Foreword

The 13th IMA International Conference on Modelling in Industrial Maintenance and Reliability (MIMAR2025) was held in Nancy, France, from 8 to 10 July 2025. Since its inception, MIMAR has become a leading international forum for the exchange of ideas, methods, and experiences in reliability modelling, risk assessment, maintenance optimization, and their applications across diverse industrial sectors.

This year's edition brought together around 90 participants from 18 countries, reflecting the truly international and interdisciplinary character of the conference. The programme featured keynotes addresses by distinguished experts, technical sessions covering a wide spectrum of topics - from stochastic modelling and optimization to AI-driven predictive maintenance and digital twins as well as stimulating discussions that fostered collaboration between academia and industry.

The contributions collected in this volume represent the state of the art in modelling approaches for reliability, risk and maintenance. With 3 keynotes, 30 full papers and 31 extended abstracts, the proceedings offer valuable insights into current challenges, methodological advances, and future directions for research and practice.

We extend our sincere gratitude to the authors for their valuable contributions, to the reviewers for their thorough evaluations and constructive feedback, and to the chairs and members of the Technical Programme Committee for their commitment to shaping a high-quality scientific programme. We also warmly thank the IMA team, together with the local organizing team and volunteers, whose support and dedication made this conference possible.

We hope that readers will find in these proceedings both inspiration and practical knowledge to advance the field of reliability, risk and maintenance engineering.

On behalf of the scientific committee

Christophe Bérenguer & Phuc Do

MATHEMATICS & ITS APPLICATIONS

Management team

Organizing committee

Phuc Do, Chair (Université de Lorraine, France)
Rasa Remenyte-Prescott, Co-Chair (University of Nottingham, UK)
Nicolae Brinzei, Co-Chair (University of Lorraine, France)

Scientific committee

Christophe Bérenguer (Grenoble INP, France)
Philip Scarf (University of Swansea, UK)

Technical Programme Committee

Babakalli Alkali (Glasgow Caledonian University, UK)
John Andrews (University of Nottingham, UK)
Fazel Ansari (TU Wien, Austria)
Anne Barros (CentraleSupelec, France)
Maria Dolores Berrade Ursúa (Universidad de Zaragoza, Spain)
Emannuel Borgonovo, Bocconi University, Italy
Inma Castro (Universidad de Extremadura, Spain)
Cristiano Cavalcante (Federal University of Pernambuco, Brazil)
David Coit (Rutgers University)
Frank Coolen (Durham University, UK)
Rommert Dekker (Erasmus University, Netherlands)
Sarah Dunnett (Loughborough University, UK)
Richard Dwight (University of Wollongong, Australia)
Paolo Gardoni (University of Illinois Urbana-Champaign, US)
Antoine Grall (University of Technologies de Troyes, France)
Geert van Houtum (Technische Universiteit Eindhoven, Netherlands)
Lisa Jackson (Loughborough University, UK)
Renyan Jiang (Changsha University of Science & Technology, China)
Abdelhakim Khatab (University of Lorraine, France)
Vytis Kopustinskas (Joint Research Centre, Italy)
Chi-Guhn Lee (University of Toronto, Canada)
Yan-Fu Li (Tsinghua University, China)
Mitra Fouladirad (Centrale Marseille, France)
Marcos Orchard (University of Chile)
Edoardo Patelli (University of Strathclyde, UK)
Ed Pohl (University of Arkansas, USA)
Darren Prescott (University of Nottingham, UK)
Elena Zaitseva (University of Žilina, Slovakia)
Adriano de Polpo (University of Western Australia, Australia)
Jin Wang (Liverpool John Moores University, UK)
Ming J Zuo (University of Alberta, Canada)
Shaomin Wu (University of Kent, UK)
Vitaly Levashenko (University of Žilina, Slovakia)
Alexandre Voisin (University of Lorraine, France)
Min Xie (City University, Hong Kong)

Model-Based Load-Dependent Degradation Modeling For PEM Fuel Cells: A Multi-Health Index Approach Toward Energy Management in Multi-Stack Systems

A. M. Houjayrie*, C. Cadet**, and C. Berenguer***

**Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: mouhamad.houjayrie@.grenoble-inp.fr*

***Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: catherine.cadet@grenoble-inp.fr*

****Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: christophe.berenguer@grenoble-inp.fr*

Abstract. Multi-stack proton exchange membrane fuel cells (PEMFCs) offer a promising pathway for high-power, carbon-free energy systems, yet durability and cost challenges hinder their commercialization. While degradation is influenced by operating parameters, optimizing dynamic power allocation across stacks could mitigate performance loss. However, this requires robust degradation models that account for complex, load-dependent mechanisms. This work introduces a degradation model for multi-stack PEMFCs, integrating two critical health indices: electrochemical surface area (ECSA) loss and rising internal resistance. The model links these degradation pathways to operational voltage, enabling real-time assessment of performance decay. By coupling degradation dynamics with load allocation strategies, the framework supports lifespan extension and efficiency optimization in multi-stack architectures, advancing scalable PEMFC deployment for sustainable energy applications.

Keywords: PEMFCs, multi-stack, stochastic degradation, catalyst degradation, ECSA, durability.

Proton exchange membrane fuel cells are energy conversion devices that generate electricity through an electrochemical reaction between hydrogen and oxygen, producing water and heat as byproducts. A typical PEMFC consists of a membrane electrode assembly (MEA), including a proton-conducting polymer membrane, catalyst layers (commonly using platinum), and gas diffusion layers (GDLs). During operation, hydrogen is oxidized at the anode, releasing protons and electrons. The protons migrate through the membrane to the cathode, while electrons travel through an external circuit, generating electrical power. At the cathode, oxygen reacts with the protons and electrons to form water. Over time, various degradation mechanisms affect the fuel cell's components, reducing performance and lifespan. These include catalyst degradation, membrane thinning, and mechanical failures.

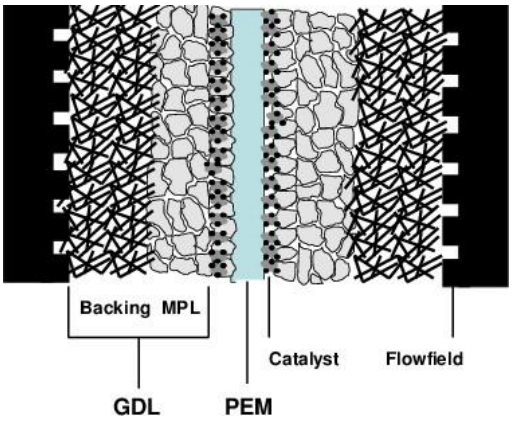


Figure 1: PEMFC structure.[1]

Platinum (Pt) catalysts in PEMFCs are susceptible to multiple degradation mechanisms that compromise both performance and durability. Platinum dissolution, driven primarily by operating conditions such as voltage cycling, is a dominant factor [2]. This process reduces the electrochemical surface area (ECSA), leading to performance degradation and, ultimately, system failure. While prior studies have proposed physical models for ECSA loss, this work correlates ECSA degradation with dynamic power demand. Additionally, internal resistance, a key health indicator for PEMFCs, increases over time due to degradation in components such as the

membrane, gas diffusion layer (GDL), and catalyst layer. This rise in resistance is load-dependent and has been modeled in previous work as a stochastic gamma process with load dependency [3].

To capture the combined impact of ECSA loss and rising internal resistance, we propose a fuel cell potential model that links voltage output to health status, expressed as $V = f(I, T, R, ECSA)$, where I is current, T is temperature, R is internal resistance, and ECSA is the electrochemical surface area. Fr om this model, we derive a health-dependent power-voltage relationship that directly ties ECSA degradation to power demand, bypassing sole reliance on voltage variations.

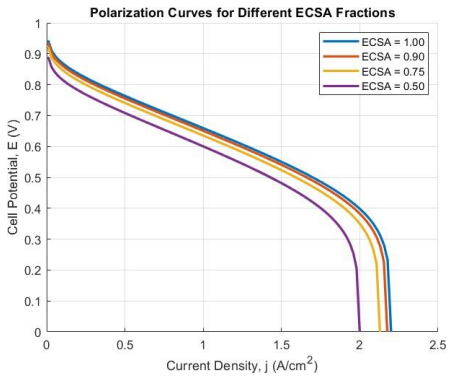


Figure 2: Modelled performence loss due to decrease in ECSA in PEMFC

By incorporating the stochastic nature of degradation and variability across stacks, this multi-health-index model serves as a prognostic tool to estimate the remaining useful life (RUL) based on projected load demand and current state of health. This framework supports health-conscious energy management and informs maintenance decisions in multi-stack PEMFC systems, laying the foundation for future work on optimized load allocation and system scalability.

Acknowledgment

The authors acknowledge the ANR – FRANCE (French National Research Agency) for its financial support of the MFC-DEPLOY project n° ANR-23-CE05-0006.

References

[1] SIGRACET® Gas Diffusion Layers for PEM Fuel Cells, Electrolyzers and Batteries (White Paper) (2025) *ResearchGate*. Accessed 15 April 2025. Available at: https://www.researchgate.net/publication/295859224_SIGRACETR_Gas_Diffusion_Layers_for_PEM_Fuel_Cells_Electrolyzers_and_Batteries_White_Paper

[2] Touil W, Li Z, Outbib R, Hissel D, Jemei S (2025) A system-level modeling framework for predicting Pt catalyst degradation in proton exchange membrane fuel cells. *Journal of Power Sources* 625, 235628. https://doi.org/10.1016/j.jpowsour.2024.235628.

[3] Zuo J, Cadet C, Li Z, Bérenguer C, Outbib R (2024) A deterioration-aware energy management strategy for the lifetime improvement of a multi-stack fuel cell system subject to a random dynamic load. *Reliability Engineering & System Safety* 241, 109660. https://doi.org/10.1016/j.ress.2023.109660.

Joint optimization of operation & maintenance for navy ships

M. Maraha*, B. Westerweel**, T. Tinga***

** Thales Nederland B.V, Mirel.Maraha@thalesgroup.com & University of Twente M.Maraha@utwente.nl*
*** Netherlands Defence Academy, B.Westerweel@mindef.nl*
**** Netherlands Defence Academy, T.Tinga@mindef.nl & University of Twente T.Tinga@utwente.nl*

Abstract. This research explores the concept of optimizing both operation and the maintenance of navy ships simultaneously. It will look into an example of a navy ship that is deployed on several missions, containing a system that degrades at a rate depending on the mission the ship is deployed on. A Markov Decision Process (MDP) is created, yielding some interesting preliminary results. In the example given, changing the order of missions shows a change in costs. These preliminary results are an interesting starting point for further research.

