of robustness; (ii) measurement of resourcefulness; and (iii) prospective estimations of the likelihood of occurrence and effects of supervening events.

2.3 Vulnerability Assessment

Whereas risk is seen from the perspective of an external threat, vulnerability focuses on the exposure or susceptibility to a threat, that is, on the weaknesses inherent in an asset. An asset (or anything, really) is regarded as being vulnerable if the inherent weakness is without protection, which means that the asset remains exposed to the possibility of being attacked at the point of its inherent weakness. In terms of sustainability, vulnerability may be defined as the degree to which an asset is likely to underperform when its weakness is exposed to transient, intransigent, evolutionary, and vagarious change [15, 16]. Resilience and vulnerability are correlated in the sense that mitigating vulnerability can be tantamount to increasing resilience. Vulnerability assessment involves identifying weaknesses inherent in an asset, monitoring the susceptibility of the weaknesses to various internal and external threats, and providing necessary knowledge for coordinated application of resources, either to prevent particular attacks (i.e., forms of stressors), or to protect the asset from exposure to attack, or to mitigate the consequences of an unfortunate attacks on the weaknesses of the asset.

2.4 Condition Assessment

By conventional preference of the technical disciplines, condition assessment has become narrowly regarded as a technical inspection and evaluation to determine 'wear-and-tear' so as to establish the maintenance requirements for an engineered asset. Broadly speaking, condition assessment involves empirical examination of fitness for purpose, that is, the overall value of an asset. Thus, the *condition* of an asset must describe the state of the asset compared against all the reasons for the existence of the asset as determined by all stakeholders. This wider viewpoint encompasses the extent of damage, degradation, deterioration or deviation in the technical characterisation of an asset. On the basis of fitness for purpose, we posit that condition assessment of discrete assets or asset systems should concurrently embrace the following dimensions – (i) functional performance (e.g., benefit in terms of ergonomics, safety, utility, social good); (ii) technical characteristics (e.g., reliability, criticality, integrity); (iii) economic/financial performance (e.g., economic cost/profit); and (iv) environmental performance (e.g., pollution, waste, impact on ecology).

2.5 Assessment Coding

With regard to making management decisions, it is conventional to summarise the outcome of an assessment in terms of codes. Managing engineered assets encompasses multidisciplinary endeavour, hence there are numerous examples of assessment codes. Most technical condition assessment codes, especially for various types of discrete civil/structural assets and infrastructural facilities are formally standardised and in the public domain (e.g., [17]), others are proprietary and remain privately confidential. Although most assessment methods tend to be biased toward the technical characteristics of discrete assets, however, to avoid perplexed decision making it is pertinent to conduct assessment at the appropriate hierarchical levels of asset systems. Furthermore, it is vital that assessment codes facilitate unambiguous and pertinent decisions with regard to the overall condition that encompasses the risk exposure, resilience and vulnerability of a discrete asset or a system of interconnected assets.

3 Case Studies

3.1 Condition Assessment – Railway Infrastructure

This subsection provides a summary of a case study attempt to assess the condition of railway infrastructure. The study was limited to a narrow and technically biased assessment of a section of a railway track as part of the hierarchical structure of the railway infrastructure system. The assessment was premised on identifying the types of defects associated with a railway track. Identification of track defects provides a quality index [18] which can guide necessary maintenance interventions so as to improve track condition (cf: [19]). Furthermore, the location of the defects provides indication of the vulnerabilities associated with the railway track. The assessment included the agglomeration of previous track geometry data over a specific section of a heavy haul railway line. The regular monitoring of track geometry provided data for the identification of defects, as well as input for manipulating maintenance interventions in order to eliminate or mitigate risks posed by the defects. In order to assess the technical condition of the section of the railway track, measurements of track quality data over a five-year period were collated to facilitate retrospective examination, and to ascertain whether previous maintenance interventions contributed to improved track quality in subsequent years. Sample measurements of the track quality (i.e., monitoring data for years 1 & 3) are displayed in Fig. 1. It was acknowledged by the rail infrastructure managers that track quality is typically used to indicate whether the track is in a condition for the safe passage of trains.

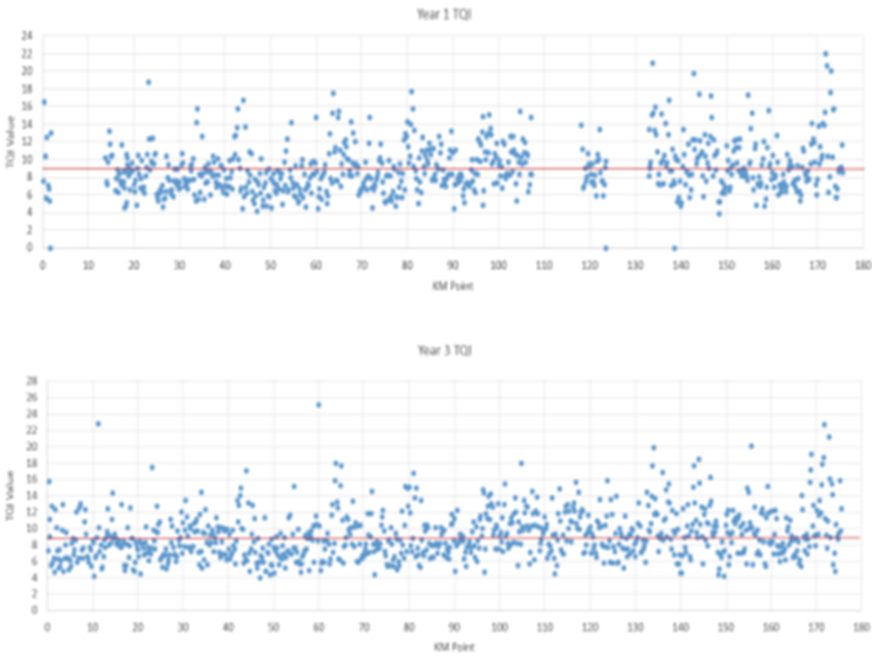


Fig. 1. Track quality data monitoring

In the top part of Fig. 1, the vertical axis represents the track quality index (TQI), while the horizontal axis represents the 175 km section of the rail track under study. Some of the missing data have been attributed to a train passing through a particular subsection of track during the time of measurement, and also due maintenance activities taking place within the area, thus forcing the recording car to be moved to an adjacent line. The variance apparent in the TQI data raises curiosity about the recorded measurements. Using TQI codes preferred by the rail infrastructure managers, the result of the condition assessment is summarised as shown in Table 1. As expected, the trend over the 5-year period indicates an overall increase in TQI value, reflecting the fact that more sections of track are showing increasing defects.

Table 1. Summary of railway track technical condition assessment over 5-year period.

TQI	'Condition'	Action	% Track condition per year*				
			Yr1	Yr2	Yr3	Yr4	Yr5
1–2.9	Good	Maintain	20	32	10	15	5
3–5.9	Safe	Treat	25	15	27	10	22
6–9	Caution	Immediate repair	28	27.5	20	18	28
>9	Danger	Emergency repair	27	25.5	43	57	45

*for confidential reasons, these percentages are relative to the actuals obtained from our assessment!

3.2 Condition Assessment – Primary Healthcare Facility

The second case study regards an assessment of a primary healthcare facility. The intriguing question was whether the composite assets were in a condition to respond to increasing burden of non-communicable diseases. The method used in [20] was adapted to partition the health facility into forty-four assessment 'bays'. The process primarily involved physical inspection of the respective bays. During the inspections, normative and statutory check sheets were used to collect and record as much data and information as was possible within the short time frame allowed for the conduct of the assessment. The check sheets covered aspects such as building structures, water and sanitation facilities, waste control and disposal facilities, air quality (heating, ventilation, air conditioning), lighting, energy supply and reticulation, machinery and equipment, pavements, security, etc. The assessment coding for the particular healthcare facility is illustrated in Tables 2 and 3. For brevity and confidential reasons, it is indicated that, of the 450 assets assessed, only 4% could be classified as 'excellent', 73% were 'good', 9% were 'fair', 8% were 'poor', and 6% were in 'very poor' condition.

Table 2. Five-point condition coding for case study health facility assets.

Code	Description
¤Excellent	Full accordance with purpose; with additional, appreciated capabilities or characteristics considering some relevant facts about the user and the environment of operation
¤Good	Full accordance with purpose; considering some relevant facts about the user and the environment of operation
¤Fair	Accordance with the purpose but not to the level required to fully satisfy the purpose because of an acceptable compromise on quality, considering some relevant facts about the user and the environment of operation
¤Poor	Accordance with purpose, but not all aspects of purpose because of some unacceptable compromises on quality or safety, considering some relevant facts about the user and the environment of operation
¤Very poor	Non-accordance with purpose because of unacceptable compromises on quality, function or safety, considering some relevant facts about the user and the environment of operation

Table 3. Assigning condition to assessment bays.

Grade	plus	minus	plus minus
¤Excellent		Excellent with some below excellent	
¤Good	Good with some excellence	Good with some not as good	Good with some excellent and some not as good
¤Fair	Fair with some better	Fair with some worse	Fair with some better and some worse
¤Poor	Poor with some fair	Poor with some unserviceable	Poor with some fair and some unserviceable
¤Very poor	A minority of assets are serviceable		

4 Discussion and Concluding Remarks

In the first case on railway infrastructure, the condition assessment was based on an index deduced solely from measurements designed to detect track defects. Whereas this technical dimension of condition assessment is typical, however, it is very biased towards reliability measurement and does not necessarily answer the question of fitness for purpose, e.g., how much payload can be put on the track and at what speeds? Nevertheless, such indices are useful to improve planning for maintenance interventions in order to mitigate risk of track failure due to track defects. The second case study was particularly focused on the real condition assessment question of fitness for purpose (i.e., overall value proposition for the healthcare facility). The contention is that the classification of the assets according to purpose facilitates higher value decision making, albeit that concerted and disciplined efforts are required to conduct the wider scope of condition assessment.

Although the first case study somewhat considered the risk of track failure due to track defects, however, neither of the two case studies involved detailed assessment of condition in terms of risk, resilience and vulnerability. The same applies to other case studies that the authors have been involved with. The salient argument from conducting several of these narrow technical condition assessments suggests that there is still a way to go to utilise the data profusion and analytic capabilities proffered by IoT and AI towards sustainable management of engineered discrete assets and systems of assets.

References

1. Salgues B (2018) Society 5.0: industry of the future, technologies, methods and tools. ISTE Ltd and Wiley. ISBN 978-1-78630-301-1
2. Bennet N, Lemoine GJ What VUCA really means for you. Harv Bus Rev 92(1/2) (2014)
3. Wilson HJ, Daugherty PR (2018) Collaborative intelligence: humans and AI are joining forces. Harv Bus Rev
4. Al-Ak'hali MS, Fageeh HN, Halboub E, Alhajj MN, Ariffin Z (2021) Quality and readability of web-based Arabic health information on periodontal disease. BMC Med Inform Decis Mak 21:41. https://doi.org/10.1186/s12911-021-01413-0
5. Asian Development Bank and Global Center on Adaptation (2021) A system-wide approach for infrastructure resilience. Technical Note ISBN 978-92-9262-694-5
6. International Journal of Condition Monitoring ISSN 2047-6426
7. Galar D, Kumar U (2017) eMaintenance - essential electronic tools for efficiency. Academic Press, Cambridge. ISBN 978-0-12-811153-6
8. Deighton MG (2016) Facility integrity management: effective principles and practices for the oil, gas and petrochemical industries. ISBN 978-0-12-801764-7. https://doi.org/10.1016/C2014-0-00795-2
9. Mital A, Desai A, Subramanian A, Mital A (2015) Product development: a structured approach to consumer product development, design, and manufacture. Elsevier, Amsterdam. ISBN 978-0-12-799945-6
10. Amadi-Echendu JE (2017) Developments in the management of engineered assets: before and beyond ISO 5500x. Keynote Address. In: Proceedings IEOM international conference, Rabat Morocco, 11–13 April 2017
11. ISO 31000:2009. Risk management – Principles and guidelines

12. Todinov MT (2007) Risk-based reliability analysis and generic principles for risk reduction. Elsevier Science. ISBN 978-0-08-044728-5. https://doi.org/10.1016/B978-0-08-044728-5.X5000-8
13. Davis R (2017) An Introduction to Asset Management. ISBN 978-0-9571508-3-6
14. Liu W (2014) The application of resilience assessment — resilience of what, to what, with what? A case study based on Caledon, Ontario, Canada. Ecol Soc 19(4):21. https://doi.org/10.5751/ES-06843-190421
15. Boyce R (2019) Vulnerability assessments: The pro-active steps to secure your organisation. SANS Institute
16. Moret W (2014) Vulnerability assessment methodologies: A review of the literature. A Report for USAID ASPIRES PEPFAR fhi360
17. Monica Watson M (2018) Documentation and coding practices for risk adjustment and hierarchical condition categories. J AHIMA 89(6). extended online version
18. Fontul S, Paixao A, Solla M, Pajewski L (2018) Railway track condition assessment at network level by frequency domain analysis of GPR data. Remote Sens 10(4):26
19. Fröhling R, Spangenberg U, Hettasch G (2012) Wheel/rail contact geometry assessment to limit rolling contact fatigue initiation at high axle loads. Veh Syst Dyn 50(sup1):319–334
20. Shohet IM (2003) Building evaluation methodology for setting maintenance priorities in hospital buildings. Constr Manag Econ 21:681–692

Model Proposal for Failure Detection and Classification of Internal Combustion Engine Operating Condition

Fernanda Mitchelly Vilas Boas[1](✉), Frederico Oliveira Assunção[2], Luiz Eduardo Borges da Silva[2], Erik Leandro Bonaldi[2], Levy de Lacerda de Oliveira[2], Germano Lambert-Torres[2], Claudio Inacio de Almeida Costa[2], Helcio Francisco Villa-Nova[1], Sandro Fernandes de Souza[3], Danilo de Souza Lima[4], Josué da Costa Lacerda[2], and Bruno Renó Gama[1]

[1] Universidade Federal de Itajubá (UNIFEI), 1303 BPS Avenue, Itajubá, MG 37500903, Brazil
d2019102116@unifei.edu.br

[2] Instituto Gnarus, 53 Jose Juliao Simoes St, Itajubá, MG 37502495, Brazil

[3] RAESA, 174 Rodovia Br, Manaus, AM 69099-899, Brazil

[4] UNIPAC, 338 Rodovia MG, Antônio Carlos, MG 36220-000, Brazil

Abstract. Internal combustion engines are intermittently operating in Thermoelectric Power Plants (TPP) and consequently suffer a lot of detriction. Due to this, these engines are subject to ostentatious controls of temperatures, pressures and physical properties of the working fluids at critical points on the machine to monitor the behavior and abnormal events of this machine. However, often these controls only indicate abnormalities after a failure, not providing an early study of the operational state of the engine. A trend analysis of failures in auxiliary systems of the internal combustion engine, based on the matrix structure, is proposed in this article in order to verify the operational conditions of the auxiliary engine lubrication system and the influence that this system is producing on electrical generation based on operating data from the lubrication system of an internal combustion engine model similar to a Wärtsilä 18V46-C2. The proposed math model is validated through comparison with the standard behavior of the internal combustion engine lubrication system. It is demonstrated that the results obtained with the proposed methodology correspond to the high compatibility of the simulated results with the real data of failure of the lubrication system of the analyzed engine model, which makes this math model a favorable mechanism to direct predictive maintenance and, consequently, reduce operating costs of electricity generation.

1 Introduction

The current stage of thermal machines, combined with the high demand of running for long periods of time and the need to guarantee an immediate supply of electricity, culminated in the development of techniques for the control and predictive detection of failures in the internal combustion engine.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 344–352, 2022.
https://doi.org/10.1007/978-3-030-96794-9_31

Several failure detection approaches that efficiently use the resources of the internal combustion engine have already been proposed in the scientific literature. For example, in [1] it is proposed a system capable of monitoring the operating conditions of the engine and detecting, in advance, engine failures through convolutional neural networks of one dimension (CNNs). In [2] a methodology is presented for the development of equipment capable of identifying and detecting early failures in an internal combustion engine, by means of current and voltage signals collected by measures external to the motor-generator group. In [3], the authors propose the development of a predictor system capable of performing the thermomechanical detection of failures in an internal combustion engine applying the electrical signature analysis of the generator's electrical signals. Already in [4] is presented the theoretical foundations for the development of a system capable of monitoring and diagnosing impending failures in marine engines in an on-line manner. In [5] the authors propose that inspection and fault detection in diesel engines be based on their operational conditions. In this proposal, the diagnostics rules are organized through the technique of failure modes and effects analysis (Failure Modes and Effects Analysis - FMEA) and of the distribution of failures. The operational parameters are established using a standard recognition technique, that is, compared a measured vector with a predefined fault vector and a standard diagnosis is arrived at. Already the fault matrix is formed by the junction of the measured vector with the predefined fault vector. In [6] a comparative bibliographic review of the diagnostics system that is based on the operational condition of diesel engines is presented. In it, the direction is given in the presentation of bibliographic sources of systems that detect flaws through knowledge and monitoring of the vibration signals. Already in [7] a system capable of organizing the fault diagnosis rules in the diesel engines through the fault tree analysis technique (Fault Tree Analysis - FTA) is proposed. As previously described, there are many references on failure detection using various techniques and models applied to both internal combustion engine and other types of engines. However, the method proposed in this work, in addition to presenting an innovative character, since, in the literature, no investigation uses auxiliary systems to detect failures in internal combustion engines, it also presents several advantages in its use. These advantages are based, especially, on the possibility of a complete analysis of the machine, involving all its subsystems, and what impact these subsystems are having on the efficiency of the engine as a whole. In addition, the variables used to carry out this analysis are, in their majority, variables already measured in a standard way in industrial machines, reducing invasive operations and programmed stops in the device under analysis for the insertion of measurement sensors. Therefore, the main objective of this work is to analyze the database of an internal combustion engine lubrication system similar to a Wärtsilä 18V46-C2 of a thermoelectric plant and, to determine if this lubrication system is more prone to failures, when compared to a standard behavior of the lubrication system of the analyzed engine model.

To fulfill the purposes of this work, the rest of it is organized as follows: Sect. 2 presents the development of the proposed system model, together with a more indepth analysis about the lubrication model. Section 3 presents the numerical results of the failures trend of the lubrication system when compared to the standard behavior of the lubrication system of the analyzed engine model. Finally, Sect. 4 presents the main conclusions of the work and some opportunities for future research.

2 System Model

This section will present the stages of a system model development that uses the lubrication system data provided by an internal combustion engine, similar to a Wärtsilä 18V46-C2, already existing in the thermal power plant.

In order to carry out the development of this system, it is initially assumed that the internal combustion engine analyzed be divided into six auxiliary systems, such as, lubrication, cooling, fuel, turbine, compressor and bearing, as shown in Fig. 1. It is worth mentioning that S1, S2,…, SN are just representations of the auxiliary system input variables.

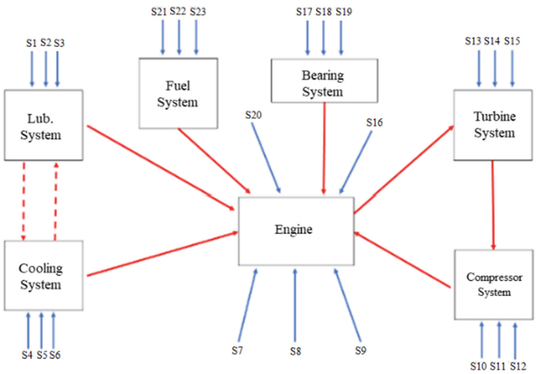


Fig. 1. Engine division into six auxiliary systems.

Then, a matrix structure is established for this engine division, that is, a submatrix is created for the analyzed engine, a submatrix for each auxiliary system of the same and a matrix that encompasses this entire division, according to Fig. 2.

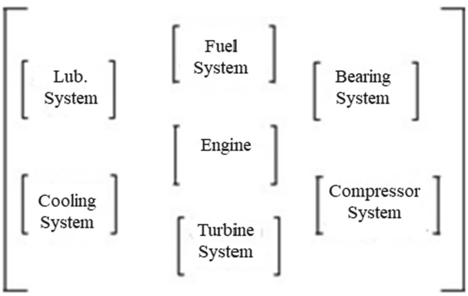


Fig. 2. Matrix structure for the system model.

Finally, for each auxiliary system of the analyzed engine, the submatrix containing the fundamental variables to characterize them is elaborated and, for the central system, the submatrix is formed by the powers and efficiencies involved in electrical generation.

According to the proposal of this work, special attention is given to the auxiliary lubrication system of the analyzed engine. Therefore, the following subsection will cover in more detail the modeling of this auxiliary system.

2.1 Lubrication System Model

The lubrication system is one of the auxiliary systems of the internal combustion engine analyzed and the main focus of this work. Fundamentally, as its name suggests, this system is lubricating the moving parts of the internal combustion engine so that metallic friction does not occur and, consequently, wear of the parts and the increase in engine temperature [8].

The lubrication system consists of crankcase, oil pump and main oil gallery. In this system, the lubricating oil is reserved in the crankcase and, only starts to circulate through the engine, when the fisherman, who is inside the crankcase, sucks and pumps this oil into the main oil gallery of the internal combustion engine. After that, this oil is sent to the filters in the engine and, then directed towards the crankshafts, the bearings, the cylinders, the rings, the heads, the tappets and the rockers [8]. Finally, when the lubricant reaches the moving parts of the engine, it returns to the crankcase through gravity, restarting the cycle. These described processes are repeated as long as the engine is running and generating energy.

The performance of the lubrication system can be affected due to several factors, such as, energy losses, irregular friction, leakage, among other factors, as shown in the fault tree of the lubrication system in Fig. 3.

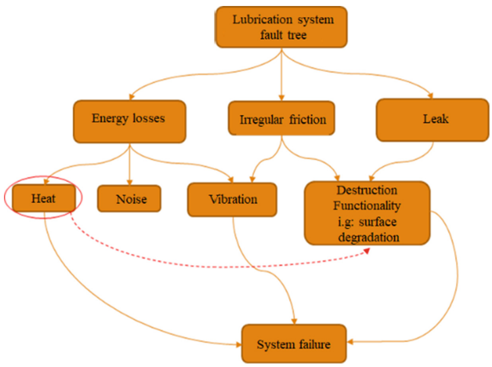


Fig. 3. Lubrication system fault tree.

One way to soften the effect of these factors is to use the lubrication system model proposed in this work and which will be described below.

In the model of the proposed lubrication system, the variables that characterize it are the magnitudes of the lubricating oil inlet and outlet temperature (T_{oi}, T_{oo}), the variation between the oil inlet and outlet pressures (ΔP), the inlet viscosities and outlet (μ_{oi}, μ_{oo}) and the respective variation of lubricating oil flow rates (Φ), as shown in Fig. 4. It is worth noting that the positions of the matrix that assume zero values can be replaced by some other variable that is fundamental for this system.

$$\begin{bmatrix} T_{oi} & \mu_{oo} & \Phi \\ \mu_{oi} & T_{oo} & 0 \\ 0 & 0 & \Delta P \end{bmatrix}$$

Fig. 4. Lubrication system submatrix.

In Fig. 5, it is presented the characteristic viscosity curve, the red curve is used to calculate the oil viscosity. For the lubricating oil flow, it is adopted the value of 1.

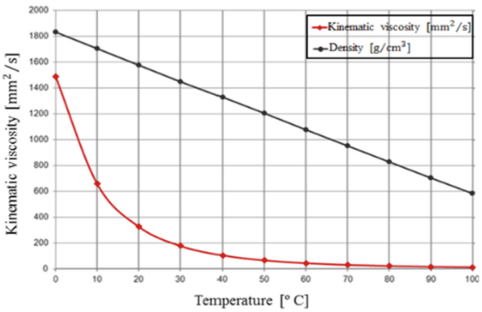


Fig. 5. Characteristic curve of viscosity as a function of temperature.

Note that it is through the resulting math model that it is possible to determine whether the auxiliary lubrication system is having a greater influence on electrical generation. Using the math model, it is also possible to identify failure tendency through comparison with standard behaviours of lubrication system. In addition, the accumulated time of the malfunction of the respective auxiliary system can also be surveyed, even if the internal combustion engine operation is intermittent and still provides a direction for predictive maintenance and, consequently, a reduction in the electric generation operating costs.

3 Experimental Results

The numerical results presented in this section were made by means of computer simulation. In these numerical results, the percentage of right matrices, the percentage of wrong matrices and the total percentage of matrices analyzed are identified, respectively by, right matrices, wrong matrices and total matrices.

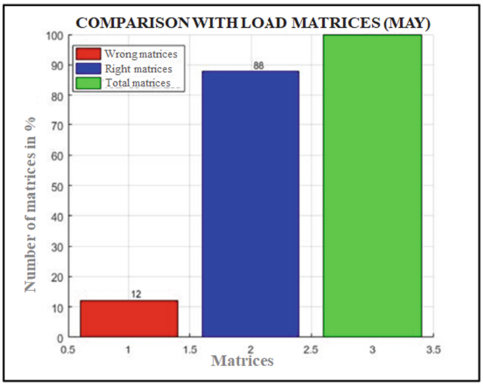
The bar graphs were obtained by the simulation of various measures of the lubrication system of the analyzed engine, each related to the generation of a matrix containing the variables, which characterize this system, imported from the databases provided by the thermoelectric power plant. After importing the data, referring to the variables that make up the matrices of this system, it is calculated, based on the characteristic curve of Fig. 5 and, added to these matrices an input and output viscosity vector. Then, the times when all these variables act together are checked. A with load/without load classification is defined by the system condition to simulate the values of the variables that correspond to the motor's active power greater than zero (with load) and the values of the variables that correspond to the motor's active power equal to zero (without load).

At this moment of the simulation, the first condition of the lubrication system is generated, the outlet temperatures with or without load must be higher than the temperatures of inlets with or without load. Consequently, for true condition the measurement matrices with or without load are then generated. From the measurement matrices with or without load, the standard average matrix structure of the data with or without load is created.

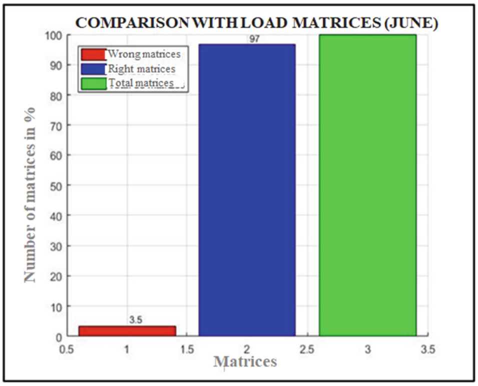
After generating the standard matrix with load or without load, a second condition is imposed on the system model, that is, the input power of the without load lubricating oil and the output power of the without load lubricating oil must be greater than zero. If this condition is true, a variation of two standard deviations is adopted, that corresponds to approximately 95% of the analysis of the data provided and the new standard matrix of the lubrication system of the internal combustion engine analyzed with load is created. Otherwise, the system will not operate without a load. It is worth mentioning that, from this point on, only the cases with load will be considered in the development of the system model.

With the measurement matrix with load and of the new standard matrix with load, the comparison of each measures matrix with the standard matrix is performed. The system model should decide by right matrices whenever the variables that make up the matrix of measures with load are within the ranges of the variables that make up the new standard matrix and should decide on the wrong matrices otherwise. The total number of matrices is calculated by adding the number of right matrices with load and the number of wrong matrices without load. Finally, the months are varied for May (a), June (b), July (c) and August (d), repeating the process to generate the bar graphs of the quantities of the right, wrong and total matrices with charge.

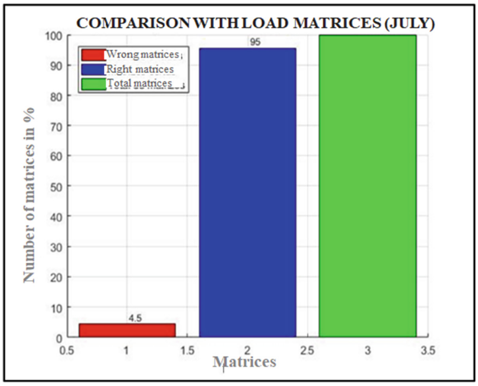
In terms of parameters used in the simulations, the data provided by the plant were considered for the variables monitored and measured at the plant. Already for non-measured variables, such as lubricating oil outlet pressure and variation of lubricating oil flow rates the value of 0 Pa and 1, respectively, is adopted. The lubricating oil inlet and outlet viscosity is calculated from the lubricating oil inlet and outlet temperatures using the characteristic red curve of Fig. 5. For the simulations of the bar graphs of the comparison of matrices with load the months were adopted May (a), June (b), July (c) and August (d).



(a)

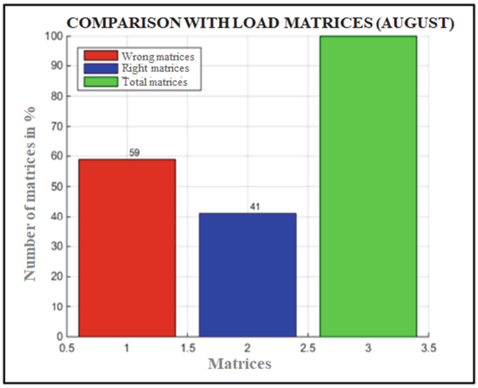


(b)



(c)

Fig. 6. Charts of quantities of matrices with load (%) for the month of May (a), June (b). Charts of quantities of matrices with load (%) for the month of July (c) and August (d).



(d)

Fig. 6. continued

In Fig. 6 are presented some bar graphs that show the quantities of right, wrong and total matrices in %, for the months of May (a), June (b), July (c) and August (d).

It can be seen in Fig. 6, that the lubrication system of the analyzed engine did not suffer stress in the months of May (a), June (b) and July (c). And as these three months pass, the performance of this engine tends to improve even more. It is also noted that the month of June (b) is the one with the best performance, since it has the highest percentage of right matrices. In addition, it is also noted that the number of right matrices from June (b) to July (c) falls, but in a very small percentage. Finally, it is also noted that, in August (d), the performance of the engine analyzed is reversed, which means that the lubrication system underwent a maximum stress level, leading to the machine stopping.

4 Conclusion

This article presented the stages of a system model development and analyzed the failure trend of the internal combustion engine lubrication system similar to a Wärtsilä 18V46-C2, based on the data of this system, supplied by the thermal power plant. The proposed model proved to be innovative, since, although there are several works in the literature that perform the detection of failures in internal combustion engines, no work performed this detection based on the auxiliary systems of these engines. Thus, as mentioned before, the math model proposed here is innovator and validated by comparing the behavior of the measurement matrices of the lubrication system found in the simulations with the standard behavior of the lubrication system of the internal combustion engine. In addition to being innovative, this model is flexible, allowing to include in the matrix of the lubrication system any other fundamental variable to such system, and comprehensive, being able to be applied to any thermoelectric with internal combustion engines. Already the analysis carried out revealed that the analyzed machine did not suffer stress in the months of May, June and July, which are shown with a higher percentage of right matrices. It was also possible to conclude that the month of June is the one that achieves the best performance, since it has the highest percentage of right matrices. Finally, there was a tendency to failure of the lubrication system of the analyzed engine, according to the inversion of the derivative of quantities from correct matrices to wrong matrices in month August, concluding, therefore, that this month there really was machine stop. These bar graph analyzes are of paramount importance as they will help both the TPP maintenance department and the engine operators to carry out predictive maintenance and, consequently, reduce operating costs of electrical generation.

As future research opportunities related to the theme of this article, mention: i) the failure trend analysis of the other internal combustion engine subsystems. ii) the adoption of a comparative scenario in which one engine had no accident and the other had.

Acknowledgments. The authors would like to thank Usina Termelétrica Rio Amazonas Energia S.A. - RAESA and the National Agency for Research and Development (ANEEL P&D) for supporting this project.

References

1. Ince T, Kiranyaz S, Eren L, Askar M, Gabbouj M (2016) Real-time motor fault detection by 1-D convolutional neural networks. IEEE Trans Industr Electron 63(11):7067–7075
2. Bonaldi EL et al (2014) Sistema para Diagnóstico Termomecânico de Falhas em Motores de Combustão Interna. https://doi.org/10.13140/RG.2.1.4713.0642
3. Mendonça P et al (2015) Desenvolvimento de Sistema Preditivo para Diagnóstico Termomecânico de Falhas em Motores de Combustão Interna. https://doi.org/10.13140/RG.2.1.1043.0487
4. Maffei ACL (2007) Monitoração e Diagnóstico de Motores de Propulsão Marítima. Doctoral dissertation, Dissertação
5. Fagerland H, Rothaug K, Tokle P (1978) Monitoring and diagnosing process deviations in marine diesel engines. Ship Res Inst Norway 90:321–340
6. Elliott C, Banisoleimanv K (1990) Advanced Health Monitoring Systems for Marine Medium Speed Diesel Engines - A Naval and Merchant View. ICMS 90 – Maritime Systems Integrity paper 1, 3–17
7. Perakis AN, Inozu B (1990) Reliability analysis of great lakes marine diesels: state of the art and current modeling. Mar Technol SNAME News 27(04):237–249
8. Tillmann CADC (2013) Motores de combustão interna e seus sistemas. Instituto Federal de Educação, Ciência e Tecnologia, 9–25

Fault Diagnosis and Isolation for Diesel Engine Combustion Chambers Based on Autoencoder and BP Neural Network

Yonghua Yu(✉), Jia Hu, and Jianguo Yang

Wuhan University of Technology, Wuhan 430063, China
yyhua@whut.edu.cn

Abstract. In order to improve the efficiency and accuracy of diesel engine combustion chamber fault isolation, a method of combining the feature dimension reduction of AutoEncoder network and the fault isolation of BP neural network was proposed based on acoustic emission signals. Taking a Z6170 diesel engine of China ZICHAI company as an example, fault simulation tests of exhaust valve and piston rings under experimental environments were carried out, and the acoustic emission signals of the cylinder head were collected, then the time-domain, frequency-domain and other characteristic parameters of different signal sections in the whole cycle were extracted. The dimension of characteristic parameters was reduced by using AutoEncoder network, BP neural network was used to fault diagnosis and fault isolation, so that a fault diagnosis and fault isolation model of combustion chamber components was established. After training and verification of the model, it shows that the proposed diagnosis and isolation method is effective with capability of identifying the faults of exhaust valve and piston ring for the combustion chamber parts of diesel engines, therefore, it is promising to detect and isolate the condition of combustion components automatically.

Keywords: Diesel engine · Fault diagnosis · AutoEncoder network · BP neural network · Acoustic emission

1 Introduction

Economy, high efficiency, long life and intelligence are the themes of the development of modern marine diesel engine. The marine diesel engine is the heart of the ship's power, its reliability directly affects the economics and reliability of related equipment and systems. The sea conditions faced by ships are complex and changeable, and the working environment of diesel engines is harsh. If a marine diesel engine breaks down during voyage and is not maintained in a timely and effective manner, it will cause economic loss in light situations, and may cause the ship to lose power due to shut down, and causes severe economic loss and even endangers the life safety of the crew on board. Therefore, the reliability requirements of marine diesel engines are higher. The working environment of the ship's diesel engine combustion chamber components is the worst, and the probability of failures accounts for the largest proportion in the entire diesel engine.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 353–363, 2022.
https://doi.org/10.1007/978-3-030-96794-9_32

Therefore, the fault diagnosis of the combustion chamber components is a prerequisite for the development of diesel engines towards intelligence. The acoustic emission signals of marine diesel engines contain abundant information of fault state. Due to its unique advantages of non-destructive and high signal-to-noise ratio, the application of acoustic emission monitoring technology in diesel engine monitoring and diagnosis has been paid more and more attention. Most of the traditional diesel engine fault diagnosis methods are "shallow learning methods", their learning ability has certain limitations, and they cannot fully dig deep into the data, for high-dimensional data samples, there is a problem of low accuracy of fault classification. Based on the above discussion, this paper proposes a method for diesel engine combustion chamber fault diagnosis and fault isolation based on AutoEncoder [1] and BP neural network. On the basis of extracting the feature parameters of multi-dimensional acoustic emission signals, the feature parameter set is normalized, then the data set is reconstructed and reduced by the AutoEncoder network, and the fault is classified by the BP neural network. The effectiveness of the method is verified by experiments.

2 Artificial Faults Simulation and Test

Taking Z6170 diesel engine as an example, the acoustic emission signals of cylinder head under normal, air leakage of exhaust valve and worn piston rings were measured at 1000 r/min and with the diesel load at 0%, 25% and 50%. and the characteristic parameters of energy in specific time domain and power in specific frequency domain of acoustic emission signals were extracted [2, 2], as shown in Table 1.

Table 1. Characteristic parameters of acoustic emission signal

Characteristic parameter	Parameter implication
n	Speed
L	Load
P_e	Signal power of combustion section 340–380° CA signal in frequency section 8–41 kHz
P_r	Signal power within 8–39.7 kHz of 425–500° CA section
r_b	Ratio of signal energy between 340–355° CA and normal value of corresponding working condition
r_s	Ratio of signal energy between 355–370° CA and normal value of corresponding working condition
r_a	Ratio of signal energy between 370–420° CA and normal value of corresponding working condition
r_o	Ratio of signal energy between 340–420° CA and normal value of corresponding working condition

The time-domain energy of the signal is shown in Eq. (1):

$$P_t = \sum_{-\infty}^{\infty} |x(n)|^2 \tag{1}$$

Besides analyzing the time-domain signal of AE in a specific crank angle, the frequency-domain signal of the time-domain waveform, corresponding to a certain working process, can also be analyzed. The area of the PSD spectrum in a certain frequency range characterizes the energy in a certain frequency range of the signal. The definition of the energy P_f in the frequency range f_a–f_b is shown in Eq. (2) [4].

$$P_f = \sum_{i=f_b}^{f_a} p(i)\Delta f \tag{2}$$

Where P_f is the energy of the AE signal (V^2); f_a and f_b are the upper and lower limits of the frequency band (Hz); $p(i)$ is the power ($V^2 \cdot Hz^{-1}$) corresponding to frequency i; Δf is the frequency range (Hz).

Failures of air leakage of the exhaust valve and worn piston rings were simulated, including exhaust valve with three levels of air leakage (one, two, or three grooves of 1 mm × 6 mm), and piston rings with three degrees of wear (the inner ring radius and gap of the first piston ring of 0.2 mm × 1 mm, 0.4 mm × 2 mm and 0.6 mm × 3 mm, respectively). From these simulations and experiment, 773 samples were formed. Encoding the category as a label set, as shown in Table 2.

Table 2. Code of fault category

Category	Code
Normal	100
Air leakage of exhaust valve	010
Worn piston rings	001

The characteristic parameters of some samples are shown in Table 3.

Table 3. A part of feature parameter set

Fault location	n	L	P_e	P_r	r_b	r_s	r_a	r_o	Category
100	1000	0	0.291	0.0295	1	1	1	1	0
100	999	25	1.053	0.0357	1	1	1	1	0
100	1000	50	1.364	0.0267	1	1	1	1	0

(*continued*)

Table 3. (*continued*)

Fault location	n	L	P_e	P_r	r_b	r_s	r_a	r_o	Category
010	1000	0	0.127	0.00250	0.0721	0.432	0.506	0.395	1
010	1000	25	0.515	0.00696	0.131	0.528	0.611	0.517	1
010	1000	50	0.443	0.00724	0.219	0.335	0.823	0.364	1
001	1000	0	0.091	0.00981	0.217	0.318	0.934	0.362	2
001	1000	25	0.474	0.0229	0.459	0.441	1.265	0.514	2
001	1000	50	0.658	0.0135	0.440	0.492	1.086	0.540	2

The six characteristic parameters under different working conditions are shown in Fig. 1. One characteristic parameter is difficult to distinguish three kinds of conditions at the same time. At the same time, each characteristic parameter overlaps under working conditions. Therefore, it is necessary to reduce and reconstruct the characteristic parameters before fault diagnosis and fault isolation.

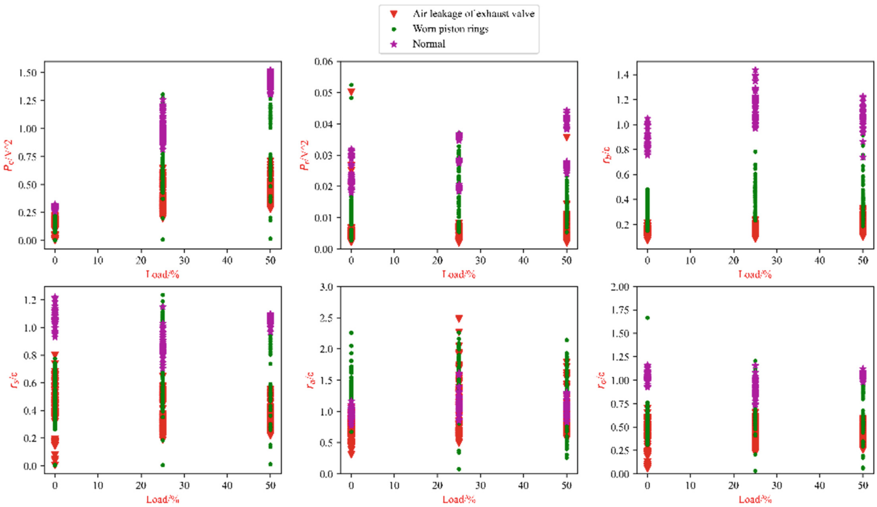


Fig. 1. Comparison of dispersion of characteristic parameters

To eliminate the dimensional influence between feature parameters, and accelerate convergence, the sample data of feature parameters are normalized, as shown in Eq. (3).

$$y = \frac{x - MinValue}{MaxValue - MinValue} \quad (3)$$

Where: x is the original sample data, y is the converted sample data, and $MinValue$ and $MaxValue$ are the maximum and minimum values in the sample respectively.