Keywords: Maintenance modeling, Condition based maintenance, Mission-oriented system, Dynamic programming

The operational performance of naval ships relies on their deployment and maintenance. Maintenance needs are affected by prescribed protocols and the condition of the ship. The deployment is the assignment of specific missions to a specific ship. The missions indicate the type of use and the type of (weather) conditions in which the naval ship is deployed. The type of use and conditions affect the system degradation and in turn this has an influence on the maintenance needed.

The field where operation/mission planning is combined with maintenance planning is relatively new. Some recent papers on this are the work by Zheng et al. (2023), developing a condition-based replacement model for a mission-oriented system. This optimization problem is formulated in a Markov decision process framework. Lee et al. (2024) introduce a mission assignment strategy aimed at actively aligning maintenance schedules to maximise opportunistic maintenance potential.

Planning the deployment in accordance with maintenance is an intricate challenge. For instance, for each mission a system can have a different performance threshold. But it is also difficult to plan the deployment while knowing that without performing interim maintenance future missions might not be feasible. When maintenance is not done in time, unplanned maintenance might arise. Unplanned maintenance is considered more expensive and more time-consuming than planned maintenance.

This research therefore focuses on ways to jointly plan maintenance and deployment, with the objective to optimize the availability of navy ships and maximize the percentage of successful missions. A mathematical optimization model will be created that allows the joint optimization of both deployment & maintenance. In the current first version of the model, the maintenance planning can be optimized given a fixed deployment.

Some interesting preliminary results have been obtained. In the model four states are considered, 1 being the worst and 4 being the best. There are six missions, each of these has a different length, performance threshold and transition probability. If the performance drops below the threshold, unplanned maintenance is needed. If maintenance is executed before that moment, the maintenance is considered planned. The costs of planned and unplanned maintenance differ. When maintenance is done, the system becomes as good as new and goes to state 4. There is also a target state at which the system should at least be at the end of the timeframe, if this is achieved, a bonus is awarded. Finally, also the start state can be specified, allowing to study the effect of imperfect systems on the optimal policy.

For a specific order of the missions, the model can calculate the optimal policy, i.e. with the lowest overall costs, for each time and state. Then a number of simulations is performed in which a trajectory is followed and the optimal policy is applied. The optimal policies of these orderings of missions are shown in Figure 1 and Figure 2, indicating the advised decision at each moment in time, given the system state at that moment. Broadly the optimal policies are quite similar; in state 1, 2 and 3 maintenance should be done and in state 4

not. However, Figure 1 shows three instances where 'no maintenance' is advised in state 3, while Figure 2 shows six instances where no maintenance should be done in state 3 and one instance where even in state 4 maintenance is advised.

Table 1 shows the costs and number of actions for the two different orderings of missions, revealing that there can be a lot to gain. The average total maintenance costs can become nearly half by simply reordering the missions. This cost reduction is attributed to the lower number of unplanned tasks. This illustrates the close relation between operation and maintenance, and the need to jointly optimize them.

Ongoing research will focus on expanding the mathematical model to include, among other things, transition probabilities that reflect realistic degradation, multiple (or redundant) systems, and the option to review all possible set variations of missions. Further, the mathematical model will be tested and demonstrated on a case study for the Royal Netherlands Navy. Results will ultimately be used to advise the on-board crews on planning of maintenance and deployment.

Table 1. Results 10000 trajectory simulations

Ordering of missions	[1, 2, 3, 4, 5, 6]	[5, 1, 6, 4, 2, 3]
Average number of maintenances	3.71	3.99
Average number of unplanned maintenances	1.36	0.38
Average number of planned maintenances	2.34	3.61
Average total maintenance costs	25133	12974
Average end state	3.74	3.81

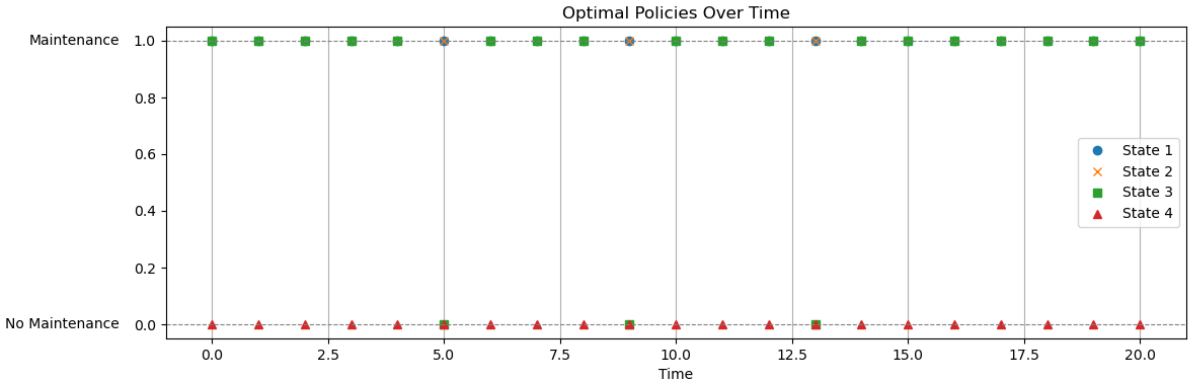


Figure 1. Optimal policies for mission set [1, 2, 3, 4, 5, 6]

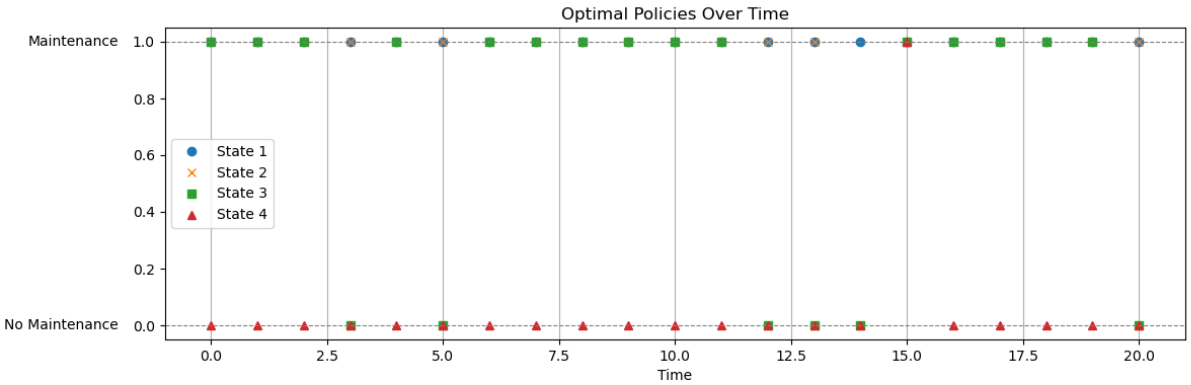


Figure 2. Optimal policies for mission set [5, 1, 6, 4, 2, 3]

References

Puterman, M. L. (1994). *Markov Decision Processes Discrete Stochastic Dynamic Programming.*

Zheng, R., Xing, Y., & Ren, X. (2023). Multilevel preventive replacement for a system subject to internal deterioration, external shocks, and dynamic missions. *Reliability Engineering and System Safety*, 239.

Lee, J., Yu, J. S., & Hong, Y. S. (2024). Active mission assignment for improving the effectiveness of opportunistic maintenance in a fleet of mission-critical systems. *International Journal of Production Research*.

A Petri Net approach to predict the lifespan of a railway track asset

Emily Buttriss*, John Andrews**, and Rasa Remenyte-Prescott***

**Resilience Engineering Group, University of Nottingham, UK: pmyelbu@nottingham.ac.uk*

***Resilience Engineering Group, University of Nottingham, UK: evzjda@nottingham.ac.uk*

****Resilience Engineering Group, University of Nottingham, UK: evzrr@nottingham.ac.uk*

Abstract: This study aims to predict the lifespan of an asset on the railway line. This is achieved by developing a Petri net model that connects the inspection, degradation, maintenance and renewal processes via sub-models. Renewal requirements are defined by five separate criteria, with strategic placing of tokens used to select a renewal strategy based on a combination of the criteria. The movement of tokens around the net is used to collect information regarding the asset's lifespan. The model was demonstrated on two components of the S&C, the switch and crossing panel.

Keywords: Petri nets, lifespan prediction, renewal criteria, railway track asset

It is vital in any efficient management strategy that an accurate estimation of a track asset's lifespan is known, so that appropriate funding can be provided for renewal when needed. This paper aims to develop a methodology that mimics the degradation and maintenance dynamics of a track asset to predict when a renewal is necessary. The methodology has been applied specifically to two components of the switches and crossings (S&C), the crossing panel and the switch panel.

Previous lifespan prediction research has largely focused on a reliability analysis of the S&C failure history, including exploring the likelihood of component failure by fitting Weibull distributions to describe the lifetimes between maintenance (Litherland and Andrews, 2023a), and drawing the estimated lifetime distribution curves based on the renewal history (Zwanenburg, 2008). Research developing an adaptable life-cycle model is rare for the S&C but has typically been achieved using Petri nets (PN) for other track assets such as bridges or overhead lines. In both (Litherland and Andrews, 2023b) – expanding on the previously mentioned reliability analysis – and (Sachan and Donchak, 2020) the entire life-cycle of a S&C component has been modelled via a PN, however neither paper considered a renewal strategy beyond a fixed lifetime. No literature was found which considers the impact of different renewal strategies on the expected lifetimes of an S&C component.

The adaptable and stochastic nature of a Petri net means it was the chosen modelling method for this research. Alongside the four essential components of a PN – arcs, tokens, transitions and places – additional features such as inhibitor arcs, place dependent transitions, and reset transitions have been exploited to better represent the complex railway procedures. The developed PN can be divided into four interconnected sub-models: inspection, degradation, maintenance, and renewal. Degradation is only noticed when an inspection occurs and requires maintenance to resolve. The frequency and impact of maintenance may trigger a renewal depending on five different criteria:

1. Maintenance limit. If the number of maintenance activities completed on the asset has exceeded a threshold, a renewal is scheduled.
2. Set period. If the number of faults occurring within a time interval has reached a threshold, a renewal is scheduled.
3. Time duration. If two faults occur within a time interval, and this happens consecutively for a given number of times, a renewal is scheduled.
4. Associated cost. If the total cost of operating an asset – including costs from staff labour, financial penalties for reduced service, and equipment hiring costs – exceeds a threshold, a renewal is scheduled.

5. Condition based. If a threshold for material removed by grinding or percentage of the component covered by weld repairs has been reached, a renewal is scheduled.

Strategic placing of tokens was used to inhibit transitions that would otherwise trigger a replacement if the renewal criteria threshold were met, so that different renewal strategies could be considered within a single model. Railway standards have been utilised to develop the structure of the PN's by defining the various faults, their stages of degradation, and recommended maintenance action. Each fault was modelled by an individual degradation sub-model, with 17 faults defined in the standards as affecting the switch panel and 24 affecting the crossing panel.

The models have been created using a bespoke code written in C++. Each PN model was demonstrated using firing times derived from statistical analysis of a UK high-speed line for the component degradation transitions, and the recommended timeframes in the railway standards for the transitions related to engineering procedures. Repeated simulation of the nets was used to observe the length of time before a token moved to the 'complete renewal' place, to analyse the average lifespan of the component. Convergence was achieved when the mean lifespan did not change by more than 1%. The results, shown in Table 1, can be used to increase the efficiency of renewal scheduling, but also to highlight the need for the two components to be replaced separately, to obtain the maximum lifetime out of both.

Table 1. Mean lifespan for the switch panel and crossing panel per renewal strategy with given parameters

Renewal strategy	**Maintenance Limit**	**Set Period**	**Time Duration**	**Associated Cost**	**Condition Based**
Implemented renewal parameters	6 faults	2 faults within a 2-year review period	3 consecutive faults within 2.5 years of the previous fault	20,000 units spent on operations	8mm material removed by grinding or 100% weld coverage
Mean lifespan switch panel (years)	16.21	11.19	11.16	21.34	25.70
Mean lifespan crossing panel (years)	25.26	23.15	26.03	22.36	28.87

Acknowledgements

The authors would like to express their gratitude to HS1 for providing the financial support to enable the research reported in the paper to be carried out.

References

Litherland J, Andrews J (2023a) A reliability study of railway switch and crossing components. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* **237**, 205-217.

Litherland J, Andrews J (2023b) A Petri net asset management framework for railway switches and crossings. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* **237**, 322-334.

Sachan S, Donchak, N (2020) Generalized stochastic Petri-net algorithm with fuzzy parameters to evaluate infrastructure asset management policy. *2020 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE),* 1-8

Zwanenburg W (2008) A model for the life expectancy of railway switches and crossings for maintenance and renewal planning in asset management systems. *WIT Transactions on The Built Environment* **103**, 765-773.

Real-Time Monitoring of Nozzle Clogging in Cold Spray Process Using Airborne Acoustic Emission and Data-Driven Prognostics

S. Koufis*, P. Dinglinger**, N. Eskue*, D. Zarouchas*, J-A Pascoe*

**Faculty of Aerospace Engineering, Delft University of Technology, Delft, The Netherlands*

***Faculty of Production Engineering, University of Bremen, Bremen, Germany*

Abstract. Cold spray is a solid-state deposition technique that can suffer from nozzle clogging, affecting coating quality and process reliability. This study investigates a real-time monitoring approach using airborne acoustic emission (AAE) to detect clogging progression. Run-to-failure experiments with single-track depositions and 3D profilometry revealed acoustic features correlated with nozzle degradation. These findings form the basis for defining a health indicator (HI) linked to coating performance and estimating the remaining useful life (RUL), showing strong potential for early clogging detection and enabling condition-based maintenance strategies.

Keywords: cold spray, nozzle clogging, airborne acoustic emission, real-time monitoring, health indicator, remaining useful life

The cold spray process is an emerging solid-state deposition technique that accelerates metallic powder particles to supersonic speeds through a converging-diverging nozzle using a carrier gas. Upon impact with a substrate, the particles form a dense, adherent deposition, making cold spray highly suitable for coating and repair applications. Its key advantages include low-temperature operation, minimal oxidation, and reduced thermal degradation of the substrate, making it particularly attractive for aerospace, automotive, and other industries.

A common challenge in cold spray is nozzle clogging, which arises from the gradual buildup of powder inside the nozzle. This accumulation restricts the gas-particle flow, compromising deposition quality and process stability. Detecting clogging is thus essential for quality assurance. Moreover, the ability to anticipate clogging in advance would allow corrective action before part quality is affected. Direct visual access to the spray plume is often limited due to the close spray distance, making non-visual sensing approaches more practical. In this context, airborne acoustic emission (AAE) is a promising solution, as acoustics have already been successfully applied for monitoring in other high-noise manufacturing processes (Kumar et. al., 2024). Prior studies highlighted the potential of AAE as a real-time, non-intrusive monitoring technique, offering valuable insights without requiring a direct view of the spray plume. Preliminary experiments indicated that acoustic signals contained information related to particle velocity, nozzle positioning and nozzle wear (Koufis et al. 2024, Arkhipov et al., 2025). In this work, the analysis of AAE signals was further advanced toward enabling both detection and prognostics of nozzle clogging.

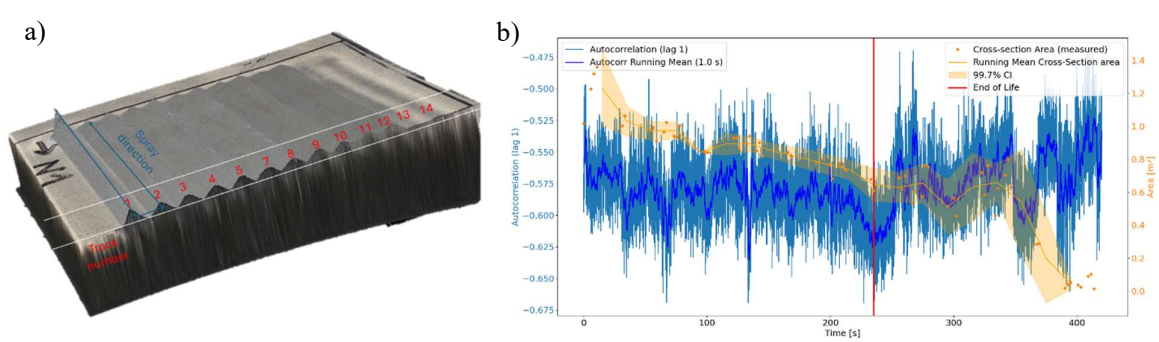


Figure 1. (a) Evolution of track geometry over time obtained via 3D profilometry, showing progressive degradation due to nozzle clogging. (b) Acoustic signal features and corresponding track area measurements over time. A correlation is observed between signal changes and deposition instability.

Run-to-failure experiments were conducted to analyze the progression of nozzle clogging during the spray process, capture its stochastic behavior, and extract relevant signal features capable of characterizing different stages of clogging: healthy condition, clogging initiation, clogging buildup, and end-of-life. To trace the degradation process during the experiments, spraying was performed continuously on a flat substrate using single-track depositions until complete nozzle clogging, identified by a significant reduction or complete absence of deposition. This single-track strategy was critical, as it allowed each deposited track to be individually inspected and analyzed, offering high spatial resolution and a clear view of the evolution of deposition quality over time. This approach enabled a systematic and objective definition of end-of-life, using deposition quality metrics as a proxy to determine when the nozzle could no longer produce acceptable tracks. Throughout the experiment, 3D profilometry was used to measure the geometry of each track, providing detailed insight into how deposition evolved with clogging (Figure 1).

In the experiment shown in the figure, the first ~200 seconds of spraying were characterized by slow and gradual degradation. This trend was evident in both the AAE signal features and the track measurements (Figure 1). A sharp shift was effectively captured by the autocorrelation feature, coinciding with a marked increase in the variability of area and track measurements—suggesting that clogging had begun to affect deposition uniformity. As spraying continued, the severity of clogging increased. By around 400 seconds, deposition efficiency had declined significantly. Tracks 13 and 14 showed minimal material buildup, indicating that deposition efficiency had been critically compromised (Figure 1). Microscopic analysis further revealed increasing porosity as clogging progressed. In practice, the end-of-life threshold should be set earlier—at the onset of instability—to avoid reduced coating quality and to protect the cold spray equipment from damage.

The physical changes were closely associated with distinct features extracted from the acoustic emission (AE) signals. This relationship supports the development of a health index (HI) to track the process from stable operation, through the gradual onset of clogging, to the final failure.

Building on these insights, the goal is to leverage the HI for prognostics, to estimate the nozzle's remaining useful life (RUL). The combination of AE-based monitoring and predictive algorithms holds promise for minimizing unplanned downtime, reducing material waste, and improving the overall efficiency of cold spray applications. Ongoing experiments under varying process parameters are being used to assess and validate the methodology's effectiveness in early clogging detection and RUL estimation.

This work highlights the potential of condition-based maintenance in cold spray applications. Early detection of clogging and linking HI to deposit quality ensure consistent coating performance and optimized operational efficiency.

Acknowledgments:

This work was made possible by funding from the Netherlands Enterprise Agency (RVO), in the framework of the 'Mobiliteitsfonds—Brightsky' project and the support of EPCOR BV.

References:

Kumar, R., Singh, A., & Sharma, A., (2024). Acoustic-Based Machine Condition Monitoring—Methods and Challenges. *Acoustics* **4(1)**, pp.42–62.

Arkhipov I, Kokal U, Özdemir OÇ (2025) Aeroacoustic process monitoring and anomaly detection in cold spray additive manufacturing. *Journal of Thermal Spray Technology* **34(1)**, 97–119.

Koufis S, Eskue N, Zarouchas D, Pascoe J-A (2024) Monitoring the cold spray process: real-time particle velocity monitoring through airborne acoustic emission analysis. *Journal of Thermal Spray Technology* **33(8)**, 2657–2671.

Risk Assessment of Condition-Based Maintenance Contracts

Stijn Loeys*, Melvin Drent**, Robert Boute***, Collin Drent****, and Katrien Antonio*

**Faculty of Economics and Business, KU Leuven, stijn.loeys@kuleuven.be*
*** Tilburg University, Tilburg School of Economics and Management*
**** Technology and Operations Management, Vlerick Business school*
***** Eindhoven University of Technology, School of Industrial Engineering,*
****** Faculty of Economics and Business, University of Amsterdam*

Abstract. We assess the financial risk incurred by maintenance service providers responsible for delivering preventive and corrective maintenance under a fixed upfront fee agreement. We propose a method to jointly estimate the machine-specific failure distribution and optimize the preventive maintenance policy using maintenance data captured during the contract or operational sensor data that captures the machine's health. Bayesian updating is used to sequentially estimate a contract's risk and tailor the preventive maintenance policy and the associated maintenance costs. An extensive simulation study shows the value of jointly estimating the failure distribution and optimization of the preventive maintenance policy.
Keywords: Condition-Based Maintenance, Service Contracts, Bayesian Learning, Condition Monitoring

The risks associated with unplanned downtime of heavy machinery due to machine failures. To address these challenges, service providers increasingly provide full-service contracts. Under a full-service contract, the customer pays a contract fee, while the service provider assumes responsibility for all maintenance costs incurred throughout the contract period. These costs typically include preventive maintenance (PM) visits, scheduled according to a specific PM policy, and more expensive, unplanned failure visits. The contract's fee and the PM policy should reflect its maintenance risk. Contracts for maintenance-light customers should be lower priced than contracts for maintenance-heavy customers. If that is not the case, maintenance-light customers in a portfolio subsidize maintenance-heavy ones. Moreover, the PM policy should also be tailored to the customer's maintenance risk profile: if their failure rate is higher, they should receive more preventive maintenance.