3 The AutoEncoder Network and BP Neural Network

The AutoEncoder network belongs to the field of unsupervised learning [5], which is a forward neural network aiming at reconstructing input signals. It can give a better feature description than the original data, and has a strong feature learning ability. In learning, the AutoEncoder network is often used to reduce the dimension of original data to get better results, it takes the feature parameters as input, projects the high-dimensional feature data into the low-dimensional space through the hidden layer, and reconstructs the high-dimensional data [6]; Then fault diagnosis and isolation are carried out by the BP neural network.

3.1 The AutoEncoder Network

The basic model of the AutoEncoder network can be seen as three layers of neural network, i.e. input layer, hidden layer and output layer. The sample of the input layer acts as the label of the output layer. The network structure is shown in Fig. 2.

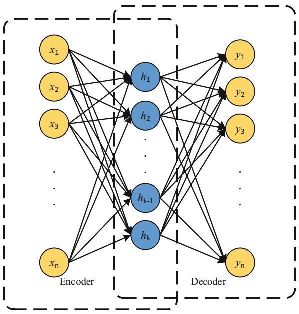


Fig. 2. The structure of AutoEncoder network

In the AutoEncoder network, the process from input to intermediate state is called encoder, and the process from intermediate state to output is called decoder. Set the training sample as X = {x(1), x(2), x(3), x(N)}, and each sample x is an n-dimensional vector.

In the encoder, the n-dimensional vector is nonlinear mapped to the k-dimensional vector h by Eq. (4).

$$h = f(wx + b) \tag{4}$$

In the decoder, the k-dimensional vector h is reconstructed to the input n-dimensional sample data y by Eq. (5).

$$y = f(w'h + b') \tag{5}$$

The sigmoid function is selected as the activation function, as shown in Eq. (6), w, b represents the weight and bias of the encoder respectively; w' and b' represent the weight and bias of the decoder respectively. The parameter of the AutoEncoder network is recorded as θ. The goal of the AutoEncoder network is to optimize the parameter θ of the model so as to minimize the reconstruction error, and achieve the purpose of dimension reduction. The loss function is the mean error function, as shown in Eq. (7).

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{6}$$

$$Loss = \frac{1}{n}\sum_{t=0}^{n} \left(y_t - \widehat{y_t}\right)^2 \tag{7}$$

3.2 The BP Neural Network

The BP neural network belongs to supervised learning, and its training process is completed by two stages: the forward propagation of input parameters and the backward propagation of errors. It constantly modifies the weight and bias between each layer, so that the optimization of the network is carried out along the direction of the fastest error reduction, and finally the error is minimized, then the network optimization stops and the training ends [7].

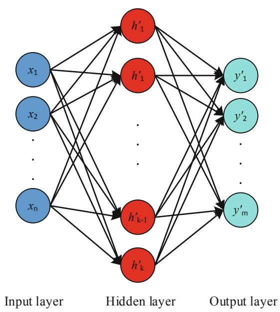


Fig. 3. The structure of BP neural network

(1) Forward propagation: the feature parameter set enters the network from the input layer, and is weighted and summed by neurons. Then, the output is generated through the activation function. In this way, the hidden layer is passed down layer by layer, and finally the output value is generated in the output layer. Sigmoid and Softmax functions are used as activation functions in the model, as shown in Eq. (9) and Eq. (10):

$$y = f\left(\widehat{w} \cdot x + \widehat{b}\right) \tag{8}$$

$$sigmoid(x) = \frac{1}{1 + e^{-x}} \tag{9}$$

$$softmax(x) = \frac{e^{x_i}}{\sum_{i=1}^{m} e^{x_i}} \tag{10}$$

(2) Back propagation: the basic idea of back propagation is to calculate the loss between the output value and the expected value through the loss function, and adjust the network parameters through the back propagation to reduce the error. The loss calculation of the BP neural network is shown in Eq. (11):

$$Loss = \frac{1}{n}\sum_{t=0}^{n}\left(y_t - \widehat{y_t}\right)^2 \tag{11}$$

(3) The accuracy calculation is shown in Eq. (12):

$$Accuracy = \frac{number(Correct\ classification)}{total\ number} \tag{12}$$

3.3 Adam Optimizes the AutoEncoder-BP Neural Network

In order to optimize the AutoEncoder-BP neural network, update the weight and bias, the Adam optimization algorithm [8] is used to replace the gradient descent (GD) algorithm to minimize the loss function. The gradient descent algorithm keeps a single learning rate (ε) to update all weights, and the learning rate does not change in the process of network training, while the Adam optimization algorithm designs independent adaptive learning rates for different parameters by calculating the first and second moment estimates of the gradient. Adam algorithm combines the advantages of the AdaGrad [9] algorithm and RMSPro algorithm. It has high computational efficiency and low memory requirements. Moreover, the diagonal scaling of Adam algorithm gradient is invariant. The algorithm principle can be expressed as [10]:

$$g_t = \nabla J(w_t) \tag{13}$$

$$m_t = \beta_1 * m_{t-1} + (1-\beta_1) * g_t \tag{14}$$

$$v_t = \beta_2 * v_{t-1} + (1-\beta_2) * g_t^2 \tag{15}$$

$$m_t^{'} = \frac{m_t}{1-\beta_1^{'}} \tag{16}$$

$$v_t^{'} = \frac{v_t}{1-\beta_2^{'}} \tag{17}$$

$$w_{t+1} = w_t - \frac{\varepsilon}{\sqrt{v_t^{'}+\alpha}} * m_t^{'} \tag{18}$$

m_t and v_t are the first and second order momentum terms respectively, and β_1 and β_2 are the exponential decay rates of the first and second moment estimates respectively, generally 0.9 and 0.999 respectively, $m_t^{'}$ and $v_t^{'}$ are respective correction values. w_t represents the weight matrix of the t time step, g_t represents the gradient of the loss function of the t time step to the weight matrix; ε is the learning rate, α is a small value (generally 1e−8) to avoid denominator 0.

Therefore, the adopted AutoEncoder-BP neural network structure is shown in Fig. 4.

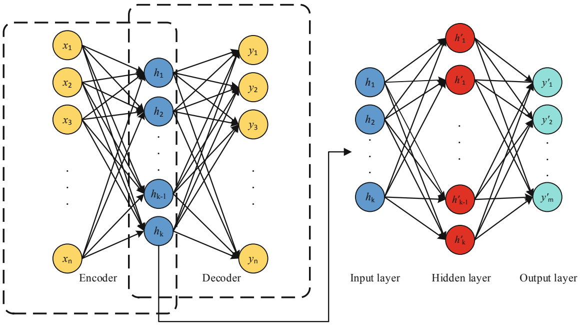


Fig. 4. The structure of AutoEncoder-BP neural network

The fault diagnosis and isolation method based on the AutoEncoder-BP neural network can be divided into four stages: feature parameter extraction stage, data normalization stage, the AutoEncoder network dimensionality reduction stage, the BP neural network fault diagnosis and isolation stage. The step flow chart is shown in Fig. 5.

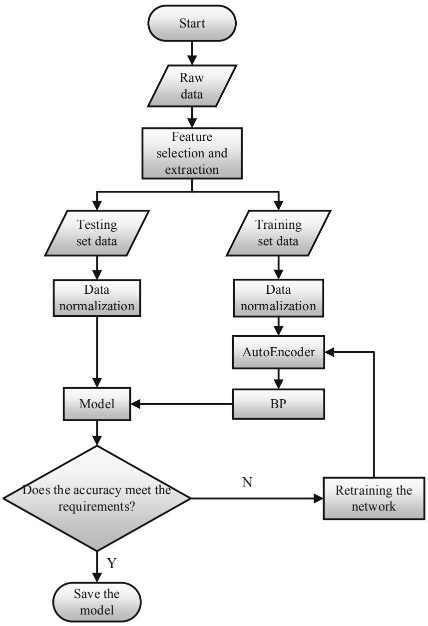


Fig. 5. Flow chart of fault diagnosis method

4 Experiment Verification

4.1 Network Parameter Setting

The AutoEncoder network uses the training set for dimension reduction training, and then reduces the dimension of the test set to obtain the 3D characteristic parameter distribution as shown in Fig. 6, where X, Y and Z are the first, second and third dimensions of the reduced dimension data set respectively.

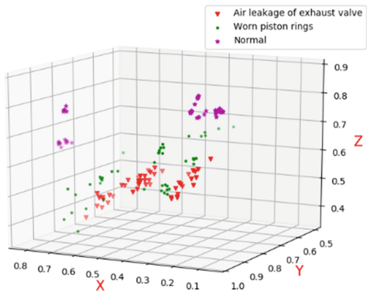


Fig. 6. Characteristics parameters dispersion after dimension reduction and reconstruction

After dimensionality reduction and reconstruction, the boundaries between the data under the three conditions of normal, air leakage of exhaust valve and worn piston rings are very clear from Fig. 6, and then use the BP neural network to classify, higher accuracy of fault classification can be obtained.

In the training process of the AutoEncoder-BP network, the setting of network layers will affect the network performance. The optimal network layers are determined by experiments. It is necessary to optimize the network layers of the BP network. In order to study the influence of different network layers on the accuracy of fault classification, the layers of the BP neural network are set as 1–4 layers respectively. It is found through experiments that the fault identification rate corresponding to different layers of the BP neural network is shown in Table 4.

Table 4. Experimental results of different layers of the BP neural network

Layers of BP neural network	Number of nodes corresponding to network layer	Failure identification rate
1	3-3	85.82
2	3-30-3	96.45
3	3-30-12-3	97.87
4	3-30-24-6-3	95.04

The single-layer BP neural network is a linear classification, so the recognition effect is poor. With the increase of layers, the processing ability of BP neural network for non-linear data is enhanced, and the fault recognition rate is increased, reaching the highest level in the third layer. However, the deepening of the network layer by layer will challenge the back propagation ability of the network. In the back propagation, the gradient of each layer is calculated on the basis of the previous layer, and the number of layers will cause the gradient to become smaller and smaller in the multi-layer propagation until the gradient disappears. So it can be seen from the table that the recognition rate of the 4-layer BP neural network is lower than that of the 3-layer BP neural network. Therefore, the BP neural network is set as three layers.

4.2 Result Analysis

After training the training set data through the BP network, the loss value iteration diagram is obtained as in the loss curve shown in Fig. 7. The test set data is used to verify after each iteration, and the accuracy iteration diagram is obtained as shown in the accuracy curve in Fig. 7.

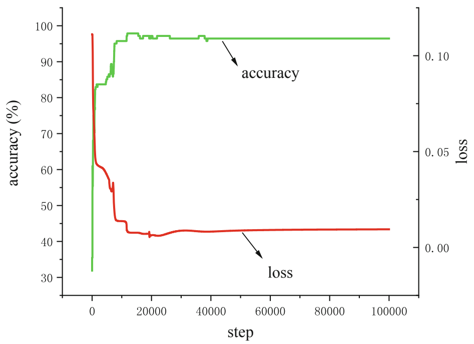


Fig. 7. Loss value iteration of training set and test set accuracy iteration chart

The loss value of the BP neural network after training is close to 0 from Fig. 7. the accuracy of the test set reaches 97.87% after training, and the recognition rate under normal diesel engine, air leakage of exhaust valve and worn piston rings is shown in Table 5.

Table 5. Accuracy rate of diesel engine under three conditions

Condition of diesel engine	Accuracy rate
Normal	100%
Air leakage of exhaust valve	98%
Worn piston rings	95.56%

It can be seen from Table 5 that the accuracy rate of the AutoEncoder-BP neural network for fault diagnosis of diesel engine in normal condition reaches 100%, so as to avoid misjudgment in normal condition of diesel engine, and the accuracy rate of fault isolation for air leakage of exhaust valve and worn piston rings reaches more than 90%, which proves the effectiveness of this fault diagnosis and isolation method.

5 Summary and Conclusion

(1) The AutoEncoder network can effectively reduce the dimension of the extracted acoustic emission characteristic parameters, prevent over fitting, and accelerate the training speed of the BP neural network. It is feasible to apply it to the acoustic emission fault diagnosis and isolation of diesel engine combustion chamber.

(2) The characteristic parameters of AE signals in different working conditions of diesel engine are extracted, the AE fault diagnosis and isolation model of diesel engine combustion chamber based on the AutoEncoder-BP neural network is established, and its classification performance is analyzed. The results show that the model is suitable for multi classification, and it can effectively realize the isolation of air leakage of exhaust valve and worn piston rings in different working conditions of diesel engine, and accurate diagnosis can be made for the normal conditions of different working conditions of diesel engine to avoid misjudgment.

References

1. Li Y, Jiang W, Zhang G et al (2021) Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. Renewable Energy 171(3):103–115
2. Xie WQ, Cao H, Yang JG (2019) Fault diagnosis study on combustion chamber components of a marine diesel engine by using acoustic emission technology. In: 29th CIMAC congress, Vancouver
3. Douglas RM, Steel JA, Reuben RL (2006) A study of the tribological behaviour of piston ring/cylinder liner interaction in diesel engines using acoustic emission. Tribol Int 39(12):1634–1642
4. Zhang S, Yu YH, Hu L, et al (2016) Parameter for valve leakage diagnosing fault for diesel engines based on acoustic emission. Trans CSICE (34):379–383
5. Yuan F-N, Zhang L, Shi JT, et al (2019) Theories and applications of auto-encoder neural networks: a literature survey. Chin J Comput (15):203–230
6. Hu ZH, Song YL (2009) Dimensionality reduction and reconstruction of data based on autoencoder network. Dianzi Yu Xinxi Xuebao/J Electron Inf Technol 31(5):1189–1192
7. Li J, Yao X, Wang X et al (2019) Multiscale local features learning based on BP neural network for rolling bearing intelligent fault diagnosis. Measurement 153:107419
8. Kingma D, Ba J (2014) Adam: a method for stochastic optimization. Computer Science
9. Hadgu AT, Nigam A, Diaz-Aviles E (2015) Large-scale learning with AdaGrad on spark. In: 2015 IEEE international conference on big data (big data). IEEE, pp 2828–2830
10. Chen TS, Computers SO (2019) An intrusion detection method for SCADA based on gate recurrent unit optimized by adam. Modern Computer (15):7

Check for updates

Asset Management and Energy Improvements in a Critical Environment – The Case of a University Bioterium

Pedro Barandier(✉) and Antonio J. Marques Cardoso

CISE - Electromechatronic Systems Research Centre, University of Beira Interior, Calçada Fonte do Lameiro, 6201-001 Covilhã, Portugal

Abstract. This paper addresses the asset management and energy efficiency of a critical environment - the Bioterium of the Faculty of Health Sciences (FCS) of the University of Beira Interior, Portugal – a facility where the environmental conditions are required to be uninterruptedly maintained at a temperature of 21 ºC and with an air-humidity of 50%. Such requirements demand a constant utilization of the Heating, Ventilation, and Air Conditioning (HVAC) system, in particular the use of chillers and propane boilers for cooling and heating purposes, respectively. Indeed, due to the significative local weather variation over the year, a system failure breakdown may result in drastic consequences for the facility assets and its critical activities. Hence, a large amount of energy is constantly required, thereby increasing the facility operating costs.

In order to improve the energy efficiency, reduce the operating costs, and guarantee good asset management, this work aims to assess opportunities for improvements that may enable a more profitable and reliable asset management of the bioterium facilities, mainly with the advent and evolution of new technologies, which besides being more efficient, provide more useful information as well.

1 Introduction

A bioterium is an area for raising and maintenance of laboratory animals used for research and/or teaching activities. Its main goal is to ensure the required environmental, nutritional, and health-controlled conditions [1, 2]. Therefore, a bioterium is a critical facility where the environmental conditions are required to be uninterruptedly maintained in terms of temperature and air-humidity parameters. These requirements demand a continuous utilization of the HVAC system, and any system malfunction may lead to serious issues for the living beings and the on-going research activities.

Additionally, due to the aforementioned incessant use of this critical environment, the energy consumption rises up and consequently, the energy costs as well. Meanwhile, this energy consumption may present significant opportunities for consumption reduction and improvement of energy efficiency. For that, the energy costs must be dealt with in the very same way as the other outlays, such as labour and materials, for example, that are considered as variable costs [3, 4].

Furthermore, such opportunities for improvements may likewise enable a more profitable and reliable asset management of the bioterium facilities, mainly with the advent

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 364–373, 2022.
https://doi.org/10.1007/978-3-030-96794-9_33

and evolution of new technologies, which besides being more efficient, provide more useful information as well.

2 Description of the Bioterium, External Conditions and Internal Conditions Setup

The bioterium of the Faculty of Health Sciences (FCS) of the University of Beira Interior (UBI) is located in the city of Covilhã, Portugal. It presents an area of 630 m^2 and its main purpose is to house some species of animals, like mice and rabbits, used for teaching and research activities. Covilhã is located in a region of Portugal where there is a significative weather variation over the year, with a maximum mean temperature of 34 ºC and a minimum mean temperature of −4 ºC.

Regardless of the external weather and temperature, the bioterium requires its environmental conditions to be uninterruptedly maintained at a temperature of 21ºC and an air-humidity of 50%. As a result of these external and internal conditions, there is an intense utilization of the HVAC system, and in particular the use of chillers and propane boilers, for cooling and heating purposes, respectively.

2.1 Bioterium Energy Consumption

The bioterium consumption per square meter is more than the triple of the rest of the FCS facilities, with about 218 kWh/m^2 and 70 kWh/m^2, respectively. The data regarding the facilities electricity and propane consumptions are collected on a monthly basis and added to the consumptions database of the current year. In this paper, the data regarding the years of 2017, 2018 and 2019 were analyzed, as it is shown in Tables 1, 2, and 3.

Table 1. Bioterium consumption (kWh).

Year	2017		2018		2019	
Month	HVAC	Other equipment	HVAC	Other equipment	HVAC	Other equipment
January	10362	3715	7149	3097	6702	3271
February	9233	2917	7524	3088	6587	2706
March	13182	3585	7301	3402	7242	3076
April	12443	2722	6852	2715	7311	3065
May	12373	3233	6798	3333	7386	3355
June	13031	3070	6562	3008	7675	2937
July	11139	2718	6529	2806	7781	2645
August	13559	2976	7105	2426	7675	1805
September	13099	3129	5465	2102	5639	1875
October	10213	2592	7804	2932	8498	3133
November	9922	2601	7318	3006	5916	2504

(*continued*)

Table 1. (*continued*)

Year	2017		2018		2019	
Month	HVAC	Other equipment	HVAC	Other equipment	HVAC	Other equipment
December	8927	2778	6625	3207	8482	2550
Total	137483	36036	83032	35122	86894	32922

Table 1 presents the energy consumptions of the bioterium, that are categorized into two main groups: HVAC and Other Equipment. The first group comprises the electricity consumed by Air Handling Units, Air Treatment Units, Fan Coil Units, among others. The second group comprises the electricity consumed by several equipment used for the undertaken activities. As it is shown in the Pareto Chart presented in Fig. 1, HVAC represents at least 70% of the bioterium consumed electricity.

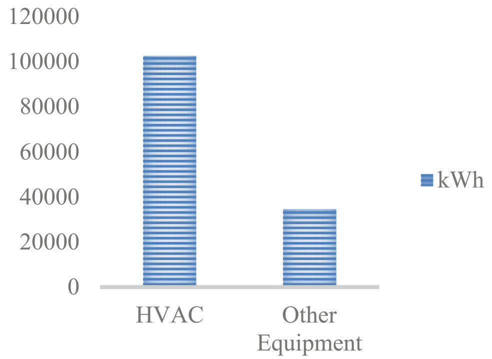


Fig. 1. Electricity average consumption in the bioterium (2017–2019).

Table 2. Chillers consumptions (kWh).

Year	2017		2018		2019	
Month	GAA1	GAA2	GAA1	GAA2	GAA1	GAA2
January	267	507	249	484	245	451
February	231	432	263	509	224	435
March	275	526	253	488	236	462
April	237	465	242	473	257	500
May	5974	496	257	524	3915	1089
June	31401	513	5116	12069	246	23775
July	34687	451	8165	31738	14137	25154
August	30904	505	7335	45734	17338	479

(*continued*)

Table 2. (*continued*)

Year	2017		2018		2019	
Month	GAA1	GAA2	GAA1	GAA2	GAA1	GAA2
September	22549	533	3213	44612	8754	7345
October	5854	472	489	12230	296	11533
November	269	524	267	519	225	440
December	252	488	284	567	250	496
Total	132900	5912	26133	149947	46123	72159
PCT%	95,74%	4,26%	14,84%	85,16%	38,99%	61,01%

Table 2 presents the energy consumption of the chillers, denominated GAA1 and GAA2, which are of the same model, *Wesper SLS 4202 BLN/LN series.* While one of them operates the other one stands for a backup in case of an eventual breakdown, and/or a scheduled preventive maintenance. These chillers cool all the facilities, including the bioterium. However, during the summer vacations period the chillers remain working only for the bioterium, what represents a waste of energy, once an equipment sized to cool the whole building works for the refrigeration of just a small part of it.

Table 3. FCS propane consumption (m^3).

Month	2017	2018	2019
January	16501	15858	17580
February	12312	15973	13164
March	9486	14982	9942
April	4012	6322	5317
May	2400	466	361
June	167	210	293
July	150	128	121
August	0	15	3
September	291	210	290
October	233	1314	4377
November	9410	14445	11495
December	11766	13771	14353
Total	66728	83694	77296

Table 3 presents the propane consumption of the FCS facilities. This expenditure refers to the propane used by the boilers for heating, and by the teaching and research activities at the laboratories. However, the laboratories consumption is practically insignificant.

Similarly to the cooling system of the building, the heating system has two propane boilers, one for operation and the other one for a backup in case of an eventual breakdown, and/or, a scheduled preventive maintenance. However, such system may present a waste of energy for the same reason pointed out before in relation to the chillers. An equipment that is meant to heat the whole building, in vacation periods operates only for the bioterium climatization.

As it is shown in Fig. 2, the HVAC electricity consumption in the bioterium is regular over the year, with just a little increase in the months with a higher thermic load, due to higher external temperatures and a more intense activity in the bioterium itself. Despite of this regularity, the electricity and propane, used for chillers and boilers, respectively, vary throughout the year according to the external temperature, as it can be seen in Fig. 3. Although propane is measured in cubic meters (m^3), the consumption was converted to kilo-watt-hours (kWh) in Fig. 3 in order to standardize the analyzed units.

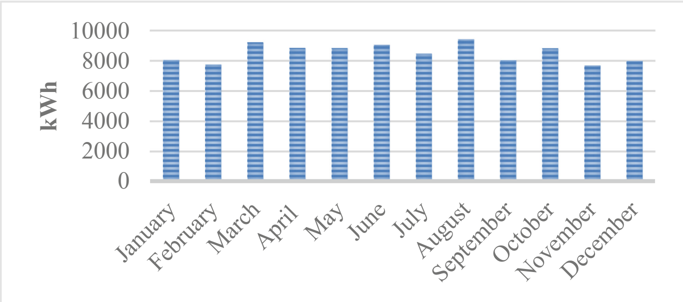


Fig. 2. HVAC average consumption in bioterium (2017–2019).

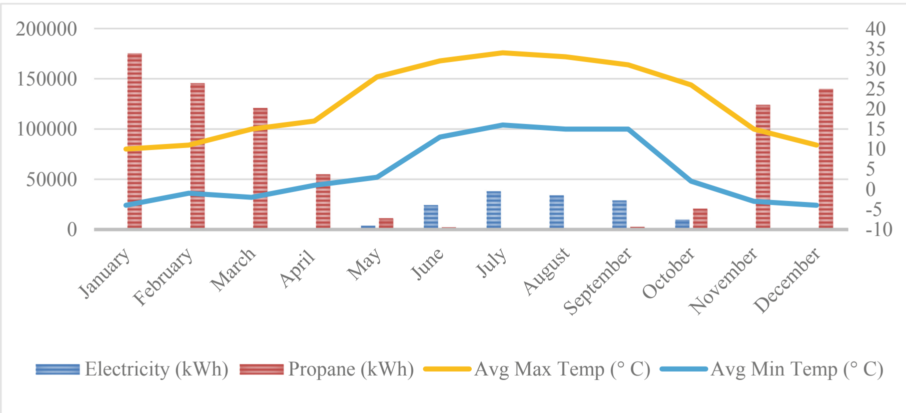


Fig. 3. Chillers and boilers average consumption in FCS throughout the year (2017–2019).

3 Improvements Opportunities

Based on the analysis performed, it was possible to identify the *Significant Energy Uses* (SEUs) of the facilities. The SEUs represent the areas which result in a substantial energy consumption, and/or, may provide a significant potential for improvements [5]. Although these opportunities are presented in different areas, each one of them has a significant potential for an increase in the energy efficiency, and a subsequent reduction in costs and wastes.

3.1 HVAC Equipment Replacement

Due to the fact that the chillers and boilers are meant to climatize the whole building, in periods when only the bioterium demands climatization, it results in an oversized and less efficient system, and consequently, in more costs and expenses. Therefore, it would be relevant to consider an independent climatization system only for the bioterium.

Such system may be achieved by heat pumps, equipment that present a higher efficiency when compared to chillers and boilers. A properly dimensioned heat pump system may return the investment within 5 years due to its consumption. After the payback, the system provides a thermal comfort with low costs and minimal environmental impacts. Furthermore, heat pumps may unify the cooling and heating systems once it performs both functions [6, 7].

According to Controls and Building Management-European Standard [8], this unification is a characteristic of high energy performance of building automation and control system (BACS) and technical building management (TBM).

Another advantage of heat pumps is the possibility that hot water may be obtained during cooling operations. This is done by a heat exchanger placed after the compressor, where the heat removed, resultant from cooling process, may be reused.

Additionally, according to the European Union package of directives, arising from Paris Agreement for the reduction of greenhouse gases emissions (GHG), several resolutions were taken, like the National Energy and Climate Plan 2030 [9], and the Roadmap for Carbon Neutrality 2050 [10], and thus, equipment that uses energy sources which result in the emission of any kind of GHG should be replaced in the coming years.

However, the acquisition of new equipment consequently implies the adoption of a good strategy to guarantee its maintainability. Thus, a careful analysis and assessment must be undertaken to assess the most adequate asset management strategy for each equipment, according to its criticality.

3.2 Adoption of New Technologies

The implementation of smart building solutions demonstrates a great potential for increase in efficiency, as well as, to implement a good strategy in the asset management.

3.2.1 Smart Meters

Smart meters are electronic meters that enable a more detailed measurement of the consumed energy and provide, in real time, data regarding the energy consumption. Such data may be visualized remotely through a computer or smartphone [11].

Furthermore, these devices grant an interface with the measured environment, even enabling the remotely activation and deactivation of the supply, and also, emitting an alert if it happens a non-notified interruption or an energy consumption above the expectable parameters [11].

Thus, the data acquisition is simplified and more precise, making possible the utilization of Quality Control Tools, such as Control Charts and Ishikawa Diagrams. Control Charts may be defined by the average consumption over time, and in real time. As a result, a better control, and the opportunity for the implementation of an on-condition based maintenance strategy may be achieved. Also, potential failures may be avoided even before they occur. Ishikawa Diagrams may provide, through a data analysis, the identification of the factors which lead to a higher energy consumption and the ones leading to faults and/or failures, that may compromise the assets. In this manner, the root causes of many problems may be treated before such issues happen.

3.2.2 KNX System

KNX control technologies have been largely employed for obtaining energy efficiency improvements. The installation of a control and automation smart system may reduce the energy consumption in several areas, mainly in lighting and climatization, which represent the largest expenses in the analyzed case [12].

The automation of lighting and climatization systems may be achieved through some control strategies, such as dimming control, occupancy control, daylight harvesting for lighting automation, and occupancy and temperature control for climatization automation [13, 14].

- **Dimming Control**

  This strategy is used to control the luminescence of a light bulb and provide varying illumination levels in a space. However, this strategy presents a higher investment due to the additional costs of components, such as dimming ballasts for fluorescents lamps and/or dimming drivers for LEDs lamps [13].

- **Occupancy Control**

  This strategy consists in the use of occupancy or vacancy sensors, which are intended to control lighting based on the presence of people. The difference between these two sensors is that an occupancy sensor automatically turns luminaires on when it detects presence in the room and off when the room becomes vacant again. The vacancy sensor requires that luminaries be manually switched on and automatically turned off when there is no more presence detected in the room [13].

- **Daylight Harvesting**

  This strategy consists in utilizing the available daylight and adjust the amount of artificial lighting needed to provide the adequate luminosity in a room. In the case of the bioterium, it is recommended a luminosity of 500 lux. Daylight harvesting are often utilized in conjunction with the aforementioned strategies, through a luminosity sensor built-in the occupancy/vacancy sensors. The amount of available daylight is measured and through an actuator, the remaining light required is provided by means of the dimming control [13, 14].

One possibility for optimizing the harvest of daylight is the implementation of an automatic control of the curtains. According to [14], through blind control it is possible to adjust the incidence of the external light into a room by controlling the opening angle of the shutters, as shown in Fig. 4, by taking into account the position of the sun. In this way, a direct light in the working area (which can cause heat stroke and consequently an increase in temperature) is avoided.

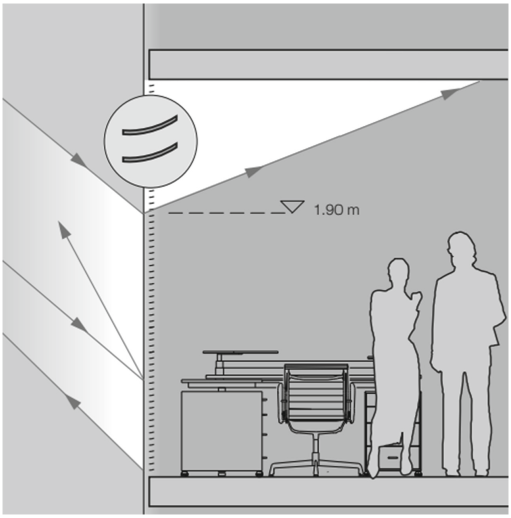


Fig. 4. External light incidence in automatic blind control [14].

- **Temperature Control**

Temperature control is accomplished by a room controller connected to a KNX network through a gateway. This controller takes into account the measured temperature, obtained by temperature sensors, to support the cooling and/or heating processes. Some devices also include CO_2 and humidity sensors as well [15]. Likewise, there are also electrical control valves in the hot and cold sources connected to the network [14, 15].

The temperature sensors constantly assess the room current temperature and compare it to a predefined setpoint. In the occurrence of a deviation, the temperature is regulated noiselessly via electronic switch actuators, that automatically adjust the electrical control valves, either from heating or cooling sources, so that only the necessary thermal load is provided for an adequate temperature adjustment [14, 15]. Such strategy presents a great opportunity to grant a trustworthy low-cost asset management in the bioterium as it provides the adequate local temperature with lower energy consumption.

This strategy can be utilized in conjunction with the occupancy control as well, and thus the occupancy/vacancy sensor for lighting control may also turn the thermostat on or off according to the presence or absence in a room [14]. However, due to the already explained reasons, this strategy cannot be applied in the bioterium once it demands a constant climatization, regardless of occupancy or vacancy.

4 Summary and Conclusion

Several possibilities of energy improvements in a critical environment (bioterium), are analyzed in this paper with the main objective of ensuring the assets management with a better control and lower costs. Since the bioterium demands a constant utilization of the HVAC system to maintain adequate environmental conditions, a large amount of energy is required, presenting thus, opportunities for consumption reduction and increase in efficiency without compromising such mandatory conditions.

Based on the electricity and propane consumption during the period of 2017–2019, it was possible to determine the *Significant Energy Uses*. Consequently, opportunities for energy improvements in different areas were identified, such as the substitution of chillers and boilers for heat pumps, because of their higher efficiency, and the implementation of new solutions, such as smart meters and KNX technologies.

However, the substitution of chillers and boilers for heat pumps for the whole building would require a large investment, and therefore, it would be reasonable to consider an independent HVAC system just for the bioterium once it operates regardless the considered period of the year. Furthermore, an independent and correctly sized system would provide an adequate climatization without energy wastes.

The use of smart meters may provide a better system control and consequently, the opportunity for establishing an on-condition based maintenance strategy that enables the possibility of monitoring, analysis, and quantification of the various parameters. Hence, it is possible to prevent failures that may compromise the assets before they happen. From this, it is also possible to estimate the system energy inputs and outputs, assessing thus, the global efficiency of the bioterium. Furthermore, smart meters may also support a root cause analysis, which identifies the origins of some problems in order to avoid them.

With KNX technologies, an automatic and smart system can be achieved. Its costs reduction depends on the adopted strategies. For this case-study, a great potential for savings is presented by lighting and climatization control, in particular the temperature control due to the fact that the HVAC system operates uninterruptedly. Although such technologies require a relatively high investment, there is the possibility of payback in a short period of time.

In conclusion, better energy systems efficiency, automation and control, the possibility of adopting on-condition based maintenance strategies and a root cause analysis for some issues were some of the accomplished results. Such results present the possibility to achieve an asset management with higher reliability and lower costs.

Acknowledgments. This work was supported by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalization (COMPETE 2020), under Project POCI-01-0145-FEDER-029494, and by National Funds through the FCT - Portuguese Foundation for Science and Technology, under Projects PTDC/EEIEEE/29494/2017, UIDB/04131/2020 and UIDP/04131/2020.

References

1. University of Coimbra. https://www.uc.pt/fmuc/bioterio. Accessed 10 Dec 2020
2. Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes
3. Capehart BL, Turner WC, Kennedy WJ (2011) Guide to energy management. The Fairmont Press, Inc.
4. International Organization for Standardization (2014): ISO 50001 Energy management systems — Requirements with guidance for use
5. Peña AC, Sánchez JMG (2012) Management of energy efficiency: calculation of consumption, indicators and improvement (in Spanish). AENOR
6. MacKay D (2008) Sustainable Energy-without the hot air. UIT Cambridge
7. Langley BC (1989) Heat Pump Technology: Systems Design, Installation, and Troubleshooting. Pearson College Division
8. CEN, T (2012) 247 Energy performance of buildings–Impact of Building Automation. Controls and Building Management–European Standard EN, 15232
9. National Energy and Climate Plans for 2021–2030 under the EU Energy Union, pp. 1–11. https://ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/national-energy-and-climate-plans-2021-2030-under-eu-energy-union. Accessed 21 Mar 2021
10. Roadmap for Carbon Neutrality 2050 (RNC 2050) Long-Term Strategy for Carbon Neutrality of the Portuguese Economy by 2050. Portuguese Republic. https://unfccc.int/sites/default/files/resource/RNC2050_EN_PT%20Long%20Term%20Strategy.pdf. Accessed 10 July 2020
11. Canhoto BAR (2018) Cálculo das perdas de redes de distribuição com as medidas dos Smart-Meters. MSc thesis. Department of Electrical and Computer Engineering, Faculty of Engineering of the University of Porto, Portugal
12. KNX (2020) The Legacy of KNX. https://www.knx.org/knx-en/for-professionals/What-is-KNX/KNX-History/. Accessed 28 Aug 2020
13. Sabourin NT (2017) The effects of connected lighting on lighting controls and design. Kansas State University, Manhattan, USA, Department of Architectural Engineering and Construction Science College of Engineering
14. ABB (2009) Smart Home and Intelligent Building Control. ciency in Buildings with ABB i-bus® KNX, Germany
15. KNX. (2018). Journal. https://knxportugal.pt/images/jor-nais/KNX-Journal-2018.pdf. Accessed 28 Aug 2020

Determination of Water Content in Heavy Fuel Oil Using a Relative Permittivity Meter

Daniel de Almeida Arantes[1](✉), Mateus M. Campos[1], Luiz Eduardo Borges da Silva[1], Wilson Cesar Sant'Ana[2], Carlos E. Teixeira[2], Germano Lambert-Torres[2], Erik Leandro Bonaldi[2], Levy de Lacerda de Oliveira[2], and Germando A. Costa[3]

[1] Av. BPS, 1303 – Pinheirinho, Itajubá, MG, Brazil
daniel_arantes@unifei.edu.br
[2] R. José Julião Simões, 53 – Santa Rita de Cássia, Itajubá, MG, Brazil
wilson@institutognarus.com.br
[3] Via de Ligação 3, S/N – Distrito Industrial de Fortaleza III, Maracanaú, CE, Brazil
germando.costa@maracanaugeradora.com.br

Abstract. The measurement of relative permittivity of fluids is a convenient way to identify the amount of each element in a two-component mixture. This approach applies especially to cases in which the permittivities of the two components are far apart from each other, such as determining the water content in heavy fuel oil (HFO). The latter is a high-viscosity and high-density fuel, obtained from residual portions in the distillation process of crude oil. The presence of water in HFO is generally unwanted, and can be a major concern due to equipment degrading, decrease in heat transfer capabilities and loss of burning efficiency. This work addresses the determination of water content in HFO samples, benefiting from the great difference between the relative permittivities of both fluids. A relative permittivity meter designed specifically for this purpose was employed. The meter uses capacitance as its working principle, and comprises a capacitive sensor that is in direct contact with the oil, and a capacitance meter circuit that connects to the sensor. This paper describes both components, as well as the calibration procedures involved in their usage. HFO samples with different amounts of water were prepared and probed in order to obtain a relationship between water content and relative permittivity. The collected data provides enough information to determine the amount of water present in other HFO samples, by measuring its relative permittivity and using adequate interpolation methods.

1 Introduction

Thermal power plants may operate from different sources of heat, including fossil fuels such as crude oil derivatives. The quality of these fuels will reflect into the quality of the burning process, including its efficiency, residue generation and equipment degrading. One of the main concerns of companies that use inputs such as Heavy Fuel Oil (HFO) is the amount of water present in the fuel when it reaches the power station. The presence of water will have an impact on the cost of the generated electricity. In addition, a water percentage greater than that supported by the equipment inside the plant will cause damage to them.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 374–381, 2022.
https://doi.org/10.1007/978-3-030-96794-9_34

The water content in HFO can be determined based on its relative permittivity. The relative permittivity of a material quantifies its tendency to electrical polarization when an external electrical field is applied to it. Its value will be near 1 for materials that present a weak polarization in response to an external electrical field (for example, the air) and may reach the order of hundreds in case of ceramic powders [1].

As the relative permittivity is related to the molecular structure of the material, its analysis can determine if a fluid is pure, adulterated, or has some contaminants, by means of comparison of a sample of this fluid with a pure sample [2]. Even the amount of each component in the mixture can be determined, provided that their individual relative permittivities are known [3, 4].

The authors have previously presented (in [5]) a system to measure the relative permittivity of fluids, based on the Howland current source topology [6]. The current paper presents an application of the system of [5] in the determination of water content in HFO samples, based on the comparison between the relative permittivities of samples.

Section 2 presents the developed permittivity system applied to water content measurement, including the capacitive sensor and the measurement circuit. Section 3 presents the experimental results, including a calibration procedure. Section 4 concludes the paper.

2 Water Content from Capacitance Measurement

The water content of an oil sample can be determined through the measurement of the capacitance between two metallic plates immersed in the sample and the estimation of its permittivity. The capacitance of the sample depends, mainly, on its format and the material. Generally, the capacitance value of a capacitor is given by (1), where ε is the electric permittivity of the dielectric material (hence, dependent on the material) and K is a parcel dependent on size and geometry of the capacitor.

$$C = \varepsilon \cdot K, \tag{1}$$

The measurement of water content in an oil sample is based on the comparison with a baseline sample of pure oil, as the permittivity of water is about 40 times higher than that of a fuel oil [7, 8]. From (1), it can be noticed that the water content in the sample will produce an increase in ε (hence, in C), in relation to the values that would be obtained with a pure oil sample.

2.1 Capacitive Sensor

Figure 1 presents a photograph of the capacitive sensor developed in [5]. It is composed of disks of Printed Circuit Board (PCB), whose copper surfaces correspond to the metallic plates of the capacitor. The sample to be evaluated fills the spaces between the disks, forming an array of capacitors in parallel.

Fig. 1. Prototype of the capacitive sensor developed in [5].

2.2 Meter Circuit

In order to charge and discharge the disks of the sensor, a circuit based on current sourcing has been developed in [5]. The capacitances can be measured according to their time of charge or discharge. Figure 2 presents the circuit, formed mainly by a voltage controlled current source (stage concerning U1A, R1, R2, R3 and R4) and a comparator (U2A). The comparator acts switching the polarity of the current whenever the capacitive sensor (connected in J1) is charged. The output TP1will produce a square wave, resulting from the output of comparator U2A trough buffer U2B. The output TP2, whenever a capacitive sensor is connected to J1, will produce a triangular wave whose slope time is related to the capacitance value. The resistor R controls the intensity of the current source – hence it has an effect on how fast the capacitor will charge and discharge.

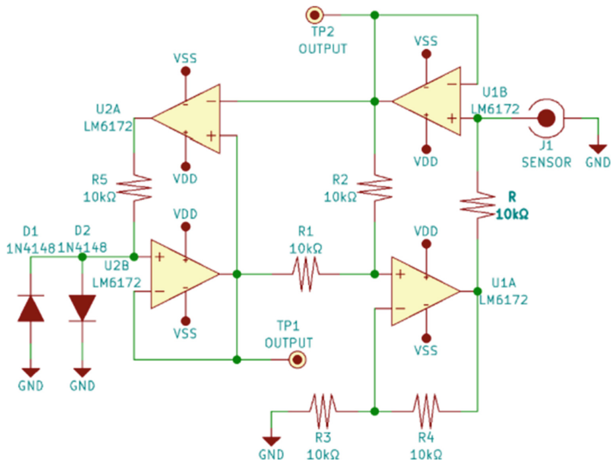


Fig. 2. Schematic circuit of charge/discharge of the capacitive sensor [5].