The operational conditions also impact the optimal preventive maintenance (PM) policy. Two machines with similar risk factors at the contract's start might have different operational conditions unknown at the contract's start translating into divergent actual risk profiles. By dynamically learning the machine's specific failure distribution, the timings of the preventive maintenance can be tailored per machine. This machine-specific failure distribution can be learned from historical maintenance data and updated using maintenance data observed during the contract. Operational sensor data can also be incorporated into estimating said failure distribution when available. This paper assumes a preventive maintenance policy where PM is performed when the machine's hazard rate surpasses a predefined threshold. We learn the failure distribution in an online, Bayesian manner throughout the contract and use it to optimize the PM policy. By updating beliefs about the failure behaviour of the machine under contract, the preventive maintenance policy is made more accurate, resulting in lower expected maintenance costs compared to contracts with a sub-optimal PM policy fixed at the contract's start.

Existing methods for online learning of the failure distribution

The parameters that govern the failure distribution and the degradation behaviour under a CBM approach can be learned offline or online (Keizer et al., 2017). In an offline approach, the parameters are first estimated based on historical data, followed by optimizing maintenance decisions. Estimating the degradation process is complex, with a substantial body of literature dedicated to this task (Lei et al., 2018). Unlike offline CBM, online CBM circumvents this sequential approach by jointly estimating the failure distribution and degradation process and optimizing the maintenance policy. The main advantage of an online approach is that it leverages both real-time condition data and historical data collected during operation, whereas an offline approach relies solely on historical data. Examples of such online approaches include Drent et al. (2023), which consider a discrete-state degradation process and learn the parameters per machine in a Bayesian manner using the real-time degradation signal. Drent et al. (2020) derive the optimal policy for a finite sequence of age-replacement problems by learning the unknown parameter of a general lifetime distribution in an online manner.

Online learning based on maintenance observations and operational sensor data

This paper assumes that a meaningful, one-dimensional health indicator is available that represents the machine's

condition in real-time. This health indicator is included in a machine's hazard rate via the proportional hazard rate model as a dichotomous covariate. Because of machine-to-machine variability, the actual parameters of the hazard rate differ from the initial estimates at the contract start, resulting in uncertainty about this hazard rate. The initial belief about a machine's hazard rate is represented by placing prior distributions on all the failure parameters. By incorporating data in the prior distribution, its variance is reduced, i.e., we become more certain of the actual value of the individual's machine hazard rate's parameter and can tailor the PM moment accordingly.

Instead of solely considering a point estimate of the expected maintenance cost, we build an expected maintenance cost distribution that reflects the hazard rate's parameter uncertainty. We incorporate each new observation, either a maintenance visit or a health indicator measurement, registered on a specific machine in the prior distributions of the hazard rate's parameters via prior-to-posterior updating (Gelman et al., 1995). To quantify the parameter uncertainty, we draw a large number of hazard rate parameters from these posterior distributions. For each sample, the preventive maintenance policy is optimized with an associated expected maintenance cost. Combining the expected maintenance costs from all samples results in a distribution of expected maintenance costs.

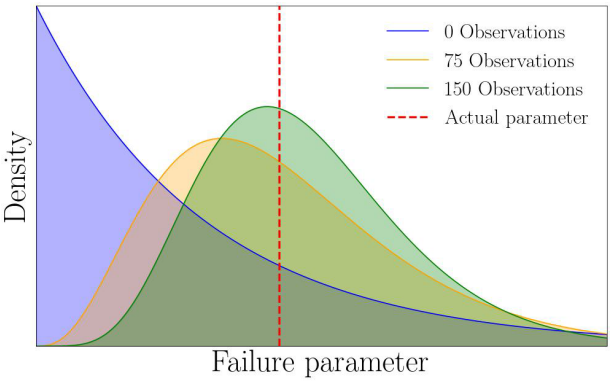


Figure 1: posterior distribution of a failure distribution parameter

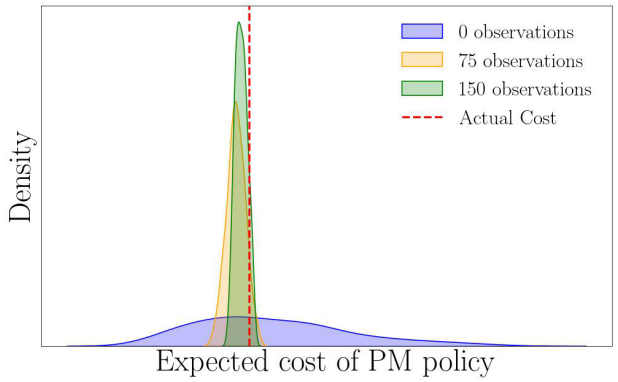


Figure 2: Expected maintenance cost distribution given current belief about failure distribution

Figure 1 and Figure 2 illustrate the posterior distribution of a failure parameter and the machine's expected maintenance cost distribution. At the start of a contract (0 observations), parameter uncertainty is at its maximum, i.e., large variance of the prior (see Fig. 1), and there is little certainty about the cost structure, i.e., a wide maintenance cost distribution (see Fig. 2). Increasing the number of observations reduces the variance of the posterior distribution and narrows the maintenance cost distribution. Both converge to their actual values.

Conclusion

In this paper, we propose a methodology to jointly improve the cost estimation of condition-based maintenance contracts and optimize the PM policy by incorporating operational data accrued during the contract period. Traditional approaches to service contract cost estimation rely on static risk factors known at the contract's start, which fail to capture the revealed machine heterogeneity in failure behaviour during the contract. By dynamically updating the cost estimate through Bayesian learning of the failure distribution and optimizing the preventive maintenance policy, we demonstrate that considering operational data leads to more accurate cost estimates and an improved preventive maintenance policy.

References

Keizer, M. C. O., Flapper, S. D. P., & Teunter, R. H. (2017). Condition-based maintenance policies for systems with multiple dependent components: A review. *European Journal of Operational Research*, *261*(2), 405-420.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., & Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, *104*, 799-834.

Drent, C., Drent, M., Arts, J., & Kapodistria, S. (2023). Real-time integrated learning and decision making for cumulative shock degradation. *Manufacturing & Service Operations Management*, *25*(1), 235-253.

Drent, C., Kapodistria, S., & Boxma, O. (2020). Censored lifetime learning: Optimal Bayesian age-replacement policies. *Operations Research Letters*, *48*(6), 827-834.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Data-driven Prognostics under uncertainty: A comparative study on the state-of-the-art HMMs

T. Kontogiannis[*,**], M. Salinas-Camus[*,**], D. Zarouchas[**], N. Eleftheroglou[*]

[*] *Intelligent Sustainable Prognostics Group, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands*

[**] *Center of Excellence in AI for Structures, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands*

Abstract. This study compares state-of-the-art Hidden Markov Model (HMM) variants for data-driven prognostics under uncertainty. Vanilla HMM, Hidden Semi-Markov Model (HSMM), Adaptive HSMM (AHSMM), and Similarity-Learning HSMM (SL-HSMM) were evaluated utilizing strain field data collected during fatigue testing of composite structures. A new prognostic measure is proposed, capable of capturing time-dependent degradation. Models are assessed using RMSE and a weighted confidence interval coverage metric. Results show AHSMM consistently outperforms others, especially for anomalous cases, by dynamically adapting to unexpected behavior, thus improving both prediction accuracy and uncertainty quantification.

Keywords: Prognostics and Health Management, Remaining Useful Life, Hidden Markov Models, Hidden semi Markov Models, adaptive modes, comparative study, uncertainty quantification, stochastic modelling

As systems grow more complex, ensuring their safety and reliability while minimizing resource waste becomes increasingly important and, at the same time, challenging. Prognostics and Health Management (PHM), originally introduced by the US Air Force (Fallon et al., 2009), offers predictive modelling solutions for such challenges. Recent developments in PHM emphasize data-driven approaches over analytical models, driven by system complexity and the limitations of traditional methods (Hu et al., 2022).

Artificial Neural Networks (ANNs) have been extensively explored in PHM applications, yet their deterministic nature limits their ability to capture the inherently stochastic progression of system degradation where predictions need to be modelled as random variables (Salinas-Camus M. et al., 2024). While uncertainty-aware ANN modifications exist, they often incur high computational costs or provide incomplete uncertainty quantification (He et al., 2023).

This work investigates stochastic models – Hidden Markov Models (HMMs) (Rabiner, 1989) – which probabilistically model latent degradation processes and naturally incorporate uncertainty. Beyond the vanilla HMM, we examine amendment variants: Hidden Semi-Markov Models (HSMMs) (Kontogiannis et al., 2025), Adaptive HSMMs (AHSMMs) (Eleftheroglou et al., 2020), and Similarity-Learning-based HSMMs (SL-HSMMs) (Eleftheroglou et al., 2024).

HMMs assume geometric (memoryless) sojourn time distributions. HSMM extends HMM by explicitly modelling state duration using arbitrary distributions, allowing for more realistic modelling in scenarios where the time spent in each state is not memoryless. AHSMMs enhance HSMMs with an adaptation mechanism during testing. They detect deviations in sojourn times using the Most Likely State (MLS) measure (Eleftheroglou and Loutas, 2016), adjusting future state durations accordingly. SL-HSMMs further improve prediction by weighing training trajectories based on similarity (e.g., Euclidean distance) to the test case.

These models are evaluated on experimental data for Remaining Useful Life (RUL) prediction from carbon-fiber-reinforced polymer specimens under tensile-tensile fatigue using full-field strain data (Eleftheroglou, 2023). The pre-processed input trajectories are presented in Figure 1. A novel, time-aware RUL formulation – developed in our prior work (Kontogiannis et al., 2025)– replaces the traditional, time-invariant state-based RUL estimation (Dong et al., 2006). The RUL prediction utilizing the vanilla HMM and the AHSMM for the left outlier trajectory, are presented in Figure 2. Prediction performance is assessed using RMSE (Table 1) and the Weighted Coverage metric (Table 2):

$$Weighted_Coverage = \frac{1}{T}\frac{\sum_{t=1}^{T} C_t}{SU}, \qquad where\ C_t = \begin{cases} 1, & if\ l_t \le y_t \le u_t \\ 0, & otherwise \end{cases},$$

$$SU = \sum_{t=1}^{T-1}\left[\frac{u_{t+1}+u_t}{2} - \frac{l_{t+1}+l_t}{2}\right],$$

where $[l_t, u_t]$ the lower and upper bounds of the Confidence Intervals (CIs).