By the time when experiments of [5] have been performed, the meter circuit had been implemented in a bread-board. Figure 3 presents a photograph of the newly developed printed circuit board, connected to the capacitive sensor.

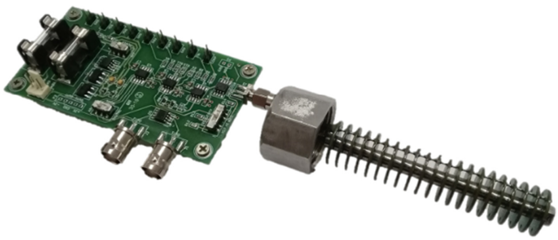


Fig. 3. Capacitive sensor connected to the meter board.

Figure 4 presents an example of a typical behavior observed at the outputs TP1 (blue plot) and TP2 (yellow plot). Considering Δt the time taken by TP2 to raise from its zero crossing to its maximum value (as indicated in the figure), the capacitance can be calculate according with (2). For the example shown in the figure, each time division is 2 μs, hence, $\Delta t = 3\mu s$. With $R = 10\,k\Omega$ (as shown in Fig. 2), the capacitance measurement for this case would be $C = 300\,pF$.

$$C = \Delta t / R. \quad (2)$$

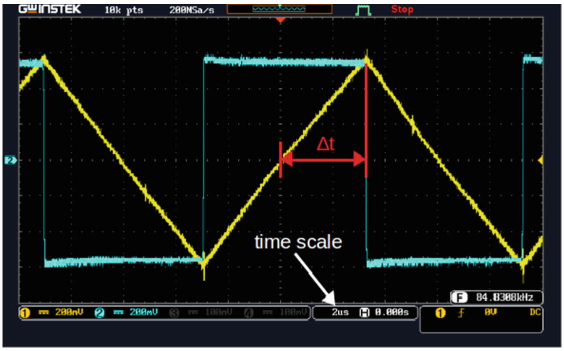


Fig. 4. Example of capacitance calculation from the slope of TP2.

In practice, there is an automated system to measure the time Δt. This system is based on an FPGA that detects the zero crossing of TP2 and counts the time until the inversion of the slope. The FPGA then connects to a database located on a remote PC via Ethernet. The description of the FPGA firmware and the remote software are outside the scope of this paper.

3 Experimental Results

The experimental results have been divided in two parts. Subsection 3.1 presents the calibration procedure, that correlates the capacitance values with the water content obtained with the well know Karl Fischer method. Subsection 3.2 presents the already calibrated system installed in a thermal power plant, in order to perform an online analysis of the oil being pumped from the tank truck to the storage system of the plant.

3.1 Calibration

For the calibration of the system, four samples were used, containing different concentrations of water. For each sample, measurements were taken with both the developed system and with a Karl Fischer, shown at Fig. 5. The concentrations of water in the samples, obtained with the Karl Fischer method were: 0.19%, 0.40%, 1,01% and 2.20%.

Fig. 5. Karl Fischer titration of one sample of oil.

These four samples have also been tested with the developed system, in order to calibrate it, according with the Karl Fischer. A sequence of 100 measurements have been performed for each sample. Figure 6 presents the calibration setup. The capacitive sensor is immersed in an oil sample.

Fig. 6. Calibration setup for the capacitive sensor (immersed in the sample).

Table 1 presents the results obtained with the samples. The first column presents the concentrations obtained with the Karl Fischer method (KF). The second column presents the temperature of the samples at each measurement. The third column presents the capacitance value obtained with the developed system.

Table 1. Results of the calibration measurements

Water concentration – KF	Temperature	Capacitance
0.19%	46.5 °C	261.8 pF
0.40%	49.1 °C	286.1 pF
1.01%	49.3 °C	353.2 pF
2.20%	47.4 °C	482.9 pF

From the values obtained in Table 1, a relationship between capacitance (in pF) and the water content (in %) can be determined as (3).

$$Water_{\%} = 0.0087 \cdot C_{pF} - 2.0712. \quad (3)$$

Using (3), the system installed at a pipe in a real power plant is able to perform online measurements of water content in the plant's fuel oil.

3.2 Online Measurements on a Power Plant

With the capacitive sensor previously calibrated (according with the Karl Fischer analysis), the whole system has been installed in a thermal power plant, located in the city of Maracanaú, Brazil. This thermal power plant operates either with diesel and HFO. The developed system has been installed at the inlet for HFO storage. Figure 7 presents a photograph of the whole system installed on a pipe just after the point where a tank truck feeds oil to the plant storage system.

The capacitive sensor is inserted inside the pipe. All circuit boards (the board that charges/discharges the sensor, the FPGA system that calculates the sensor charge time, as well as some other circuits that are outside the scope of this paper) are located in the box over the pipe.

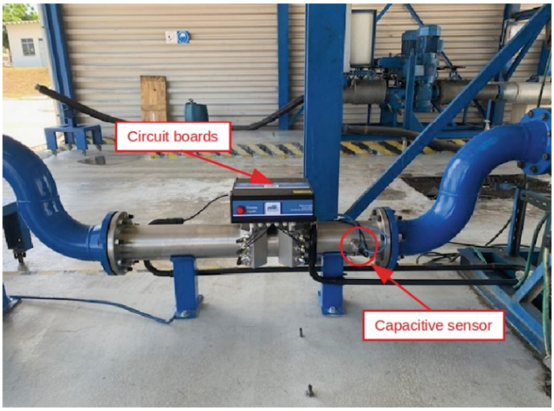


Fig. 7. Prototype for oil analysis installed on a thermal power plant

Figure 8 presents a trend curve of the measurements performed for about 15 min on the oil that remains in the pipe when there is no pumping from the truck. The blue dots are each individual water concentrations (left axis), while the red dots are the temperature (right axis) at which of these measurements were taken. It can be noticed that the concentration of water at the standing oil in the pipe was about 0.4%.

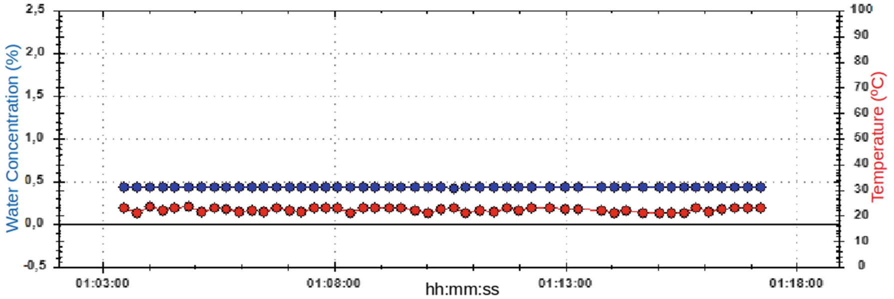


Fig. 8. Water content of the standing oil in the pipe (measurements taken for about 15 min).

Figure 9 presents a trend curve of the measurements performed for about 1 h with the oil being pumped from the truck to the storage, passing through the capacitive sensor. One can observe that the oil being pumped at that time also had around 0.4% water content.

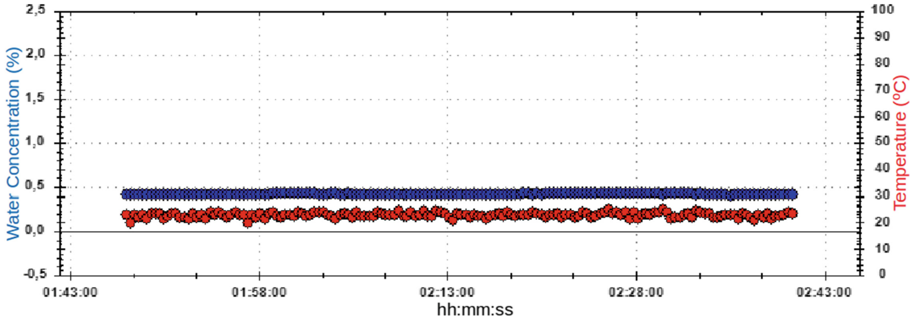


Fig. 9. Water content of the oil being pumped through the pipe (measurements taken for about 1 h).

4 Summary and Conclusion

This paper presented the use of a capacitive sensor in order to determine the water concentration in heavy fuel oil. This sensor has been manufactured using conductive disks, resulting in an array of capacitors, whose dielectric material is the oil in which it is immersed. In order to charge and discharge the capacitive sensor, a circuit board with operational amplifier has been presented. The time of charge/discharge of sensor is used to estimate the electric permittivity of the oil.

The method to detect the water content in the oil in which the sensor is immersed is based on the fact that the electric permittivity of the contaminated oil greatly increases in relation to a pure sample. A procedure to calibrate the sensor, in relation to the well known Karl Fischer method, is presented in order to correlate the measured capacitance values with the actual water content of the samples.

The presented prototype has been installed in a thermal power plant, in order to online test the quality of the oil, while it is being pumped from tank trucks to the plant's storage system.

Acknowledgments. The authors would like to thank ANEEL R&D and Maracanaú Geradora de Energia SA for the financial support provided for this project.

References

1. Sheen J, Wang Y-L (2013) Microwave measurements of dielectric constants for high dielectric constant ceramic materials by mixture equations. IEEE Trans Dielectr Electr Insul 20:932–936
2. Khan TT (May 2012) A method for measuring dielectric properties of nonmagnetic liquids and predicting their contamination level. In: 2012 international conference on informatics, electronics & vision (ICIEV), Dhaka, Bangladesh
3. de Graaf G, Lacerenza G, Wolffenbuttel R, Visser J (May 2015) Dielectric spectroscopy for measuring the composition of gasoline/water/ethanol mixtures. In: 2015 IEEE international instrumentation and measurement technology conference (I2MTC) proceedings, Pisa, Italy
4. Teixeira CE et al (2019) An ultrasound-based water-cut meter for heavy fuel oil. Measurement. https://doi.org/10.1016/j.measurement.2019.106907
5. Arantes DA, et al (2019) Relative permittivity meter using a capacitive sensor and an oscillating current source. In: 45th annual conference of the IEEE industrial electronics society, 2019, pp 806–811. https://doi.org/10.1109/IECON.2019.8927762
6. Chen DX, Deng X, Yang WQ (2010) Comparison of three current sources for single-electrode capacitance measurement. Rev Sci Instr 81:034704
7. Malmberg CG, Maryott AA (1956) Dielectric constant of water from 0° to 100°C. J Res National Bureau Stand. 56
8. Warren WJ (1962) BS&W measurement – principles and practices. In: SPE production automation symposium

Particulate Matter Monitoring in Joinville, Santa Catarina, Brazil

Marianna Gonçalves Dias Chaves(✉) and Emílio Graciliano Ferreira Mercuri

2 Cel. Francisco H. dos Santos Av., Curitiba 100, Brazil
mariannag@id.uff.br

Abstract. Air pollution is directly related to the increased risk of acute respiratory infections, and it was estimated by the World Health Organization that 6.5 million deaths in 2012 were caused by air pollution-related illnesses. In Brazil, there are few cities that have air quality data collection, although legislation has been recently updated and quality standards have been planned with promising targets, according to CONAMA Resolution 491/2018. This research presents an alternative air quality monitoring station, focused on the use of Internet of Things (IoT) resources, low-cost equipment and real-time acquisition via the internet. The purpose of the research is to use the SDS011 sensor for particulate matter (PM_{10} and $PM_{2.5}$) concentration monitoring in Joinville, Santa Catarina, Brazil, connected to the Raspberry Pi computer. The code for data acquisition of the sensor was developed in Python and the data cloud storage was done through the use of Dropbox API. The system aims to reduce the costs of equipment and monitoring stations along with their energy consumption, in addition to enabling real-time monitoring, with availability of data for the scientific community and the city population, also representing an easy-to-install and cheap maintenance. The results show the time series of particulate matter as well as the daily profile and the data completeness calculated for the period of monitoring. It is intended as future work to relate the measurements of PM_{10} and $PM_{2.5}$ with other methodologies and compare data with other equipment. Joinville is a highly industrialized city and has no official monitoring, so the continuity of this work with the expansion of monitoring stations is vital to make the data available to the population and to understand the relationship between local atmospheric stability and air quality.

1 Introduction

Air pollution has been a concern in large urban centres, notably in highly industrialized areas [1–3]. The gases and the particulate matters emitted by industries and traffic spreads throughout cities, affecting their population by being accumulated and inhaled, generating health problems related to the penetration of particles into the respiratory and circulatory system, with drastic side-effects and consequences on the functioning of lungs, heart and even brain [1, 3–6]. Within this scenario, it is estimated that 4.2 million people die each year as a result of the effects of air pollution, a number that may increase if there is no expansion in the use of clean technologies and greater control of emissions [7].

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 382–389, 2022.
https://doi.org/10.1007/978-3-030-96794-9_35

The city of Joinville has a predominant mesothermal, humid climate, with no dry season, an annual average temperature of 21 °C and an annual relative humidity of 76%. This city is the most industrialized in the state of Santa Catarina and is one of the twenty most industrialized in the country, having as main sectors the chemical and plastic products, metalworking and metallurgy [8]. Despite its industrial and economic development, Joinville lacks air quality monitoring, since the state of Santa Catarina has not implemented air pollution control programs as established by the National Environmental Council (CONAMA).

A large part of the lack of monitoring problem and emission control is related to the high cost of the equipment currently used by fixed stations and their limitations in relation to their spatial arrangement and data dissemination. As in the case of Joinville, in Brazil there are around 250 monitoring stations, in a total of more than 5000 cities, and the monitoring networks, which have a set of fixed and mobile stations, are present in only 1,7% of cities. Aiming to reduce the costs of monitoring equipment, increase the number of existing stations and enable the creation of new monitoring networks, low-cost sensors have emerged to be implemented with the help of other resources in the area of information technology, such as the Arduino and Raspberry Pi platforms [1, 11–17].

There are sensors for the analysis of parameters such as temperature, humidity and concentration of particulate material (MP) in the air, which are classified according to their particle size, being $PM_{2.5}$ for particles up to 2.5 μm in diameter and PM_{10} for particles up to 10 μm in size. The functioning of low-cost sensors for particulate matter is based on the optical principle, measuring the light dispersed by the particles and outputting a signal that must then be converted into a concentration value [6, 10, 11]. These sensors must then be connected to a processor (e.g. desktop computers, notebook, Raspberry Pi) and to storage and/or connectivity resources (e.g. flash card, SD card, Bluetooth, Wi-Fi, etc.) [10, 11].

The expansion of interest and use of low-cost sensors has been evident in recent work, in which many compare the various models and existing manufacturers, and justified mainly by the fact that, in addition to presenting a cheap and low energy consumption alternative, it enables real-time monitoring and easy installation and maintenance. On the other hand, few studies include the simultaneous use of different sensors for monitoring environmental and meteorological parameters, such as gas concentration, particulate matter, temperature and local humidity, and study the relationship between particulate matter concentrations and the increase or decrease in temperature and humidity, focusing on the use of connectivity tools that make the data available to the scientific community and the society.

Thus, this work aims to allow the monitoring of particulate matter concentrations ($PM_{2.5}$ and PM_{10}) in the city of Joinville, using the SDS011 sensor. Another objective of the study is to develop computer codes in python, using Raspberry Pi, which performs the reading of the sensors saving in its internal memory and a methodology for cloud storage in real time, allowing its remote access, its download and further creation of informative graphics to be openly shared.

2 A Description of the Analyzed Site and Sensors for Data Acquisition

The sensor was installed in a school that is located within the Joinville industrial park, in the Tupy industrial area, located in the Boa Vista neighborhood. This industrial area covers approximately 1,208,000 m^2 and is occupied by the foundry industry of same name, which has a production capacity of 409,660 tons per year and is one of the main contributors to the city's occupation and development [18]. In Fig. 1 it is possible to see the proximity of the monitoring site and the Tupy industrial area.

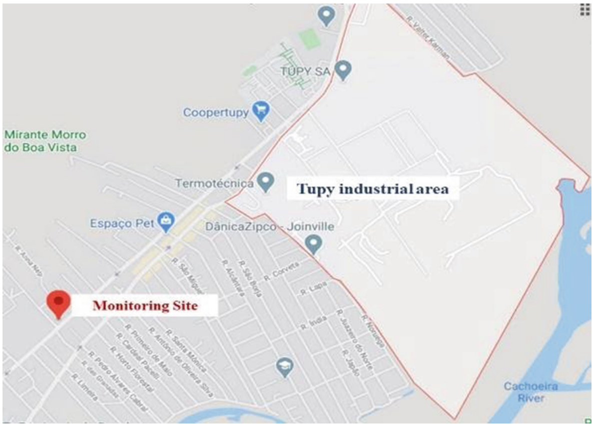


Fig. 1. Industrial zone and the school's location in the Boa Vista neighborhood.

To data acquisition it was chosen to connect the sensor to the Raspberry Pi model 3B+ so the data could be stored. As mentioned in the introduction section, this board represents a cheap alternative compared to desktop and laptop computers and allows the development of python codes. The sensor for particulate matter concentration monitoring was chosen considering their use and results in previous studies and their cost.

The code developed for particle material sensor reading generates texts files (.txt) with the date and time of the measurement. Since the SDS011 sensor can read $PM_{2.5}$ and PM_{10} simultaneously, the code was written in a way that reads these two concentrations, within 5 s, and then generates two text files that are stored in the internal memory of the board.

The particulate matter sensor has been coupled to a rod that exists at the monitoring point, using a rigid box for its protection that contains a hole for the air intake. Raspberry Pi was installed about 3 m from the SDS011 sensor, in an open but covered location (in a marquee near a window), since its electrical part would demand greater protection and easy access in case of modifying any code or saving the files in an external storage, and the sensor is connected by an USB cable. The Table 1 below shows the price of the board and sensors utilized and the total cost of the implemented system, which may vary depending on the supplier.

Table 1. Price of the sensors, board and components employed in the system and its total cost.

Components or gadgets	Unit price	Quantity used	Final cost
Raspberry Pi model 3B+	R$ 339,90	1	R$ 339,90
SDS011 sensor	R$ 224,26	1	R$ 224,26
USB cable	R$ 10,90	5	R$ 54,50
Total cost of implemented system			**R$ 618,66**

The monitoring site granted the internet available for use in the research through Wi-Fi or network cable, having used cable for connection since the Wi-Fi signal in the outdoor area was weak. The Dropbox storage system was then synchronized through its Application Programming Interface (API), enabling remote access and data download. In this way, a folder is created for the station using a Dropbox account, centralizing all the data and facilitating its access and download. Thus, for each folder other folders are created to store the text files generated during the reading period, considering each day and each hour of monitoring. It is noteworthy that the whole system of data acquisition and storage in the internal memory of the Raspberry Pi and Dropbox was configured to start automatically when the microcomputer is turned on, thus avoiding that the system stops working completely in case of power outages. The Fig. 2 shows the flowchart scheme of the devices and codes used to generate the text files containing the information obtained and make them available in Dropbox.

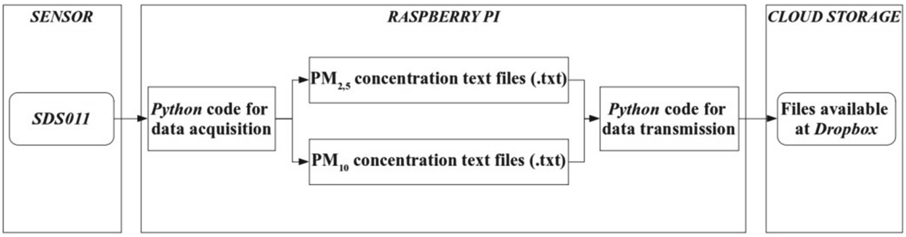


Fig. 2. Flowchart showing the devices and tools used to make the data available on cloud storage.

2.1 Data Analysis

For the data analysis of the data obtained in the measurements described, 2 codes were developed in Python language. The first code, to calculate the daily averages, reads the files generated by the SDS011 sensor and calculates the daily averages according to each day and time recorded in the headings. The averages are adjusted according to the set interval desired (averages per minute, 5 min, 10 min, etc.) and the daily (24 h) average is calculated. The developed code is also used for plotting the raw data from the SDS011 sensor and its daily averages.

3 Data Completeness

The second code was developed to analyze the data completeness from Eq. 1, since the raw data from sensor SDS011 would need to be analyzed. The text file is read, selecting the date and time format, and organized according to hourly averages. It is then verified if there were hours without monitoring, calculating the completeness based on the expected hours and the hours obtained after the verification. The calculation of monitored hours was done taking into account only the number of days (multiplying the hours obtained by 24), however, this code can be adjusted for the completeness analysis considering other intervals (months, years, minutes, etc.).

$$data\ completeness\ (\%) = \frac{quantity\ of\ data\ obtained}{quantity\ of\ data\ expected} \times 100 \qquad (1)$$

4 Results and Discussion

4.1 Generated Files and Storage Capacity

The text files containing the concentration values were generated correctly, without overwriting, and the size of the files with the hourly data of Particulate Matter did not exceed 25 kilobytes (kB). This means that, considering the two text files generated, for 24 h of monitoring the size of the folder in which the files were stored did not exceed the corresponding 1.3 Megabyte (MB). Thus, the files and folders generated do not require much storage memory, and it would be feasible to keep the system running for a longer time using only the internal storage, if necessary or if there is no Internet access. As for cloud storage, the byte limits that the free account options offer are suitable for long-term monitoring. Using Dropbox, 2 Gigabytes (GB) of free storage is available (which corresponds to approximately 2000 MB) guaranteeing more than 3 years of data storage for each monitoring performed in this research, considering the total of approximately 40 MB of data expected per month and 520 MB of data expected per year.

When running the code that calculates the daily averages and generates the graphs, it was noticed that some files had heading repetitions, which can be attributed to possible power outages at the monitoring site or to the fact that the Raspberry Pi was suddenly turned on/off, which caused the codes to start running and abruptly end. In this case the headings in the middle of the files were removed. However, this shows that the automatic startup configuration of data acquisition and transmission code worked correctly and prevented losing files as well as interrupting the monitoring.

4.2 Data Consistency Analysis

Upon starting the analysis of data some peaks of values exceeding the sensor detection limit - which is 999 $\mu g.m^{-3}$ for $PM_{2.5}$ and 1999 $\mu g.m^{-3}$ for PM_{10} - were eventually detected. Therefore, to check for possible errors that would interfere with the result of the comparison, a data “filter” was added to each code whose criterion was to classify values in the range or above the sensor’s detection limit as inconsistent.

Since the stipulated reading frequency for the SDS011 sensor is small the establishment of this criteria does not negatively affect the comparison, since the data set remains large enough to calculate the averages per minute that can then be adjusted according to the desired interval. In addition to applying the filter, the data consistency analysis allowed the identification of missing or overwritten data, attributed to failure of readings or sensor problems.

4.3 Daily Averages and Data Completeness

The first result analysis allowed the initial evaluation of raw data from the SDS011 sensor during the monitoring period. The raw data of $MP_{2.5}$ and MP_{10}, represented respectively in Figs. 3 and 4, allow us to observe the behavior of the sensor readings, identifying the peaks and demonstrating the sensor's sensitivity to the particulate matter oscillations. In addition, the daily averages show how the observed peaks affect the calculated daily average, providing higher values. It is still necessary to investigate the possible causes of the oscillations and high peaks of concentrations in the region, since in many cases it exceeded the maximum value allowed by current legislation. In addition, it is important to compare these data with those of reference equipment to ensure their reliability.

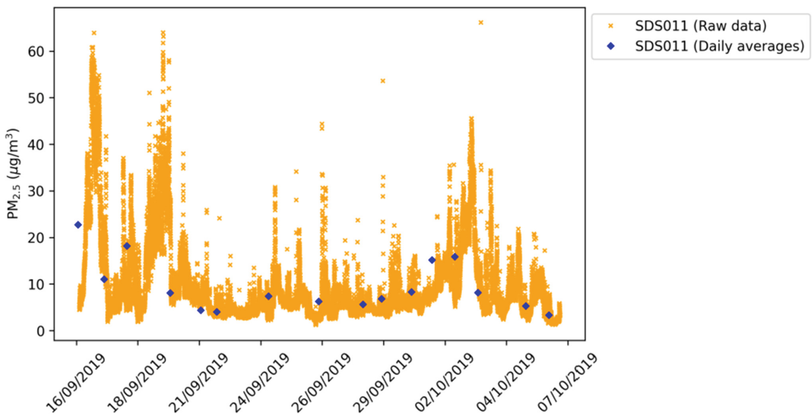


Fig. 3. Raw data and daily averages of $PM_{2.5}$ concentrations.

Despite some failures observed in the files initially obtained, the data completeness calculated for the period was 100%, which indicates that the automatic startup system guaranteed a sufficient volume of data to calculate the daily averages, which usually does not occur with reference equipment as any problem or failure requires technical intervention to re-establish and configure the system.

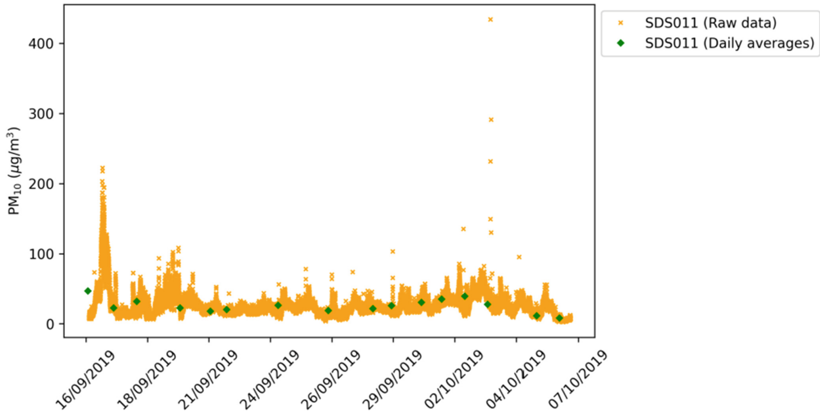


Fig. 4. Raw data and daily averages of PM_{10} concentrations.

5 Summary and Conclusion

The system implemented using Raspberry Pi represents a cheap and efficient alternative for the use of sensors, especially applied to the data acquisition and transmission using low cost sensors. Its total cost is much lower than that of reference equipment or conventional monitoring networks associated with efficient and fast data availability. The SDS011 sensor proved to be sensitive to the oscillations of particulate matter, identifying concentration peaks and values that reached the maximum detection limit. It is intended to compare the sensor measurements with the measurements of reference equipment installed at the monitoring site to propose an alternative monitoring system that suits the city' needs, creating a network capable of sharing the data with the population.

Acknowledgments. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

References

1. Aguiar EFK, Roig HL, Mancini LH, De Carvalho ENCB (2015) Low-cost sensors calibration for monitoring air quality in the federal District—Brazil. J Environ Prot 6:173–189
2. Cohen AJ et al (2005) The global burden of disease due to outdoor air pollution. J Toxicol Environ Health A 68:1–7
3. Barthwal A, Acharya D (2018) An internet of things system for sensing, analysis & forecasting urban air quality. In: 2018 IEEE international conference on electronics, computing and communication technologies (CONECCT). IEEE, pp 1–6
4. Ng KM et al (2018) Remote air quality monitoring system by using MyRIO-LabVIEW. In: 2018 9th IEEE control and system graduate research colloquium (ICSGRC). IEEE, 2018, pp 105–109
5. Binsy MS, Sampath N (2019) User configurable and portable air pollution monitoring system for smart cities using IoT. In: Smys S, Bestak R, Chen J-Z, Kotuliak I (eds) International Conference on Computer Networks and Communication Technologies, vol 15. LNDECT. Springer, Singapore, pp 345–359. https://doi.org/10.1007/978-981-10-8681-6_32

6. Badura M et al (2018) Optical particulate matter sensors in PM2.5 measurements in atmospheric air. In: E3S web of conferences. EDP Sciences, p 00006
7. WHO (2019) Ambient air pollution: Health impacts. https://www.who.int/airpollution/ambient/health-impacts/en/.
8. IBGE (2011) Censo Demográfico 2010, Área territorial brasileira. IBGE, Rio de Janeiro
9. Pope FD et al (2018) Airborne particulate matter monitoring in Kenya using calibrated low-cost sensors. Atmos Chem Phys 18(20):15403–15418
10. Johnston SJ et al (2019) City scale particulate matter monitoring using LoRaWAN based air quality IoT devices. Sensors 19(1):209
11. Badura M et al (2018) Evaluation of low-cost sensors for ambient PM2.5 monitoring. J Sens 2018
12. Macedo RYN Vargas-Cuentas NI (2018) Low cost air quality measurement system using light scattering sensors. In: 2018 IEEE sciences and humanities international research conference (SHIRCON). IEEE, pp 1–4
13. Parmar J et al (2018) IOT based weather intelligence. In: 2018 international conference on smart city and emerging technology (ICSCET). IEEE: pp 1–4
14. Pu'ad MFM et al (2018) Development of air quality measurement system using raspberry pi. In: 2018 IEEE 5th international conference on smart instrumentation, measurement and application (ICSIMA). IEEE, pp 1–4
15. Liu HY et al (2019) Performance assessment of a low-cost PM2. 5 sensor for a near four-month period in Oslo, Norway. Atmosphere: 10, 2, 41
16. Caya MVC, et al (2017) Air pollution and particulate matter detector using raspberry Pi with IoT based notification. In: 2017 IEEE 9th international conference on humanoid, nanotechnology, information technology, communication and control, environment and management (HNICEM). IEEE, pp 1–4
17. Gerboles M, Spinelle L, Signorini M (2015) AirSensEUR: an open data/software/hardware multi-sensor platform for air quality monitoring. Part A: sensor shield. EUR Technical Report JRC97581
18. TUPY (2021) Instalações: Parque Industrial de Joinville. https://www.tupy.com.br/portugues/empresa/estrutura_jlle.php

Failure Detection and Isolation by LSTM Autoencoder

Tauno Toikka(✉), Jouko Laitinen, and Kari T. Koskinen

Tampere University, 33014 Tampere, Finland
tauno.toikka@tuni.fi

Abstract. Failure diagnosis on some system is often preferred even the data of the system is not designed for the condition monitoring and does not contain any or contains little example cases of failures. For this kind of system, it is unrealistic to directly observe condition from single feature or neither to build a machine learning system that has been trained to detect known failures. Still if any data describing the system exists, it is possible to provide some level of diagnosis on the system. Here we present an LSTM (Long Short Term Memory) autoencoder approach for detecting and isolating system failures with insufficient data conditions. Here we also illustrate how the failure isolation capability is effected by the choice of input feature space. The approach is tested with the flight data of F-18 aircraft and the applicability is validated against several leading edge flap (LEF) control surface seizure failures. The method shows a potential for not only detecting a potential failure in advance but also to isolate the failure by allocating the anomaly on the data to the features that are related to the operation of LEFs. The approach presented here provides diagnostic value from the data than is not designed for condition monitoring neither contain any example case failures.

1 Introduction

Failure diagnosis is an active area of research and increasing target of application in industry due to increasing interest of condition based maintenance over a scheduled maintenance. Also the evolution of computing power, censoring technology and machine learning algorithms have been boosting the development of failure diagnostics approach in reliability engineering in recent years.

Failure diagnosis can be divided by several subcategories such like: failure detection, failure isolation, failure identification and failure classification [1]. In practice it is desirable to have diagnostics system that is capable of achieving all levels of diagnosis but the goal is cumbersome to achieve when diagnosing a complex system. With an insufficient data conditions, it is typical that the achieved level of diagnosis is only an isolation. One approach to elude the problem is to use autoencoder type neural network. In the following literature autoencoder, or variants of it, are used to perform diagnostics in various conditions.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 390–399, 2022.
https://doi.org/10.1007/978-3-030-96794-9_36

In [2] a deep learning method for fault classification and degradation assessment was presented. In the study a vibration data of rotating machinery was used and the results were validated by injected failures. The method was compared against a conventional methods and was proven to be superior. In [3] a reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems was introduced. In the method faults were isolated based on the network reconstruction. An applicability of the method was illustrated on a gas turbine process. The author claimed the method to be robust and not requiring a prior knowledge. In [4] MPL and RBF was used for detecting and isolating faults of the Tennessee Eastman benchmark process. As a novelty they transferred a time domain data to 2D image data. In [5] LSTM network was applied for fault detection and isolation task on electro-magnetical actuators of aircraft. In [6] used vibrational autoencoder (VAE) for failure detection in case of TFT-LCD manufacturing process. In [7] a stacked convolutional sparse denoising auto-encoder (SCSDAE) was used for defect defection in wafer maps in semiconductor manufacturing process. In [8] a deep transfer learning autoencoder was used for predicting remaining useful life of drilling tool. In the method a failure data was used. In [9] a stacked sparse autoencoder was used for steel grinding burn detection in supervised manner. In [10] a stacked long short-term memory autoencoder for anomaly detection in rotary machine was proposed. In [11] sparse autoencoder with PCA and SVM was proposed for power system fault diagnosis. In [12] a stacked denoising autoencoder was proposed for health state identification. In the study the diagnosis method was applied on rolling bearings.

In order to achieve all levels of failure diagnosis by data driven model, the life time data of a set of systems is needed. Life time data of a sets of systems is not available until all systems of some fleet have reached the end of their life and will be discarded. Many times a great diagnostics results have been achieved with the high quality life time data monitored in laboratory by using carefully selected sensors. On the other hand, there is a need for diagnosing systems that are in their early life state and do not yet have any failures in their history. Thus there is a need for tools that can provide failure diagnosis based on the data that is truncated, not life time data, not data from set of similar systems and does not contain necessarily any example cases of failures. This data condition here we call simply as insufficient data conditions. There exists a little study considering the failure diagnosis in insufficient data conditions.

Deep autoencoder is a special type neural network that maps its inputs to its outputs. Here we will demonstrate that carefully constructed LSTM autoencoder neural network can not only provide failure detection but also some low level isolation with insufficient data conditions. This is valuable since failure isolation is important for maintenance decision support since it provides some hint about the location from where the potential failure is developing. The following sections are organized as follows. In Sect. 2 the construction of autoencoder, LSTM neuron type, a nature of the data, training procedure and identification metrics that is a reconstruction error are described. In Sect. 3 are the results of the study followed by discussion on Sect. 4.

2 Methods

2.1 Autoencoder

In this study we applied autoencoder for detecting and isolating failures. Auto encoder is a special type of artificial neural network that has a capability of capturing the core structure of data without copying the actual data, so called representation learning. It can be also seen as a data compression tool. Autoencoder has a two parts, encoder and decoder. Encoder compresses the data by its internal structure of shrinking number of free parameters of network layers. Decoder decodes the encoded data by its structure of expanding number of free parameters of layers. Final layer of decoder part has an output that has dimensionality same as input dimensionality. When training the autoencoder then the target values are the input values.

2.2 LSTM Network

Long Short Term Memory (LSTM) neural network is constructed by using LSTM neurons. LSTM neuron has a capability of remember of its previous output state. The output state is stored until some other neuron gives the activation signal that frees the memory, so called forgot gate. Due to the memory of a LSTM network, the network is useful for modelling the data that has a temporal characteristic, that is the case with the data of this study.

2.3 Network Setup

In this study the autoencoder neural network have been build up by using Keras [13] libraries. One example network structure used here is described in Fig. 1.

```
Layer (type)                 Output Shape              Param #
=================================================================
lstm_1 (LSTM)                (None, 20, 13)            1404
_________________________________________________________________
lstm_2 (LSTM)                (None, 4)                 288
_________________________________________________________________
repeat_vector_1 (RepeatVecto (None, 20, 4)             0
_________________________________________________________________
lstm_3 (LSTM)                (None, 20, 13)            936
_________________________________________________________________
dropout_1 (Dropout)          (None, 20, 13)            0
_________________________________________________________________
lstm_4 (LSTM)                (None, 20, 13)            1404
_________________________________________________________________
time_distributed_1 (TimeDist (None, 20, 13)            182
=================================================================
Total params: 4,214
Trainable params: 4,214
Non-trainable params: 0
_________________________________________________________________
```

Fig. 1. Autoencoder structure.

Several other network configurations were also tested. The aspects configured while constructing a variety of networks were:

- Feature space size
- Network depth
- Shrinking layer size between encoder and decoder
- Dropout layers and dropout rate

3 Results

3.1 Data

The dataset used in this study is composed of 43 consequent flight data from single aircraft. Three flights were containing LEF seizure failures. Sample monitoring frequency was 1/10 s and total $1.5 * 106$ samples were available.

Several feature space configurations were used and they are described in the Table 1.

Table 1. Features (feature space 1 (FS1) feature space 2 (FS2) feature space 3 (FS3)) used for testing different autoencoder setups. Marker * shows included features.

Flight parameter	FS1	FS2	FS3
Left power leaver angle	*		
Right power leaver angle	*		
Left engine inlet temperature	*		
Right engine inlet temperature	*		
Left compression pressure	*		
Right compression pressure	*		
Left drain air temperature	*		
Right drain air temperature	*		
Left low rotor speed	*		
Right low rotor speed	*		
Left high rotor speed	*		
Right high rotor speed	*		
Dynamic pressure (I)	*	*	
Dynamic pressure (II)	*	*	
Static pressure	*	*	
Ambient temperature	*	*	
Barometer corrected pressure	*	*	
Pressure altitude	*	*	
Air speed	*	*	
Left leading edge flap position command	*	*	
Rigt leading edge flap position command	*	*	
Left inner leading edge flap position	*	*	error position
Right inner leading edge flap position	*	*	error position
Left outer leading edge flap position	*	*	error position
Right outer leading edge flap position	*	*	error position

The feature spaces of the Table 1 were constructed by the following intuition:

- FS3 contained only four error positions of the four LEF's of the aircraft thus describing only the behaviour of the LEF's and interrelationship between each other. The motivation of FS3 was to see if the actual source of failure could be isolated among the four LEF's.
- FS2 contained all directly LEF related data plus some features that we assumed to be related indirectly to the LEF behaviour. The aim of the construction of this dataset was to have data that as it maximum amount describes the behaviour of the LEF's, without aiming any obvious accessories.
- FS3 contained the data of SF2 plus some extra data that was assumed to be non-phenomenon related. The aim of this data was to construct a dataset that was hard do classifier since containing non phenomena related information, and this way to test how diagnostic results might be effected if proper feature extraction cannot be done.

The time window for LSTM was selected to be 20 samples (due to computing capacity and especially memory reasons), and thus having the window length of 2 s. The window length of 20 samples and feature space FP2 from Table 1 having 13 features, can be seen in Fig. 1 as both as an input and output.

3.2 Training

The data for training was further treated in order match to input and output shape of the LSTM autoencoder. By applying a sliding window principle, the temporal length of the data did increase by the multiplication of the window size. For example, with feature space of 13 features and window size of 20 the original data was extended from $1.5 * 10^6 \times 13$ to $1.5 * 10^6 \times 20 \times 13$.

As a training data it was selected first 35 consequent flights, thus leaving 4 consequent healthy flight before failure for validating the system, since first failure did occur during flight 40. The training data was further separated to training and test data with radios of 2/3 and 1/3. The test data was used for preventing the overfitting during training. As a validation data it was used 8 last flights from which 4 first were healthy flights and 4 last contained a LEF seizure failure. The data train, test and valid split and some other additional training information are listed in Table 2.

Table 2. Training parameters.

Lookback	20
Epochs	5
Batch	10000
Learning rate	0.0001
Train test split	0.33
Train samples	883082
Test samples	434951
Valid samples	197226

3.3 Reconstruction Error

Turing the training phase, the data that was presenting the normal behaviour of the system, was used for training autoencoder. Turing the training phase of autoencoder, the internal parameters of autoencoder network will be learned so that the network will present the structure of the data that has been used for training. By monitoring also, the loss rate during the training of autoencoder, it was ensured the trained autoencoder is presenting a normal system. During the diagnosis phase, a reconstruction error of the trained autoencoder was calculated. Reconstruction error of autoencoder is a feature vice error between the input and the output of the autoencoder. Error space size is same as the feature space size, and thus there is an error related every feature. A large reconstruction error signals that the data diagnosed does not correspond the data used earlier for training. A large reconstruction error is hinting about an abnormal behaviour and thus providing a potential failure detection. Since there is a reconstruction error related with each feature, an isolation is also provided.

3.4 Failure Cases

The target failure for autoencoder to detect and isolate is Leading Edge Flap (LEF) seizure failure demonstrated in the Fig. 2. In the Fig. 2 it can be seen that in the middle of the flight actual LEF position does not correspond to the control position and that state is defined here as a failure.

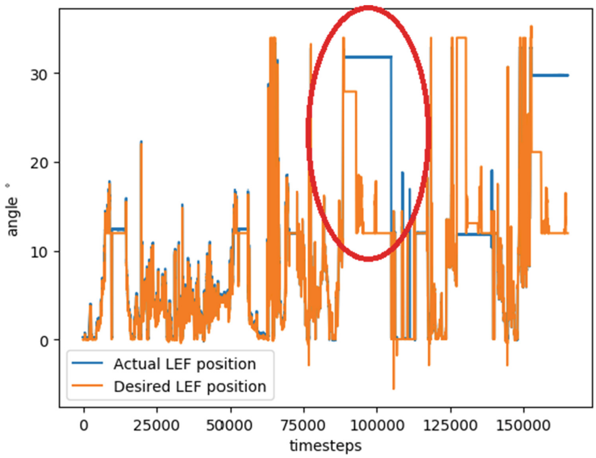


Fig. 2. Leading Edge Flap (LEF) seizure failure during the flight no. 40.

The LEF seizure failure occurred during the three flights (fights 40, 41, 43) which all started as a healthy flights (see Fig. 3). Before the failure flights there were 39 healthy flights. Healthy flights and the failure flights are presented in Fig. 3 in terms of the error position of all four LEF's of the aeroplane. The data of the Fig. 3 is the same data as the data FP3 in Table 1 except the sliding window.

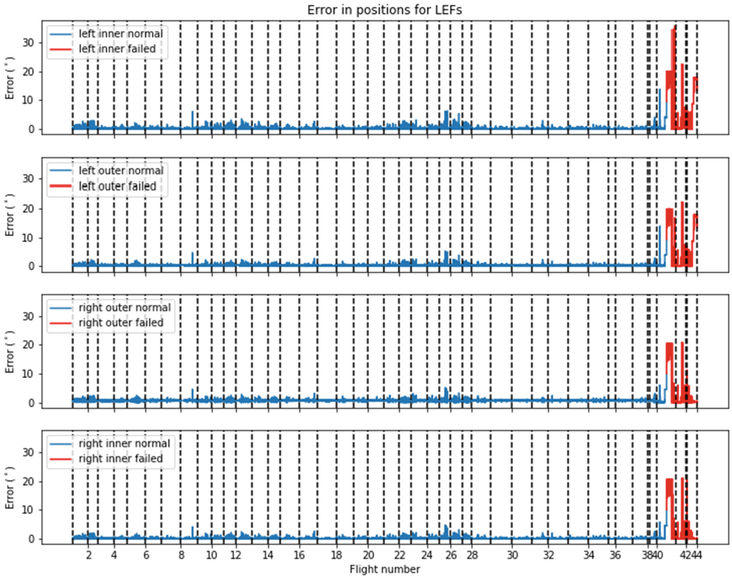


Fig. 3. Error positions for all four LEF's of F-18 aircraft during the 43 consequent flights (separated by dashed vertical lines).

3.5 Diagnosis Results

From the data of FS3 that was the data with only LEF error positions we were not able to construct a system that would reveal the source of the failure among the four LEF's, neither isolate the individual failure source. When considering a reason for this and explanation may be that in practice the LEF's of the same wing side are physically jointed. On the other hand, the interrelation between left and right side LEF's are computationally corrected by the system automation.

With the data of FS1 a failure detection was achieved since the trained autoencoder did produce large reconstruction during a healthy part of the failure flight and a one flight before. On the other hand, the reconstruction error was large on all channels and thus no meaningful isolation was achieved. The most applicable result was achieved with the data FP2 of the Table 1 and with the autoencoder of Fig. 1. The result is presented in Fig. 4.

From the figure it can be seen that the reconstruction error is small during the flights 36, 37 and 38 on all channels. This is correct since those flights have been normal flights also in practice. Also there can be seen a large reconstruction errors related to the LEF seizure (red curvature), but this is not interesting since the failure is obvious and known at that point. What is notable is that during the flight 39 that is the flight before the LEF seizure failure flight, there is large reconstruction error on all channels that are related to LEF operation. Also on the healthy part of the failure flights there can be seen large reconstruction errors.

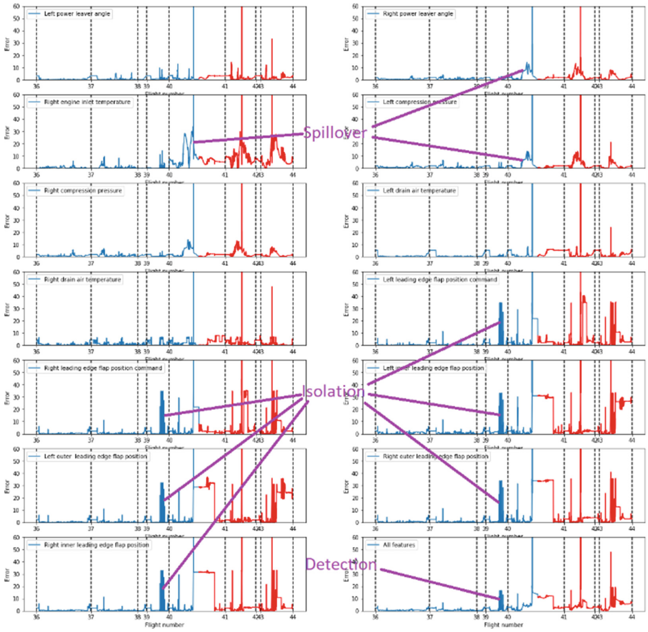


Fig. 4. Reconstruction errors for 8 last consequent flights. Blue graph indicates healthy system and red indicates system after first failure.

A large reconstruction error means that the diagnosed data does not correspond the autoencoder model and thus does not correspond the data of previous flights. These large reconstruction errors can be interpreted many ways: potential failure, abnormal flying style, abnormal environmental conditions, use of some rare functionalities of system and so on. Thus system expert is further required to analyse the result. If the system expert does not conclude an ab-normal flying style or environmental conditions one may conclude the potential failure and start further investigations.

Since the reconstruction error is large on specific channels but not on all channels, the isolation is provided among the channels. The isolation is obscured by the spillover effect. The spillover can be seen for example on channel “Right engine inlet temperature”. The temperature cannot fail, thou the sensor can. Still here the observed large reconstruction error does not present temperature sensor failure, but rather a spillover effect. Due to the internal construction of autoencoder, the autoencoder is forced to compress data. This compression leads to the creation of internal relations that do not necessarily exist in real world. From the result 4 it may be concluded that the “Right engine inlet temperature” is reconstruction from the LEF operation parameters. In practical diagnosis this incorrect behaviour need to be judged by the system expert.

4 Discussion

In the field of condition based maintenance decisions are based on a system diagnosis. A diagnosis can be done in many levels and a data analytic level is a one of them. On the other hand, many times data available is not optimal for system failure diagnosis, but rather designed for other purposes. In this type of so called insufficient data conditions the failure diagnosis requires more careful algorithmic choices.

In this study it was demonstrated that LSTM autoencoder is capable for failure diagnosis even with insufficient data conditions. The level of diagnosis achieved here was failure detection and isolation. Here it was also demonstrated how LSTM autoencoder failure isolation capability is effected by the choice of input feature space. The level of failure isolation is notable achievement when considering the limitations of the data used here. In general, the model build on data cannot present more than the nature of the original data. Since the basis here was that a data does not contain failures, the model cannot present failures. On the other hand, the model can present a normal system and thus anomalous behaviour can be detected against the model of normal system. The significance of the method presented here is that the anomalies will be allocated on features and thus providing isolation.

It is generally known that a knowledge based data pre-processing in an important preliminary step when applying machine learning, which did also apply here. The method proposed here brings up another challenge when interpreting the results, since the system knowledge turns out to be vital also on this site. The method provides only allocated anomalies, so called isolation, but it is the task of system expert to further interpreted if the anomaly actually presents a potential failure or something else.

Benefits and practical application of the methodology are that many real world system providing data have a characteristic of data similar of our data. In practice any system without failure history or with data designed directly for condition monitoring has data conditions similar to this study.

Limitations of this study are that it was done on one system. Reason here was that we had one system with data sufficient to validate our methods. In future more studies with similar methods for different domains would be needed.

References

1. Barua PS, Khorasani K (2009) A diagnostic tree approach for fault cause identification in the attitude control subsystem of satellites. IEEE Trans Aerospace Electron Syst 45(3):983–1002
2. Li Z, Wang Y, Wang K (2019) A deep learning driven method for fault classification and degradation assessment in mechanical equipment. Comput Ind 104:1–10
3. Ren S, Si F, Zhou J, Qiao Z, Cheng Y (2018) A new reconstruction-based auto-associative neural network for fault diagnosis in nonlinear systems. Chemom Intell Lab. Syst 172:118–128
4. Hajihosseini P, Anzehaee MM, Behnam B (2018) Fault detection and isolation in the challenging tennessee eastman process by using image pro- cessing techniques. ISA Trans 79:137–146
5. Yang J, Guo Y, Zhao W (2019) Long short-term memory neural network based fault detection and isolation for electro-mechanical actuators. Neurocomputing 360:85–96
6. Lee S, Kwak M, Tsui K-L, Kim SB (2019) Process monitoring using variational autoencoder for high-dimensional nonlinear processes. Eng Appl Artif Intell 83:13–27
7. Yu J, Zheng X, Liu J (2019) Stacked convolutional sparse denoising auto-encoder for identification of defect patterns in semiconductor wafer map. Comput Ind 109:121–133

8. Sun C, Ma M, Zhao Z, Tian S, Yan R, Chen X (2018) Deep transfer learning based on sparse autoencoder for remaining useful life prediction of tool in manufacturing. IEEE Trans Ind Inf 15(4):2416–2425
9. Guo W, Li B, Shen S, Zhou Q (2019) An intelligent grinding burn detection system based on two-stage feature selection and stacked sparse autoencoder. Int J Adv Manuf Technol 103:1–11
10. Li Z, Li J, Wang Y, Wang K (2019) A deep learning approach for normaly detection based on sae and lstm in mechanical equipment. Int J Adv Manuf Technol 103:1–12 (2019)
11. Wang Y, Liu M, Bao Z, Zhang S (2018) Stacked sparse autoencoder with pca and svm for data-based line trip fault diagnosis in power systems. In: Neural computing and applications, pp 1–13
12. Lu Z-Y, Wang W-L, Qin MaJ (2017) Fault diagnosis of rotary ma- chinery components using a stacked denoising autoencoder-based health state identification. Signal Process 130:377–388
13. Keras: The python deep learning library. https://keras.io/

Condition-Based Inspection Grouping Policy for Boiler Heat Exchanger Tubes

Huy Truong Ba[1,2](✉), Michael E. Cholette[1], Lin Ma[2], and Geoff Kent[2]

[1] International University, Ho Chi Minh City, Vietnam
[2] Queensland University of Technology, Brisbane, QLD, Australia
h.truongba@qut.edu.au

Abstract. Boiler heat exchanger tubes lose thickness over time, resulting in costly ruptures and losses in capacity as thin tubes are taken out of operation. To mitigate the costs associated with thickness loss, boiler tubes are inspected and can be preventively taken out of service (i.e. "plugged") to avoid in-service ruptures. Moreover, there is an economic dependence among thousands of tubes of heat exchangers in inspection and maintenance activities where a large setup cost may be induced when inspection is conducted for any tube. This paper extends the authors' previous study on boiler tube inspections to include a dynamic inspection and preventive maintenance grouping policy to pursue additional savings on setup costs. The inspection policy is said to be condition-based since the time for next inspection is based on the current inspected state. A heuristic thickness loss threshold for plugging is computed for each tube by balancing the risk of an in-service rupture with the lost revenue due to capacity loss and an optimal inspection grouping strategy is developed using the Markov Decision Process paradigm. The policy is applied to a case study of a boiler operating in an Australian sugar factory and benchmarked with policies that represent current practices and do not consider grouping. The results show that the proposed condition-based grouping inspection policy yields significant savings compared to tube individual inspection policies.

1 Introduction

Boilers are key components of energy production systems including the traditional fossil fuels and other sustainable alternatives such as biomass [1]. For solid fuel boilers, particles in the hot flue gases can cause the heat exchanger tubes to erode over time [1, 2] and this resulting erosive wear has been identified as a key issue in their operation and maintenance [2]. Tubes that lose sufficient thickness are generally preventively maintained by one of three methods: plugging the high risk tubes, augmenting/shielding the critical area of tubes, and/or replacing tube sections [3]. Among these, inspecting and plugging the high-risk tubes is a common choice, but each plugging decreases the effective heat transfer area and results in a capacity loss for the boiler. However, the inspection activities for thousands tubes is usually costly and thus inspections and maintenance decision should be optimized to minimize operational costs [4]. Thus, developing effective inspection/maintenance strategies for boiler tubes are crucial in enhancing the efficiency of the energy production in energy power plant operations.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 400–410, 2022.
https://doi.org/10.1007/978-3-030-96794-9_37

Many existing studies have sought to determine optimal inspection schedules. Some studies optimized the periodic intervals for inspection [5], while other studies have sought additional savings through the use of non-periodic or dynamic inspection policies [6, 7]. Dealing with the problem of economic dependency among the components of systems, some other studies have developed opportunistic (or grouping) inspection and maintenance policies for [8, 9]. However, there are only few articles in literature regarding inspections for boiler and heat exchanger systems, with most focused on methods for inspection [10, 11] or the risk-based inspection (RBI) approaches [11]. Yet, there are three major inspection and maintenance issues for heat exchanger systems that are underexplored in literature. Firstly, unlike most other systems, preventive plugging constitutes a "worse maintenance" action since this plugging diminishes the capacity of the boiler. Second, there is possibility to apply the dynamic inspection policies for boilers where the inspection can be delayed according to detected conditions of tubes. Lastly, the economic dependence among the tubes in inspection should be considered in inspection optimization. The authors' previous study [12] developed optimal condition-based inspection and plugging policies for the tubes of a boiler. However, this study assumed that the inspection cost for each boiler tube is independent, thus the consideration of the economic dependence among the tubes has still not been considered. This is both suboptimal and does not align well with practice — there is usually a shared setup cost for conducting *any* inspection activity, regardless of the number of tubes inspected. The aim of this study is to extend the authors' previous work in developing a tube-grouping inspection policy that considers the risks of tube rupture, performance loss of boiler due to plugged tubes, and the economic dependence among the tubes when conducting inspection activities. The inspection policy in this paper is also condition-based, where the time for next inspection is based on the current inspected thickness losses of tubes. A grouping policy is developed in this study, which consists of multiple thickness thresholds that are set to balance the risk of tube failures and costs saved by grouping inspections. The optimal parameters of the inspection grouping strategy are found using the Markov Decision Process paradigm. The developed policy is evaluated via Monte Carlo simulation with a real case study of a boiler system operating in sugar industry. The remainder of the paper is organized as follows. Section 2 describes the modelling approach, formulating the condition-based inspection/plugging optimization problem and grouping strategy. Section 3 discusses the evaluation of the identified optimum via a Monte-Carlo simulation procedure. The case study of a boiler operating in a sugar factory are presented in Sect. 4. The last section summarizes the main conclusions and proposes future studies.

2 Model Description

Similar to authors' previous study [12], the boiler tube inspection in this paper is a dynamic condition-based inspection where the next time inspection is determined according known thickness loss. A tube may be preventively plugged during the inspection or may be plugged reactively in response to a tube rupture. Inspections and subsequent preventive plugging for N_{tubes} of the investigated boiler may occur during the production outages to prevent future ruptures during subsequent high-production periods. Whether or not to preventively plug a tube depends on a pre-determined threshold on the tube thickness loss, which is only revealed by inspection. Moreover, inspections indirectly affect boiler capacity since any inspection may lead to a subsequent plugging action. The inspection policy must balance the risk costs of rupture against the costs of capacity loss while also considering the expensive setup cost for conducting the outsourced inspections. Hence, a tube grouping strategy is developed to balance the setup cost, individual inspection, and risk costs of failure in the intensive production.

The inspection grouping policy is developed based on the Markov Decision Process (MDP) approach. The timeline is discretized into K intervals $\left[t_{k-1}, t_k\right)$ with an equal length Δt where $k = 1, 2, \ldots, K$ and $t_k = k\Delta t$. It is assumed that: 1) the thickness loss of tubes at the beginning of the time horizon is known (e.g. just after renewal of the heat exchanger) and 2) the state of a tube is perfectly detected when inspected. The decision maker will decide whether or not to inspect an at-risk tube at the beginning of each time interval $\left[t_{k-1}, t_k\right)$, but the decision is executed at the end of the interval during the planned stoppage (if the tube survives). All tubes inspected at the same time will share a common set up cost. Hence, the number of inspected tubes will affect the savings—inspecting more tubes will lead to a lower setup cost per tube. Given a specific state of a tube q at a time epoch k, it may not be economical to inspect it alone due to high (setup) cost, but it may be beneficial to inspect it if there are at least $z - 1$ other tubes inspected at the same time. Let $G_z, z = 1, 2, \ldots, N_{tubes}$ be the set of tubes whose inspection is economical if $z - 1$ other tubes are inspected as well. At each decision point, a tube q will be assigned to these groups according to its last known thickness and the time since the last inspection (i.e. state) [12]. The inspection grouping strategy for each tube q at each time epoch k employs a matrix of thickness loss thresholds $s_{i,k}$ and the time indices of the last inspection $n_{i,k}$ to assign the tube to different inspection groups G_z. It is worthy to note that the thresholds are monotonically increasing $n_{i,k} \leq n_{j,k}$ and $s_{i,k} \leq s_{j,k}$ $\forall i \leq j$ and that $q \in G_Z \Rightarrow q \in G_{Z+1}$ since the setup cost per tube is monotonically decreasing in z. A conceptual illustration of the group assignment matrix is presented in Fig. 1 for tube q at time index k. The two axes correspond to the two elements of the tube state—the last inspection time (x-axis) and the thickness loss measured at that inspection (y-axis). If the measured thickness loss was low, then the tube will be near the bottom of the figure and inspection is preferable if the time since the last inspection is large (placing the state in the bottom right). Here, "preference" means that the tube q is assigned to G_z with a lower z value, indicating that it is economical to inspect the tube with fewer other tubes. Eventually, a tube may be assigned to G_1, meaning that it is economical to inspect it now, regardless of the inspection decisions for other tubes. Similarly, if the measured thickness was high (placing the state near the top of the plot), inspection again becomes preferable (decreasing z), even if the time since the last inspection is smaller. After assigning all at-risk tubes to groups, a decision is made on the groups to be inspected. A group is inspected if its set cardinality reaches the critical level $|G_z| \geq z$, i.e. there are sufficient tubes in the group to share the inspection cost, making the inspection economical. The inspected tubes are then denoted as:

$$\mathcal{G}_k = \bigcup_{z:G_z \geq \sim z} G_z$$

The computational strategy for assigning a tube to the set of possible inspection group is developed for every tubes according to MDP paradigm. The same model in previous

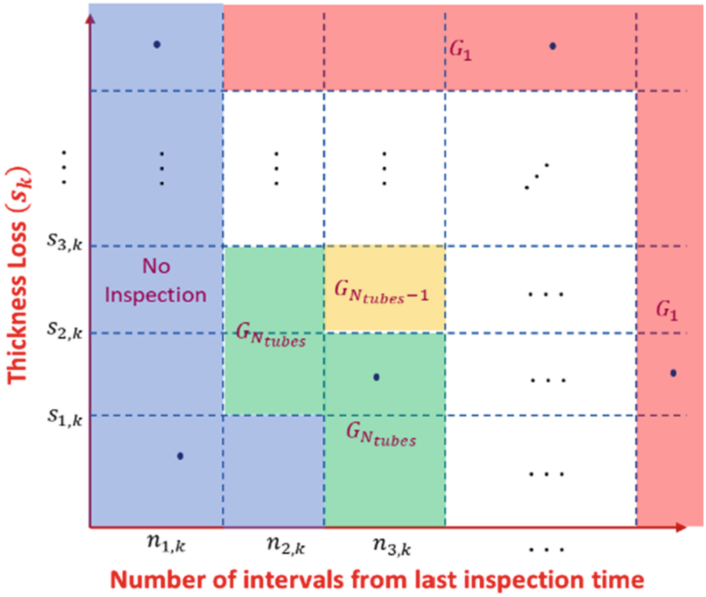


Fig. 1. Illustrative grouping assignment matrix.

study [12] is used and modified in this paper to adapt the strategy described above. The next subsections summarize the MDP framework for the sake of completeness.

2.1 Degradation Model and Transition Probability

The degradation model was developed using a physically motivated law and a Gamma Process to estimate the distribution of the thickness loss [12, 13]:

$$\Delta\theta_k = \theta_k - \theta_{k-1} \sim \mathcal{G}\left(b \cdot \mu\left(x, y, \dot{m}_{steam,k}, T_{flue,k}\right), b\right) \tag{1}$$

where θ_k is the thickness loss at time t_k, b is the scale parameter and $\mu\left(x, y, \dot{m}_{steam,k}, T_{flue,k}\right)$ is a function of boiler operation condition and location of tube in boiler. The thickness loss between zero and the rupture threshold, h_r is discretized into N bins of equal size with the following edges $\{0, \Delta s, 2\Delta s, \ldots, N\Delta s\}$, where $\Delta s = \frac{h_r}{N}$. The centers of the bins i are $s_i = i\Delta s + \frac{\Delta s}{2}$ and therefore, a discrete thickness loss θ^d can be defined as:

$$\theta_k^d = s_i \text{ if } \theta_k \in [i\Delta s, (i+1)\Delta s). \tag{2}$$

To describe the situation where a tube is taken out of service, an absorbing state value $s_N \geq h_r$ is introduced. The inspection decision is not only based on the thickness loss but also the time (index) of last inspection. Let n_k be the index of the last inspection time t_{n_k} and let the state be defined as:

$$X_k = \left(\theta_{n_k}^d, n_k\right) \tag{3}$$

where $\theta^d_{n_k}$ is the result of the thickness loss at last inspection n_k. The inspection decision at t_k is denoted as $a_k \in \{0, 1\}$, where 1 indicates that the tube is to be inspected at time t_{k+1}. The transition probabilities can be computed from the degradation model, which are denoted as:

$$p_k\big(s_j, m|s_i, \ell, a\big) \triangleq \mathbb{P}\Big[\theta^d_m = s_j, n_{k+1} = m|\theta^d_\ell = s_i, n_k = \ell, a_k = a\Big] \tag{4}$$

In the case where an inspection is decided, there are two non-trivial transition probabilities: 1) the transition between two in-service thicknesses, and 2) the transition from an in-service thickness to the tube being out-of-service (via rupture or preventive plugging upon inspection). The transition for the case of inspection is:

$$p_k\big(s_j, m|s_i, \ell, 1\big) = \begin{cases} \frac{F_{\ell:k+1}((j+1)\Delta s - s_i) - F_{\ell:k+1}(j\Delta s - s_i)}{F_{\ell:k}(h_r - s_i)} & s_i, s_j < h_p, m = k+1 \\ 1 - \frac{F_{\ell:k+1}(h_p - s_i)}{F_{\ell:k}(h_r - s_i)} & s_i < h_p, s_j = s_N, m = k+1 \\ 1 & s_i \geq h_p, s_j = s_N \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

here $F_{\ell:k}(x)$ is the CDF $\mathcal{G}\Big(b \cdot \sum_{j=\ell+1}^{k} \mu\big(x, y, u_j\big), b\Big)$, i.e. it is the CDF of the thickness losses since the last inspection (i.e. from t_ℓ to t_k).

For the case where no inspection is to occur at t_{k+1}, there are two non-trivial probabilities of interest: 1) the probability of not rupturing (which leaves the state unchanged) and 2) the transition to a thickness exceeding the rupture thickness, which will take the tube out of service. The transition probabilities for the no inspection case can be summarized as:

$$p_k\big(s_j, m|s_i, \ell, 0\big) = \begin{cases} \frac{F_{\ell:k+1}(h_r - s_i)}{F_{\ell:k}(h_r - s_i)} & s_j = s_i < h_p, m = \ell \\ 1 - \frac{F_{\ell:k+1}(h_r - s_i)}{F_{\ell:k}(h_r - s_i)} & s_i < h_p, s_j = s_N, m = \ell \\ 1 & s_i \geq h_p, s_j = s_N \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

2.2 Cost Model and Plugging Threshold

For a particular time period, there are three possible cases for each tube, each of which induce different costs:

1. The tube ruptures and must be plugged reactively, leading to a production loss cost due to downtime and a capacity loss cost for the remaining life of the heat exchanger.
2. The tube is inspected, which incurs an associated inspection cost including setup cost and individual inspection cost. If the inspection shows thickness loss above the threshold, the tube is plugged and a cost of plugging and capacity loss cost are incurred for the remaining life of the heat exchanger.
3. The tube survives and no inspection occurs, which doesn't incur any cost.

These costs can be estimated for the time period $\left[t_k, t_{k+1}\right)$ as a function of the state (s, n), the selected action a, and number of tubes in inspection group z:

$$C_k(s, n, a, z) = a \cdot \left[\frac{c_{set}}{z} + c_{ins,k} + C_k^{plug}(s, n)\right] + C_k^{gen}(s) + C_k^{rupture}(s, n) \quad (7)$$

Where c_{set} is the setup cost of inspection activity and $c_{ins,k}$ is the inspection cost (per tube). In addition, $C_k^{plug}(s, n)$ is the expected cost of plugging, $C_k^{gen}(s)$ is the expected cost of lost electricity generation and $C_k^{rupture}(s, n)$ is the expected rupture cost. Each of these costs is now described:

$$C_k^{gen}(s) = \begin{cases} c_{gen,k} & s = s_N \\ 0 & otherwise \end{cases} \quad (8)$$

$$C_k^{plug}(s, n) = c_{plug,k} \cdot \frac{F_{n:k+1}(h_r - s) - F_{n:k+1}\left(h_p - s\right)}{F_{n:k}(h_r - s)} \quad (9)$$

$$C_k^{rupture}(s, n) = c_{down,k} \cdot \frac{1 - F_{n:k+1}(h_r - s)}{F_{n:k}(h_r - s)} \quad (10)$$

where $c_{gen,k}$ represents the tube's contribution to electricity generation revenue, which could be time-varying due to fluctuating electricity prices, $c_{plug,k}$ is the individual plugging cost and $c_{down,k}$ is the production loss cost due to failure.

A tube in inspection can be preventively plugged if its thickness loss is larger than threshold h_p which is optimally computed as a balance of plugging cost, the lost production cost due to the decreased capacity of the boiler and the risk of the tube rupture in service. This threshold is approximately determined according to the balance between plugging cost and risk cost of rupture in the next season [12].

$$c_{plug,k} \le c_{down,k+1} \cdot \mathbb{P}\left[\text{rupture}\right] = c_{down,k+1} \cdot \left[1 - F_{k:k+1}\left(h_r - h_p\right)\right] \quad (11)$$

Hence, the plugging thickness loss should be:

$$h_p = h_r - F_{k:k+1}^{-1}\left(1 - \frac{c_{plug,k}}{c_{down,k+1}}\right) \quad (12)$$

2.3 Optimization via a MDP and Grouping Procedure

The inspection cost for a tube depends on how many tubes are inspected. For each inspection group size z, the optimal inspection decision, which depends on the time index of last inspection and the thickness loss, is determined as MDP paradigm:

$$V_{q,z,k}(s, n) = \min_{a \in \{0,1\}} C_{q,k}(s, n, a, z) + \sum_{j=0}^{N} \sum_{\ell \in \{k+1,n\}} p_q\left(s_j, \ell | s, n, a\right) \overline{V}_{q,k+1}\left(s_j, \ell\right)$$

$$k = 0, 1, 2, \ldots, K - 1 \quad (13)$$

Where $\overline{V}_{q,k}(s,n) = \frac{\sum_z^{N_{gr}} V_{q,z,k}(s,n)}{N_{gr}}$ is the average of all $V_{q,z,k}(s,n)$ for a particular group, which is necessary because it is not known what group a tube will be assigned to in the future. and $V_{q,z,K}(s,n) = 0 \forall s, n, z$ by definition. Note that $C_{q,k}(s,n,a,z)$ is the stage cost for tube q as Eq. (13) above. The optimal policy for tube q is denoted as $\pi^*_{q,z} = \left(\mu^*_{q,z,0}\left(\theta^d_{n_0}, n_0\right), \mu^*_{q,z,1}\left(\theta^d_{n_1}, n_1\right), \ldots, \mu^*_{q,z,K-1}\left(\theta^d_{n_{K-1}}, n_{K-1}\right)\right)$ and can be found as

$$\mu^*_{q,z,k}(s,n) = \underset{a\in\{0,1\}}{\arg\min}\, C_{q,k}(s,n,a,z) + \sum_{j=0}^{N} \sum_{\ell\in\{k+1,n\}} p_q\left(s_j, \ell|s,n,a\right)\overline{V}_{q,k+1}\left(s_j,\ell\right)$$
$$k = 0, 1, 2, \ldots K = 1 \tag{14}$$

The solutions for each tube q from MDP optimization in Eq. (14) above is a collection $\{s^*, n^*\}_{z,q,k}$ for each group G_z with size z at each time epoch k that relates to the decision of do inspection or not. In other words, given a thickness loss s and time index of last inspection n, the tube q can be "not inspected" in some group and be "inspected" in the other groups. Let $\mathcal{I}_{q,k}(s,n)$ and $\overline{\mathcal{I}}_{q,k}(s,n)$ be the set of inspection and not inspection groups, respectively of tube q at time epoch k.

$$\begin{aligned} \mathcal{I}_{q,k}(s,n) &= \left\{z \middle| \mu^*_{q,z,k}(s,n) = 1\right\} \text{ and} \\ \overline{\mathcal{I}}_{q,k}(s,n) &= \left\{z \middle| \mu^*_{q,z,k}(s,n) = 0\right\} \end{aligned} \tag{15}$$

The grouping procedure for N_{tubes} at a time interval k is as follows:

Given: the current states of all N_{tubes} tubes $\left(s_{q,k}, n_{q,k}\right)\forall q$ at current time k.
Step 1: Determine the inspection groups for each tube $\mathcal{I}_{q,k}\left(s_{q,k}, n_{q,k}\right)\forall q$.
Step 2: Identify the set of all possible inspection group $\mathcal{G}_k = \bigcup_q^{N_{tubes}} \mathcal{I}_{q,k}$.
Step 3: For each group $G_z \in \mathcal{G}_k$, the number of tubes in this group $|G_z|$ is:

$$|G_z| = \sum_{q=1}^{N_{tubes}} \mathbb{I}\left(z \in \mathcal{I}_{q,k}\left(s_{q,k}, n_{q,k}\right)\right) \text{ where } \mathbb{I}(x) = \begin{cases} 1 & x - \text{true} \\ 0 & \text{otherwise} \end{cases}$$

Step 4: Update the set of inspection group $\mathcal{G}_k$ as keeping only groups with $|G_z| \geq z$[1]:

$$\mathcal{G}_k = \{G_z \in \mathcal{G}_k || G_z| \geq g\}$$

Step 5: Select the inspection group $G_{z^*} \in \mathcal{G}_k$ that has the largest size, $G_{z^*} = \underset{G_z\in\mathcal{G}_k}{\operatorname{argmax}}|G_z|$, and do inspection for all tubes in this group.

3 Policy Evaluation via Simulation

Once a policy is known, Monte-Carlo Simulation is used to evaluate its total cost. Figure 2 shows the details of the procedure for a single simulation.

[1] Note that $\mathcal{G}_k$ can be null, and in this case, there is no inspection.

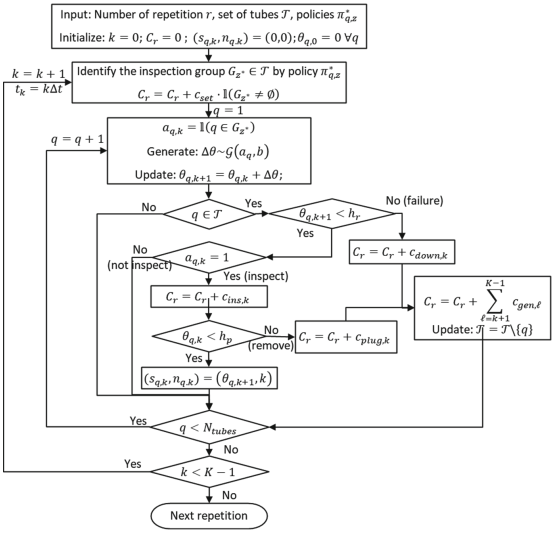


Fig. 2. Simulation flowchart

Finally, the average simulated operation cost of the boiler is:

$$\overline{Cost}_{boiler} = \frac{1}{N_{sim}} \sum_{r=1}^{N_{sim}} C_r$$

4 Numerical Examples

4.1 Case Study

An evaluation of the proposed inspection policy on a sugar factory bagasse boiler as described in [12, 13] is presented. The boiler produces steam for sugar production and the surplus is used for the generation of electricity. Due to the peculiarities of sugar production, there are two "seasons": the crushing season where production occurs and the maintenance season where no production occurs and renewals, inspections, and plugging can be done without production losses. There are 1404 tubes at risk and therefore should be in the inspection regime. The rupture and plugging thickness loss were assumed to be 3.5 mm and 2.9 mm, respectively.

The inspection cost includes individual inspection cost and setup cost, which are set at \$5000 per an inspection and \$100 per inspected tube. If a tube ruptures during the production season, it causes a downtime cost $c_{down,k}$ that is computed as:

$$c_{down,k} = \left(c_{prod} \cdot hr^{prod}_{down} + c_{elec,k} \cdot P \cdot hr^{cogen}_{down} + c_{plug,k} \right) \tag{16}$$

where P is the boiler power generation capacity (MW), $c_{elec,k}$ is the electricity price (\$/MW.h) in interval k, c_{prod} is the production loss cost, and hr^{prod}_{down} and hr^{cogen}_{down} are the numbers of downtime hours due to a rupture for production and cogeneration, respectively. If a tube is plugged or ruptures, it will cause a reduction in boiler efficiency until the end of the boiler life. This efficiency loss relates to cogeneration loss cost $c_{gen,k}$:

$$c_{gen,k} = \frac{P \times c_{elec,k} \times hr_k}{N_{cogen}} \tag{17}$$

where hr_k is the number of operating hours in the k^{th} interval. It was assumed that the boiler is running continuously for 24 h per day, 7 days per week and 25 weeks per year. N_{cogen} is the equivalent number of tubes that are responsible for cogeneration.

Table 1 summarizes the cost and parameters used in this case study.

Table 1. Cost and parameters for case study [13]

Variable	Quantity	Variable	Quantity
P	37 MW	c_{prod}	\$3750/*hour*
N_{cogen}	1244	$c_{plug,k}$	\$580/*plug*
c_{set}	\$5000/inspection	hr_{down}^{prod}	56 h
$c_{ins,k}$	\$100/inspected tube	hr_{down}^{cogen}	64 h
$c_{elec,k}$	\$60/*MW.h*	hr_k	4200 h

4.2 Results

The proposed inspection grouping policy were evaluated for this boiler operating over a finite time horizon of 15 years, which corresponds to the approximate renewal age of the heat exchanger. This policy is benchmarked with the one without considering tube grouping and another practical policy that conduct the inspection for a sample of 300 top high-risk tubes every year from year 7 [12]. For the not grouping policy, thickness loss of each tube is recorded at each year and it is used to decide of doing inspection or not for each tube. For the grouping inspection policy, the thickness loss of each tube is used to assign the tube into potential inspection groups, then the suitable tube group (i.e. the most economic benefit group) is selected for inspection. For the sample policy, 300 highest risk tubes are selected for inspection at each year since year 7.

Figure 3 shows the cumulative costs of three benchmarked policies. The grouping inspection can yield a saving up to 20% versus the policy without grouping and 16% compared with the fixed sampled inspection. This savings is significant enough to consider the grouping policy in practice of boiler inspection. The savings of grouping inspection to not grouping one is due to the lower number of ruptures even though it leads to the higher numbers of inspections and plugs versus the policy without grouping (Fig. 4). It is reasonable because the grouping inspection results earlier inspections and more inspections as well as plugs for tubes, and therefore reduces the chances of ruptures. In comparison between sampled inspection policy and grouping one, the number of ruptures is almost the same for both, however, the grouping inspection policy induces significant deduction in numbers of inspection and plugs.

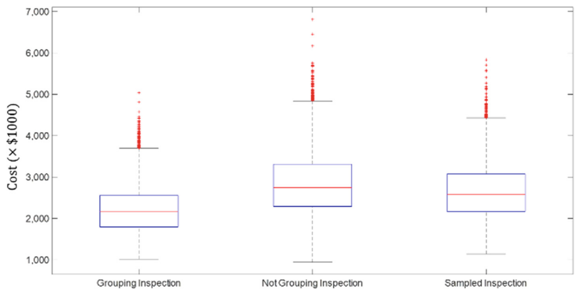


Fig. 3. Total costs of benchmarked policies

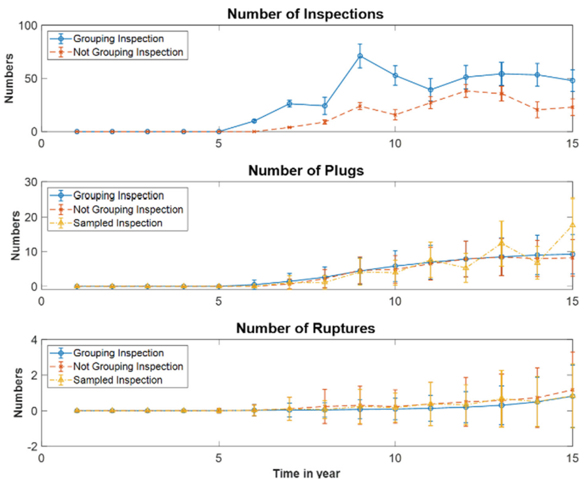


Fig. 4. Numbers of inspections, plugs and ruptures of benchmarked policies

5 Conclusion

This paper has proposed a dynamic inspection and grouping strategy for a heat exchanger system consisting of thousands tubes over a finite asset life with the next inspection depending on the condition from the previous inspection. The total maintenance cost was minimized through a Markov Decision Process where tube thickness loss thresholds and the time of last inspection for the next inspection decision as well as the heuristic grouping mechanism are jointly optimized. Numerical results of a case study for a boiler in a sugar factory showed that the proposed inspection grouping policy yields some cost saving versus the dynamic inspection policy applied for individual tube. In particular, the numerical results confirmed the intuitive hypothesis that savings are larger when setup cost of inspection activities is dominant.

Future work will focus on developing joint operation, inspection and replacement planning and consider opportunistic maintenance policies.

Acknowledgement. This study has been undertaken as part of the Australian Research Council (ARC) Linkage Project LP150100545.

References

1. Nunes LJR, Matias JCO, Catalão JPS (2016) Biomass combustion systems: a review on the physical and chemical properties of the ashes. Renew Sustain Energy Rev 53:235–242. https://doi.org/10.1016/j.rser.2015.08.053
2. Arjunwadkar A, Basu P, Acharya B (2016) A review of some operation and maintenance issues of CFBC boilers. Appl Therm Eng 102:672–694. https://doi.org/10.1016/j.applthermaleng.2016.04.008
3. Truong-Ba H, Cholette ME, Picotti G, Steinberg TA, Manzolini G. (2020) Sectorial reflectance-based cleaning policy of heliostats for Solar Tower power plants. Renew Energy 166:176–89. https://doi.org/10.1016/j.renene.2020.11.129
4. Golmakani HR, Fattahipour F (2011) Optimal replacement policy and inspection interval for condition-based maintenance. Int J Prod Res 49:5153–5167. https://doi.org/10.1080/00207543.2010.505935
5. Zhang MM, Gaudoin O, Xie M (2015) Degradation-based maintenance decision using stochastic filtering for systems under imperfect maintenance. Eur J Oper Res 245:531–541. https://doi.org/10.1016/j.ejor.2015.02.050
6. Castro IT, Landesa L (2019) A dependent complex degrading system with non-periodic inspection times. Comput Ind Eng 133:241–252. https://doi.org/10.1016/j.cie.2019.04.053
7. Zhang Z, Yang L. (2020) Postponed maintenance scheduling integrating state variation and environmental impact. Reliab Eng Syst Saf 2020:107065. https://doi.org/10.1016/j.ress.2020.107065
8. Wu T, Ma X, Yang L, Zhao Y (2019) Proactive maintenance scheduling in consideration of imperfect repairs and production wait time. J Manuf Syst 53:183–194. https://doi.org/10.1016/j.jmsy.2019.09.011
9. Babishin V, Taghipour S (2016) Optimal maintenance policy for multicomponent systems with periodic and opportunistic inspections and preventive replacements. Appl Math Model. https://doi.org/10.1016/j.apm.2016.07.019
10. Kapayeva SD, Bergander MJ, Vakhguelt A (2016) Ultrasonic and EMAT-important tools to analyze a combined effect of multiple damage mechanisms in boiler tubes. Ultrason Nondestruct Test 2016:2–9
11. Yang L, Ma X, Peng R, Zhai Q, Zhao Y (2017) A preventive maintenance policy based on dependent two-stage deterioration and external shocks. Reliab Eng Syst Saf 160:201–211. https://doi.org/10.1016/j.ress.2016.12.008
12. Truong-Ba H, Cholette ME, Borghesani P, Ma L, Kent G (2020) Condition-based inspection policies for boiler heat exchangers. Eur J Oper Res 2020. https://doi.org/10.1016/j.ejor.2020.09.030
13. Cholette ME, Yu H, Borghesani P, Ma L, Kent G (2019) Degradation modeling and condition-based maintenance of boiler heat exchangers using gamma processes. Reliab Eng Syst Saf 183:184–96. https://doi.org/10.1016/j.ress.2018.11.023

Developing a Lubrication Oil Age Prediction Model

Najat Mohammad Nazari(✉) and Masdi Muhammad

Universiti Teknologi PETRONAS, 32610 Seri Iskandar, Perak Darul Ridzuan, Malaysia
najat_19001016@utp.edu.my, najatmn@gmail.com

Abstract. In this study, lubrication oil age is predicted based on selected monitoring indicators. The information that was extracted from the oil analysis report are the TBN, oxidation, kinematic viscosity (100 °C), contaminants and elemental analysis. Correlation analysis was applied to the data to assess the relationship between the lubrication parameters and oil age. Based on the analysis, oxidation was identified to have high correlation with oil age. Mahalanobis-Taguchi Gram Schmidt (MTGS) method was applied to identify the critical variable to predict oil age. Based on the MTGS analysis, TBN, oxidation, Pb and Mo have a positive SN ratio gain and were selected to be included in the lubrication oil age prediction model. The study demonstrates the lubrication oil age prediction model based on Artificial neural network (ANN) with TBN, oxidation, Pb and Mo as predictor variables with an R squared of 0.8176, mean square error (MSE) and mean absolute deviation (MAD) of 1191 and 26 respectively. Based on the available sample data and threshold value, it can also be observed that readings of the lubrication oil parameters are still within limits after the recommended duration for lubrication oil to be in service. These findings are beneficial for future works to predict the remaining useful life of lubrication oil.