Results indicate that AHSMM consistently outperforms other models, accurately handling both inlying and outlying trajectories due to its dynamic adaptation capability. This highlights the potential of stochastic models

– especially AHSMMs – for robust, uncertainty-aware prognostics in real-world settings.

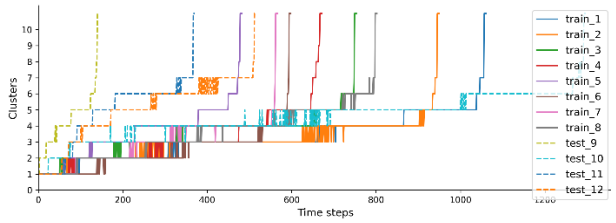


Figure 1 Pre-processed input trajectories

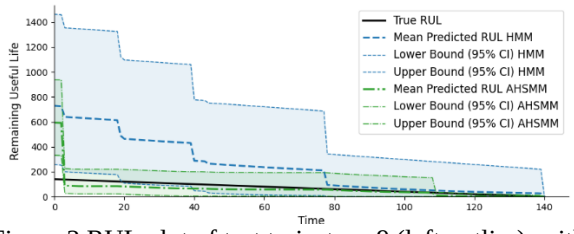


Figure 2 RUL plot of test trajectory 9 (left outlier) with HMM and AHSMM

Table 1 RMSE of RUL for the test trajectories

Trajectory	HMM	HSMM	AHSMM	SLHSMM
1	242.30	308.53	**70.92**	515.17
2	575.11	509.70	**386.45**	536.14
3	188.11	161.71	**155.20**	197.45
4	173.66	139.37	177.67	**93.45**
Average	294.80	279.83	**197.56**	335.55

Table 2 Weighted_Coverage of RUL for the test trajectories

Trajectory	HMM	HSMM	AHSMM	SLHSMM
1	0.000175	0.001914	**0.008190**	0.000000
2	0.000374	0.000000	**0.001281**	0.000000
3	0.003231	0.003146	**0.004292**	0.002872
4	0.002354	**0.004272**	0.003796	0.003980
Average	0.0015336889	0.002333	**0.004390**	0.001713

References

Dong, M., He, D., Banerjee, P., Keller, J., 2006. Equipment health diagnosis and prognosis using hidden semi-Markov models. The International Journal of Advanced Manufacturing Technology 30, 738–749. https://doi.org/10.1007/s00170-005-0111-0

Eleftheroglou, N., 2023. Adaptive Prognostics: A reliable RUL approach. Annual Conference of the PHM Society 15. https://doi.org/10.36001/PHMCONF.2023.V15I1.3495

Eleftheroglou, N., Galanopoulos, G., Loutas, T., 2024. Similarity learning hidden semi-Markov model for adaptive prognostics of composite structures. Reliab Eng Syst Saf 243, 109808. https://doi.org/10.1016/J.RESS.2023.109808

Eleftheroglou, N., Loutas, T., 2016. Fatigue damage diagnostics and prognostics of composites utilizing structural health monitoring data and stochastic processes. Struct Health Monit 15, 473–488. https://doi.org/10.1177/1475921716646579

Eleftheroglou, N., Zarouchas, D., Benedictus, R., 2020. An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. Compos Struct 245, 112386. https://doi.org/10.1016/J.COMPSTRUCT.2020.112386

Fallon, T., Mahal, ;, Hebden, ;, Iain, 2009. F-35 Joint Strike Fighter Structural Prognosis and Health Management an Overview. ICAF 2009, Bridging the Gap between Theory and Operational Practice 1215–1215. https://doi.org/10.1007/978-90-481-2746-7_68

He, W., Jiang, Z., Xiao, T., Xu, Z., Li, Y., 2023. A Survey on Uncertainty Quantification Methods for Deep Learning. ACM Comput Surv 37, 2155. https://doi.org/https://doi.org/10.48550/arXiv.2302.13425

Hu, Y., Miao, X., Si, Y., Pan, E., Zio, E., 2022. Prognostics and health management: A review from the perspectives of design, development and decision. Reliab Eng Syst Saf 217, 108063. https://doi.org/10.1016/J.RESS.2021.108063

Kontogiannis, T., Salinas-Camus, M., Eleftheroglou, N., 2025. Hidden Markov model applications: Aviation prognostics, in: Stochastic Modeling and Statistical Methods: Advances and Applications. Academic Press, pp. 191–213.

Rabiner, L.R., 1989. A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proceedings of the IEEE 77, 257–286. https://doi.org/10.1109/5.18626

Salinas-Camus, M., Eleftheroglou, N., 2024. Uncertainty in Aircraft Turbofan Engine Prognostics on the C-MAPSS Dataset. PHM Society European Conference 8, 10–10. https://doi.org/10.36001/PHME.2024.V8I1.4007

Optimal maintenance planning for offshore wind farms considering time-varying costs and limited manpower

Roby Cremers*, Rommert Dekker**

*Erasmus School of Economics: cremers@ese.eur.nl

**Erasmus School of Economics: rdekker@ese.eur.nl

Abstract This study addresses the challenge of scheduling multi-component maintenance for wind turbines while accounting for seasonal cost fluctuations and limited personnel availability. When the number of planned maintenance activities exceeds available manpower, one approach is to delay certain planned preventive maintenance tasks to a later period, prioritizing components where postponement has the least expected cost impact. Another strategy is to bring maintenance forward, performing tasks earlier when future workforce shortages are expected. We present heuristic-based strategies that strategically delay or advance maintenance tasks when planned activities exceed workforce capacity. A simulation-based evaluation on a set of 80 identical components highlights the perfomance of these heuristics under varying levels of seasonal cost fluctuations and workforce constraints. Our results show that these heuristics generally outperform naïve scheduling approaches and provide insights into which strategies are most effective under which conditions.

1. Introduction

To meet sustainable energy targets, wind energy plays an important role alongside solar power. These sources complement one another well, as solar generation is confined to daylight hours only, especially in summer, while wind energy is also prevalent at night and typically more during the winter. Yet, wind turbines all across the country side faces public resistance due to visual and environmental concerns. As a result, offshore wind has emerged as a preferred alternative, particularly in North-West Europe, where countries are embarking on large scale development of wind farms at sea.

Operating in the offshore environment, however, comes with significant logistical and financial challenges. Conditions on sea are much harder than on land and logistics costs are very high, driving up maintenance costs considerably. The maintenance consists of a mix of corrective maintenance, time-based preventive maintenance and condition-based maintenance. Maintenance is ideally performed during summer months, when weather conditions are better and wind speeds and production losses are lower. However, effective planning is complicated by the high number of tasks, long travel times from the onshore base to the windfarm, and limited availability of qualified personnel.

In this contribution, we present a mathematical model and optimization approaches to assist in operational maintenance planning for offshore wind farms under manpower constraints. We consider both corrective and preventive tasks with time-varying downtime costs. The latter is typical for offshore production and is not considered in other papers on multiple component maintenance planning.

2. Literature

Preventive maintenance optimization has been studied extensively. Classical approaches assume stationary cost structures and unlimited resources. Age replacement policies and block replacement policies are commonly used frameworks for maintenance (Chen and Savits 1992). Recent literature introduces dynamic aspects such as cost seasonality (Schouten et al. 2022) and logistic delays (Zhu et al. 2019).

The recent literature on the topic of offshore wind turbine maintenance has been reviewed by George et al. (2022). They emphasize that the lack of models that account for both limitations on availability of personnel resources and the impact of seasonal variations on scheduling decisions. They emphasized the lack of studies that incorporate personnel constraints. As the former restricts the number of employees that can be deployed, this can be seen as a limitation on the available staff.

3. Problem

We consider an offshore windfarm consisting of many identifical wind turbines. Each wind turbine consists of a number of components which are subject to failure. The lifetime of each component follows a known lifetime

distribution (typically Weibull), indicating a failure rate. A component is assumed to fail independently of its turbine, the time of year, or the condition of other components.

The problem we consider is to schedule the maintenance of multiple identical components. When a component fails, corrective maintenance (CM) is performed immediately. To prevent failures, preventive maintenance (PM) can be scheduled. In addition, each component has a maximum allowable age, after which PM is mandatory. Both CM and PM are assumed to take a fixed amount of time, and PM restores the component to an as-good-as-new condition. Maintenance costs vary seasonally due to fluctuating production losses, with downtime in winter being significantly more expensive than in summer.

The problem is to schedule the PM, when the number of maintenance activities that can be performed each month is limited, reflecting constraints in manpower or spare parts. We assume there is enough manpower to execute CM and PM at the maximum age, but other planned PM may need to be delayed or advanced to meet workload constraints. The objective is to determine a PM schedule that minimizes the long-term average maintenance costs, while given a starting condition and a certain amount of manpower available.

4. Methods

4.1. *Basic single component model*

We first consider a model from Schouten et al. (2022). They considered a single component on a wind turbine, yet with time varying downtime costs, depending on the average windspeed during a year. These costs are incurred both for CM as well as PM as both induce downtime of the wind turbine.

Schouten et al. (2022) model the problem in discrete-time as a Markov Decision Chain (MDC) and provide Mixed-Integer-Programming formulations for optimal Age Replacement and Block Replacement policies with period depending costs (indicated by p-ARP and p-BRP). The first schedules PM at a specific age, the latter when a specific time since the last PM has passed, while both critical levels depend on the time of the year (the states are tuples $i = (i_1, i_2)$ with the first being the period of the year and the second the age / time). This implies that most preventive maintenance is preferably done during the month with the lowest downtime costs: July. In the sequel, we mainly focus on p-ARP.

Our main idea for the multiple component case is to use a decomposition approach. This is not necessarily optimal, but feasible for large numbers of components and moreover, interpretable.

The decomposition approach is based on an idea from Wildeman et al. (1997), where we start with determining an optimal PM moment for every component individually when ignoring any manpower limitation. Next, we determine the expected cost of rescheduling the PM (CM can not be postponed). By comparing those costs for different components, we arrive at a priority ordering of PM in case of a manpower restriction.

4.2. *Delaying PM*

The costs of delaying PM for a component are based on the interpretation of the value function in an MDC. For any two states i, j and policy R in a unichain MDC with average costs criterion, $v_i(R) - v_j(R)$ measures the difference in the total costs over an infinite period when starting in state i rather than starting in state j, if we follow policy R (see Tijms (2003)). Note that we can obtain these value functions from the following set of equations

$$v_i(R) = c_i(R_i) - g(R) + \sum_{j \in \mathscr{I}} \pi_{ij}(R_i) v_j(R), \quad \forall i \in \mathscr{I} \tag{1a}$$

$$v_q = 0, \quad \text{for some state} \quad q. \tag{1b}$$

where $\mathscr{I}$ indicates the state space, $c_i(R)$, the direct costs in state i and $g(R)$ the long run average costs under policy R with action R_i in state i. Now, comparing PM in state i with delaying it for one period (action 0) induces expected costs equal to

$$\Delta(i,0,R) = c_i(0) + \sum_{j \in \mathscr{I}} \pi_{ij}(0) v_j(R) - g(R) - v_i(R) = \sum_{j \in \mathscr{I}} \pi_{ij}(0) v_j(R) - g(R) - v_i(R), \tag{2}$$

where π_{ij} indicates the transition probabilities in the MDC. In case of delaying PM, we do not have direct costs: $c_i(0) = 0$ and the transitions are either to a higher age, $i+1$, or to a failure state. We use this expression to determine the cost for each component separately to defer the execution of PM with one period. Components with the lowest $\Delta(i,0,R)$ values are selected for deferral, while others are maintained. To address cascading delays in highly seasonal scenarios, we extend this logic: when a component's PM is delayed, we also determine the optimal future month to reschedule the task. Specifically, we compute a new planned maintenance time based on the single-component MDP, assuming that PM is no longer allowed in the current period. If the component has not failed before that time, we execute the rescheduled PM. This extension ensures feasibility under seasonal peaks and avoids indefinite postponement of preventive actions. This leads to the algorithm, called DMD (Dynamic Maintenance Delay). During the operational planning, we perform this algorithm repeatedly and it may be that PM for some components is delayed multiple times, depending on the tightness of the manpower constraints.

4.3. *Advancing and delaying PM*

When you know beforehand that you are dealing with a shortage of manpower, it could be beneficial to consider bringing forward PM instead of only delaying. We focus on making a maintenance planning for multiple components with the constraint that there can be no more than a certain number r of PM activities per period. To avoid violation of this constraint, we can either delay or move up PM. Although we keep the value r fixed, everything presented in this section can also be applied to a situation in which r varies over the year. A year consists of N periods, but for ease of notation, the modulo operator (mod) to indicate different years is omitted everywhere. This means that for example $i_1 + v(+x)$ should actually be $i_1 + v(+x) \pmod N$.

In case of using p-ARP, we propose an ongoing process of rescheduling PM. The optimal number of PM activities in a certain period depends on the ages of the different components and is therefore unknown in advance. Hence, we present an algorithm that should be applied at the start of each period. We first determine the optimal maintenance planning for each component $s \in \mathscr{S}$ separately. Denote this optimal policy by $R^{s*}, \forall s \in \mathscr{S}$. At the beginning of a period $i_1 \in \mathscr{I}_1$, we compute the expected number of PM activities for each of the next ℓ periods (ℓ indicating the planning horizon). If this number exceeds r for any of these periods $i_1 + v \in \mathscr{I}_1, v = 0,1,...,\ell$, we reschedule certain activities where we bring them forward or postpone them by a maximum of ℓ months. We implement all changes that relate to the current period i_1 and move to the next period.

Let $\mathbb{1}_{\mathrm{A}}$ be an indicator function that equals one if condition A is true. For a component $s \in \mathscr{S}$ currently in state $i = (i_1, i_2)$, we compute the expected penalty cost of deviating from the original execution date $i_1 + v$ by $x \in \{\text{-}v, \text{-}v+1, ..., \text{-}1, 1, ..., \ell\}$ periods, such that we would maintain when the component is in state $j = (i_1 + v + x, i_2 + v + x)$ instead of in $z = (i_1 + v, i_2 + v)$ as

$$\begin{aligned}\delta^s(i,v,x) = &-\pi_{iz}^{s\,(v)}\Delta^s(z,1,R^s) + \mathbb{1}_{x<0}\cdot\pi_{ij}^{s\,(v+x)}\sum_{k=x+1}^{0}\pi_{(i_1+v+x,0),(i_1+v+k,k-x)}^{s\,(k-x)}\Delta^s((i_1+v+k,k-x),0,R^s)\\ &+\mathbb{1}_{x>0}\sum_{k=0}^{x-1}\pi_{i(i_1+v+k,i_2+v+k)}^{s\,(v+k)}\Delta^s((i_1+v+k,i_2+v+k),0,R^s) + \pi_{ij}^{s\,(v+x)}\Delta^s(j,1,R^s)\end{aligned} \tag{3}$$

The first term is needed in case the PM in period $i_1 + v$ is a rescheduled PM, such that it is not optimal to maintain in this period. We add for each period in between the original and new execution date (including the former) the probability that the component does not fail before we reach this period times the additional cost of not maintaining in this period. If we move up maintenance, this means that it should not fail before and not after the PM. Note that we restrict the action for the periods in between the original and new date to 0 (do nothing). We need the last part of the first line if ℓ is at least the minimum critical maintenance age, as in this case PM in between the two dates could be optimal but we wish to prohibit this. The last term is the probability that the component does not fail before it reaches the period to which we reschedule times the additional cost of PM in this period.

Finally, given a set $\mathscr{I}_1^A$ of possible periods to which we can reschedule the PM, we define the optimal period to perform the PM as the one that yields the lowest difference in costs. We can compute this optimal period

and the corresponding expected additional costs respectively as

$$i_1^{s*} = i_1 + v + x^{s*} = (i_1 + v +_{x\,:\,i_1+v+x\in\mathscr{I}_1^A} \{\delta^s((i_1,i_2),v,x\}) \tag{4}$$

$$\delta^s(i=(i_1,i_2),v) = \min_{x\,:\,i_1+v+x\in\mathscr{I}_1^A} \{\delta^s((i,v,x)\} = \delta^s(i,v,x^{s*}) \tag{5}$$

Given the current period $i_1 \in \mathscr{I}_1$ and the component ages $i_2^s \in \mathscr{I}_2, \forall s \in \mathscr{S}$, we introduce the following algorithm for rescheduling multi-component maintenance for p-ARP under restricted manpower:

DMR (Dynamic Maintenance Reschedule)

Step 0.Let r be the maximum number of PM activities per period and Q the set of periods to which we are not allowed to reschedule PM. Initialize $Q = \emptyset$. Define $\mathscr{S}^D_{i_1+v}, v = 0,1,...,2\ell-1$ as the set of components for which we delayed PM to period i_1+v and that have not failed in the meantime and $\mathscr{S}^B_{i_1+v}, v = 0,1,...,\ell-1$ as the set of components for which we brought forward PM such that we already performed the PM and should not maintain in period i_1+v.
Using these sets, we define $\mathscr{S}^P_{i_1+v}, v = 0,1,...,\ell$ as the sets of components for which we expect to perform PM in period i_1+v. That is for $v = 0,1,...,\ell$, $\forall s \in \mathscr{S}$, add s to $\mathscr{S}^P_{i_1+v}$ if $s \notin S^B_{i_1+v}$, $i_2^s + v \leq \mathbb{E}(X|X \geq i_2^s)$, $s \notin S^P_{i_1+k}$, $k < v$ and either $i_2^s + v \geq t_{i_1+v}$ or $s \in \mathscr{S}^D_{i_1+v}$. Note that we only look at the first planned PM for each component. Here, the expected failure time is

$$\mathbb{E}(X|X \geq i_2^s) = \sum_{j_2=0}^{M} j_2 P(X = j_2+1|X \geq i_2^s) = \sum_{j_2=0}^{i_2^s-1} j_2 + \sum_{j_2=i_2^s}^{M-1} j_2 \frac{P(X\leq j_2+1)-P(X\leq j_2)}{1-P(X\leq i_2^s)} + M\frac{1-P(X\leq M)}{1-P(X\leq i_2^s)}$$

Step 1.Find the smallest $v = 0,1,...,\ell$ for which $\mathscr{S}^P_{i_1+v} > r$ and add period i_1+v to Q. Define the set of available periods to which we can move PM activities $\mathscr{I}_1^A = \{i_1, i_1+1, ..., i_1+v+\ell\} \setminus Q$.
For each component $s \in \mathscr{S}^P_{i_1+v}$, compute the expected additional cost of rescheduling PM, that is $\delta^s((i_1,i_2^s),v)$, using (4) and (5).Sort the components in list $\mathscr{S}^P_{i_1+v}$ in non-decreasing order of their values $\delta^s((i_1,i_2^s),v)$ and go to Step 2. If such a v does not exist, go to Step 3.

Step 2.Subtract component s from the top of list $\mathscr{S}^P_{i_1+v}$. Update $\mathscr{S}^P_{i_1+v} = \mathscr{S}^P_{i_1+v} \setminus \{s\}$ and $\mathscr{S}^P_{i_1^{s*}} = \mathscr{S}^P_{i_1^{s*}} \cup \{s\}$, with i_1^{s*} from equation (4). Repeat this $|\mathscr{S}^P_{i_1+v}| - r$ times. Go back to Step 1.

Step 3.STOP. There are no periods for which the number of planned PM activities exceeds the maximum number r anymore. In the current period i_1, we will maintain the components $\mathscr{S}^P_{i_1}$.

Note that the algorithm terminates after a finite number of steps. Each time we execute steps 1 and 2, we select a period that is not yet in the list Q and we add that period to Q. We ensure that the number of planned PM activities for that period does not exceed r and only add PMs to periods not in Q, which means the algorithm does not cycle. Since the number of periods is finite, this procedure must terminate.

Something to bear in mind is the fact that the order in which we visit the different periods will most likely influence the obtained planning. We stick to starting in the period closest to the current period for which the number of planned PM exceeds the threshold. The reason being that there exist $\ell!$ possible orders in which to visit all periods, which means it is very time-consuming to consider all possibilities.

5. Experiments and Results

In this section, we show how our *DMD* and *DMR* heuristics for rescheduling PM perform using simulation. We assume that the downtime costs vary with sinusoid pattern through the year with a maximum increase/decrease of Δ percentage. This is different from the function for the additional costs of deviating from the optimal policy R^{s*} by taking action a, which we denote by $\Delta^s(i,a,R^{s*})$.