1 Introduction

Lubricant condition monitoring (LCM) is one of the important condition monitoring (CM) techniques applied to a scheduled or unscheduled maintenance activity. It reveals the particles in the fluid representing the mechanical wear of the machine. When implemented correctly, it can provide great information about the machine in operation. Lubricant reduces or eliminate wear between rotating members and aid the reduction of heat and friction which is essential to ensure proper operability of equipment. LCM plays a vital role in maintaining the operating conditions of machinery and plants. Maintenance personnel will have access to gather information on equipment, lubricant's condition and state through this monitoring programme [1]. The importance of LCM in maintenance decision validate the importance for an accurate and reliable remaining useful life (RUL) framework to be established [2].

A substantial amount of oil is wasted annually because of premature and unnecessary oil changes. Lubrication oil condition should always be known for optimization and extension of oil change interval without sacrificing the life of the equipment. Users or maintenance personnel are usually interested in two output from predictive maintenance.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 411–421, 2022.
https://doi.org/10.1007/978-3-030-96794-9_38

They are the condition of the oil and the time that oil requires changes. Current practice is lack of practicality and usefulness when implemented in the industry [3]. The Engine manufacturers (OEM) pre-determined oil change intervals were designed to deliver maximum engine protection when operating under a wide variety of conditions. Despite the majority of equipment owners that comply with these guidelines, there is a rise in trend to extend the oil change intervals beyond OEM recommendation [4]. Oil change based on predetermined standard interval is a conservative approach and results in inefficient usage of oil [5]. Oil condition measured by mileage and time intervals does not reflect the existing condition of oil or accurate level of decay [6].

In industry, the selected monitoring indicators to inspect physical and chemical properties of lubricating oil are total base number (TBN), oxidation and kinematic viscosity (100 °C) [7]. TBN represents the oxidation degree of lubrication oil. Viscosity is considered an objective mean of oil degradation detection. Kinematic viscosity is the parameter used to measure resistance of fluid flow [6, 8].

The assessment of remaining useful life (RUL) on previous researches were based on a few or a combination of limited lubricant characteristics [5]. The selection of iron and lead applicability as material wear indicators to estimate soft failure time point and optimize coefficient for hard time preventive maintenance were based on technical expert judgement, regression correlation assessment among variable and several traditional clustering methods [9]. The authors [10] stated that statistical method specifically correlation-regression analysis to be a relevant approach to study the process taking place in lubricating oil due to the complexity of physiochemical processes of oil and their interactions. Statistical analysis was applied for different purposes including development of predictive maintenance algorithm to classify analysis, multiple regression analysis and factor analysis [11].

This study aims to predict lubrication oil age using ANN model. The data used in this research were from the available oil analysis report from the routine lubrication oil analysis activities of rotating equipment installed at offshore. The machine learning application for predictive model applied was ANN. Correlation analysis was conducted to identify the relationship between oil age with the oil analysis test parameter and to identify the parameters that have high correlation with oil age. MTGS was implemented for feature selection. Lubrication oil age was predicted by ANN model based on the feature selected. The models were evaluated based on its coefficient of determination (R squared) and prediction error expressed in terms of MSE and MAD.

2 Methodology

In this study, the prediction of the lubrication oil age for gas engine that is driving the pump was conducted. Data was taken from the available lubrication oil analysis reports where the lubrication oil sample were taken and analyzed every 2 months based on Fourier-transform infrared spectroscopy (FTIR) method. Lubrication oil data for this study were taken from equipment that was operational most of the time. Among the information that were extracted from the oil analysis report are the oil change date, sampling date, total base number (TBN), oxidation, kinematic viscosity (100 °C), contaminants and elemental analysis. Lubrication oil age was calculated as per Eq. 1.

$$\textit{Oil age} = \text{Oil sampling date} - \text{Oil change date} \tag{1}$$

As a general rule, the maximum duration for lubrication oil to be in service for the equipment in this study is 6 months even if the equipment is on standby. This is to avoid microorganism growth which will stick to the wall of reservoir and piping which will cause further contamination and oxidation. Whereas the recommended oil change by lubrication oil manufacturer is 2.5 months for continuous running.

Correlation analysis was conducted to evaluate the relationship between oil age and lubrication oil analysis test parameters [12]. The type of correlation can be categorized into three types whose values range from −1 to 1. A positive correlation indicates the variables are changing in the same direction, negative correlation indicates the variables are changing in the opposite direction and no correlation indicates there is no linear relationship between the variables [13, 14]. A correlation coefficient between 0 and 0.3 implies a weak relationship, values of 0.3 to 0.7 suggests moderate relationship and correlation coefficient of 0.7 to 1 signifies strong relationship [15].

Mahalanobis-Taguchi system (MTS), designed by Taguchi is a systematic method in using Mahalanobis distance (MD). It aims to develop and optimize a diagnostic system with a measurement scale of abnormality. Effectiveness of the system is assessed by Signal-to-noise (SN) ratio [16]. The method can be used to reduce the number of variables for diagnosis and prediction of a system. Methods for handling multicollinearity was presented by using Gram-Schimdt orthogonolization [17]. Multidimensional present in a system by nature, therefore it is important to recognize sets of variables through discriminant analysis of the system. The critical variables identified can be used for further diagnosis and prediction [17]. Mahalanobis space (MS), known as the reference group is obtained from the standardized variable of normal data. MS can be used to distinguish between normal and abnormal objects. Once MS is established, orthogonal array (OA) and signal to noise (SN) ratio are computed to reduce the number of attributes. MTS has been has been used in various application such as medical diagnosis, strategy formulation and multi attribute decision making [18].

In this paper, the MTGS method was implemented for feature selection. MTGS is an improvement on MTS where MD is calculated by Gram Schimdt Orthogonalization Process [19, 20]. The steps of MTGS method are summarized below:

Task 1: Generation of Normal Space

First, define the normal space and k variables ($1 < k < 1000$). Next, collect data from normal group on k variables with sample size, n ($n \gg k$). Then, calculate MD for each sample in normal group. The variables were normalized as per Eq. 2.

Gram-Schmidt orthogonalization is calculated based on Eq. 3 through 5.

$$Z = \frac{x - \overline{\mathrm{x}}}{\boldsymbol{\sigma}_x} \tag{2}$$

$$U_1 = Z_1 \tag{3}$$

$$U_k = Z_k - c_{k,1}U_1 - c_{k,2}U_2U_3 - \cdots - c_{k,k-1}U_{k-1} \tag{4}$$

Where $c_{2,1}, c_{3,1}, \ldots, c_{k,k-1}$ are Gram-Schimdt vector coefficients

$$c_{k,j} = \frac{Z_k' U_j}{U_j' U_j} \tag{5}$$

Where
k = kth variable
j = variable from 1 to k−1
MD corresponding to jth observation is given by Eq. 6:

$$MD_j = \frac{1}{k}\left(\frac{u_{1j}^2}{s_1^2} + \frac{1_{2j}^2}{s_2^2} + \cdots + \frac{u_{kj}^2}{s_k^2}\right) \tag{6}$$

Task 2: Confirmation of Discrimination Power
In this step, data from outside the normal group on k variables is collected with sample size, r. MD for each sample outside the normal group is calculated and the discrimination power is evaluated.

Task 3: Identify Critical Variables and Optimization
The MTGS is optimized. Critical variables will be evaluated, number of variables will be optimized and the optimum system will be defined. SN ratio is used as an assessment criterion for the discrimination power is calculated per Eq. 7:

$$\eta = -10\log\left[\frac{1}{t}\Sigma_{j=1}^{1}\frac{1}{MD_j}\right] \tag{7}$$

The gain in SN ratio is calculated per Eq. 8. If the value is positive, the variable is stored, if not, it is removed from the analysis.

$$Gain = \overline{{}^{S}/_{N}\ ratio_{level-1}} - \overline{{}^{S}/_{N}\ ratio_{level-2}} \tag{8}$$

ANN comprised of a few simple and highly interconnected processing elements just like neurons in the brain. It can be interpreted as a black box that consists of a series of equations to calculate for the output with the given inputs [21]. ANN is a useful tool for function approximation, classification and prediction [22]. A regular neural network comprised of n neurons in the input layer, m neurons in the hidden layer and i neurons in output layer [23]. Figure 1 shows sample of ANN model which comprised of input layer, hidden layer and output layer.

The goodness of fit of the models were evaluated using performance indicators. Performance indicators used to compare the model were the coefficient of determination (R squared), MSE and MAD [21, 22, 24]. The equation for MSE and MAD are as in (9) and (10) respectively.

$$MSE = \frac{\sum_{i=1}^{n}(Yactual - Ypredicted)^2}{n} \tag{9}$$

$$MAD = \frac{\sum_{i=1}^{n}|Yactual - Ypredicted|}{n} \tag{10}$$

Where

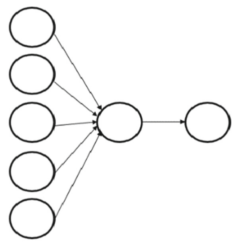


Fig. 1. Sample of ANN model

Yactual	Actual oil age
Ypredicted	Predicted oil age
n	Number of samples

3 Results and Discussion

3.1 Correlation Coefficient

Based on the data that we have obtained, the longest oil age for this equipment was 244 Days. Correlation between oil age with parameters from oil analysis was conducted to observe the relationship between them and identify lubrication oil parameters that have high correlation with oil age. The result of the correlation between the parameters is shown in Table 1. Based on the guideline by [15], oxidation was identified as highly correlated with oil age with correlation coefficient of 0.763. TBN, sodium, (contaminants), copper and lead (wear metals) are moderately correlated while kinematic viscosity (100 °C), silicon (contaminants), iron, aluminum, antimony and molybdenum (wear metals) have low correlation with oil age. The correlation coefficient between each parameter are presented in Table 1.

Table 1. Correlation analysis

	OA	TBN	Oxi	Kv	Na	Si	Fe	Cu	Al	Pb	Sb	Mo
OA	1.000	−0.666	0.763	0.213	0.357	0.212	0.220	0.343	0.221	0.678	−0.009	0.017
TBN	−0.666	1.000	−0.922	−0.386	−0.520	−0.398	−0.270	−0.548	−0.511	−0.804	−0.130	0.110
Oxi	0.763	−0.922	1.000	0.215	0.476	0.302	0.220	0.541	0.498	0.800	0.095	0.130
Kv	0.213	−0.386	0.215	1.000	0.248	−0.024	−0.076	−0.004	−0.013	0.336	−0.056	−0.492
Na	0.357	−0.520	0.476	0.248	1.000	−0.077	−0.027	0.263	0.388	0.714	0.329	−0.288
Si	0.212	−0.398	0.302	−0.024	−0.077	1.000	0.873	0.764	0.648	0.095	−0.083	−0.187
Fe	0.220	−0.270	0.220	−0.076	−0.027	0.873	1.000	0.885	0.706	0.120	0.064	−0.252

(*continued*)

Table 1. (*continued*)

	OA	TBN	Oxi	Kv	Na	Si	Fe	Cu	Al	Pb	Sb	Mo
Cu	0.343	−0.548	0.541	−0.004	0.263	0.764	0.885	1.000	0.894	0.482	0.199	−0.207
Al	0.221	−0.511	0.498	−0.013	0.388	0.648	0.706	0.894	1.000	0.569	0.289	−0.331
Pb	0.678	−0.804	0.800	0.336	0.714	0.095	0.120	0.482	0.569	1.000	0.326	−0.252
Sb	−0.009	−0.130	0.095	−0.056	0.329	−0.083	0.064	0.199	0.289	0.326	1.000	−0.103
Mo	0.017	0.110	0.130	−0.492	−0.288	−0.187	−0.252	−0.207	−0.331	−0.252	−0.103	1.000

3.2 Mahalanobis-Taguchi Gram Schmidt

Lubrication oil parameters in Table 1 were included in this analysis. The normal group (n = 17) is defined as sample data with parameters that does not exceed the threshold values (alarm and limit boundaries). Two sample were identified as abnormal group as the TBN and oxidation readings exceed the threshold values. Based on the equations from Sect. 2, the MD for normal group and abnormal group is shown in Table 2 and Table 3 respectively.

Table 2. Mahalanobis distance normal group

MD1	0.9374986	MD10	0.8386148
MD2	0.5652648	MD11	0.9387015
MD3	0.6849957	MD12	1.1260174
MD4	1.4503904	MD13	0.9918782
MD5	1.2553525	MD14	1.0398626
MD6	1.1393792	MD15	1.0034314
MD7	1.0603971	MD16	0.9561493
MD8	1.2906585	MD17	1.2103363
MD9	0.4649282		

Table 3. Mahalanobis distance for abnormal group

MD1	7.7349371
MD2	8.4216756

The variables were assigned to a two-level orthogonal array. Level 1 is defined as to use variable and level 2, not to use variable. Table 4 presents the two-level array, L_12 (2^11).

MTGS was recalculated for each run of orthogonal array. MD was calculated with abnormal sample as data, then the SN ratio was calculated as an assessment criterion for

Table 4. Two level orthogonal array

L12	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10	X11
1	1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	2	2	2	2	2	2
3	1	1	2	2	2	1	1	1	2	2	2
4	1	2	1	2	2	1	2	2	1	1	2
5	1	2	2	1	2	2	1	2	1	2	1
6	1	2	2	2	1	2	2	1	2	1	1
7	2	1	2	2	1	1	2	2	1	2	1
8	2	1	2	1	2	2	2	1	1	1	2
9	2	1	1	2	2	2	1	2	2	1	1
10	2	2	2	1	1	1	1	2	2	1	2
11	2	2	1	2	1	2	1	1	1	2	2
12	2	2	1	1	2	1	2	1	2	2	1

discrimination power. The results of the response data and SN ratio gain are presented in Table 5 and Table 6. Variables with positive SN ratio gain are included in future analysis, while variable with negative SN ratio gain are excluded. From Table 6, it can be observed that the SN ratio gain for TBN, oxidation, lead (Pb) and Molybdenum (Mo) are positive. Thus, based on the MTGS method, these variables will be included in future analysis.

Table 5. Response data

	1	2	SN ratio
1	7.7349371	8.4216756	9.0653498
2	15.537342	17.577968	12.173466
3	14.85331	18.827408	12.202624
4	14.805886	19.22906	12.234974
5	15.021915	18.259926	12.17047
6	15.161572	17.997792	12.163873
7	30.564832	34.758867	15.122469
8	30.515852	35.395602	15.155439
9	31.185162	33.53008	15.09406
10	1.8216618	4.0408323	3.9988615
11	7.835169	7.1837629	8.7479131
12	7.5429577	8.1159647	8.9314995

Table 6. SN ratio gain

Variable	TBN X1		OXI X2		KV X3		Na X4		Si X5		Fe X6	
	X1_1	X1_2	X2_1	X2_2	X3_1	X3_2	X4_1	X4_2	X5_1	X5_2	X6_1	X6_2
SN Ratio	11.668	11.175	13.136	9.708	11.041	11.802	10.249	12.594	10.212	12.632	10.254	12.584
SN Ratio Gain	0.493		3.428		-0.761		-2.345		-2.420		-2.330	
Variable	Cu X7		Al X8		Pb X9		Sb X10		Mo X11			
	X7_1	X7_2	X8_1	X8_2	X9_1	X9_2	X10_1	X10_2	X11_1	X11_2		
SN Ratio	10.213	12.630	11.044	11.799	12.083	10.761	11.285	11.558	12.091	10.752		
SN Ratio Gain	-2.417		-0.755		1.322		-0.273		1.339			

3.3 Oil Age Prediction

Table 7 shows the sample of lubrication oil data with its oil age calculated as per Eq. 1. Table 8 presents the alarm and limits of lubrication oil parameters for TBN, oxidation, Pb and Mo. The alarm and limits of the parameters are the standard set by the plant, considering recommendation by the engine manufacturer and plant responsiveness towards lubrication oil fault detection. Figure 2, 3, 4 and 5 illustrate the degradation level of these four parameters with oil age for different cycle of lubrication oil with its alarm and limits boundaries. Based on the figures, TBN decreases as oil age increases while oxidation, Pb and Mo increases as oil ages. In this paper, these four variables are included in the proposed prediction model for oil age.

Table 7. Sample of lubrication oil data

OA	19	45	29	94	151	41	103	191	244	8	64	46	101	157	276	22	42	128	216
TBN	12.4	11.8	12.4	10.6	10.1	11.5	9.8	8.7	8.5	12.6	11.2	10.8	10.5	11.5	10.7	10.8	13.1	11.4	10.8
OXI	9.1	10.8	9.1	14	17.4	11	17.1	22.4	22.8	7.8	12.1	12.4	12.3	11.7	13.6	12.3	7.9	14.7	19
PB	0.82	1	0	5	20	2	9	24	31	5	6	3	3	8	11	2	1	2	3
MO	1.99	1	3	0	1	0	1	1	1	1	0	3	1	0	0	1	1	3	5

Table 8. Alarm and limits of lubrication oil

Parameter	Alarm	Limits
TBN (mgKOH/g)	8.775	6.75
Oxidation (abs/0.1 m)	20	25
Pb (ppm)	35	50
Mo (ppm)	35	50

The neural network comprised of four neurons, (TBN, oxidation, Pb and Mo) in the input layer, three neurons in hidden layer and one neuron (oil age) in the output layer. Figure 6 and 7 shows the ANN model and the graph of the actual oil age against the predicted oil age with an R squared value of 0.8176 respectively. Actual oil age is calculated as per Eq. 1.

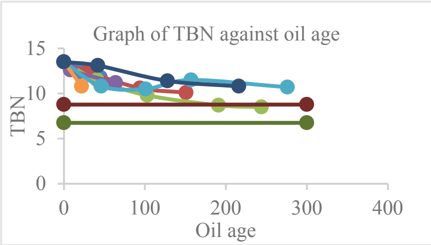


Fig. 2. Degradation of TBN with time

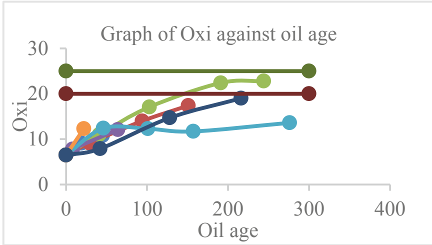


Fig. 3. Degradation of oxidation with time

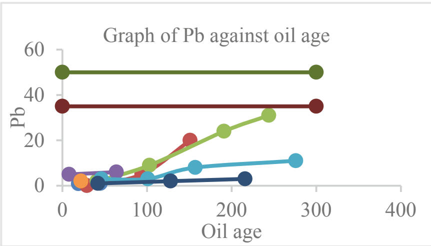


Fig. 4. Degradation of Pb with time

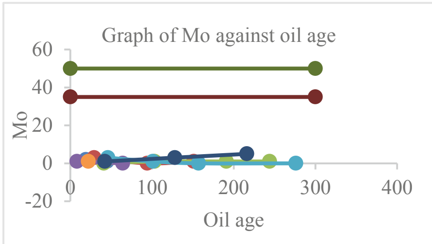


Fig. 5. Degradation of Mo with time

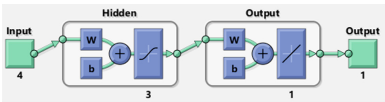


Fig. 6. ANN model

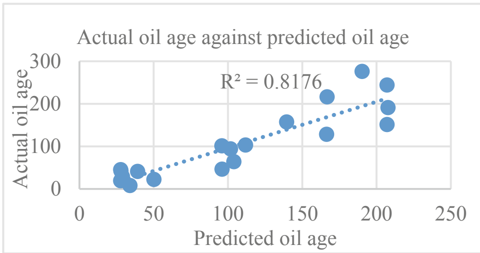


Fig. 7. Graph of actual oil age against predicted oil age

3.4 Model Evaluation

The prediction model was evaluated based on its R squared value, MSE and MAD as per Eq. 9 and Eq. 10. Table 9 summarizes the results. The recommended duration for lubrication oil to be in service is for 6 months (180 days). Referring to Table 7, Table 8 and Fig. 2, 3, 4 and 5 the recorded parameters are within its threshold values after the recommended duration. These data show that the duration for lubrication oil can be extended as the readings of parameters are still within limits after the 6 months due.

Table 9. Model evaluation

Analysis	R squared	MSE	MAD
Result	0.8176	1191	26

4 Summary and Conclusion

The primary aim of this paper is to predict lubrication oil age with its parameters as predictor based on ANN. MTGS was applied to identify the critical variable to predict the oil age. Based on the MTGS method, TBN, oxidation, Pb and Mo were selected to be included in the lubrication oil age prediction model. The ANN model presented an R squared of 0.8176, MSE and MAD of 1191 and 26 respectively. The recommended duration for lubrication oil to be in service is 6 months. Based on the available lubrication oil sample data and threshold value, it can be observed that the lubrication oil parameters are still within limits after the 6 months due. These finding is beneficial for future works to predict the remaining useful life of lubrication oil. The study proves that lubrication oil age can be predicted with TBN, oxidation, Pb and Mo as predictor variables. MTGS can be applied to identify critical variables for further analysis. Moving forward, RUL prediction model will be developed based on these parameters.

Acknowledgments. The authors would like to acknowledge the financial support provided by the Universiti Teknologi PETRONAS.

References

1. Wakiru JM, Pintelon L, Chemweno PK, Muchiri PN (2017)Analysis of lubrication oil contamination by fuel dilution with application of cluster analysis. In: IS 2017: presented at the XVII international scientific conference on industrial systems
2. Tanwar M, Raghavan N (2020) Lubricating oil remaining useful life prediction using multi-output Gaussian process regression. IEEE Access 8:128897–128907
3. Tic V, Tasner T, Lovrec D (2014) Enhanced lubricant management to reduce costs and minimise environmental impact. Energy 77:9
4. Council FM (ed) (2017) Extended Oil Drain Intervals
5. Sharma BG, Gandhi OP (2006) RUL assessment of lube oil using AHP and vector projection approach. Ind Lubr Tribol 58:8
6. Hashim H et al (2010) Artificial Neural Networks based algorithm for identifying engine oil parameters. Presented at the 2010 2nd international conference on computational intelligence, communication systems and networks
7. Du Y, Wu T, Zhou S, Makis V (2019) Remaining useful life prediction of lubricating oil with dynamic principal component analysis and proportional hazards model. Proc Inst Mech Eng Part J J Eng Tribol 234(6):964–971
8. Zhu J, Yoon JM, He D, Qu Y, Bechhoefer E (2013) Lubrication oil condition monitoring and remaining useful life prediction with particle filtering. J Prognostics Health Manage 4:15
9. Valis D, Zak L, Pokora O (2015) System condition estimation based on selected tribodiagnostic data. Qual Reliab Eng 32(2):10

10. Sharma GK, Chawla OP (1988) Modelling of lubricant oil alkalinity in diesel engines. Tribol Int 21(5):6
11. Idros MFM, Razak AHA, Junid SAMA, Sulliman SI, Halim AK (2018) FPGA implementation of predictive degradation model for engine oil lifetime. Presented at the international conference on applied electronic and engineering
12. Wakiru J, Pintelon L, Muchiri P, Chemweno P (2020) A data mining approach for lubricant-based fault diagnosis. J Qual Maint Eng 27(2):264–291
13. Gogtay N, Thatte U (2017) Principles of correlation analysis. J Assoc Phys India 65:4
14. Kumar S, Chong I (2018) Correlation analysis to identify the effective data in machine learning: prediction of depressive disorder and emotion states. Int J Environ Res Pub Health 15(12):24
15. Ratner B (2009) The correlation coefficient: Its values range between $+1/-1$, or do they? J Target Meas Anal Mark 17:4
16. Taguchi G, Chowdhury S, Wu Y (2005) Taguchi's quality engineering handbook
17. Cudney EAF, Ragadell KM (2006) Forecasting using the Mahalanobis-Taguchi system in the presence of collinearity
18. Chang ZP, Li YW, Fatima N (2019) A theoretical survey on Mahalanobis-Taguchi system. Measurement 136:10
19. Zhou Z-H, Zheng R, Liu J-F, Ding X-B (2018) Anomaly detection for sleep EEG signal via Mahalanobis-Taguchi-Gram-Schimdt method. In: 2018 4th annual international conference on network and information systems for computers, p 5
20. Souza GFM, Melo IS, Michalski MAC (2018) Applying Mahalanobis-Taguchi method to detect faults in rotating machinery. In: ESREL 2018: presented at the 28th European safety and reliability conference
21. Tosun E, Aydin K, Bilgili M (2016) Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. Alex Eng J 55:9
22. Ul-Saufie AZ, Yahya AS, Ramli NA, Hamid HA (2011) Comparison between multiple linear regression and feed forward back propagation neural network models for predicting PM10 concentration level based on gaseous and meteorological parameters. Int J Appl Sci Technol 1:8
23. Vacaro A et al (2019) Artificial neural network modelling in the prediction of bananas' harvest. Sci Hortic 257:108724
24. Sheela KG, Deepa SN (2013) Review on methods to fix number of hidden neurons in neural networks. Math Prob Eng 2013:1–11

Railway Track Geometry Degradation Modelling and Prediction for Maintenance Decision Support

Sinda Rebello[1](✉), Michael E. Cholette[1], Huy Truong-Ba[1], Venkat Reddy[2], Alan Rosser[2], and Tina Watkin[2]

[1] Queensland University of Technology (QUT), 2 George St, Brisbane, Qld 4000, Australia
sinda.rebello@qut.edu.au

[2] Queensland Rail, Rail Centre 2, 309 Edward St, Brisbane, QLD 4001, Australia

Abstract. This paper describes methods for predicting track geometry degradation to provide support for maintenance planning. The track geometry data recorded by the Track Recording Car (TRC) and the maintenance work orders were used for degradation modelling. The degradation indicator considered in this study was the deviation of the longitudinal level from the design value sampled every 100 m-long track segment. A Wiener process model was built to model the time-evolution of the geometry condition indicator for a track segment and the parameters were estimated via standard maximum likelihood estimation techniques. It was found that the proposed degradation model provided good fit to the data, except in some extreme cases where the TRC data exhibits large changes in the degradation indicator between two runs. Nevertheless, even in these cases the Wiener process model "responds" to this lack of confidence about the slope in an intuitively appealing manner: the variance on the slope is increased to reflect a lack of confidence in the degradation rate. Finally, the utility of the model in maintenance decision making is demonstrated by predicting the expected number of different maintenance interventions and policies.

1 Introduction

Rail operators manage many kilometres of railway infrastructure, which represent a large and costly investment. As with any asset, the deterioration of track infrastructure is unavoidable, and maintenance and renewal (M&R) of rail infrastructure is necessary to ensure an efficient, comfortable and safe rail network.

One of the key degradation modes is the distortion of track geometry from repeated loading, which is a critical factor for rail maintenance decisions since it has a direct impact on the ride quality and derailment risks [1, 2]. In fact, the inspection and maintenance of track geometry is often the greatest component of total maintenance costs, so the timing of inspections and maintenance tasks must be set to carefully balance track closures against the risks of poor geometry. It is therefore of paramount importance to develop engineering methods that enable the balancing of risk and direct costs of inspection and maintenance actions across these linear assets.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 422–432, 2022.
https://doi.org/10.1007/978-3-030-96794-9_39

Several track geometry degradation and maintenance modelling methodologies have been developed in literature to predict the track geometry condition. However, modelling and predicting track geometry condition is quite challenging due to its spatially distributed (linear) nature and the large number of influential physical factors (e.g. climate, track slope, curvature, track component types and materials) that also vary spatially [3] and the variable effect of maintenance actions (e.g. tamping vs. lift and pack).

The modelling approaches for track geometry degradation can be broadly classified into mechanistic models and data-driven models. While mechanistic models attempt to model the physical interactions among track components and their influence on the track geometry degradation, data-driven models rely only on monitored data to "learn" the degradation behavior of a system [4]. In general, most of the mechanistic models are focused on track settlement modelling. Some of the popular mechanistic models for track settlement include Sato model [5] and the ORE model (Office for Research and Experiments of the International Union of Railways, UIC) model. An important advantage of mechanistic models is that the known relationships between track responses and parameters of traffic can be properly represented. However, there are two key drawbacks to these models. Firstly, a precise quantification of track and vehicle properties remains a challenge. Secondly, and more importantly, mechanistic models are incapable of coping with the inherent uncertainty of the track degradation behavior, which occurs due to the large number of heterogeneous (and often unmeasured) influential factors. Thus, while mechanistic models lead to a detailed understanding of the influential factors, they are insufficient for planning purposes since they only provide deterministic predictions (i.e. a violation occurs, or it doesn't) and cannot provide insight into risk (i.e. probability and consequence) of an unnoticed violation between inspections.

In order to develop effective maintenance plans, uncertainty in the degradation modelling process needs to be considered. This requires employing concepts from probability theory and stochastic processes. However, these methodologies and approaches require sufficient track geometry data. A few researchers have applied stochastic processes such as Wiener process and Gamma process [6, 7]. However, real-world large datasets have a number of data quality challenges that can lead to poor predictive performance of these models. This paper presents a method for track geometry degradation modelling that seeks to address two key challenges in real track data: 1) the synthesis of a condition indicator of a track section under noisy measurements and 2) the identification of maintenance (e.g. tamping) events for situation where work orders are insufficiently specific. The degradation indicator considered in this study is the deviation of the longitudinal level (TOP) from the design value sampled at every 100 m-long track segment and the subsequent degradation is modelled with a Weiner Process. Using Monte Carlo simulation and the model parameters estimated from the data, the expected number of different maintenance interventions is predicted under different maintenance policies.

2 Data Collection and Processing

The maintenance and inspection data obtained from the track recording car (TRC) for one particular corridor (called "*C1*" here) was used in this study. Since a railway track is a linear asset, it is necessary to divide the track into similar length segments and model them independently. Hence, after the data is extracted into the programming platform and prior to degradation modelling, two key data processing steps were performed: track segmentation and identification of maintenance actions for each segment.

The corridor considered in this study is approximately 44 km long and was divided into consecutive segments of 100 m for modelling purposes. A single value for the geometry index is typically selected to summarize the condition of each segment, e.g. mean absolute deviation, standard deviation of the absolute deviation, and extreme values of isolated defects (maximum or 95th percentile). All of these indicators have advantages and drawbacks. Using mean tends to smooth out occasional high excursions of indexes, while selecting the maximum will yield high sensitivity to outliers and over-emphasize single high values. A good compromise was found to be the 95th percentile value, which removes some of the sensitivity to single outlier points, but still yields an estimate of the high deviations present in the track segment (albeit not the highest). The histogram (Fig. 1) illustrates the trade-off for two segments. On the left-hand plot, it is clearly seen that the mean is not a good representation of the worst-case deviation in the segment, while the max and 95th percentile values are quite close. On the other hand, the right-hand plot shows the drawback of using the maximum value: the track condition is overly sensitive to a single high value (which might be an outlier). After comparing the results obtained using these 3 statistics, 95th percentile was selected for track segmentation in this study.

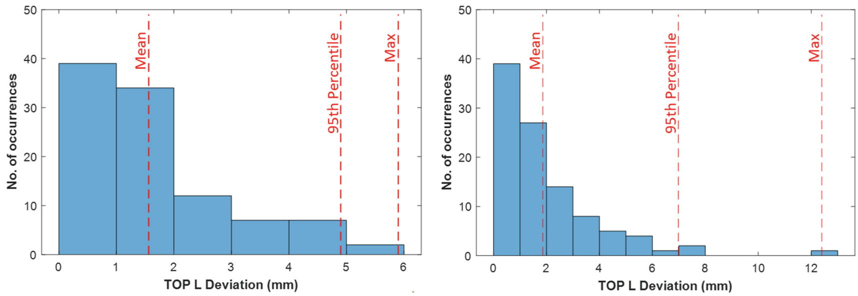


Fig. 1. Histogram showing TOP L deviations over a segment of 100 m

After track segmentation, the maintenance actions performed on each track segment need to be identified. A precise identification of the maintenance actions and the time when maintenance was done on a segment is crucial; maintenance actions (e.g. tamping) restore the track condition, and degradation starts anew from that restored point.

After analysing the work orders, it was found that the execution of some work orders did not seem to significantly impact track geometry condition. This can be perhaps attributed to the incorrect/non-specific location and execution time of the work, which is a common issue with work order databases [8]. In certain segments, significant improvement in track geometry condition in between successive TRC runs were obvious, but there were no work orders associated with those dates. These issues led to the development of an alternative methodology for identifying maintenance times that uses a statistical approach and was used later for degradation model development. In this approach, the maintenance times are identified directly from the TRC data by assessing the improvements in track geometry condition between successive TRC runs. It is outlined below:

Step 1: Find the difference in track geometry measures between successive TRC runs.

Step 2: Find significant improvements in segment condition (which are expected to be negative outliers). The significant improvements are identified as those samples that are m times the interquartile below the lower quartile boundary ($m = 1.5$ is typical and used in this study). Alternatively, a fixed threshold may be used if good knowledge is available on the typical improvement due to a maintenance event.

Step 3: If a negative outlier is found, it is assumed that a maintenance event has happened in between the TRC runs. The maintenance time is assumed to be the date halfway between the two respective TRC run dates.

3 Degradation Modelling and Prediction

In this section, the degradation model is developed using the pre-processed TRC data and identified maintenance times. In this work, a Wiener process is used to model the degradation process. The structure of the model and identification of the parameters are described in this section.

3.1 Wiener Process Model and Parameter Estimation

Wiener processes have been extensively used in the literature to model the reliability of products that are affected by degradation processes such as wear, corrosion, crack growth etc. In this study, Wiener process is used to represent the track geometry condition evolution over time due to its ability to deal with non-monotonic behaviour of the track geometry deviations obtained from TRC data. A Wiener process is particularly well-suited for modelling the evolution of a degradation mechanism characterised by a linear increase over time with random noise [9].

A Wiener process-based model has two parameters, one related to the expected value of the degradation rate and one that represents the magnitude of the random noise [7]. The degradation measure $Z(t)$ can be represented by [10]:

$$Z(t) = z_0 + \sigma W(t - t_0) + \mu(t - t_0) \quad t \geq t_0 \tag{1}$$

Where z_0 is the initial degradation ($z_0 \in \mathbb{R}$), t_0 represents the beginning of the degradation ($t_0 \in R$), μ is the drift parameter ($\mu \in \mathbb{R}$), σ is the variance ($\sigma > 0$), and $W(t)$ is the standard Wiener process representing the stochastic dynamics of the degradation process $(0, \infty)$ [9]. The maintenance actions such as tamping will improve the track geometry condition; however, it cannot fully restore the geometry condition to an as-good-as-new state. Therefore, z_0 can be considered as the average deviation from the design value that persists even after a maintenance action. Note that z_0 is assumed to be normally distributed in this study. Typically, a Wiener process is characterized by continuous sample paths which have independent, stationary and normally distributed increments. The distribution of $Z(t)$ can be expressed as:

$$Z(t) \sim \mathcal{N}\left(z_0 + \mu t, \sigma\sqrt{t}\right) \quad (2)$$

which simply means that modelling a stochastic degradation process as a Wiener process implies that the mean degradation path is a linear function of time, i.e. $\mathbb{E}(Z(t)) = z_0 + \mu t$, and the variance is $\sigma^2 t$.

Given a set of inspection data, the parameters of the model can be identified via Maximum Likelihood Estimation (MLE) as follows. Consider n realizations of the degradation process $\{Z_i(t)\}$, $i = 1, \ldots, n$ which are observed, where for each realization there are m_i observations $Z_{ij}\left(t_{ij}\right)$ at times $t_{ij}, j = 1, \ldots, m_i$. Furthermore, let $Y_{ij} = Z_i\left(t_{ij}\right) - Z_i\left(t_{ij-1}\right)$ be the observed increments of the process and let $s_{ij} = t_{ij} - t_{ij-1}, j = 2, \ldots, m_i$ be the time elapsed between observations. The MLE estimator for the Wiener process parameters (μ, σ) can be shown to be:

$$(\mu, \sigma) = \underset{\mu,\sigma}{\operatorname{argmax}} \prod_{i=1}^{n} \prod_{j=2}^{m_i} \frac{1}{\sqrt{\sigma^2 s_{ij}}} \phi\left(\frac{Y_{ij} - \mu s_{ij}}{\sqrt{\sigma^2 s_{ij}}}\right) \quad (3)$$

Where ϕ is the density function of the standard normal distribution. For this study, the initial degradation value z_0 of a segment is estimated by averaging the z_0 obtained for each degradation process related to a specific number of neighbouring track segments[1].

Usually, the times where the degradation measures exceed a certain boundary degradation level (e.g. maintenance thresholds[2]) are of interest. The distribution of these times can be found analytically from the model. For a given threshold h (e.g. a D1 threshold), the lifetime T_h of the track is the instant at which the degradation process $Z(t)$ exceeds the level h for the first time. Mathematically, this can be written as:

$$T_h = \inf\{t \geq t_0 : Z(t) \geq h\} \quad (4)$$

Consequently, the lifetime distribution is

$$F_{T_h}(t) = \Pr(Z(t) \geq h) \quad (5)$$

The probability density function (PDF) and the cumulative density function (CDF) of lifetime T_h at time t_i can be directly obtained as [9]:

$$f_{T_h}(t) = \frac{h - z_0}{\sigma\sqrt{2\pi t^3}} \exp\left(-\frac{(h - z_0 - \mu t)^2}{2\sigma^2 t}\right) \quad (6)$$

$$F(t|h, z_0, \mu, \sigma) = \int_0^t \frac{h - z_0}{\sigma\sqrt{2\pi x^3}} \exp\left(-\frac{(h - z_0 - \mu x)^2}{2\sigma^2 x}\right) dx \quad (7)$$

---

[1] This was done due in part to the small sample size, which can lead to significant bias in MLE.

[2] The rail operators follow some standards which define the maintenance threshold level for each geometry index. For e.g., D1, D7, M1 levels determine the threshold when exceeded, the maintenance should be done within 1 day, 1 week, 1 month respectively.

3.2 Predicting the Expected Number of Maintenance Requirements Using Monte Carlo Simulation

With the Wiener model parameters estimated for each segment of the track, it is also possible to predict the expected number of maintenance interventions (D1, D7, M1 etc.) required in the future. In this study, Monte Carlo (MC) simulation is used. Under each simulation, a degradation path for the geometry index is simulated for each track segment using estimated model parameters and Eq. (2). While a number of maintenance policies are possible, a combined corrective maintenance (CM)/preventive maintenance (PM) policy is simulated in this study to demonstrate the utility of the model. Aside from the degradation model, there are three inputs to the simulation:

1. The inspection interval Δt (e.g. three months);
2. The index threshold that triggers preventive resurfacing, y_{PM} (e.g. the M1 threshold);
3. The index threshold that triggers corrective resurfacing, y_{CM} (e.g. D7 threshold).

Note that a corrective-maintenance-only policy can be simulated as $y_{PM} \to \infty$.

When the track is maintained, the degradation measure is set to z_0 and the degradation path is simulated henceforth till the next inspection point. If y_{CM} has not been exceeded at the inspection time, the track continues its degradation path until the next inspection point where it is again compared with the threshold value. This process continues over a fixed time horizon and the CM and PM events are counted at each time. Subsequently, the expected value is determined by taking the mean value of the number of exceptions that resulted from each simulation. In addition, a confidence interval was obtained by taking the 2.5 and 97.5 percentile values.

4 Results

4.1 Degradation Modelling and Prediction Based on Wiener Process

This section shows results of the Wiener process model fitted to Top L (Left rail) of corridor *C1* as an illustrative example. The model parameters (μ, σ) estimated for TOP L for every segment of the corridor are shown in Fig. 2 and Fig. 3.

As can be seen in these plots, the degradation rates for the track segments are quite varied and certain segments degrade at a faster rate than others, e.g. segment 214.7. In order to illustrate each of these degradation rates, the model fitting of 2 segments which exhibit a low and high degradation rate (μ) for TOP L are shown in Fig. 4 and Fig. 5 respectively. As can be noted from Fig. 4, the segment shows a reasonably flat degradation trend and the 95% confidence interval (indicated by the red dashed line) is well below the M3 threshold. On the other hand, the degradation trend in Fig. 5 appears to be strongly driven by a few large jumps that drive up the average degradation rate. Due to the wide range of degradation rates, the standard deviation is also quite large, resulting in the wide 95% confidence intervals displayed. Two possible explanation for this result are as follows: 1) the increases appear to occur just before maintenance (tamping) events, which suggests that there may be some error in the TRC readings right after the tamping event; 2) The linear fit degradation trend is not a good candidate for this segment. In both cases, note that the highly variable degradation rate observed in the data results in a large estimated variance (and large 95% confidence interval). This can be seen as an acknowledgement that the average linear trend is a poor fit for this segment and the model fitting "responds" with a high estimated uncertainty.