We consider 80 identical components with independent lifetimes that follow a Weibull distribution with scale parameter α and shape parameter β. We express the time in months and for the maximum component age M we use the rounded up 99.9th percentile of the lifetime distribution. All heuristics are tested using a batch of 10 simulation runs, consisting of 50.000 periods, as we are interested the long-term behavior. We measure significance of results using t-tests. Finally, we did not make use of a warm-up period to remove the

Table 1: Results of delay of PM for p-ARP (*DMD*), with $\alpha = 3$ years, $\beta = 2$ and $\ell = 1$. Yearly costs, in 1000 €, difference in costs when including restrictions on manpower and number of times we delay PM for the set of 80 components and percentage of times this directly leads to a failure.

	costs	additional costs			#PM	# delayed PM / % of times delay leads to CM					
Δ	$r = 80$	$r = 60$	$r = 40$	$r = 20$	$r = 80$	$r = 60$		$r = 40$		$r = 20$	
0%	13.516	-0.00%	-0.00%	-0.00%	173 128	7	0%	18	0.6%	37	2.4%
10%	13.240	0.00%	0.00%	0.04%**	185 521	7	0%	21	0.5%	13 918	2.5%
20%	12.713	0.03%	0.05%	0.04%	296 039	17 236	1.9%	47 019	1.9%	78 207	2.0%
30%	11.784	0.17%**	0.46%**	4.03%**	297 566	48 327	1.8%	132 903	1.8%	194 598	3.0%
40%	10.852	0.26%**	0.69%**	8.37%**	301 088	51 725	1.8%	136 010	1.8%	222 541	3.1%
50%	9.905	0.38%**	1.01%**	11.80%**	303 654	54 167	1.7%	138 321	1.8%	224 492	3.1%

* and ** denote the cost difference is significantly different from 0, using a two-sided t-test, *p<0.05, **p<0.01.

Table 2: Results of rescheduling PM for p-ARP (*DMR*) with $\alpha = 3$ years and $\beta = 2$. Yearly costs, in 1000 €, difference in costs when including restrictions on manpower and the difference in comparison with *DMD*. The results are for $\ell = 3$ but are exactly the same for $\ell = 5, 7$.

	Costs	additional costs			additional costs compared to *DMD* (with $\ell = 1$)		
Δ	$r = 80$	$r = 60$	$r = 40$	$r = 20$	$r = 60$	$r = 40$	$r = 20$
0%	13.516	0.00%	0.00%	-0.00%	0.00% (0.000)	0.00% (0.000)	0.00% (0.000)
10%	13.240	0.00%	0.00%	0.01%	-0.00% (0.003)	-0.00% (0.005)	-0.03%† (0.035)
20%	12.713	0.46%**	1.37%**	0.40%**	0.43%§§ (0.082)	1.32%§§ (0.165)	0.36%§§ (0.131)
30%	11.784	0.83%**	2.29%**	3.04%**	0.66%§§ (0.028)	1.83%§§ (0.019)	-0.96%†† (0.073)
40%	10.852	1.28%**	3.39%**	4.56%**	1.02%§§ (0.030)	2.68%§§ (0.028)	-3.51%†† (0.061)
50%	9.905	1.90%**	4.75%**	6.42%**	1.51%§§ (0.038)	3.70%§§ (0.034)	-4.82%†† (0.103)

* and ** denote the cost difference is significantly different from 0, using a two-sided t-test, *p<0.05, **p<0.01.
† and †† denote that the *DMR* heuristic performed significantly better than *DMD* heuristic and § and §§ denote that the *DMR* heuristic performed significantly worse, based on a one-sided t-test, †,§ p<0.05, ††,§§ p<0.01.

effect of the initial conditions as the number of periods is quite large and we are not interested in the outputs for a single simulation run, but rather in how the outcomes differ amongst policies and heuristics. We started in the first month of the year, let each component have age one month and used common random numbers: for each combination of parameter values and heuristic we drew the exact same sequence of failure times for each component for the same simulation run.

The results for *DMD* for $\alpha = 3$ years and $\ell = 1$ are summarized in Table 1 and those for *DMR* for $\alpha = 3$ years and $\ell = 3$ (same as for $\ell = 5, 7$) can be found in Table 2. The additional costs are computed as the relative difference under a restricted number of permitted PM operations per period when applying the heuristic, compared to the costs without this constraint. The results for *DMD* with $\ell = 3, 5$ are similar.

Tables 1 and 2 show that the additional costs become larger when the constraint on available manpower is more restrictive and when there are larger cost fluctuations over the year (i.e. larger Δ). Furthermore, in Table 1, it can be seen that the number of times we delay PM increases with Δ. This is because all planned maintenance in the optimal policy is concentrated around one month a year when cost fluctuations are large. The number

Table 3: Results of delaying and rescheduling PM for p-ARP, with $\alpha = 3$ years and $\beta = 2$. Percentage additional costs when including restrictions on manpower with respect to the optimal planning without this constraint, divided into a part due to more/less maintenance activities and a part due to maintenance in more/less expensive periods. For delay we used $\ell = 1$ and for reschedule $\ell = 3$ but for the latter results are exactly the same for $\ell = 5, 7$.

	additional costs due to more or less maintenance / more or less expensive periods					
	DMD (delay)			*DMR* (reschedule)		
Δ	$r = 60$	$r = 40$	$r = 20$	$r = 60$	$r = 40$	$r = 20$
0%	-0.00% / 0.00%	-0.00% /0.00%	-0.00% /0.00%	0.00% / 0.00%	0.00% / 0.00%	-0.00% / 0.00%
10%	0.00% /0.00%	0.00% /0.00%	-0.02%* /0.06%**	0.00% /-0.00%	0.00% /-0.00%	0.11%**/-0.09%**
20%	-1.06%**/1.09%**	-2.93%**/2.98%**	-4.92%**/ 4.96%**	-0.10%**/ 0.56%**	-0.07%**/ 1.44%**	-4.08%**/ 4.47%**
30%	-0.03%**/0.19%**	-0.07%**/0.53%**	-2.05%**/6.08%**	-0.09%**/ 0.91%**	-0.02%**/ 2.32%**	-0.14%**/ 3.18%**
40%	-0.07%**/0.32%**	-0.16%**/0.85%**	-0.41%**/8.77%**	-0.04%**/ 1.32%**	0.05%**/ 3.34%**	-0.21%**/ 4.77%**
50%	-0.09%**/0.47%**	-0.21%**/1.22%**	-0.57%**/12.37%**	-0.02%**/ 1.91%**	0.11%**/ 4.64%**	-0.26%**/ 6.68%**

* and ** denote the cost difference is significantly different from zero, based on a two-sided t-test, *p<0.05, **p<0.01.

of delays are counted in such a way that each component can be delayed only once in a lifetime. Note that we never have significant cost savings, which makes sense as we compare to an optimal policy.

In Table 3, we made a distinction between the additional costs due to the fact that more or less maintenance is needed and the difference due to the fact that we maintain in more or less expensive periods. We measure the first part by calculating the difference in costs when using a constant cost function ($\Delta = 0\%$). The second part is then the remainder of the difference. It turns out that on average we maintain less often but in more expensive periods when we delay PM. This result is in line with Table 1, which shows that a component only fails in less than 3% of the times that we delay PM. Delaying thus causes limited additional CM while delaying without failure leads to less maintenance. Furthermore, the increasing additional costs when Δ is large are mainly caused by the fact that we maintain in periods with high costs.

6. Conclusions

Two heuristics are presented for scheduling PM for multiple components under time-varying costs and manpower restrictions. One which only delays (DMD) and the other (DMR) which can also advance. DMD works better than DMR for not too strict manpower restrictions. Our findings highlight the importance of incorporating seasonality and capacity limits into maintenance planning.

References

Chen CS and Savits TH (1992) Optimal age and block replacement for a general maintenance model., *Probability in the Engineering and Informational Sciences*, 6.1, 81-98.

George B, Loo J and Jie W (2022) Recent advances and future trends on maintenance strategies and optimisation solution techniques for offshore sector, *Ocean Engineering*, 250, 110986.

Schouten TN, Dekker R and Eruguz, AS (2022) Maintenance optimization for a single wind turbine component under time-varying costs, *European Journal of Operational Research*, 300, 979–991.

Wildeman RE, Dekker R and Smit, ACJM (1997) A dynamic policy for grouping maintenance activities, *European Journal of Operational Research*, 99, 530-551.

Zhu W, Castanier B and Bettayeb B (2019) A dynamic programming-based maintenance model of offshore wind turbine considering logistic delay and weather condition, *Reliability Engineering & System Safety*, 190, 106512.

Redefining Prognostic Essentials: Focus on Reliability, Robustness and Feasibility

N. Eleftheroglou*

**Intelligent Sustainable Prognostics Group, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, Netherlands, N.Eleftheroglou@tudelft.nl*

Abstract. Prognostics play a pivotal role in predictive maintenance by forecasting the future health and performance of assets based on their current condition and operational context. While traditional research has primarily focused on improving the accuracy of Remaining Useful Life (RUL) predictions, this work argues that feasibility, robustness, and reliability are equally critical, especially given the increasing complexity of modern assets in both their design and operational use. Motivated by these considerations, this study compares various state-of-the-art prognostic methodologies using condition monitoring data from the aerospace sector, highlighting current limitations, advantages, and future directions in the field.

Keywords: Prognostics and Health Management, Remaining Useful Life, Reliability, Robustness, Feasibility

In predictive maintenance, prognostics is essential for estimating an asset's future condition and performance based on its current state and operating environment. This study argues that, to effectively support maintenance and operational decision-making, prognostic methodologies must satisfy three key criteria: feasibility, robustness, and reliability.

Feasibility entails the ability of a prognostic methodology to be trained with limited run-to-failure data, as acquiring extensive degradation data is often cost-prohibitive. For instance, in aviation structures, only a single full-scale fatigue experiment is typically conducted during the certification process, providing just one degradation history for training prognostic models. Secondly, prognostics must demonstrate robustness by ensuring reliable performance across a broad spectrum of operational conditions not seen during the training process. Thirdly, prognostics must be reliable, particularly given their inherent uncertainties. These uncertainties stem from various sources, including past, present, and future operating conditions, as well as modeling limitations, all of which contribute to the overall uncertainty in prognostic outputs (Salinas-Camus et al., 2024). In the case of Remaining Useful Life (RUL) prediction, this means treating RUL as a random variable. To ensure reliability, the predicted mean RUL should closely match the actual RUL, and the estimated uncertainty should provide decision-makers with actionable insights.

This study explores the aforementioned prognostic criteria using state-of-the-art methodologies, including the Non-Homogeneous Hidden Semi-Markov Model (NHHSMM) (Eleftheroglou et al., 2016), the Adaptive NHHSMM (ANHHSMM) (Eleftheroglou et al., 2020), and the Similarity Learning Hidden Semi-Markov Model (SLHSMM) (Eleftheroglou et al., 2024). The performance and behavior of these methodologies will be evaluated using condition monitoring data from aerospace structures.