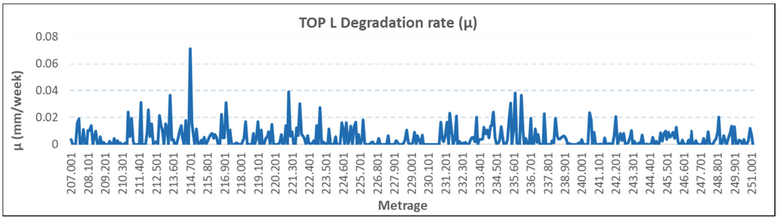


Fig. 2. TOP L Degradation rate (***mm/week***) for the entire corridor

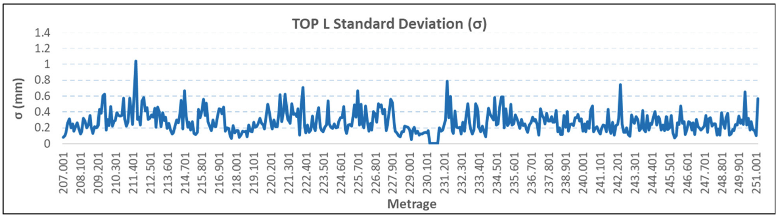


Fig. 3. TOP L Standard Deviation (**mm**) for the entire corridor

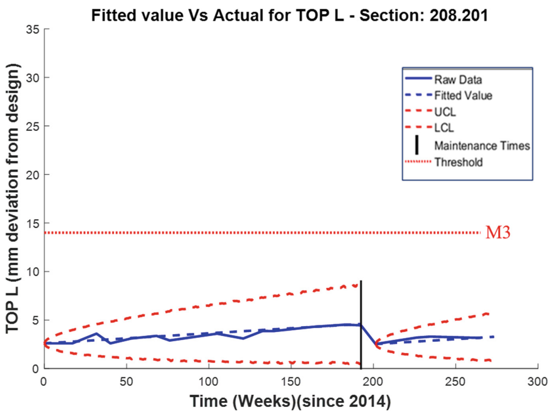


Fig. 4. Model fitting to TOP L (low degradation rate)

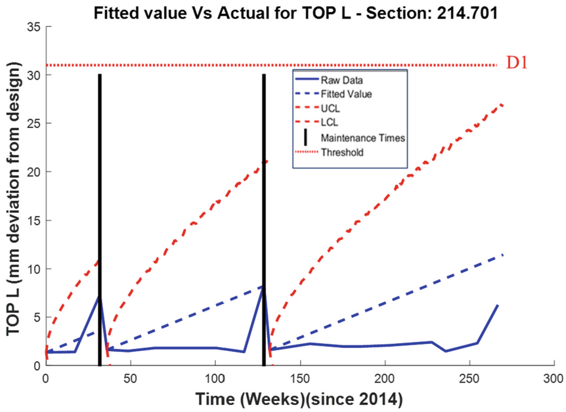


Fig. 5. Model fitting to TOP L (high degradation rate)

4.2 Expected Number of Failures and What-If Analysis

This section shows the result of what-if analysis that was described in Sect. 3.2, which is aimed at providing decision support for inspections (TRC runs) and maintenance. Three different scenarios are described in this section to show how the expected number of failures (D1 & D& exceptions) change with different maintenance policies. They are listed below:

1. Corrective Maintenance (CM) Only: inspect every 3 months and repair if exceeding D7 threshold
2. Preventive + CM: inspect every 3 months and do maintenance if exceeding M1 threshold
3. More Aggressive Preventive + CM: inspect every 3 months and do maintenance if exceeding M3 threshold

- **Scenario 1: CM Only**

Under this policy, the track is periodically monitored every 3 months and the track is correctively restored if any segment has exceeded the D7 maintenance limit. Figure 6 shows the predicted number of TOP L exceptions for track under consideration over 15 years. It can be noted that 57 exceptions in TOP L are expected to arise from corridor *C1* in 15 years, under this corrective maintenance policy.

- **Scenario 2: CM and PM on M1 thresholds**

  Figure 7 illustrates the expected number of exceptions predicted over 15 years if preventive maintenance is performed in addition to corrective maintenance with a periodic inspection of 3 months. In particular, at the inspection time, the track is correctively repaired if it exceeds D7 threshold and preventively maintained if M1 threshold is exceeded. As can be noted from Fig. 7, the number of D1 & D7 exceptions has considerably decreased with preventive maintenance. However, this

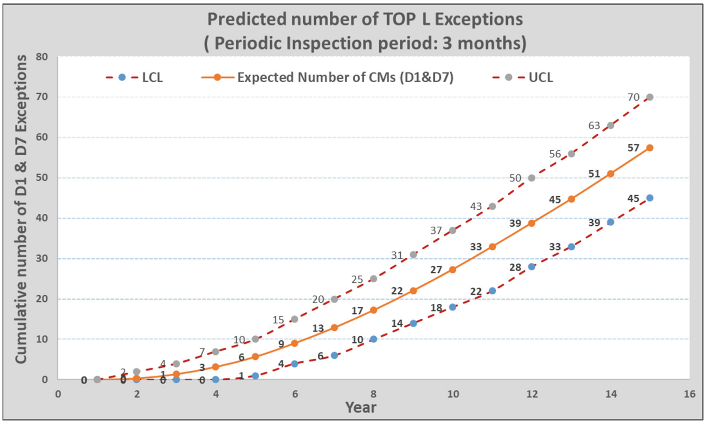


Fig. 6. Predicted number of TOP L exceptions for *C1* under CM only

improvement is achieved at the cost of preventive maintenance. For example, considering the entire corridor *C1*, the number of D1 & D7 exceptions reduced by 54 with preventive maintenance, however, it incurs 85 preventive maintenances over the span of 15 years.

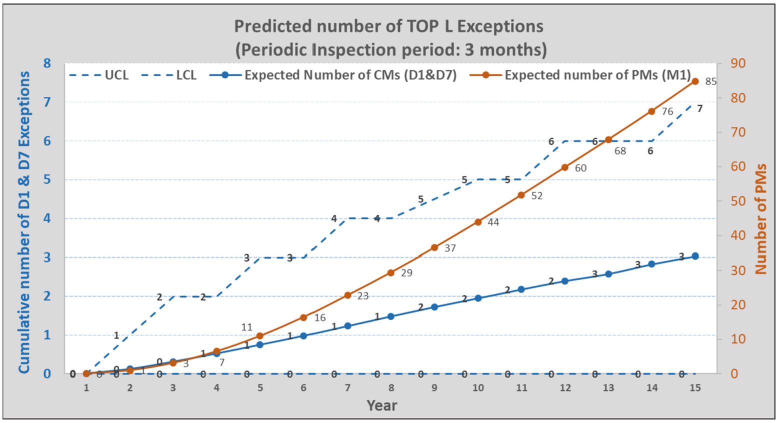


Fig. 7. Predicted number of TOP L exceptions for *C1* under scenario 2

- **Scenario 3: CM and PM on M3 thresholds**

Figure 8 illustrates the expected number of exceptions, predicted over 15 years if the track is preventively maintained at M3 threshold. The expected number of exceptions for *C1* is nil under this policy, however, the number of required preventive maintenance has doubled compared to Scenario 2. It is also noteworthy that repeated tamping actions can in fact result in more damage to underlying ballast and sub-ballast, which is not yet accounted for in our model. This may lead to a faster degradation in track geometry. Hence for choosing the most effective maintenance policy, more detailed study needs to be conducted and the effect of excessive tamping needs to be incorporated into the model.

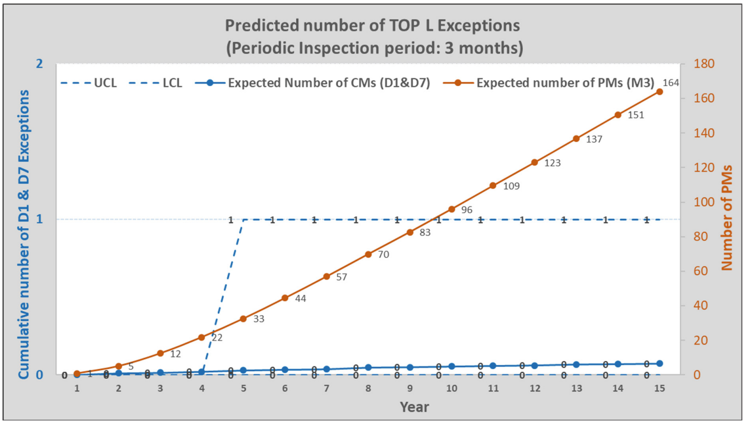


Fig. 8. Predicted number of TOP L exceptions for *C1* under scenario 3

5 Conclusion

A method for modelling the degradation in track geometry is presented in this paper. The track geometry data recorded by the Track Recording Car and the maintenance work orders were analysed and used for track degradation modelling and prediction. The degradation model was developed based on Wiener process in order to incorporate the random noise observed in the monitored data. The track was divided into multiple homogenous segments for analysis and the 95th percentile index was used as a measure of the condition of the segment. The effective maintenance actions carried out on a particular segment were found via work order and statistical analysis. Unfortunately, the work orders alone were insufficient, likely due to lack of accurate location information, and hence the statistical method was employed in further modelling.

With the maintenance actions in hand, the degradation model parameters of each segment were estimated using maximum likelihood estimation. It was found that the proposed linear (on average) degradation model provided good fits to the data, except in some extreme cases where the TRC data exhibits large changes in the degradation indicator between two runs. Nevertheless, even in these cases the Wiener process model "responds" to this lack of confidence about the slope in an intuitively pleasing manner: the variance on the slope is increased to reflect a lack of confidence in the degradation rate. Using the estimated model parameters, the expected number of different maintenance interventions was estimated using Monte Carlo simulation under different maintenance policies. The simulation-based prediction of maintenance interventions support maintenance decision making. For instance, the predicted number of D1 and D7 exceptions and the number of preventive maintenances can be associated with various cost factors to determine the expected maintenance cost. Henceforth, the maintenance cost associated with different inspection schedules and maintenance policies can be compared to find the optimum inspection interval and maintenance policy.

References

1. Khajehei H, Ahmadi A, Soleimanmeigouni I, Nissen A (2019) Allocation of effective maintenance limit for railway track geometry. Struct Infrastruct Eng 15:1597–1612
2. Prescott D, Andrews J (2015) Investigating railway track asset management using a Markov analysis. Proc Inst Mech Eng Part F J Rail Rapid Transit 229(4):402–416
3. Soleimanmeigouni I, Ahmadi A, Kumar U (2018) Track geometry degradation and maintenance modelling: a review. Proc Inst Mech Eng Part F J Rail Rapid Transit 232(1):73–102
4. Mishra M, Odelius J, Thaduri A, Nissen A, Rantatalo M (2017) Particle filter-based prognostic approach for railway track geometry. Mech. Syst. Sign. Process. 96:226–238
5. Sato Y (1995) Japanese studies on deterioration of ballasted track. Veh Syst Dyn 24(sup1):197–208
6. Meier-Hirmer C, Senée A, Riboulet G, Sourget F, Roussignol M (2006) A decision support system for track maintenance. books.google.com
7. Letot C, Soleimanmeigouni I, Ahmadi A, Dehombreux P (2016) An adaptive opportunistic maintenance model based on railway track condition prediction. IFAC-PapersOnLine 49(28):120–125
8. Arif-Uz-Zaman K, Cholette ME, Ma L, Karim A (2017) Extracting failure time data from industrial maintenance records using text mining. Adv Eng Inform 33:388–396
9. Si X-S, Wang W, Hu C-H, Chen M-Y, Zhou D-H (2013) A Wiener-process-based degradation model with a recursive filter algorithm for remaining useful life estimation. Mech Syst Signal Process 35(1–2):219–237
10. Kahle W, Lehmann A (2010) The wiener process as a degradation model: modeling and parameter estimation. In: Advances in Degradation Modeling. Springer, pp 127–146. https://doi.org/10.1007/978-0-8176-4924-1_9

Efficient Implementation of Artificial Neural Networks for Sensor Data Analysis Based on a Genetic Algorithm

André D'Estefani[1], Raymundo Cordero[1](✉), and João Onofre[2]

[1] Cidade Universitária, Campo Grande, MS 79070-900, Brazil
raymundo.garcia@ufms.br
[2] Oak Ridge National Laboratory - ORNL, Oak Ridge, USA
pintoj@ornl.gov

Abstract. The reliability of many industrial processes depends on the sensor system. However, these sensors can be affected by noise, perturbations and failures. Hence, sensor monitoring and diagnosis are fundamental to guarantee the quality of an industrial process. Nowadays, artificial neural networks (ANN) are widely used in sensor signal processing and diagnosis. However, those ANNs usually require many artificial neurons, being difficult to implement in software and hardware due to their high computational costs. This paper presents an optimized implementation of artificial neurons in ANNs for sensor data analysis using a Genetic Algorithm (GA). The objective of GA is to find an adequate segmentation to reduce the activation function approximation error. One of the advantages of the proposed approach is that the cost function used in GA considers the effect of factors such as the ANN architecture or the number of bits used in arithmetic operations. The proposed ANN implementation technique aims to get the best possible approximation for a specific ANN architecture, making easier its implementation in software and hardware. Simulation and experimental results using FPGA (Field Programmable Gate Array) prove the advantages of the proposed approach for implementing sensor data analysis systems based on ANNs.

1 Introduction

Nowadays, many real-time applications such as image/speech processing, sensor data processing and industrial process control are based on artificial neural networks (ANNs) [1–3]. ANN-based techniques such as Deep Learning have complex structures composed by a great quantity of artificial neurons. As a result, those ANNs execute a great number of arithmetic operations. The greater the number of arithmetic operations, the greater the complexity and cost of digital processors such as DSP (Digital Signal Processor) or FPGA (Field Programmable Gate Array) required to calculate the ANN outputs with the required processing speed [1–4]. This problem is relevant in the application of Deep Learning in online sensor signal processing required in electric and hybrid vehicles, where fast analysis is required to guarantee the good performance and safety of those vehicles [4, 5].

The calculation of the nonlinear activation function is one of the most difficult tasks in the implementation of ANNs [6]. Sigmoid activation function is widely used in ANNs.

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 433–442, 2022.
https://doi.org/10.1007/978-3-030-96794-9_40

However, the division and the exponential in this function are difficult to implement in digital processors such as FPGAs. For that reason, approximation techniques are used to calculate this function. Many techniques were proposed to approximate the sigmoid function: look-up tables (LUT), Taylor series, piecewise linear method (PWL), SPLINE techniques and others [7–12]. Taylor series gives accurate estimations, but it requires many mathematical operations. A LUT uses a great quantity of memory to give accurate estimations. SPLINE and PWL have good compromises between accuracy and computational cost. In both techniques, the input range of the function is segmented, and each segment is approximated by a linear function (PWL) or by a higher-order polynomial (SPLINE) [7, 8]. The points used to divide the function input range are called knots. The segmentation of this range (i.e., the selection of the knots) has great influence in the approximation accuracy [13]. Other important topic is the effect of constraints, such as the number of bits used in arithmetical operations, in the approximation accuracy. These constraints are difficult to express analytically.

On the other hand, genetic algorithm (GA) is a heuristic technique widely applied to find an optimal solution. This technique, based on natural selection theory, combines the strategies of survival of the fitness among individuals (possible solution) with information exchange [14]. GA is suitable to find optimal solutions for non-linear problems whose solution cannot be expressed analytically.

This paper proposes the application of a genetic algorithm (GA) and SPLINE technique in the implementation of the nonlinear activation function. A genetic algorithm is used to find the optimal segmentation of the activation function to reduce the approximation error when SPLINE is applied. The optimal segmentation can be adjusted to considering different error measurement methods (e.g., maximum absolute error or mean square error) and other problems such as the limited number of bits used in arithmetic operations, which are difficult to express analytically. The proposed approach can be applied to other activation function such as radial basis functions. Simulation and experimental tests performed in an FPGA show the advantage of the proposed approximation method for ANNs.

2 Theoretical Foundations

2.1 Activation Function

Let $f(x)$ represents the activation function of an artificial neuron. In this case, the sigmoid function will be considered. However, the proposed approximation technique can be applied to other activation functions. The sigmoid function, shown in Fig. 1, is defined by (1), where x is the function input.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

The sigmoid function has symmetry at the point (0, 0.5) [8, 12]. Thus, it is possible to prove that the sigmoid function satisfies the property expressed in (2):

$$f(-x) = 1 - f(x), 0 \leq x \tag{2}$$

Equation (2) shows that only the approximation of the function for $x > 0$ is required, as the function value for negative inputs $-x < 0$, i.e., $f(-x)$, can be deduced from $f(x)$. Besides, in practical applications, it is considered that $f(x) \approx 1$ for $x > 8$ [7, 8]. Therefore, the approximation techniques of the sigmoid function only consider the interval [0, 8] in practical applications.

2.2 SPLINE Approximation Technique

Let $f(x)$ be a real function defined in the range $x \in [a, b]$. In SPLINE technique, the function input range $[a, b]$ is divided into n sub-intervals $I_j = [k_{j-1}, k_j]$, $j = 1, \ldots, n$, where $k_0 = a$ and $k_n = b$. The terms $k_0, k_1, \ldots, k_n$ are called knots. Note that, $I_1 \cup I_2 \cup \ldots \cup I_n = [a, b]$. Thus $f(x)$ can be divided into n segments, where the j-th segment is defined in the interval I_j. The SPLINE approximation technique consists in approximate each function segment by a r-order polynomial function $p_{jr}(x)$ [12, 13]:

$$f(x) \approx p_{jr}(x), \;\; x \in I_j = [k_{j-1}, k_j]. \tag{3}$$

where

$$p_{jr}(x) = a_{jr}x^r + a_{j(r-1)}x^{r-1} + \ldots + a_{j1}x^1 + a_{j0}. \tag{4}$$

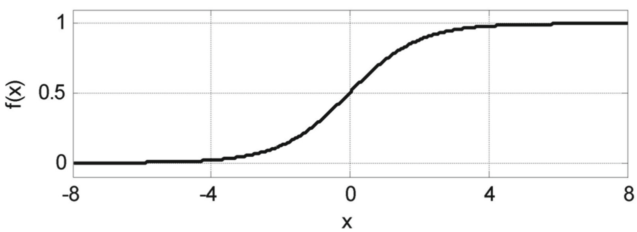


Fig. 1. Sigmoid activation function.

Equation (4) can be expressed in matrix notation:

$$p_{jr}(x) = A_j X^T, A_j = [\, a_{jr} \ldots a_{j1} \; a_{j0} \,], X = [\, x^r \ldots x \; 1 \,]. \tag{5}$$

In PWL approximation technique, $r = 1$ (line), i.e., the activation function is approximated through a set of linear functions. The approximation error depends on the value of r (the order of the polynomials) and how the function is segmented, i.e., the location of the knots [13].

2.3 Genetic Algorithm

A genetic algorithm (GA) is a heuristic technique, based on natural selection, used to find a solution that reduces a fitness function. The fitness function allows evaluating the performance of a possible solution. Initially, a set (population) of n_p possible solutions (individuals) is generated. An individual is composed by chromosomes, i.e., assumptions about the solution parameters. The individuals of the initial population are generated randomly [14]. The GA uses three iterative procedures to create a new population (new possible solutions):

- Selection: the objective of this procedure is to select individuals as parents of the next population. Usually, the roulette technique is used for the selection: the lower the fitness function of an individual, the higher its probability to be selected as parent. However, some individuals with high fitness function are selected to avoid that the GA converges to a local minimum.
- Crossover: parents are mated in pairs to generate children (new solutions) that will create a new population. The crossing ratio $c_r \in [0, 1]$ defines the fraction of the new

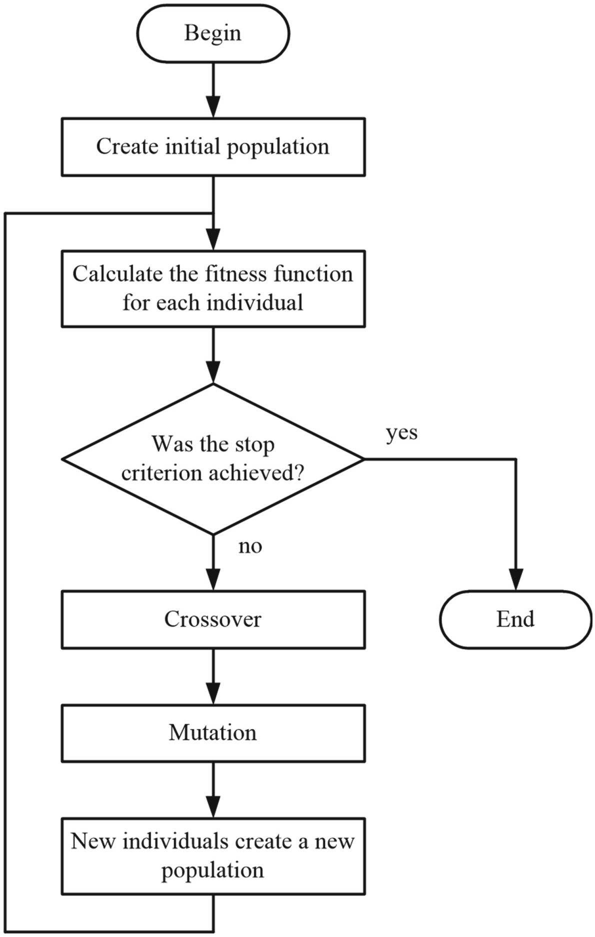


Fig. 2. Structure of a Genetic Algorithm (GA).

children that are obtained through crossover: the combination of the chromosomes of the parents. The other children of the new generation will be a copy of their parents.

- Mutation: the chromosomes of the children are randomly changed (mutated). The mutation ratio $m_r \in [0, 1]$ defines the probability of a children to be mutated. Mutation is applied to get new children, i.e., new possible solutions.

These three procedures are repeated until a stop criterion is achieved, e.g., the fitness function of an individual has an adequate value, or when the number of generations is achieved. Figure 2 illustrates the structure of a GA [14].

3 Proposed Approximation Approach

The approximation error in SPLINE depends on the function segmentation process, i.e., on the number and location of knots. Let consider that the activation function is divided into n segments. As $k_0 = 0$ and $k_n = 8$ in the sigmoid function input range, a GA is applied in this work to set the values of the knots $k_1, k_2, \ldots, k_{n-1}$ that minimize the approximation error. Thus, each individual V (possible solution) of the GA population is composed by the knots $k_1, k_2, \ldots, k_{n-1}$(chromosomes):

$$V = \left[k_1 \; k_2 \; \ldots \; k_{n-1} \right]. \tag{6}$$

Observe that $k_{i-1} < k_i$. To satisfy the aforementioned condition, the chromosomes in each individual are sorted in increasing order before the estimation of the fitness function.

The initial population is composed by random numbers. However, the function approximation of a small interval means that the approximation of the other segments will be done considering greater segments, increasing the approximation error. Hence, a minimum distance between the knots is defined:

$$k_j - k_{j-1} > b > 0, \;\; for \; j = 1, \ldots, n \tag{7}$$

where b is the minimum distance between two consecutive knots. Thus, the initial population is created by generating vectors of random numbers until getting n_p individuals that satisfy (7). In this paper $b = 0.5$.

The crossover and mutation procedures create new individuals V, i.e., new sets of knots. In the fitness function, the sigmoid function is calculated considering a set of points $x \in [0, 8]$. The knots that compose the individual V are used to divide the interval [0, 8] into n sub-intervals I_j. Least-square polynomial regression is used to estimate the approximation polynomials $p_{jr}(x)$ for each sub-interval. Thus, as $f(0) = 0.5$ and based on (2), the sigmoid approximation function $f_a(x)$ is defined as follows:

$$f_a(x) = \begin{cases} 0.5, x = 0; \\ p_{jr}(x), x \in \left]k_{j-1}, k_j\right], j = 1, \ldots, n; \\ 1, x > 8 \end{cases} \tag{8}$$

$$f_a(-x) = 1 - f_a(x), x > 0 \tag{9}$$

where $k_0 = 0$ and $k_n = 8$ Eq. (9) is applied to get the value of sigmoid function for a negative input $-x < 0$.

The fitness function for the individual V, called $c(V)$, is defined according to the desired error measurement indicator (e.g., maximum absolute error or mean square error). In this paper, the maximum absolute error will be used:

$$e(V) = \max\left[|f(x) - f_a(x)|\right], 0 \leq x \leq 8 \quad (10)$$

4 Results

4.1 Simulations

Simulations were done using MATLAB/SIMULINK. To make a comparison with the approach described in [12], a first simulation test was performed considering $n = 3$ (three segments, two knots k_1 and k_2 to be estimated) and $r = 5$ (SPLINE approximation using fifth-order polynomials). The maximum absolute error was used to measure the approximation error. In [12], $k_1 = 1$ and $k_2 = 3$. In this first test, a 64-bit float number arithmetic was used. The parameters of the GA were $n_p = 50$, $c_r = 0.8$, and $m_r = 0.01$. The knots obtained through GA were $k_1 = 1.487$ and $k_2 = 3.918$. Figures 3 and 4 show the approximation errors for the approach in [12] and for the proposed approach based on GA. The maximum errors are, respectively, 7.41×10^{-5} and 2.27×10^{-5}.

A second simulation test was performed, also considering 64-bit float number arithmetic, $n_p = 50$, $c_r = 0.8$, $m_r = 0.01$, and $n = 3$, but making $r = 4$ (approximation using fourth-order polynomials). The knots obtained through GA were $k_1 = 1.777$ and $k_2 = 4.772$. Note that the segmentation defined by the GA changes according to the available computational resources. Figures 5 and 6 show the approximation errors for the approach in [12] and for the proposed approach based on GA. The maximum errors are, respectively, 4.8×10^{-4} and 4.91×10^{-5}.

Observe that the maximum error obtained through the proposed approach for $n = 3$ and $r = 4$ is less than the error obtained with the knows in [12] for $r = 5$. That means that the proposed approximation technique allows getting a better approximation using a lower computational cost.

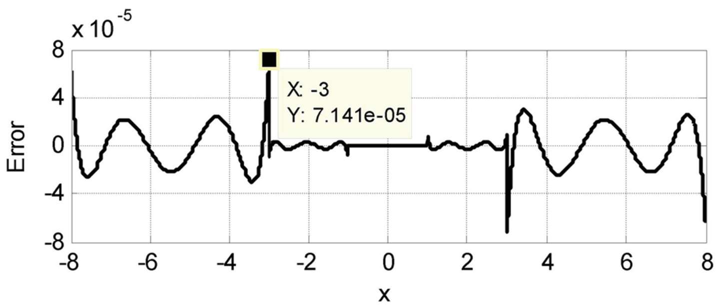


Fig. 3. Approximation error for the approach in [12] ($r = 5$, $n = 3$).

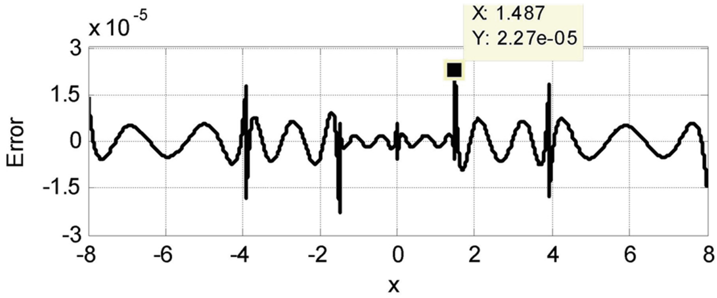


Fig. 4. Approximation error for the proposed approach ($r = 5$, $n = 3$).

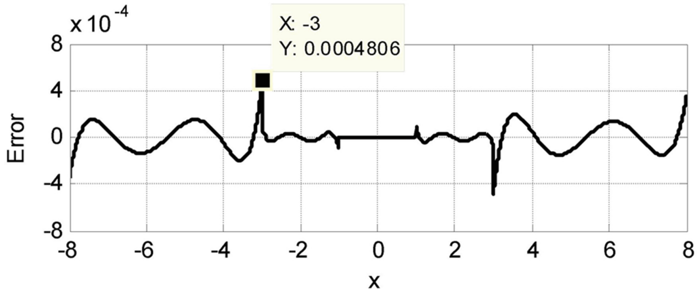


Fig. 5. Approximation error for the approach in [12] ($r = 4$, $n = 3$).

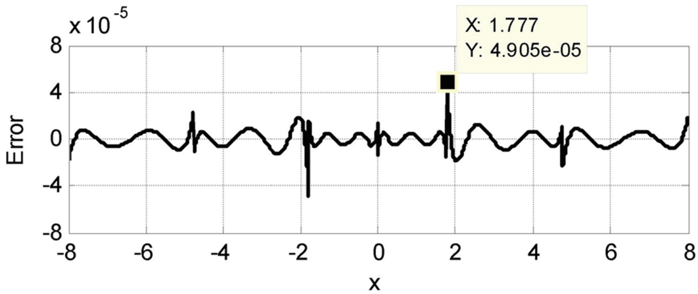


Fig. 6. Approximation error for the proposed approach ($r = 4$, $n = 3$).

4.2 Experimental Results

The proposed sigmoid function approximation approach was implemented in the FPGA EP4CE115F29C7, through the Development and Educational Board DE2115 of ALTERA. In order to send the function input to the FPGA, and to get the value of the approximation function, the Hardware in-the-Loop (HIL) approach was used: the main algorithm (the proposed approximation technique) is implemented in a digital processor (FPGA), while a PC executes the testing algorithm [15]. The testing algorithm, developed in SIMULINK, sends the function input to the FPGA, gets the estimated value of the sigmoid function calculated by the FPGA and makes a comparison between the actual and the estimated sigmoid function. The communication between the PC and the FPGA is done through an Ethernet cable. Figure 7 shows the experimental setup.

Based on the simulation results, the approximation function using $n = 3$ and $r = 4$ was implemented in the FPGA to get an approximation with less computational cost and higher accuracy than in [12]. The input is represented as a 21-bit fixed-point number with 16 fractional bits. The estimated sigmoid function will be represented as 18-bit fixed-point number with 16 fractional bits. A 32-bit fixed point arithmetic was used for the internal arithmetic operations done in the FPGA. The knots obtained through GA were $k_1 = 1.747$ and $k_2 = 4.144$. Figure 8 shows the error using $k_1 = 1$ and $k_2 = 3$(the knots used in [12]), while Fig. 9 shows the experimental result for the proposed approach. The proposed approach allows obtaining a less approximation error. Those results also state that the value of the knots should depend on the available hardware resources.

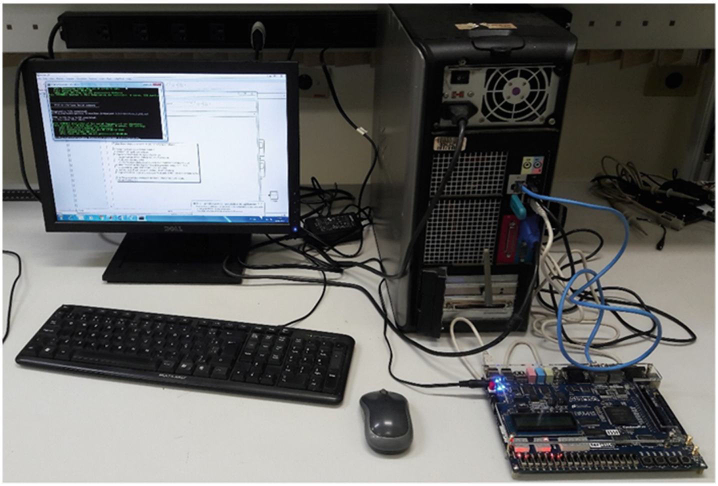


Fig. 7. Experimental setup.

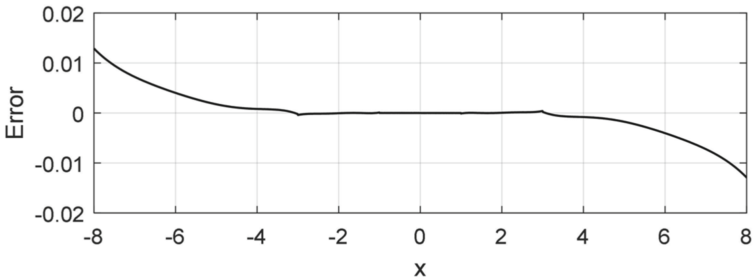


Fig. 8. Experimental approximation error for the approach in [12] ($r = 4$, $n = 3$).

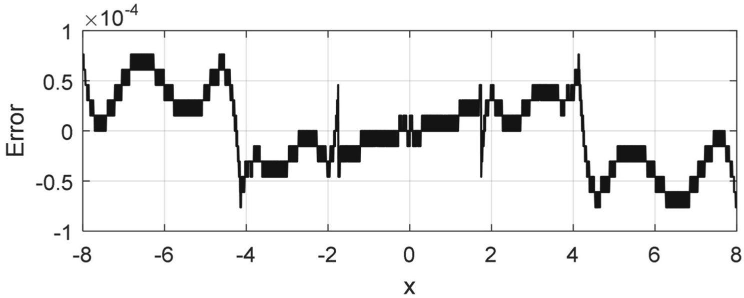


Fig. 9. Experimental approximation error for the proposed approach ($r = 4$, $n = 3$).

5 Summary and Conclusion

This paper proposes an efficient implementation of the activation functions used in ANNs based on SPLINE technique and GA. The proposed approximation technique considers the available computational resources, and allows a better approximation while using a lower computational cost than other approaches. The proposed approach can be easily adapted to consider other activation functions, any number of knots or the use of approximation polynomials with different orders. Other error measurement techniques, such as the minimization of the mean square approximation error can be also used in the fitness function to define the accuracy of the SPLINE approximation. The proposed approach can be easily adapted to be used in other approximation techniques such as PWL.

Acknowledgments. Authors want to thank the BATLAB Laboratory and the Graduation Program in Electrical Engineering of Federal University of Mato Grosso do Sul – UFMS, for the support to this research.

References

1. Fan X, Zhang S, Gemmeke T (2020) Approximation of transcendental functions with guaranteed algorithmic QoS by multilayer Pareto optimization. IEEE Trans Very Large Scale Integr (VLSI) Syst 28:2495–2508
2. Obla S, Gong X, Aloufi A, Hu P, Takabi D (2020) Effective activation functions for homomorphic evaluation of deep neural networks. IEEE Access 8:153098–153112
3. Dong H et al. (2020) PLAC: Piecewise linear approximation computation for all nonlinear unary functions. IEEE Trans Very Large Scale Integr (VLSI) Syst 28:2014–2027
4. Chandwani A, Dey S, Mallik A (2020) Cybersecurity of onboard charging systems for electric vehicles—review, challenges and countermeasures. IEEE Access 8:226982–226998
5. Li Y, Yin G, Zhuang W, Zhang N, Wang J, Geng K (2020) Compensating delays and noises in motion control of autonomous electric vehicles by using deep learning and unscented Kalman predictor. IEEE Trans Syst Man Cyber Syst 50:4326–4338
6. Ortega-Zamorano F, Jerez JM, Muñoz DU, Luque-Baena RM, Franco L (2016) Efficient implementation of the backpropagation algorithm in FPGAs and microcontrollers. IEEE Trans Neural Netw Learn Syst 27:1840–1850

7. Sai CH, Chih YT, Wong WH, Lee CY (2015) A hardware-efficient sigmoid function with adjustable precision for a neural network system. IEEE Trans Circ Syst II Exp Briefs 62:1073–1077
8. Youssefi B, Leigh AJ, Mirhassani M, Wu J (2019) Tunable neuron with PWL approximation based on the minimum operator. IEEE Trans Circ Syst II Exp Briefs 66:387–391
9. Qin Z et al (2020) A novel approximation methodology and its efficient VSLI implementation for the sigmoid function. IEEE Trans Circuits Syst, II, Exp Briefs 67:3422–3426
10. Sun H et al (2020) A universal method of linear approximation with controllable error for the efficient implementation of transcendental functions. IEEE Trans Circ Syst I Reg Papers 67:177–188
11. Zamanlooy B, Mirhassani M (2014) Efficient VLSI implementation of neural networks with hyperbolic tangent activation function. IEEE Trans Very Large Scale Integr (VLSI) Syst 22:39–48
12. Soares AM, Pinto JOP, Bose BK, Leite LC, da Silva LEB, Romero M. Field Programmable Gate Array (FPGA) based neural network implementation of stator flux oriented vector control of induction motor drive. In: Proceedings of 2006 IEEE International Conference on Industrial Technology (ISIE), Mumbai, India, 15–17 December 2006, 31–34
13. Jupp DLB (1978) Approximation to data by SPLINES with free knots. SIAM J Numer Anal 15:328–343
14. Goldberg DE (1989) Genetic Algorithm in Search Optimization & Machine Learning. Addison-Wesley Longman Publishing Co., Boston
15. Rafiei M, Khooban MH, Igder MA, Boudjadar J (2019) A novel approach to overcome the limitations of reliability centered maintenance implementation on the smart grid distance protection system. IEEE Trans Circuits Syst II Exp Briefs 67:320–324

Application of Frequency Division Multiplexing and Neural Networks in the Operation and Diagnosis of the Stator Current and Shaft Position Sensors Used in Electric/Hybrid Vehicles

Raymundo Cordero[1(✉)], Polynne Modesto[1], Thyago Estrabis[1], and João Onofre[2]

[1] Cidade Universitária, Campo Grande, MS 79070-900, Brazil
raymundo.garcia@ufms.br
[2] Oak Ridge National Laboratory - ORNL, Oak Ridge, USA
pintoj@ornl.gov

Abstract. Fast, precise and robust sensing of currents and motor shaft angle is essential for the excellent performance of electric and hybrid vehicles (EV/HEV). Multiplexing techniques are commonly applied in data acquisition systems (DAQs) to digitize the signals sensed in EV/HEV drives. Frequency-division multiplexing (FDM) applied to get the signals from current sensors and resolver angular position sensor has advantages over conventional multiplexing approaches. However, problems such as aging and mechanical imperfections distort the outputs of those sensors, producing measurement errors of the angular position and currents. Conventional techniques designed to compensate for those errors cannot be applied in signals multiplexed in frequency. This paper proposes online techniques to detect and compensate for the distortions in the resolver sensor and current sensors. The demultiplexing process was adjusted to allow distortion detection and compensation. An auto-associative neural network (ANN) compensates for the current measurement error, while an energy-based technique is applied to compensate for the distortions in the resolver outputs. The obtained results show that the distortions were compensated, allowing a more accurate estimation of stator currents and angular position when FDM is applied in EV/HEV DAQs.

1 Introduction

Electric and hybrid vehicles (EV/HEV) are promoted to reduce CO2 emissions and improve energy consumption efficiency in transportation [1, 2]. The propulsion system, usually based on three-phase motors, is one of the main parts of an EV/HEV that defines the vehicle performance and safety operation [3]. Precise, fast and robust current and angular position/speed sensing is required for efficient operation of the propulsion system [3, 4]. Measurement errors of currents or angular position may increase the torque ripple (causing mechanical problems), the EV/HEV energy consumption and affect the speed control performance [4–6].

Resolver sensor is widely applied in EV/HEVs as angular position transducer as it has more accuracy and robustness than encoders in harsh conditions [7–9]. This sensor produces two high-frequency voltages. Besides, two stator currents are sensed in

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 443–454, 2022.
https://doi.org/10.1007/978-3-030-96794-9_41

modern motor drives [10]. Sensing the third motor stator current increases the current measurement robustness through signal redundancy [11]. Analog-to-digital converters (ADCs) with a multiplexing technique are commonly applied to acquire those signals and reduce the complexity and cost of the EV/HEV data acquisition system (DAQ). Most multiplexed ADCs apply time-division multiplexing (TDM): sensor outputs are sent sequentially to a single ADC through an analog switch [12]. However, TDM has drawbacks: the sampling rate for the signal is inversely proportional to the number of switch inputs (channels), sampling synchronization errors, and crosstalk (interference between channels) [13]. Frequency-division multiplexing (or frequency domain multiplexing, FDM) was proposed to overcome those problems when sensing resolver outputs and stator currents. A resolver output and a current signal can be multiplexed using an analog adder if the motor inverter switching frequency is twice the frequency of the excitation voltage that feeds the resolver [14, 15]. FDM allows a faster sampling of each signal than TDM. Besides, FDM replaces the analog switch used in TDM with a simple analog adder.

Problems in the resolver, current sensors, and the conditioning circuits (e.g., aging, component tolerance, amplifier imbalances, and resolver mechanical problems) may distort the outputs from the current sensor and resolver, producing measurement errors of the angular position/speed and the stator currents [16–19]. Techniques were proposed to compensate for those distortions [16–19]. However, those techniques cannot be directly applied when FDM is applied in DAQs as signals from the resolver and current sensors continuously interfere each other in FDM.

For that reason, this paper proposes online techniques to compensate for the incorrect measurement of stator current and angular position due to sensor faults when FDM is applied. The demultiplexing process was adjusted to reduce the effect of the multiplexing process in fault detection and compensation. An auto-associative neural network (AANN) is applied to compensate for the incorrect current measurement. In contrast, an energy-based technique is applied to compensate for the incorrect measurement of the resolver signals. The results obtained in this paper prove that the faults were detected and compensated, reducing the estimation error of the stator currents and the angular position.

2 Acquisition of Current Signals and Resolver Outputs Through FDM

Figure 1 shows the schematics of the resolver sensor. This sensor has an excitation winding and two perpendicular output windings. Equation (1) defines the high-frequency voltage $\upsilon_e(t)$ sent to the excitation winding. As a result, two signals ($\upsilon_s(t)$ and $\upsilon_c(t)$) are induced in the output windings [7–9]:

$$\upsilon_e(t) = a_e \cos(\omega_r t), \tag{1}$$

$$\upsilon_s(t) = k_e \sin(\theta) \upsilon_e(t), \tag{2}$$

$$\upsilon_c(t) = k_e \cos(\theta) \upsilon_e(t), \tag{3}$$

where t denotes time, a_e is the excitation amplitude, $\omega_r = 2\pi f_r$, f_r is the resolver excitation frequency, k_e is a transformation ratio, and θ is the angular position. Resolver outputs are high-frequency signals with narrow bandwidth [13]. On the other hand, when a PWM technique with a fixed switching frequency (f_{sw}) is used to control the motor inverter, the stator current spectrum is located at the fundamental frequency and at integer multiples of the switching frequency [13–15].

The application of FDM in the motor drives data acquisition system is shown in Fig. 2. Let $s_{ia}(t)$, $s_{ib}(t)$, $s_{vs}(t)$ and $s_{vc}(t)$ represent the current signals ($i_a(t)$, $i_b(t)$) and the resolver outputs after the signal conditioning stage. The resolver outputs and the stator currents have non-overlapping spectra if the resolver excitation frequency has a value between the fundamental current frequency and the inverter switching frequency [13]. Hence, according to FDM theory, a resolver output and a current signal can be multiplexed by combining them through an analog adder. Thus, the signals $s_1(t)$ and $s_2(t)$ are generated [13]:

$$s_1(t) = s_{ia}(t) + s_{vs}(t), \tag{4}$$

$$s_2(t) = s_{ib}(t) + s_{vc}(t). \tag{5}$$

The signals $s_1(t)$ and $s_2(t)$ are sent to the DAQ. Techniques to get the current signals and the angular position from $s_1(t)$ and $s_2(t)$ are found in [13–15].

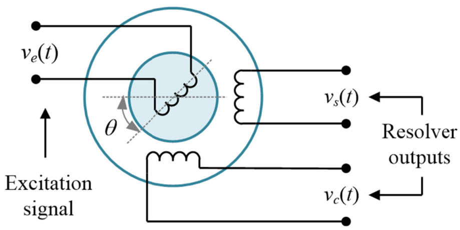


Fig. 1. Schematics of a resolver sensor.

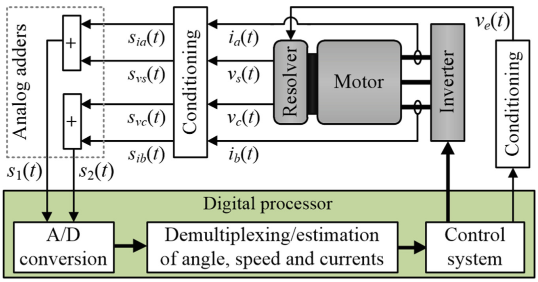


Fig. 2. Application of FDM in the acquisition of stator currents and resolver signals.

3 Distortions in the Signals from Current Sensors and Resolver

Problems in the resolver sensor, the current sensors and the conditioning circuits (e.g., aging, temperature, gain imbalance and mechanical imperfections) distort the outputs of the resolver and the current sensors. Those distortions produce measurement errors of the angular position and the stator currents. The distorted outputs from resolver and the current sensors can be modeled as follows [16–19]:

$$s_{\upsilon s}(t) = (1+\alpha)k_e \sin(\theta)\upsilon_e(t) + o_{vs}, \tag{6}$$

$$s_{\upsilon c}(t) = k_e \cos(\theta+\beta)\upsilon_e(t) + o_{vc}, \tag{7}$$

$$s_{ia}(t) = k_a i_a(t) + o_a, \tag{8}$$

$$s_{ib}(t) = k_b i_b(t) + o_b, \tag{9}$$

$$s_{ic}(t) = k_c i_c(t) + o_c, \tag{10}$$

where $s_{ic}(t)$ represents the stator current $i_c(t)$ after the conditioning stage (acquired to get data redundance), α is the amplitude imbalance, β is the imperfect quadrant angle, k_a, k_b and k_c are gains, while o_a, o_b, o_c, o_{vs} and o_{vc} are offsets. Those distortions produce measurement errors of the angular position and the stator currents. Replacing (6), (7), (8) and (9) into (4) and (5), yields:

$$s_1(t) = (1+\alpha)k_e \sin(\theta)\upsilon_e(t) + k_a i_a(t) + o_1, o_1 = o_{vs} + o_a, \tag{11}$$

$$s_2(t) = k_e \cos(\theta+\beta)\upsilon_e(t) + k_b i_b(t) + o_2, o_2 = o_{vc} + o_b. \tag{12}$$

Equations (11) and (12) model the signals sent to the DAQ. Note that the offsets o_1 and o_2 can be considered as DC current signals. The techniques in [14, 15] reject the current signal interference in the angle estimation. Hence, o_1 and o_2 mainly affect the stator current measurement.

4 Proposed Estimation of Stator Currents and Angular Position with Distortion Compensation

The proposed approach compensates the distortions in the sensor signals and estimates the fundamental component of stator currents and the angular position based on the signals in (10), (11), and (12). Many motor control techniques use only the fundamental current components as current harmonics in the control loops may affect the drive performance [14]. It is considered that the three stator currents of the three-phase motor are sensed to get data redundancy, while a PWM technique with a fixed switching frequency (f_{sw}) is applied to control the motor.

4.1 Estimation of the Stator Currents

Figure 3 shows the current demultiplexing process applied in this work, which is based on [14]. Let t_i be the i-th instant of time when the PWM carrier reaches its valleys. The PWM carrier, the DAQ and the resolver excitation signal are synchronized so that resolver outputs are zero when the PWM reaches its valleys, i.e., $v_e(t_i) = 0$. Making $t = t_i$ and $v_e(t_i) = 0$ into (10), (11) and (12), yields:

$$s_1(t_i) = k_a i_a(t_i) + o_1, s_2(t_i) = k_b i_b(t_i) + o_2, s_{ic}(t_i) = k_c i_c(t_i) + o_c. \quad (13)$$

This synchronization requires that the inverter switching frequency is twice the resolver excitation frequency: $f_{sw} = 2f_e$ [15]. According to the synchronous sampling theory, the current samples taken at the PWM valleys, i.e., $i_a(t_i)$, $i_b(t_i)$ and $i_c(t_i)$, represent the fundamental component of the stator currents [14].

The offsets o_1, o_2 and o_c can be considered as constants or having a slow variation. Therefore, an offset is a low-frequency signal respect to the resolver outputs and the current signals. Thus, the estimated offsets o_{1e}, o_{2e} and o_{ce} can be obtained by applying a low-pass filter into $s_1(t_i)$, $s_2(t_i)$ and $s_{ic}(t_i)$ when the motor operates at steady-state speed operation:

$$o_{1e} = LPF\{s_1(t_i)\}, o_{2e} = LPF\{s_2(t_i)\}, o_{ce} = LPF\{s_{ic}(t_i)\} \quad (14)$$

where $LPF\{\}$ denotes the use of a low-pass filter. Note that the offset estimation technique in (14) has more accuracy when it is applied before the motor start-up, i.e., when the motor stator currents are zero ($i_a(t_i) = i_b(t_i) = i_c(t_i) = 0$). The time interval between the DAQ activation and the motor start-up can be exploited to estimate the offset through (14). From (13) and (14), it is possible to deduce (15):

$$k_a i_a(t_i) \approx s_1(t_i) - o_{1e}, k_b i_b(t_i) \approx s_2(t_i) - o_{2e}, k_c i_c(t_i) \approx s_{ic}(t_i) - o_{ce}. \quad (15)$$

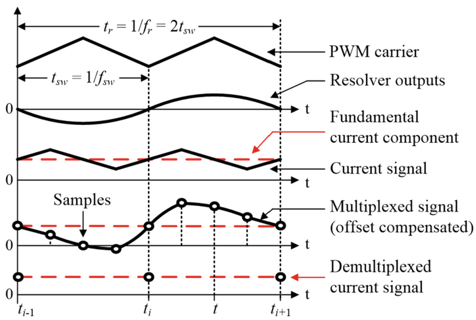


Fig. 3. Current demultiplexing based on synchronous current sampling.

This paper applies an auto-associative neural network (AANN) to compensate the current measurement errors and estimate the fundamental components of the stator currents. AANN is trained to model the relationship between inputs and outputs signals.

The inputs must have a relationship between them, like the three-phase motor stator currents. AANN is composed by three hidden layers: mapping layer, bottleneck, and de-mapping layer, as shown in Fig. 4. The bottleneck layer has the smallest number of neurons, while both mapping and de-mapping layers have more neurons than the input and output layers. Differently to [19], where the AANN estimates the three-phase currents, the AANN in this paper uses the signals $k_a i_a(t_i)$, $k_b i_b(t_i)$ and $k_c i_c(t_i)$ obtained in (15) to estimate the fundamental components of the stationary dq stator currents $i_d(t_i) = [2i_a(t_i) - i_b(t_i) + i_c(t_i)]/3$ and $i_q(t_i) = [i_b(t_i) - i_c(t_i)]/\sqrt{3}$ used in vector control [20]. The AANN training data are the expected values of $k_a i_a(t_i)$, $k_b i_b(t_i)$, $k_c i_c(t_i)$, $i_d(t_i)$ and $i_q(t_i)$.

4.2 Estimation of the Angular Position

Let $x_1(t) = [s_1(t) - s_1(t_i)]\upsilon_e(t)$ and $x_2(t) = [s_2(t) - s_2(t_i)]\upsilon_e(t)$, where $s_1(t)$, $s_2(t)$, $s_1(t_i)$ and $s_2(t_i)$ are defined in (11), (12) and (13). Applying trigonometrical properties, it is possible to prove that:

$$x_1(t) = 0.5(1+\alpha)k_e \sin(\theta) a_e[1+\cos(2\omega_r t)] + r_1(t)\upsilon_e(t), \tag{16}$$

$$x_2(t) = 0.5k_e \cos(\theta+\beta) a_e[1+\cos(2\omega_r t)] + r_2(t)\upsilon_e(t), \tag{17}$$

where $r_1(t) = k_a i_a(t) - k_a i_a(t_i)$ and $r_2(t) = k_b i_b(t) - k_b i_b(t_i)$. Frequency-domain analysis allows proving that $\cos(2\omega_r t)$, $r_1(t)\upsilon_e(t)$ and $r_2(t)\upsilon_e(t)$ are high-frequency signals [15]. Thus, the signals $x_{1f}(t)$ and $x_{2f}(t)$ are obtained by applying a low-pass filter into $x_1(t)$ and $x_2(t)$:

$$x_{1f}(t) \approx 0.5(1+\alpha)a_e k_e \sin(\theta+\gamma), \tag{18}$$

$$x_{2f}(t) \approx 0.5 a_e k_e \cos(\theta+\beta+\gamma), \tag{19}$$

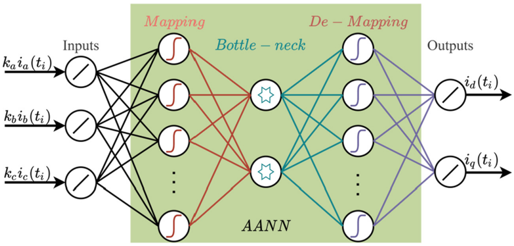


Fig. 4. Structure of an auto associative neural network (AANN).

where γ is the filter delay. Both signals $x_{1f}(t)$ and $x_{2f}(t)$ have the same period (T_x). The energies of $x_{1f}(t)$ and $x_{2f}(t)$ (denoted as $E_1(t)$ and $E_2(t)$, respectively) calculated in the interval when both signals are positive ($[0, T_x/4]$) are:

$$E_1(t) = \int_0^{T_x/4} [x_{1f}(t)]^2 dt = g_1[0.5(1+\alpha)a_e k_e]^2, \tag{20}$$

$$E_2(t) = \int_0^{T_x/4} [x_{2f}(t)]^2 dt = g_2[0.5a_e k_e]^2, \tag{21}$$

where g_1 and g_2 depend on T_x. Considering $\beta \approx 0$, then (18) and (19) show that $x_{1f}(t)$ and $x_{2f}(t)$ are in quadrature. Thus, it is possible to prove that $g_1 \approx g_2$. The estimated amplitude imbalance (α_e) is deduced from (20) and (21), as follows:

$$\alpha \approx \alpha_e = \sqrt{E_1(t)/E_2(t)} - 1 \tag{22}$$

The terms $\sin(\theta)$ and $\cos(\theta + \beta)$ can be obtained from (18), (19) and (22):

$$\sin(\theta + \gamma) \approx \frac{x_{1f}(t)}{0.5a_e k_e(1 + \alpha_e)}, \cos(\theta + \beta + \gamma) \approx \frac{x_{2f}(t)}{0.5a_e k_e}. \tag{23}$$

Trigonometrical properties and (23) allow defining $S(t)$ and $C(t)$:

$$S(t) = \sin(\theta + \gamma) - \cos(\theta + \beta + \gamma) = 2\sin\left(\frac{\beta}{2} + \frac{\pi}{4}\right)\sin\left(\theta + \gamma + \frac{\beta}{2} + \frac{\pi}{4}\right), \tag{24}$$

$$C(t) = \sin(\theta + \gamma) + \cos(\theta + \beta + \gamma) = 2\cos\left(\frac{\beta}{2} + \frac{\pi}{4}\right)\cos\left(\theta + \gamma + \frac{\beta}{2} + \frac{\pi}{4}\right). \tag{25}$$

Note that the amplitudes of $S(t)$ and $C(t)$ depend on β. Hence, the energy-based estimation technique in (20) and (21) can be also applied to get β. Let T_y be the interval when both $S(t)$ and $C(t)$ are positive. Thus, the estimated imperfect quadrant angle (β_e) is calculated as follows:

$$R_s(t) = \int_0^{T_y/2} [S(t)]^2 dt \propto \sin^2\left(\frac{\beta}{2} + \frac{\pi}{4}\right), \tag{26}$$

$$R_c(t) = \int_0^{T_y/4} [C(t)]^2 dt \propto \cos^2\left(\frac{\beta}{2} + \frac{\pi}{4}\right), \tag{27}$$

$$tan\left(\frac{\beta}{2} + \frac{\pi}{4}\right) \approx \sqrt{\frac{R_s(t)}{R_c(t)}} \rightarrow \beta \approx \beta_e = 2\left[atan\left(\sqrt{\frac{R_s(t)}{R_c(t)}}\right)\right] - \frac{\pi}{4}. \tag{28}$$

The signals $x_{1f}(t)$ and $x_{2f}(t)$ cannot be used for angle estimation as they are affected by filtering delays. Hence, it is proposed to compensate the resolver signal distortion directly to the signals $s_1(t)$ and $s_2(t)$ send to the DAQ, which are defined by (11) and (12). Equations (29) and (30) define the signals $s_1(t)$ and $s_2(t)$ after compensation, denoted by $s_{1c}(t)$ and $s_{2c}(t)$, respectively:

$$s_{1c}(t) = \frac{s_1(t) - o_{1e}}{1 + \alpha_e} \approx k_e \sin(\theta) \upsilon_e(t) + \frac{k_a i_a(t)}{1 + \alpha_e}, \tag{29}$$

$$s_{2c}(t) = \frac{(s_2(t) - o_{2e}) + \sin(\beta_e) s_{1c}(t)}{\cos(\beta_e)} \tag{30}$$

$$\approx \frac{k_e \cos(\theta+\beta_e)\upsilon_e(t) + k_e \sin(\beta_e)\sin(\theta)\upsilon_e(t)}{\cos(\beta_e)} + \frac{k_b i_b(t)}{\cos(\beta_e)} + \frac{\sin(\beta_e) k_a i_a(t)}{\cos(\beta)(1+\alpha_e)}.$$

As $\cos(\theta+\beta_e) = \cos(\theta)\cos(\beta_e) - \sin(\theta)\sin(\beta_e)$, it is possible to prove that:

$$s_{2c}(t) \approx k_e \cos(\theta)\upsilon_e(t) + \frac{k_b i_b(t)}{\cos(\beta_e)} + \frac{\tan(\beta) k_a i_a(t)}{(1+\alpha_e)}. \tag{31}$$

Note that $s_{1c}(t)$ and $s_{2c}(t)$ contain the ideal resolver outputs defined in (2) and (3), while $i_a(t)$ and $i_b(t)$ are affected by gains. Thus, $s_{1c}(t)$ and $s_{2c}(t)$ can be used, instead of $s_1(t)$ and $s_2(t)$, by the techniques described in [14, 15] to get the angular position when FDM is applied in motor drive DAQs.

5 Results

Simulations were performed in MATLAB/SIMULINK. The stator currents and position/speed signals belong to an induction motor (5 HP, 1950 RPM) with open-loop V/F control. The motor speed curve is shown in Fig. 5. The simulation parameters are listed in Table 1. The currents signals and resolver outputs were normalized for those signals to operate in the range ± 1 p.u.

The approach in [15] was applied to estimate the angular position. The mapping, bottleneck and de-mapping layers of the AANN has 8, 2 and 4 neurons, respectively. The bottleneck layer uses linear activation function, while the other hidden layers use tansig function.

Table 1. Simulation parameters

Parameter	Value
Resolver excitation amplitude (a_e)	1 (normalized)
Transformation ratio (k_e)	1(normalized)
Resolver excitation frequency (f_e)	2.5 kHz
Inverter switching frequency (f_{sw})	5 kHz
Amplitude imbalance (α)	[−0.1, 0.1] (values based on [18])
Imperfect quadrant angle (β)	[0.1, 0.2] rad (values based on [18])
Gains k_a, k_b, k_c	0.9, 1, 0.9
Offsets o_{vs}, o_{vc}, o_a, o_b, o_c,	0.01 p.u. in all cases

The offsets were estimated using a moving average filter between 0 to 0.1 s (before the motor start-up). Figure 6 shows the estimated fundamental components of the stationary dq stator currents with and without compensation. The current signals estimated by the proposed AANN are closer to the actual currents that the estimated currents without compensation, especially for the case of i_q. Hence, the proposed AANN compensate the distortions in the current signals.

Figure 7 shows the estimated amplitude imbalance and imperfect quadrant angle, while Fig. 8 shows the angle estimation error. The proposed compensation approach for the resolver distortions starts at $t = 0.8$ s to show the effect of the compensation in the angle estimation. The angle estimation error peaks at $t = 1.5$ s and $t = 2.5$ s are produced by the fast increment of the amplitude imbalance and the imperfect quadrant angle, but they are quickly compensated. These results prove that the distortions were properly detected and compensated, allowing an accurate angle estimation.

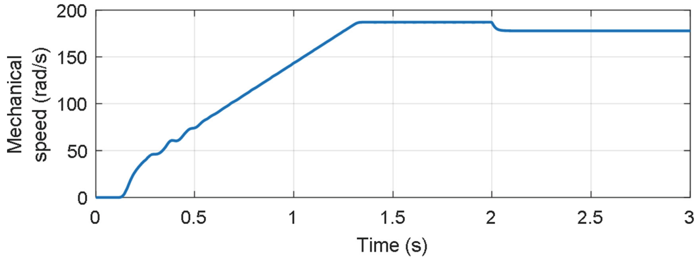


Fig. 5. Motor speed curve used in simulations.

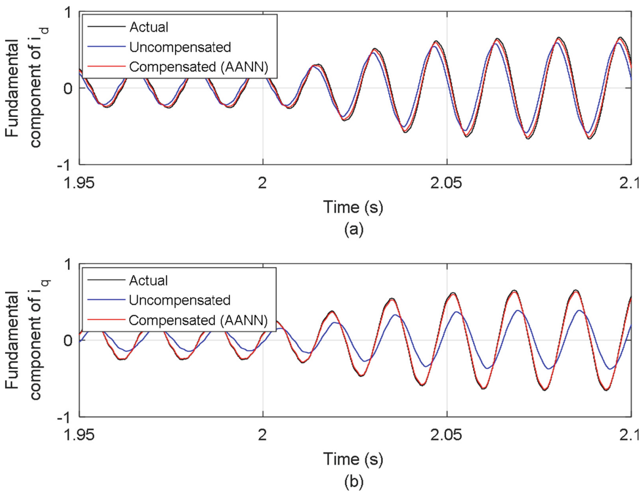


Fig. 6. Estimated fundamental components of the stationary *dq* stator currents with and without compensation. a) Estimation of i_d. b) Estimation of i_q.

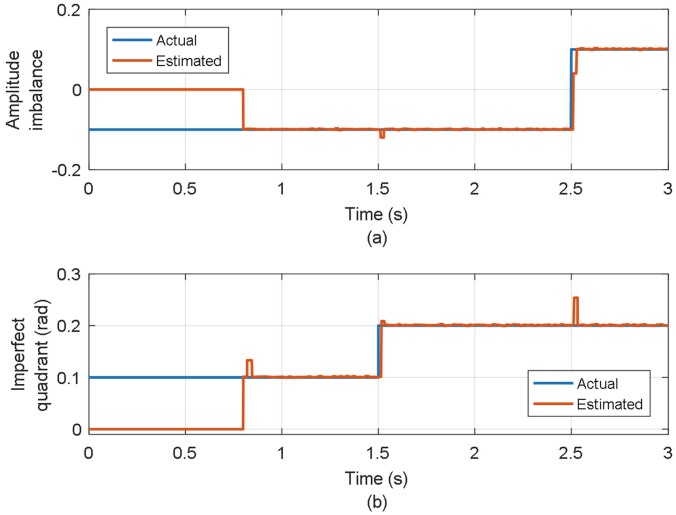


Fig. 7. Estimation of resolver distortions. a) Amplitude imbalance. b) Imperfect quadrant.

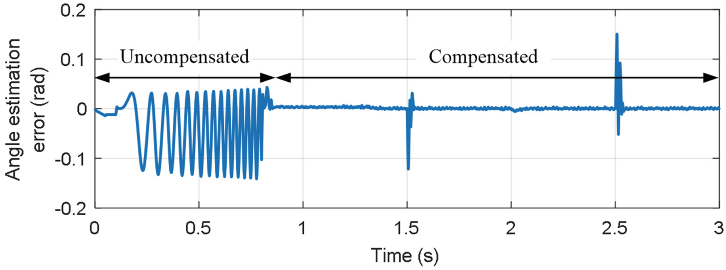


Fig. 8. Angle estimation error with and without distortion compensation.

6 Summary and Conclusion

This paper presents online detection and compensation techniques of the distortions that affect the signals from the resolver sensor and the current sensors when FDM is applied in EV/HEV DAQs. The AANN reduces the measurement error of the fundamental component of the stator currents, while the amplitude imbalance and imperfect quadrant that affects the resolver outputs were properly estimated and compensated. Thus, the fundamental component of the stator currents and the shaft angular position are estimated with accuracy and robustness. As future work, the proposed compensation techniques will be experimentally tested.

Acknowledgments. Authors thank BATLAB Laboratory and the Graduation Program in Electrical Engineering of Federal University of Mato Grosso do Sul for supporting this research.

References

1. Aymen F, Alowaidi M, Bajaj M, Sharma NK, Mishra S, Sharma SK (2021) Electric vehicle model based on multiple recharge system and a particular traction motor conception. IEEE Access 9:49308–49324
2. Wang C, Wang Z, Zhang L, Cao D, Dorrell DG (2021) A vehicle rollover evaluation system based on enabling state and parameter estimation. IEEE Trans Ind Informat 17:4003–4013
3. Kumar K, Santra SB (2018) Performance analysis of a three-phase propulsion inverter for electric vehicles using GaN semiconductor devices. IEEE Trans Ind Appl 54:6247–6257
4. Schubert M, Scharfenstein D, De Doncker RW (2021) A novel online current and voltage sensor offset adaption scheme utilizing the effect of inverter voltage distortion. IEEE Trans Ind Appl 55:6011–6017
5. Mok HS, Kim SH, Cho YH (2007) Reduction of PMSM torque ripple caused by resolver position error. Electron Lett 43:646–647
6. Kim J, Ryu S, Rafaq MS, Choi HH, Jung J (2020) Improved torque ripple minimization technique with enhanced efficiency for surface-mounted PMSM drives. IEEE Access 8:115017–115027
7. Sabatini V, Di Benedetto M, Lidozzi V (2019) Synchronous adaptive resolver-to-digital converter for FPGA-based high-performance control loops. IEEE Trans Instrum Meas 68:3972–3982
8. Bahari M, Davoodi A, Saneie H, Tootoonchian F, Nasiri-Gheidari Z (2020) A new variable reluctance pm-resolver. IEEE Sensors J 20:135–142
9. Bahari M, Nasiri-Gheidari Z, Tootoonchian F (2021) Modeling, performance analyzing, and prototyping of variable reluctance resolver with toroidal winding. IEEE Sensors J 21:4425–4432
10. Idkhajine L, Monmasson E, Naouar MW, Prata A, Boullaga K (2009) Fully integrated FPGA-based controller for synchronous motor drive. IEEE Trans Ind Electron 56:4006–4017
11. Noh Y, Kim W, Lee J (2021) The optimal current ratio control of redundant electric drive systems and diagnostic strategies for disagreement. IEEE Access 9:32115–32130
12. Kumar RSA, Behera D, Krishnapura N (2018) Reset-free memoryless delta-sigma analog-to-digital conversion. IEEE Trans Circuits Syst I, Reg Papers 65:3651–3661
13. García RC, Pinto JOP, Suemitsu WI, Soares JO (2017) Improved demultiplexing algorithm for hardware simplification of sensored vector control through frequency-domain multiplexing. IEEE Trans Ind Electron 64:6538–6548
14. García RC, Pinto JOP, Ono IES, Fahed VDS, Brito M (2018) Simplification of the acquisition system for sensored vector control using resolver sensor based on FDM and current synchronous sampling. In: Proceedings of the IEEE 4th Southern Power Electronics Conference (SPEC), Singapore, 10–13 December 2018
15. García RC, Filho FJT, Pinto JOP, Suemitsu WI, Fonseca BS, Brun ADM (2021) Improved acquisition system for sensored vector control through frequency-domain multiplexing, synchronous sampling, and differential evolution. IEEE Sensors J 21:7784–7792
16. Qamar NA, Hatziadoniu CJ, Wang H (2015) Speed error mitigation for a DSP-based resolver-to-digital converter using autotuning filters. IEEE Trans Ind Electron 62:1134–1139
17. Chen S, Zhao Y, Qiu H, Ren X (2020) High-precision rotor position correction strategy for high-speed permanent magnet synchronous motor based on resolver. IEEE Trans Power Electron 35:9716–9726
18. Noori N, Khaburi DA (2016) Diagnosis and compensation of amplitude imbalance, imperfect quadrant and offset in resolver signals. In: Proceedings of Power Electronics and Drive Systems Technologies Conference (PEDSTC), Tehran, Iran, 16–18 February 2016, pp 76–81

19. Galotto L et al (2007) Auto-associative neural network based sensor drift compensation in indirect vector controlled drive system. In: Proceeding of 33rd Annual Conference of the IEEE Industrial Electronics Society (IECON), Taipei, Taiwan, 5–8 November 2007, pp 1009–1014
20. Bose K (2001) Modern power electronics and AC drives. Prentice-Hall, Upper Saddle River

Machine Learning Based Prediction of Fatigue Events in Railway Rails

Vincent Laurent[1,2(✉)], Olivier Vo Van[3], Mathilde Mougeot[1], and Jean-Michel Ghidaglia[1,2]

[1] Centre Borelli, 4 Avenue des Sciences, 91190 Gif-sur-Yvette, France
[2] Eurobios Scientific Computing Branch, 61 Avenue du Président Wilson, 94230 Cachan, France
vlaurent@eurobios.com
[3] SNCF Innovation and Research Department, 3 Avenue François Mitterrand, 93210 St Denis, France

Abstract. In this paper we present a study that tackles the health monitoring problem of rolling contact fatigue between train wheels and railway rails. This study is focused on model exploration, and explainable Machine Learning, with the objective of predicting defect apparition.

1 Introduction

Rolling Contact Fatigue (RCF) cracks can appear at the surface or subsurface of rails due to the repeated passage of railway train wheels. These defects, such as squats and head-checks [1], can propagate and lead to rail fracture. The occurrence of rail fractures could have economic (delayed trains, maintenance costs) and passenger safety (derailment) implications. To avoid rail fractures, the French National Railway Company (SNCF) applies a strict method with both corrective and scheduled maintenance strategies which have a high cost for the company. Many mechanical studies have been carried out on RCF for initiation [2] and propagation [3] of cracks at a local scale. These models are computationally costly, which makes their direct use for maintenance difficult, meaning that less refined degradation models have been developed [4]. Moreover, models based on mechanical simulations cannot consider exogenous phenomena. The aim of this paper is to train data driven models to better predict crack initiation and then to highlight the observed phenomena that are not yet explained by mechanical models.

Related Work. For data studies related to Railway Transportation Systems the reader can refer to the review of [5]. Many studies have attempted to model track geometry defects using Markov Chains [6], and in many cases, this is used as a method to plan maintenance [7, 8]. In the latter, geometric defect apparition is modelled by Random Forest. Many studies demonstrate promising application of Artificial Neural Network to rail maintenance to predict track degradation [6], weld defects [9] or geometry defects [7]. To model crack appearance, our approach uses survival methods as in [10], where frailty models are used to predict recurrent rail defects, and the point of view provided is relevant in our context. Indeed, some cracks can be removed and then reappear on the

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 455–466, 2022.
https://doi.org/10.1007/978-3-030-96794-9_42

same rail segment due to grinding. As far as fatigue is concerned, the contribution of exogenous phenomena is expected to be difficult to capture, and this study focuses on using Machine Learning algorithms that can properly discriminate defective and healthy rail segments, and also on benchmarking and validating these models. Refined models that consider recurrence of events or monitoring will be tackled in further studies.

Key Contributions. For the first time, consequential work is presented on the analysis of SNCF crack appearance data due to RCF, and this has led to a creation of an associated modelling approach. The originality in the present work is the study of two complementary approaches; the first consists of a binary diagnosis based on independent and identically distributed observations and the second is a statistical approach which is aimed at modelling the survival probability of each rail segment. These approaches are put into perspective to answer operational team issues. For both approaches, the use of tree-based methods is justified by their ability to explain results and to compute complex interactions between covariates. The interest of using such models is demonstrated by the quality of the results they yield, and some modern methods that give insights into the predictions are provided.

2 Data Aggregation and Processing

This section presents the aggregation of the learning data. To assure data maintenance and replicability of the project, all the data is managed using a PostgreSQL database that keeps track of the different data processing steps. The input data provides a description of the network divided into segments that are indexed by two kilometer-points, for a given line number (e.g. Lyon-Paris), a track (e.g. first, second track) and an indication of the left or right rail. We use a data processing that is similar to [10]. The originality in this case is that both continuous and discrete indexation of the data is dealt with, respectively, for the kilometer points and the triplet (line, track, rail track). This implies that common database operations, such as join, are adapted to our study without loss of efficiency.

Processing Steps. The main steps for data processing are:

1. Import all data from external sources
2. Preprocessing: reformat data to store information optimally (remove redundancy, interpret data type, locate rail segment where description is constant) and various data corrections.
3. Aggregate all data using a master table, this step generates approximately 1 million segments that provide a minimalist description of the French network (see Fig. 1). This makes a dynamic description of the network possible that considers replacement operations: a description of the state of the French network at each time in its life can be provided.
4. Feature engineering: recode variables (interior or exterior rail for curved sections for instance), extrapolate tonnage with linear assumption, etc.

5. Build learning base composed of a fixed length rail segment and add target observations independently.

The data processing pipeline is optimized to take only several minutes of computation time on an ordinary PC.

Features. One can make the distinction between historical data and contextual data that does not vary with time. The contextual data helps when describing risky configurations for crack initiation. A macro description of the data used for the current study is provided in the table below. As for the defect analysis, two types of defect are studied: head-checks and squats. Both are types of contact fatigue that are caused by the pressure of the wheel on the rail. They respectively appear on the edge and on the table of the rail. Further studies will focus on describing time varying covariates, such as temperature, relative humidity, tonnage or accelerating and braking statistics.

Data types	Example
Asset data	Curvature, declivity, cant, gauge, steel grade,
Operating data	Ton gross (2017 and 2018), nominal speed…
Maintenance data	Removal and track replacement
Observations and measures	Crack occurrences

Minimal Length for Aggregation. The choice of the length (108 m) at which all data is aggregated is motivated by the regular length of rail which is approximately 36 m. The choice is motivated by the preference to have, on average, the same number of welds on each segment. It was chosen not to describe segments too finely, because it is desirable to capture the effects of having transition segments (end and start of a tunnel, a curve, etc.). The second reason is computational: a length of 108 m already results in 1 million segments.

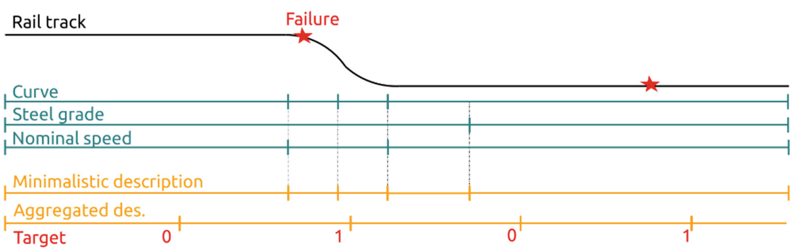


Fig. 1. Illustration of data aggregation.

3 Modelling Rail Defect Appearance

In this section, the core modelling concepts are presented, as well as the benchmarking of several models that guide our analysis.

Learning Data. The dataset is composed of 27 covariates and between 0.8 and 1.5 million observations that describe rail segments. All variables are numerical and missing values are accepted. The size of the dataset mainly depends of the length of segments and the perimeter of the study.

Scoring Setting. In this work two different kinds of problem are considered for the same target. First, the probability of crack appearance is computed with respect to contextual features in order to compute a heatmap, where similar configurations have the same level of risk. This point of view leads us to estimate $\eta(x) = \mathbb{P}(Y = 1|X = x)$; what is of prime importance here is that the estimated value of η, namely $\widehat{\eta}$, renders the proper ordering with respect to the observed event. Receiver Operating Characteristic (ROC) curves are used throughout the paper to characterize the performance of a model and its Area Under Curve (AUC) [11]. The ROC curve (see Fig. 2) is optimal when always equal to 1 in]0, 1], which leads to an AUC of 1. When the ROC curve follows identity in]0, 1], its AUC is equal to 0.5 and the model makes random predictions.

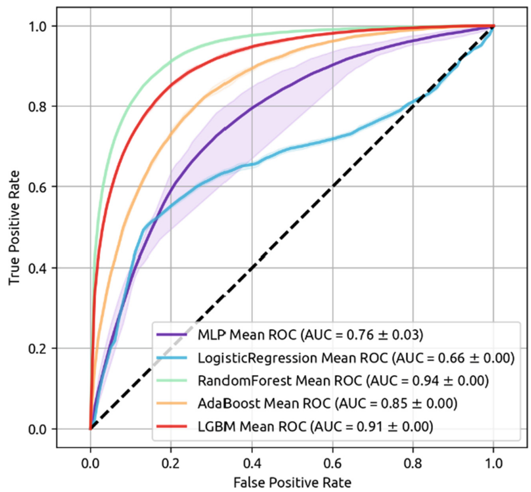


Fig. 2. Benchmark of several simple models. The shades stand for the bundle of ROC curves obtained for each model of the cross-validation.

Model Benchmark for Scoring. To obtain a better understanding of the difficulty of the problem, several algorithms from the literature must first be compared. We thus evaluate performance in Fig. 2 for Multi-Layer Perceptron Classifier (MLP), Logistic Regression (LR), Adaptive Boosting (AdaBoost) [11], Light Gradient Boosting (LGBM) and Random Forest (RF) [12]. The three latter methods seem to be far more promising, even if only default parameters are used, and no tuning is made. For the implementations, Scikit-Learn algorithms [13] are used, except for LGBM for which we refer to the implementation of Microsoft [14] (Table 1).

Table 1. Model configurations and build time, with the same 6-CPU.

Algorithm	Parameters	Timer (min:sec)
MLP	1 hidden layer, 100 neurons per layer	53:44
ADABOOST	10 decision trees as base estimators – 2 max depth	00:51
LGBM	100 trees – 31 leaf size	00:12
RF	20 trees – 15 max depth	01:20
LR	L_1 penalty	03:05

Survival Analysis. The second problem setting is more ambitious. Here, the focus is on evaluating the time T when crack initiation occurs relative to installation date. In this case, the survival function $s(t) = \mathbb{P}(T \geq t)$ is estimated, which is the probability that a defect appears after age t. The core concept of survival analysis is censorship, because the event $\{T \geq t\}$ is not necessarily observed, especially when T exceeds the time window that contains all observations. A positive random variable C is introduced, such that instead of observing T, (T', δ) are known along with $T^{'} = \inf(T, C)$ and $\delta = 1\{T \geq C\}$. In our case, C corresponds to the date of the latest maintenance visit of the rail segment. For observations where $\delta = 1$, a "crack initiation" event is observed, and when $\delta = 0$ the event has not yet occurred. In this work, it is assumed that C and T are independent[1]. Based on data analysis, especially the number of defects detected per year, it appears that the observations are only reliable since 2008. Instead of observing$T$, we observe T if $T > t_o$, where t_o is the age of the rail track in 2008. In survival analysis, this entails using truncated data. The fact that we do not know if diagnosis of a defect could have been made before truncation time drastically lowers the probability of survival as seen in Fig. 3. The estimator represented in this figure is called the Kaplan Meier estimator [16], and is a non-parametric estimate of the survival function s:

$$\hat{s}(t) = \prod_{t_i < t} \frac{n_i - d_i}{n_i}$$

where n_i are the instances at risk and d_i are the deceased just before time t_i. In this research, the relevance of many features is often questioned, and the approach is oriented toward tree-based methods that automatically carry out feature selection [13]. Many survival models like COX regression or Weibull Accelerated Failure Time (AFT) model assume a linearity relation within covariates, while previous tree-based models are well suited to extremely large datasets and noisy observations. Building on the scoring benchmark, the best lead for modelling seems to be Left Truncated Right Censored Trees

[1] For a more general approach one could look at competition risk methods [15]

(LTRC-Trees) [17], which are an adapted version of CART for survival modelling. Note that considering left truncation is compelling: to our knowledge, many non-parametric survival models only study censored data and not truncation, like Random Survival Forest [18] or survival regression with an AFT model with XGBoost [19].

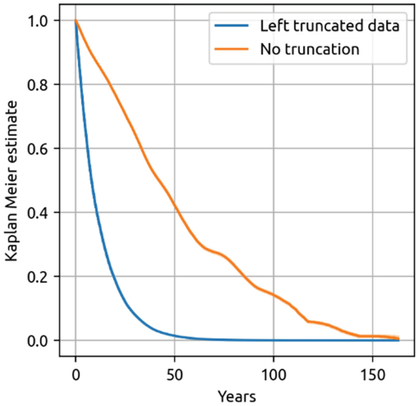


Fig. 3. Effect of left truncation on Kaplan Meier estimate

4 Validation and Business Value Assessment

In Machine Learning, the test procedure used to validate the model is of prime importance more than identifying the best models and parameters to fit the data. The following paragraph is dedicated to this question. In this section, results of the models are presented, as well as the method used to interpret them.

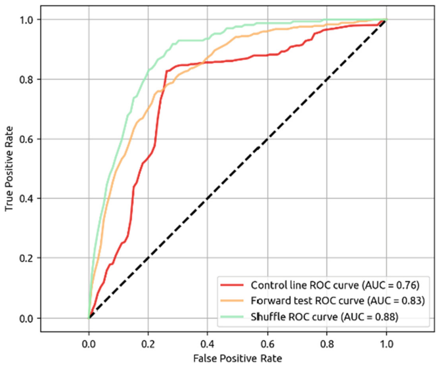


Fig. 4. Validation method applied to head-check scoring

Validation Methodologies. Several methods are identified to test which are complementary to one other. One widely used method in the Machine Learning field is K-fold cross-validation, which consists of separating training data into K samples and then using one sample as a test and the others for training. This entails running K models. This method provides several advantages, such as using all training data and providing several trained models for which metrics can be computed. It enables users to have an idea of the stability of the model (see Fig. 2). In any ML models, the observations of the couple X, Y are assumed to be independent. Yet, when two segments are close (same line, same track and a similar location) the observations, and particularly observation of the target event, are probably correlated. Thus, another method assessed here is the simple idea of using a specific and independent sample to validate the model. It is suggested that one of the most important lines (Paris-Lyon) is used as the control sample. The third method introduced in the analysis is focused on assessing the prediction. Datasets are generated at two different times, where the first dataset is used to train and the second as a test. What varies from the first dataset to the second is the defects that appear in the meantime and the increase in cumulated tonnage. Different levels of performance are observed in Fig. 4: it is clear that the time-based method has a lower performance than shuffling dataset. The overall performance of the model is good (scoring model with Random Forest for head-check defects). The model is relevant for targeting risky configurations regarding head-check initiation but using this scoring model for prediction probably calls for data enrichment, especially time varying ones.

Variable Importance. One simple way to interpret models for tree-based methods is to compute Mean Decrease Impurity in the nodes of the trees. To measure the impurity, the Gini function is used.

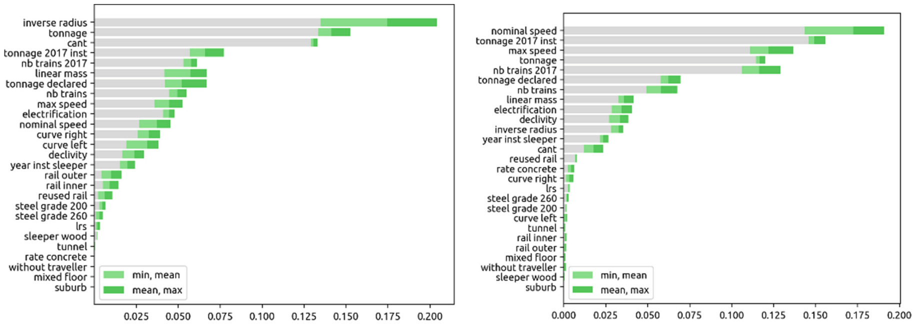


Fig. 5. Mean decrease impurity for the Gini index. The green shades are describing the variation of scores in the cross validation. The left figure corresponds to head-checks defect while the right one corresponds to squats.

The Gini variable importance associates a score to each feature which represents the capacity to divide the data into pure nodes: with only defects or only healthy rail-segments. The results given by this approach are quite convincing. They highlight some expert insights, namely the fact that tonnage is of prime importance for fatigue phenomena for both head-checks and squats. In the case of head-checking, the radius of curvature has a great impact on defect initiation (Fig. 5, left figure), and this is coherent with observations, where head-check almost only appears in curves. As for squats, factors describing speed seem to have great importance (Fig. 5, right figure) which also corroborates observations. This is particularly interesting because the SNCF is using UIC classification for monitoring maintenance, which depends of tonnage and speed.

Correlated features are used in the current model and these can have an impact on the extracted conclusion. Indeed, if several methods are used to describe tonnage (by the number of trains, tonnage since installation or for a specific year), the importance of tonnage can be distributed over several variables and thus can be underestimated. Yet, all these features need to be studied for several reasons: the noisiness and relevance of each feature is difficult to assess individually, and we do not have a clear preconception of what would be the best representation for each feature. Another limit here is that the impact of covariates on the model cannot be seen, and particularly the positive or negative influence on the output. This leads us to use SHAP values to explain the results of the models.

SHAP Analysis. For further analysis, SHAP (SHapley Additive explanation) is used to provide individual analysis for the predictions [20]. This game theory approach offers an interpretation method by modelling the contribution of each feature to the prediction. A feature plays the role of a participant of a game where the profit, here the prediction, is distributed between the players (features). Tree SHAP [21] is used in the current approach, which is a variant of the classical SHAP methods. Due to complexity, only $2000 \times n_{features}$ are computed. Note that all these result interpretation methods apply to scoring settings.

In Fig. 6 and Fig. 7, each point symbolizes one rail segment and the color is linked to the value of the associated feature. The abscissa axis represents the impact of the feature on the model output for each instance. Thus, the further that a point is from 0, the greater impact on the prediction. For instance, low inverse curvature radius (no curve) seems to have very negative effect on head-checks apparition (Fig. 7). Generally, when blue and red points are well separated, the feature has a discriminative power on the data.

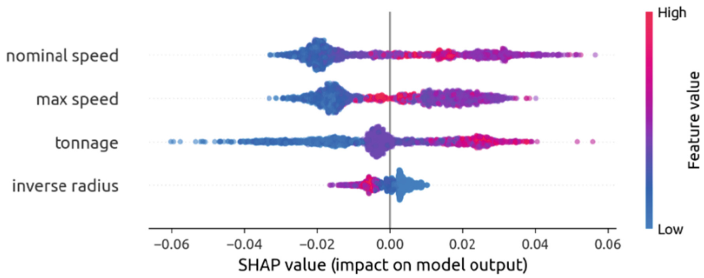


Fig. 6. SHAP values for scoring model on squats

Thus, the signed importance can be analyzed to once again corroborate the observations: high tonnage and high speed (low UIC) entail squat formation, while high tonnage

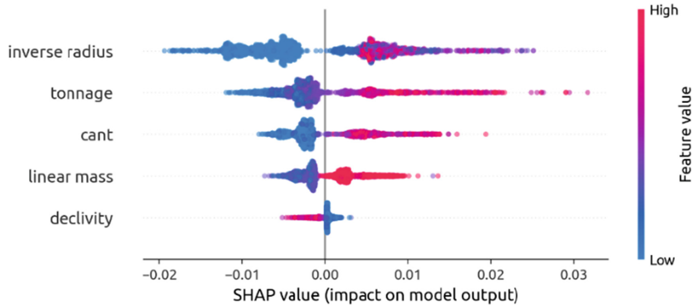


Fig. 7. SHAP values for scoring model on head-checks

and strong curvature (high inverse curvature radius) implies head-checking initiation. From our point of view, SHAP values may have a real impact on asset management: by building an explanatory model, an asset manager is provided with insight into what causes a defect to appear and this may orient renewal works (Fig. 6 and Fig. 7).

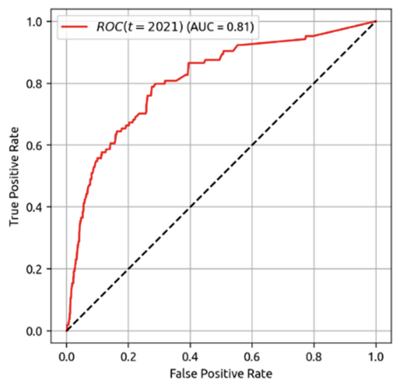


Fig. 8. ROC curve for survival approach

Inferring Residual Life and Survival Model Analysis. Several methods exist for analysis of survival models, among which one is inspired by work in medical research on the time-dependent ROC curve [22]. The idea is to offer a generic way to interpret models. Indeed, survival models like LTRC Trees yield an estimate of the survival function at time t. By interpreting the estimate $\hat{s}(t)$ as a score, we can compute the ROC curve and its AUC. The overall statement that can be made is that survival models are comparable to pure scoring models with time shifting windows, when comparison is made with the AUC based on the same test method (Table 2). One of the best possibilities for improving prediction efficiency for survival models would be to build scores using RF, then to use scoring results in survival models that tackle both truncated and censored data (LTRC-Forest, COX and AFT models). For the current application, one of the main interests of using survival analysis is to compute residual life. Indeed, using

Bayes formula, for all $t > t_f$ we have

$$\mathbb{P}\big(T > t | T > t_f\big) = \frac{\mathbb{P}\big(T > t, T > t_f\big)}{\mathbb{P}\big(T > t_f\big)} = \frac{s(t)}{s\big(t_f\big)}$$

where t_f represents the time of the last acquisition. This enables us to make predictions and assess residual life for each segment of the rail network and thus, evaluate when a segment should be replaced.

Limits of the Approaches. As usual, the approaches presented here have some limitations. 1) For now, the scoring setting can only be interpreted as an index measuring how settings and properties of a rail segment favors defect initiation. A lot more time dependent data is needed to use scoring models for prediction with time shifting windows. 2) The defect can be removed by grinding and several defects can occur on the same segment. Further study could consider recurrent risk methods to describe behavior of crack initiation.

Results Summary. Results are presented in Table 2 for all models for all validation methods. They are presented in terms of the AUC, of the C-index [23] and of the F1-score [24]. For the time method test, models are trained for data collected before May 2018 and tested on data between May 2018 and February 2021. Segments with defect in the first interval are removed in the test. The comparison between other studies of rail defects in the literature is not simple because of the wide variety of metrics is used. In many studies, accuracy is used to show the performance of models which may be misleading in the presence of unbalanced classes: predicting every instance as heathy may yield good accuracy when there are few defects in test. Nonetheless, with an accuracy between 66%

Table 2. Modelling results

Model	Defect type	Test method	Results
RF	Head-checks	Control Line	AUC = 0.76
		Shuffle	AUC = 0.88
		Time	AUC = 0.83
RF	Squats	Control Line	AUC = 0.75
		Shuffle	AUC = 0.79
		Time	AUC = 0.66
LTRC - Trees	Head-checks	Time	AUC = 0.81
			C-Index = 0.82
			F1-score = 0.6
LTRC - Trees	Squats	Time	AUC = 0.72
			C-Index = 0.70
			F1-score = 0.67

and 71% obtained in this study, results are comparable to the literature: for instance, 75% accuracy is obtained in [8] for geometry defect detection and 64.7% in [25] for broken rail prediction. AUC would be a better choice for comparison of the models, because it does not depend on the classification threshold and it is not sensitive to unbalanced classes. In the context of modelling geometric defects, the maximum AUC obtained in [26] is 0.7 with XGBoost, with our results being between 0.66 and 0.83.

5 Conclusion

In this paper, two Machine Learning algorithms have been applied to real data describing railway rail defects. Simple scoring settings have oriented our analysis towards tree-based methods, because of the capacity to discriminate healthy and deteriorated segments. This makes the study relevant for SNCF asset managers because the current monitoring policy sometimes results in rails without damage being replaced. Through analysis of the Gini Importance and the SHAP values, evidence validating the approach is given to operational teams, in the form of proof that the models are consistent with current maintenance policy. Furthermore, providing a residual lifespan for each segment makes predictive maintenance possible at several time horizons. As for performance levels, our main achievement was in data processing and aggregation. Indeed, it is during the data consolidation work that a clear improvement in performance was observed. The current results are satisfactory compared to other studies in the area.

Perspectives. Both explanatory methods attest to the relevance of the models and help prepare the ground for feature selection or an aggregation method as in [26], that may be used in future work to make more intelligible models. The methods developed here, from data aggregation to modelling, are comprehensive and can be reproduced for other railway networks, and for other target events. Future work will include study of other defects such as corrosion and geometry defects. Finally, the intention is to include more time dependent variables in the studies in order to strengthen the robustness of the models to make predictions.

References

1. Popović Z, Radovic V, Lazarević L, Vukadinovic V, Tepić G (2013) Rail inspection of RCF defects. Metalurgija 52:537–540
2. El-sayed HM, Lotfy M, El-din Zohny HN, Riad HS (2018) Prediction of fatigue crack initiation life in railheads using finite element analysis. Ain Shams Eng J 9(4):2329–2342. https://doi.org/10.1016/j.asej.2017.06.003
3. Bonniot T, Doquet V, Mai SH (2020) Fatigue crack growth under non-proportional mixed-mode I + II. Role of compression while shearing. Int J Fatigue 134:105513. https://doi.org/10.1016/j.ijfatigue.2020.105513
4. Burstow M (2004) Whole life rail model application and development–Continued development of an RCF damage parameter. T115 project report, RSSB
5. Ghofrani F (2018) Recent applications of big data analytics in railway transportation systems_ a survey. Transp Res Part C: Emerg Technol 90:226–246

6. Shafahi Y, Masoudi P, Hakhamaneshi R (2008) Track degradation prediction models, using Markov Chain, artificial neural and neuro-fuzzy network. In: 8th world congress on railway research, p. 9 (2008)
7. Guler H (2014) Prediction of railway track geometry deterioration using artificial neural networks: a case study for Turkish state railways. Struct and Infrast Eng 10(5):614–626. https://doi.org/10.1080/15732479.2012.757791
8. Sharma S, Cui Y, He Q, Mohammadi R, Li Z (2018) Data-driven optimization of railway maintenance for track geometry. Transp Res Part C: Emerg Technol 90:34–58. https://doi.org/10.1016/j.trc.2018.02.019
9. Yao N, Jia Y, Tao K (2020) Rail weld defect prediction and related condition-based maintenance. IEEE Access 8:103746–103758. https://doi.org/10.1109/ACCESS.2020.2999385
10. Ghofrani F, He Q, Mohammadi R, Pathak A, Aref A (2019) Bayesian survival approach to analyzing the risk of recurrent rail defects. Transp Res Rec 2673(7):281–293. https://doi.org/10.1177/0361198119844241
11. Schapire RE (1999) A brief introduction to boosting. Ijcai 99:1401–1406
12. Breiman L (2001) Random forests. Mach Learn 45(1):5–32. https://doi.org/10.1023/A:1010933404324
13. Pedregosa F et al (2011) Scikit-learn: machine learning in python. J Mach Learn Res 12:2825–2830
14. Ke G et al (2017) LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst 30:3146–3154
15. Zhang Z (2017) Survival analysis in the presence of competing risks. Ann Transl Med 5(3):47. https://doi.org/10.21037/atm.2016.08.62
16. Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. J Am Stat Assoc 53(282):457–481
17. Fu W, Simonoff JS (2017) Survival trees for left-truncated and right-censored data, with application to time-varying covariate data. Biostat 18(2):352–369. https://doi.org/10.1093/biostatistics/kxw047
18. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS (2008) Random survival forests. Ann Appl Stat 2(3):841–860. https://doi.org/10.1214/08-AOAS169
19. Barnwal A, Cho H, Hocking TD (2020) Survival regression with accelerated failure time model in XGBoost. arXiv preprint arXiv:2006.04920. Accessed 10 Apr 2021. http://arxiv.org/abs/2006.04920
20. Lundberg S, Lee S-I (2017) A unified approach to interpreting model predictions. arXiv preprint, 2017, Accessed 12 Apr 2021. http://arxiv.org/abs/1705.07874
21. Lundberg SM, Erion GG, Lee S-I (2019) Consistent individualized feature attribution for tree ensembles. arXiv preprint arXiv:1802.03888. Accessed 21 Apr 2021. http://arxiv.org/abs/1802.03888
22. Kamarudin AN, Cox T, Kolamunnage-Dona R (2017) Time-dependent ROC curve analysis in medical research: current methods and applications. BMC Med Res Methodol 17(1):1–19. https://doi.org/10.1186/s12874-017-0332-6
23. Harrell FE (1982) Evaluating the yield of medical tests. JAMA 247(18):2543. https://doi.org/10.1001/jama.1982.03320430047030
24. Chinchor N, Sundheim BM (1993) MUC-5 evaluation metrics
25. Schafer DH, Barkan CP (2008) A prediction model for broken rails and an analysis of their economic impact
26. Mohammadi R (2019) Exploring the impact of foot-by-foot track geometry on the occurrence of rail defects. Transp Res Part C: Emerg Technol 102:153–172

Online Temperature Estimation of Permanent Magnet Synchronous Machines (PMSM) Using Non-linear Autoregressive Neural Networks with Exogenous Input (NARX)

Thainara de Araújo(✉), Renan Aryel F. da Silva, Marcio L. M. Kimpara, and João Onofre

Federal University of Mato Grosso do Sul, Campo Grande, Brazil

Abstract. PMSMs are widely used in high-performance industry applications. This popularity is due to their high torque-to-inertia ratio, high efficiency, low maintenance, fast dynamic response, among others features. However, the construction of such machines includes some components that are highly sensitive to the temperature, hence, requiring control strategies that mitigate failures and loss management, taking the machine temperatures into account. Sensor-based temperature measurements of such parts are difficult to be implemented, and are not always well-accurate. Therefore, this paper proposes an approach based on artificial neural network model to estimate the temperature at the most critical points of a PMSM, namely, the permanent magnet, stator teeth, windings, and stator yoke. In this study, the variables, ambient and coolant temperatures, motor speed, and the stator voltages and currents in the direct and quadrature axes are taken as inputs to a Non-linear Autoregressive Neural Networks with Exogenous Input (NARX). To develop and test the proposed temperature estimator, a 140-h multivariate database from a torque-controlled 52 kW PMSM was used. The obtained results have shown that the proposed method successfully estimates the temperature at the selected points.

1 Introduction

The global concern about the environmental issues has increasingly grown up in recent years, leading the engineers and companies to seek for sustainable alternatives that satisfy the population needs. In this scenario, a technology that has been gaining notoriety is the transportation electrification. Although important advances have been reached in this field, further approaches that contribute to the use of electric motors efficiently are the subject of many researches.

For automotive applications, the electric motor that has been standing out is the permanent magnet synchronous type (PMSM), due not only to its efficiency, but also to the low need for maintenance, high torque-inertia ratio and fast dynamic response, among other advantages. Nevertheless, despite to the aforementioned attractive characteristics, the PMSM contain parts that are sensitive to excessive heat, which demands temperature monitoring and control for maintaining machine performance. The major components

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 467–478, 2022.
https://doi.org/10.1007/978-3-030-96794-9_43

susceptible to heat are: the magnets, which can be demagnetized; and the stator winding insulation that can be thermally aged and seriously threaten the safe operation of the motor.

Usually, the thermal monitoring of the PMSM stator is carried out through the incorporation of sensors, which is hard to be replaced, even if their functionality deteriorates over time. Apart from that, the sensor-based measurements in the rotor structure are technically unfeasible since it spins. An alternative is the IR (infrared) sensor or the sensors attached to the rotor with wireless communication, however there is an additional cost associated to such methods. Therefore, in order to avoid the damage caused by excessive heat, the manufacturers commonly adopt a safety margin in embedded materials, which increases the whole costs [1].

To overcome these problems, an alternative is estimating, in real time, the temperature of multiple parts of interest from the heat point of view. A satisfactory temperature estimation will contribute for designing motors with optimized size and amount of material, in addition to allowing the application of control techniques that admit the usage of the maximum capacity of the PMSM, thus assisting in asset management. From the point of view of asset management, to ensure process efficiency, temperature monitoring is important so that failure and reliability indices are created. And so, in this way, determining the probability of failures and developing a preventive maintenance program to achieve maximum system efficiency, reducing interruptions in operation, customer dissatisfaction and, consequently, market loss. Accordingly, this paper proposes a methodology using artificial intelligence to estimate the temperatures at crucial points: permanent magnet, stator teeth, winding and stator yoke. Given the complexity of modeling the PMSM, in addition to the variation tend of the variables that describes the motor, the proposed estimator is a data-driven approach rather than a model-based approach. More specifically, the method used to estimate the temperature based on an artificial neural network (ANN). The application of ANN for data estimation is quite common in the literature. An application of neural temperature estimator in was described by Oliver et al. (2017), in which they presented a model that mixes the use of ANNs with lumped-parameter thermal networks (LPTNs) in an attempt to increase the accuracy of the estimates beyond the state of the art. Thermal models of electric motors are usually complex, typically composed of multiple inputs and multiple outputs (MIMO), thus, ANNs show themselves as promising alternatives. In the study presented, they achieved decent results, especially when keeping in mind that established thermal model techniques, like LPTNs, have already been researched for decades. Based on this, they conclude that it is necessary to investigate alternatives to save computational time, such as different ANN topologies and optimization algorithms, in order to increase the robustness of network training.

Following this line of study, Kirchgassner et al. (2019) is dedicated to the study of the estimation of high dynamic temperatures of a PMSM using different ANN architectures: recurrent networks and temporal convolutional networks. This work proposes a single network capable of identifying all target temperatures and comparing the different architectures in terms of precision and computational cost of its application in real time. The network architectures of this work were defined automatically from Bayesian optimization. Although the work has achieved good results, the architectures used are highly complex, which can compromise their real-time application. Guo et al. (2020) present a methodology for estimating the stator winding temperature based on measurements of ambient temperature, coolant temperature, direct-axis voltage, quadrature-axis voltage, motor speed, torque, direct-axis cur-rent, quadrature-axis current, permanent magnet surface temperature, stator yoke temperature, and stator tooth temperature. The prediction was performed by means of a deep neural network (DNN), achieving satisfactory results, however, this method has two disadvantages. First, this architecture use two or more hidden layers, therefore has a high computational cost. Second, it demands some inputs that are difficult to measure, as torque, permanent magnet surface temperature, stator yoke temperature, and stator tooth temperature.

Literature lacks temperature estimation studies in PMSM, which makes this work of unique importance not only for academic purposes, but also for the automobile industry and other sectors that make use of the PMSM.

Therefore, the present paper explores the ability of nonlinear autoregressive neural networks with exogenous inputs (NARX) to model the functions that describe the temperatures of the PMSM at the points subject to heat stress. The validation of the estimation method was accomplished in a 52 kW PMSM.

2 Permanent Magnet Synchronous Motor

Traditionally, permanent magnet machines play an important role in a wide range of industry applications, moreover, its application in pure electric and hybrid electric vehicles have recently grown. Permanent magnet synchronous motors are AC motors with multi-phase stator winding that creates a rotating magnetic field according to the frequency of the line current. Meanwhile, a constant magnetic field arises from the permanent magnets mounted in the rotor structure.

The rotors of electric machines with permanent magnets are relatively loss-free, however, despite having advantages, the characteristics of permanent magnets depend on the temperature. The increase in temperature causes the residual flux density of the magnets to decrease, as in neodymium-iron-boron [5]. Hence, keeping the temperature in a safe range is essential for maintaining machine performance [6]. High power PMSMs, usually prevent the rotor overheating by means of cooling mechanism.

The mathematical modeling of the three-phase permanent magnet motor is done using a series circuit containing a resistor, an inductor and a voltage source representing the counter-electromotive force connected in "Y" as shown in Fig. 1.

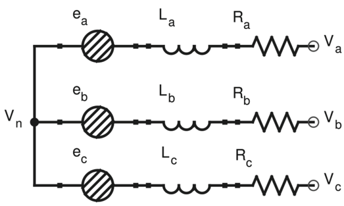


Fig. 1. Permanent magnet motor model [7].

Disregarding the currents induced in the rotor by the harmonic fields of the stator and the dispersion losses in the iron core, the voltage equation of the circuit is represented in (1).

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \begin{bmatrix} R_a & 0 & 0 \\ 0 & R_b & 0 \\ 0 & 0 & R_c \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} L-M & M_{ab} & M_{ac} \\ M_{ba} & L-M & M_{bc} \\ M_{ca} & M_{cb} & L-M \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} - \begin{bmatrix} V_n \\ V_n \\ V_n \end{bmatrix} \tag{1}$$

where:

L_a, L_b e L_c are the stator inductances of the phase *a, b* e *c* [H];
M_{ab}, M_{ac}, M_{ba}, M_{bc}, M_{ca} e M_{cb} are the mutual inductances [H];
R_a, R_b e R_c are the stator resistances of the phase *a, b* e *c* [Ω];
V_n stands for the neutral terminal voltage [V].

Assuming that the three phases are symmetrical and balanced, and the rotor presenting smooth poles, the Eq. (2) is derived:

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \begin{bmatrix} R_a & 0 & 0 \\ 0 & R_b & 0 \\ 0 & 0 & R_c \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} L & M & M \\ M & L & M \\ M & M & L \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} - \begin{bmatrix} V_n \\ V_n \\ V_n \end{bmatrix} \tag{2}$$

The sum of the currents per phase is zero and considering the isolated neutral terminal, Eq. (2) becomes:

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \begin{bmatrix} R_a & 0 & 0 \\ 0 & R_b & 0 \\ 0 & 0 & R_c \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \frac{d}{dt} \begin{bmatrix} L-M & 0 & 0 \\ 0 & L-M & 0 \\ 0 & 0 & L-M \end{bmatrix} \cdot \begin{bmatrix} i_a \\ i_b \\ i_c \end{bmatrix} + \begin{bmatrix} e_a \\ e_b \\ e_c \end{bmatrix} - \begin{bmatrix} V_n \\ V_n \\ V_n \end{bmatrix} \tag{3}$$

For a salient pole machine, we have the following equation:

$$V_a = E_a + E_d + E_q \tag{4}$$

Where:

V_a is the phase voltage;
E_a is the internal tension;
E_d is the direct axis component of the armature reaction voltage;
E_q is the quadrature axis voltage of the armature reaction voltage.

Each component of armature stress can be modeled by:

$$E_d = -jx_d I_d \tag{5}$$

$$E_q = -jx_q I_q \tag{6}$$

Where x_d and x_q are the direct and quadrature axis reactances, respectively. Adding the voltage drop across the armature resistance R_a the equation for the total stator voltage looks like this:

$$V_a = E_a - jx_d I_d - jx_q I_q - R_a I_a \tag{7}$$

Modelling the PMSM involves parameters that may change over time which represents addition complexity to find an exact model. Some conventional temperature estimation models use electrical parameters that have values varying according to the temperature, however, the estimated temperature based on the electrical parameter is the average value of the winding or magnet [6]. Thus, in contrast to the model-based estimation approaches, machine learning models adjusts itself to the machine particularities though empirical data, that detach from theorical approximations.

3 Autoregressive Non-linear Network with Exogenous Inputs

Haykin (2008) describes ANNs as processors massively parallel distributed from simple processing units, which have the natural propensity to store experimental knowledge and make it available for use. These systems are able, from the knowledge acquired about their environment, through a learning process and the forces of connection between neurons, the synaptic weights, to perform a computation useful for human beings. Therefore, these are the two degrees of freedom that must be defined by the designer, taking into account the restrictions that the problem to be solved imposes on the network.

In this paper, giving the dynamic nature of the problem to be solved, the architecture chosen was the NARX type (Nonlinear Autoregressive with eXogenous Inputs), an Multi-layer Perceptron whose input consists of the output itself fed with time delays and an exogenous input, also containing a delay. The NARX model is non recursive model, as there is no coupling between the expected output and the input; they are the outputs of the real system that serve as input to the model. In this way, the adjustment of weights follows a supervised learning algorithm. The output of the NARX network can be considered as an estimation of the output of the dynamic system is being modeled. This output is fed back to the input of a feedforward network with parallel architecture (image on the left in Fig. 2). Since the ongoing output is available during network training, one can create a parallel series architecture, in which the known output is used instead of feeding back the estimated output, as shown by the right-side image of Fig. 2 [9]. The latter architecture has two advantages. The first is that the entry to the feedforward network is more accurate, and the second is that the resulting network has a purely feedforward architecture and static back propagation can be used for training.

NARX networks have been proved to effectively address time series problems. Among the various applications of NARX networks are: predicting the next value of the input signal; non-linear filtering, where the target output is a noiseless version of the input signal; modeling of nonlinear dynamic systems, to name a few.

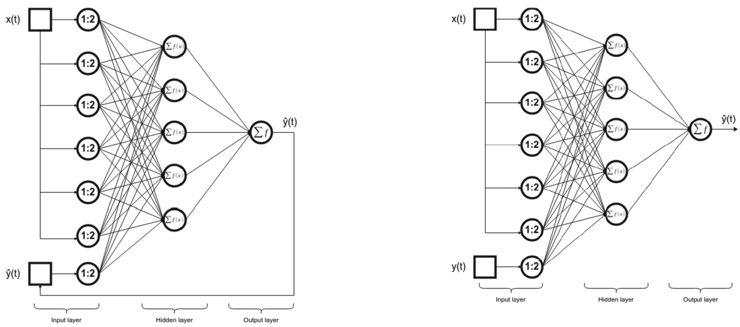


Fig. 2. Parallel network model (left) and parallel series (right).

3.1 NARX Configuration

The data basis used in this study come from [3]. These data set refer to the experimental tests performed on a 52 kW three-phase PMSM. The measurements in the database correspond to 140 h of multivariate measurements sampled at 2 Hz. These measurements correspond to 52 test profiles.

The variables available in the database correspond to the measures of: ambient temperature (t_a); coolant temperature (t_c); current and voltage on the stator direct and quadrature axes (i_d, i_q,u_d, u_q); motor speed (s_m); torque (T); and the temperature on the permanent magnet (t_{pm}); stator tooth (t_{tooth}), winding ($t_{winding}$) and stator yoke (t_{yoke}).

Since the motor temperature cannot be measured reliably and economically in automotive application, the target temperatures have been defined as t_{pm},t_{tooth}, $t_{winding}$, t_{yoke}, resulting in four temperatures to be estimated.

In order to minimize the computational cost of the estimation, has been applied though the correlation study performed between the input variables, aiming to discard those that present high correlation with the others. The resulting correlation matrix of the inputs can be seen in (8), where the entries are ordered this way: t_a, t_c, u_d, u_q, s_m, T, i_d e i_q.

$$r = \begin{bmatrix} 1,000 & 0,494 & 0,181 & 0,164 & 0,134 & -0,264 & -0,017 & -0,262 \\ 0,494 & 1,000 & 0,175 & 0,109 & 0,010 & -0,185 & 0,092 & -0,179 \\ 0,181 & 0,175 & 1,000 & -0,006 & -0,293 & -0,793 & 0,435 & -0,767 \\ 0,164 & 0,109 & -0,006 & 1,000 & 0,642 & -0,098 & -0,043 & -0,083 \\ 0,134 & 0,010 & -0,293 & 0,642 & 1,000 & -0,008 & -0,689 & -0,032 \\ -0,264 & -0,185 & -0,793 & -0,098 & -0,008 & 1,000 & -0,286 & 0,996 \\ -0,017 & 0,092 & 0,435 & -0,043 & -0,689 & -0,286 & 1,000 & -0,243 \\ -0,262 & -0,179 & -0,767 & -0,083 & -0,032 & 0,996 & -0,243 & 1,000 \end{bmatrix} \quad (8)$$

By the correlation matrix in (8) it is possible to observe a high correlation between the 6th input (T) with inputs 3 and 8 (u_d e i_q, respectively). Therefore, it is possible to remove the input T, from the analysis. Properly, this procedure not only reduces the number of inputs, but also eliminates a variable that is difficult to be accurately acquired. Notably, there is also a high correlation between the inputs 3 e 8 (u_d e i_q, respectively), thus it was also possible to disregard the input u_d. As a result, the correlation analysis allowed to reduce the number of input variables, reducing the computational cost of the estimator. Afterward, the data sets for training and test of the estimator were separated to allow the verification of the assertiveness through the cross validation. About 70% of the data profiles were selected to be used for training the estimator, and the others for testing. The proposed NARX network is formed by three layers - input, hidden and output - and is it characterized by receiving as input not only external variables (both current value and previous values), but also previous values of the variable that is intended to be predicted. Therefore, the main configuration parameters of this network are: input delay, output delay, and the number of neurons in the hidden layer [10]. Another important parameter in the configuration of an ANN is the activation function. The most common activation functions are linear and sigmoid, the first applied to the output layer and the second to the intermediate layers, which can be logistic or hyperbolic tangent [11]. In this work, the hyperbolic tangent function has been chosen for the hidden layer. Due to its architecture, the NARX network is deemed more powerful than conventional recurring networks, once the ability to converge faster and generalize better, as well as to obtain better performance in problems involving long-term dependencies [12]. Due to the presence of memory in its architecture – added by the use of delays – it allows that, during the learning process by descending gradient, it is possible to propagate the information gradient by a shorter path than by back propagation of the signal.

In order to apply this technique to the PMSM temperature estimation problem, four NARX networks were defined nn_{pm}, nn_{tooth}, $nn_{winding}$ and nn_{yoke}, one to estimate each desired temperature output: t_{pm}, t_{tooth}, $t_{winding}$ and t_{yoke}, respectively. The exogenous inputs for each neural estimator are the variables: t_a, t_c, i_d, i_q, u_q and s_m. Since there is no rule for defining the optimal parameters of an ANN, some attempts have been made considering different numbers of neurons and delay values. Thus, modifying the parameters and observing the performance after each trial, it was realized that in a certain point, the use of more neurons, as well as more delays, did not add substantial improvement to the estimators, and therefore their associated computational cost is not justified. The final configuration of each neural network is shown in Table 1.

Thirty-seven profiles were used to train the ANNs, and afterward, another 15 profiles were applied to validate their efficiency. During the training step, the feedback from the output to the input was performed based on the real values (parallel-series architecture), in order to optimize the adjustment time of the synaptic weights. However, during the test procedure, the outputs feedback into the estimator were the estimated outputs, in order to simulate a real time situation.

In order to evaluate such modification, firstly, the performance of nn_{pm}, nn_{tooth}, $nn_{winding}$ and nn_{yoke} during the training phase has been found to reach a very small mean squared error (MSE) in a few iterations, as shown in Table 2. Notably, the networks have properly adapted to the data with high assertiveness. Analyzing only the data set used for training, Fig. 3 depicts the regression curves obtained for each estimator.

Subsequently, with the training phase completed and the synaptic weights of the neural networks already established, the structure of the networks have been modified (the feedback input was changed receive the estimated data) aiming to obtain temperature estimations suitable to real case applications. Then, a verification of the adequacy of the

Table 1. ANNs architectures

Neural network	Number of neurons	Input delay	Feedback delay
nn_{pm}	15	1:5	1:5
nn_{tooth}	5	1:2	1:2
$nn_{winding}$	10	1:2	1:2
nn_{yoke}	15	1:2	1:2

Table 2. Training performance

Neural network	Epochs	MSE
nn_{pm}	43	2.968×10^{-5}
nn_{tooth}	49	1.856×10^{-6}
$nn_{winding}$	86	2.267×10^{-6}
nn_{yoke}	15	4.600×10^{-6}

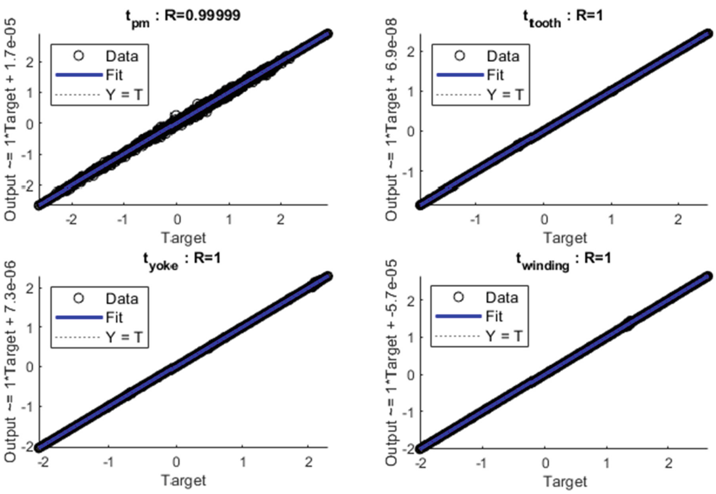


Fig. 3. Regression curve for training data.

networks to the training data after this structure modification was carried out and reveals an increase in the MSE, as shown in Table 3. Although one can note a considerable increase in the MSE, meanwhile, there was not a significant reduction in the R^2 parameter of the regression curves, as can be seen in Fig. 4. Based on that, once can conclude the estimators still present a satisfactory adequacy to the training data.

In the next step, unknown data will be presented in order to assess the efficiency and generalization capability of the temperature estimators.

Table 3. Training performance after modifying the feedback

Neural network	MSE
nn_{pm}	0.010
nn_{tooth}	0.026
$nn_{winding}$	0.055
nn_{yoke}	0.013

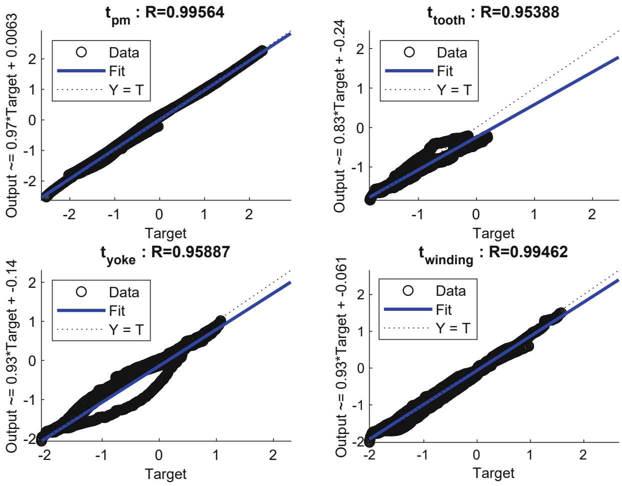


Fig. 4. Regression for the training data after modifying the feedback input.

4 Results

As aforementioned, 15 profiles from the database were separated to test the neural network estimators. The comparison between the real and estimated values for one of the tested profiles is shown in Fig. 5. The MSE and R^2 of the estimators for this specific profile are shown in Table 4.

Finally, the same analysis is accomplished for all fifteen test profiles. The MSE for all the test data is shown in Table 5, and the corresponding regression curves in Fig. 6. The regression analysis, shown in Fig. 6, confirms that the proposed ANNs are able to obtain a good estimate of the temperature in all the desired critical points, proving to be good temperature estimators.

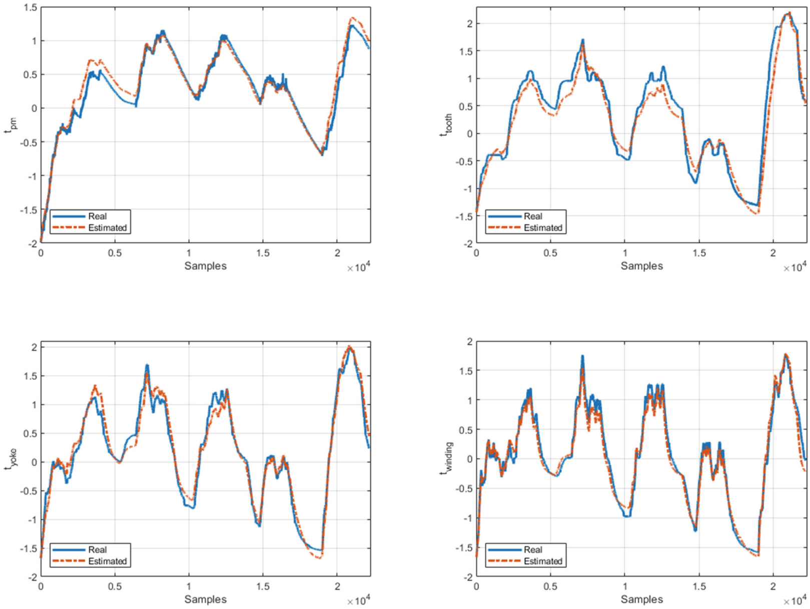


Fig. 5. Comparison between measured and estimated data.

Table 4. Estimation performance for one profile from the test data set

Neural network	MSE	R^2
nn_{pm}	0.012	0.985
nn_{tooth}	0.048	0.977
$nn_{winding}$	0.021	0.987
nn_{yoke}	0.015	0.990

Table 5. Test performance for fifteen profiles

Neural network	MSE
nn_{pm}	0.021
nn_{tooth}	0.043
$nn_{winding}$	0.035
nn_{yoke}	0.023

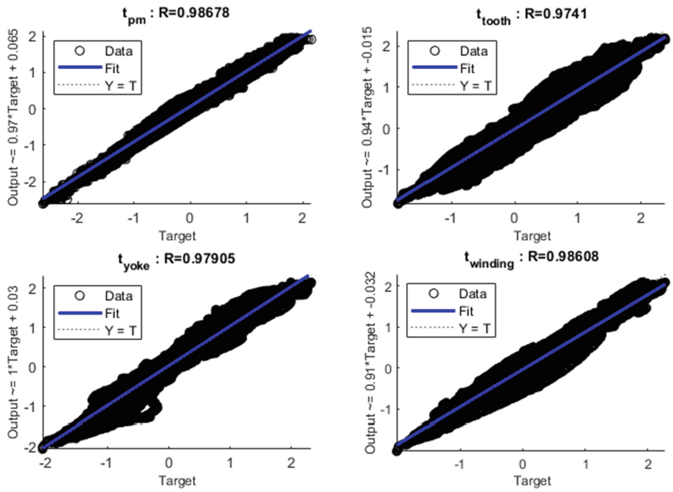


Fig. 6. Regression for the test data set.

5 Conclusion

Electric motors are an asset present in many industries, equipment and vehicles. The synchronous PM motor type plays an important role in these applications, however monitoring the temperature is essential to guarantee safe operation and motor's performance. Thus, this work focuses on determining whether the NARX neural network technique is feasible for estimating the PMSM temperature at critical points, since this variable is key information for the analyzing the motor wear, as well as for assessing the losses, which are directly linked to the reduction in the efficiency of the permanent magnet.

The NARX type neural network can be used to estimate the temperature at the critical points of the PMSM from data already commonly measured in the engine, without the need to include more sensors inside the machine. The estimators obtained proved to be functional, allowing them to be used instead of temperature sensors at these points.

References

1. Kirchgassner W, Wallscheid O, Boecker J (2021) Estimating Electric Motor Temperatures with Deep Residual Machine Learning. IEEE Trans Power Electron 36(7):7480–7488. https://doi.org/10.1109/tpel.2020.3045596
2. Wallscheid O, Kirchgässner W, Böcker J (2017) Investigation of long short-term memory networks to temperature prediction for permanent magnet synchronous motors. In: 2017 International joint conference on neural networks (IJCNN), pp 1940–1947. https://doi.org/10.1109/IJCNN.2017.7966088
3. Kirchgassner W (2019) Electric motor temperature: 140 hours recordings from a permanent magnet synchronous motor (PMSM). Retrieved from https://www.kaggle.com/wkirgsn/electric-motor-temperature. Accessed 03 December 2020

4. Guo H, Ding Q, Song Y, Tang H, Wang L, Zhao J (2020) Predicting temperature of permanent magnet synchronous motor based on deep neural network. Energies 13:4782. https://doi.org/10.3390/en13184782
5. Umans SD, Kingsley C (2014) Fitzgerald & Kingsley's electric machinery, 7th edn. McGraw-Hill, New York
6. Lee J, Ha J (2020) Temperature estimation of PMSM using a difference-estimating feedforward neural network. IEEE Access 8:130855–130865. https://doi.org/10.1109/access.2020.3009503
7. Krishnan R (2001) Electric motor drives: modeling, analysis and control. Prentice Hall, New Jersey
8. Haykin S (1999) Neural networks: a comprehensive foundation. Mc Millan, New Jersey
9. DESIGN Time Series NARX Feedback Neural Networks. MathWorks (2021). <https://www.mathworks.com/help/deeplearning/ug/design-time-series-narx-feedback-neural-networks.html>. Accessed 10 May 2021
10. Sun W, Qiu Y, Sun L, Hua Q (2020) Neural network-based learning and estimation of battery state-of-charge: a comparison study between direct and indirect methodology. Int J Energy Res 44(13):10307–10319. https://doi.org/10.1002/er.5654
11. MULTILAYER Shallow Neural Network Architecture. MathWorks (2021). <http://www.mathworks.com/help/nnet/ug/multilayer-neural-network-architecture.html>. Accessed 09 July 2021
12. Siegelmann HT, Horne BG, Giles CL (1997) Computational capabilities of recurrent NARX neural networks. IEEE Trans Syst Man Cybernet Part B (Cybernetics) 27(2):208–215. https://doi.org/10.1109/3477.558801

Author Index

J. O. P. Pinto et al. (Eds.): 15th WCEAM Proceedings, LNME, pp. 479–481, 2022.
https://doi.org/10.1007/978-3-030-96794-9