The training dataset consists of strain data from aerospace specimens subjected to fatigue loading conditions (Eleftheroglou, 2023). In contrast, the validation-testing dataset includes strain data from specimens exposed to both fatigue and in-situ impact loading, as well as one specimen with a manufacturing defect. These variations represent unforeseen conditions not present in the training data. Figure 1 illustrates the axial strain degradation histories for both the training and testing datasets. These differences serve as a basis for evaluating the robustness of the proposed state-of-the-art prognostic models. Reliability is assessed by examining the accuracy of the estimated mean RUL values and their associated uncertainty. Finally, feasibility is evaluated by progressively reducing the size of the training dataset to understand how varying data availability impacts prognostic reliability.

In Figure 2, the RUL predictions of the ANHHSMM are presented and compared with those of the NHHSMM for the specimen with the manufacturing defect, showcasing the increased robustness and reliability of the ANHHSMM.

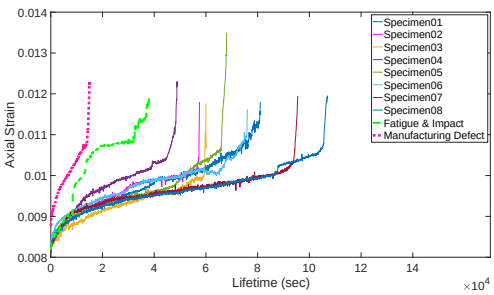


Figure 1 Axial strain condition monitoring data

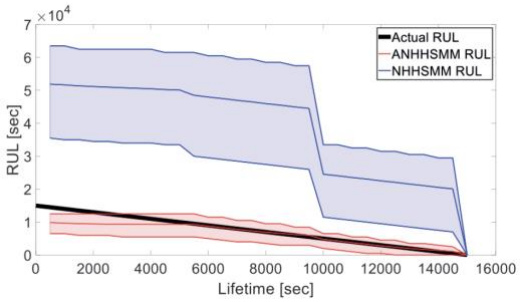


Figure 2 RUL plot of the manufacturing-defected specimen

In conclusion, this study highlights the importance of developing prognostic methodologies that are robust, reliable, and feasible, particularly given the increasing complexity of modern assets and the diverse operational conditions in which they are used. To address these challenges, state-of-the-art prognostic approaches were compared and validated using a dataset derived from aerospace specimens. By incorporating domain adaptation and similarity learning, the reliability, robustness, and feasibility of prognostics can be significantly enhanced—resulting in models that better support operational and maintenance decision-making.

References

Eleftheroglou, N., & Loutas, T. (2016). Fatigue damage diagnostics and prognostics of composites utilizing structural health monitoring data and stochastic processes. *Structural Health Monitoring, 15 (4)*. https://doi.org/10.1177/1475921716646579

Eleftheroglou, N., Zarouchas, D., & Benedictus, R. (2020). An adaptive probabilistic data-driven methodology for prognosis of the fatigue life of composite structures. *Composite Structures*, 245, 112386. https://doi.org/10.1016/j.compstruct.2020.112386

Eleftheroglou, N., Galanopoulos, G., & Loutas, T. (2024). Similarity learning hidden semi-Markov model for adaptive prognostics of composite structures. *Reliability Engineering & System Safety*, *243*, 109808. https://doi.org/10.1016/J.RESS.2023.109808

Eleftheroglou, N. (2023). Adaptive Prognostics: A reliable RUL approach. *PHM Society Annual Conference*, 15(1). https://doi.org/10.36001/phmconf.2023.v15i1.3495

Salinas-Camus, M., & Eleftheroglou, N. (2024). Uncertainty in Aircraft Turbofan Engine Prognostics on the C-MAPSS Dataset. *PHM Society European Conference*, *8*(1), 10–10. https://doi.org/10.36001/PHME.2024.V8I1.4007

Evaluation of Mission Reliability for a System in the Presence of Spare Parts

Reem Alrashed*'**, F.P.A. Coolen***, T. Coolen-Maturi****, and G.J. van Houtum*****

* Department of Mathematical Sciences, Durham University, Durham, United Kingdom: reem.alrashed@durham.ac.uk
** Department of Mathematics, Prince Sattam bin Abdulaziz University, Kingdom of Saudi Arabia: r.alsalman@psau.edu.sa

*** Department of Mathematical Sciences, Durham University, Durham, United Kingdom: frank.coolen@durham.ac.uk

**** Department of Mathematical Sciences, Durham University, Durham, United Kingdom: tahani.maturi@durham.ac.uk

***** Department Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, Eindhoven, the Netherlands: G.J.v.Houtum@tue.nl

Abstract: This study aims to evaluate the reliability of a system in the presence of spare parts using a Markov stochastic process and survival signature methodology. It presents an analysis of a system with multiple types of components, assuming an Erlang distribution function for the failure times of components. During the mission, spare parts are available to replace failed components immediately, and repairs are not possible. The study introduces the method for a system with one component type and then generalizes it to a system with multiple types. Numerical examples illustrate the application of the proposed method.

Keywords: System reliability, Spare parts, Mission reliability.

Introduction

The concept of reliability refers to an object's ability to accomplish missions and achieve requirements under specific conditions [1]. The continuous operation might be interrupted or stopped due to failures, such as equipment failure or structural collapse. Therefore, engineering system design requirements have become more complex and robust, prompting numerous scholars to focus on enhancing systems' reliability and performance efficiency. This study uses the survival signature to represent the system structure. The survival signature [2] is the probability that a system functions given the number of functioning components of each type. It enables separating the crucial information about the system structure from the components' failure times and reduces calculation complexity compared to employing the whole structure function. The survival signature and the component's failure time distributions contain all the information required to calculate system reliability metrics, such as the reliability function that determines the probability that the system survives for a certain length of time.

Methodology

This study presents an analysis of a system with multiple types of components, supposing that the distribution function of the failure time of components is an Erlang distribution. The failure times of all components of a type are independent and identically distributed (iid), while the failure times of components of different types are independent of each other. The lifetime of a component is assumed to consist of a constant number of independent phases, each following an exponential distribution with a constant rate. Since the Erlang distribution represents the parametric sum of independent exponentially distributed random variables with a fixed rate, it is used for modelling the lifetime distribution of components. During the mission, spare parts are available for some component types,

allowing for the immediate replacement of failed components [3]. However, when all available spare parts have been used, further failed components cannot be replaced or repaired. Each component has two states: operational state and failure state. System states are described as aggregate states for each type, which show the number of installed components in the system whose lifetime is in the j-th phase, the number of installed components that have failed, and the available spares. State space is classified into subsets of states relying on the number of spare parts available in stock, which are ordered lexicographically. The behaviour of all components of a type is jointly described using a Markov process.. At time 0, the Markov process starts in the optimal state where all installed components in the first phase are functioning, with a full stock of spare parts. The transition rates between states are specified by the generator matrix of a Markov process, which includes submatrices. Each submatrix represents the transitions between two subsets and within each subset. Transitions occur within a subset when one of the installed components completes a phase of its lifetime. Transitions occur between two subsets when one of the installed components fails and is replaced by a spare part. After that, the uniformization technique is applied to analyze the Markov process at transition moments, supposing that the transitions occur according to a Poisson process with a constant rate λ. The transition probability matrix P is computed, and then the probabilities of the process being in a particular state after a specific transition and at time t are determined. The mission reliability of the system at time t is computed after the system's survival signature and the probability distribution of the number of functioning components of a given type at time t are specified. This method is introduced for a system with just one component type, after which the application is presented to a system with multiple component types.

Conclusion

The proposed method provides a dynamic structure that enables its implementation in various operating scenarios. In this study, it is implemented to assess the reliability of a 6-component system of two types, a 3-component system of one type in four various configurations, and a network with 4 nodes and 7 links, incorporating the concept of spare parts into the probabilistic analysis. It accurately depicts the transitional behaviour of states within a system and offers a flexible representation of failure and replacement dynamics. Therefore, it can be used as input to make decisions about the number of spare parts available to meet some mission reliability requirements. Its applicability may extend to assessing the reliability of complex systems that consist of multiple components or configurations across various operational scenarios.

Acknowledgement

Reem Alrashed expresses gratitude to Prince Sattam Bin Abdulaziz University and the Saudi Arabian Cultural Bureau in the United Kingdom. Durham University is acknowledged for its valuable support and stimulating academic environment.

References:

[1] Birolini, A. (2007). *Reliability Engineering* (Vol. 5). Berlin: Springer.

[2] Coolen, F.P.A. and Coolen-Maturi, T. (2012) 'Generalizing the signature to systems with multiple types of components', in Zamojski, W., Mazurkiewicz, J., Sugier, J., Walkowiak, T. and Kacprzyk, J. (eds.) *Complex Systems and Dependability*. Springer, Berlin, pp. 115-130.

[3] Van Houtum, G. J., and Kranenburg, B. (2015). *Spare Parts Inventory Control under System Availability Constraints* (Vol. 227). New York: Springer.

On the Value of Predictive Spare Parts Printing as the Second Supply Mode

Ipek Tanil*, Collin Drent**, Rob Basten***

*Eindhoven University of Technology, 5612 AZ Eindhoven, *m.i.tanil@tue.nl, **c.drent@tue.nl, ***R.J.I.Basten@tue.nl*

Abstract. We investigate the use of additive manufacturing (AM) for predictive on-site spare parts printing, leveraging sensor data to optimize decision-making in spare parts provisioning. AM enables extremely short lead times but is more expensive than regular replenishments, leading to a trade-off. We examine this trade-off for a partially observable degrading system with embedded sensors providing classification data. Using a Bayesian procedure incorporating the classifier confusion matrix, a decision maker sequentially chooses to print, replenish, or do nothing. We model this as a partially observable Markov decision process (POMDP) and derive sensor guidelines to support AM integration in spare parts provisioning.

Keywords: Additive manufacturing, spare parts, predictive printing, Bayesian learning, Markov decision process

We develop a model for a system, such as a vehicle, with an embedded sensor that relays classification data on the condition of a critical component. The objective of this work is to analyze the supply decision of whether to order or print a spare part based on the sensor information. We assume that the spare part can be sourced via either ordering from a manufacturer or printing on-site using the AM capability, a setup known as dual sourcing in the academic literature. Printing is assumed to be faster but more expensive than ordering, which leads to a time and cost trade-off. In addition, printing or ordering the part and receiving it earlier than the component failure may lead to unnecessary holding cost while receiving the part later than the failure leads to downtime cost. The core research problem is to make optimal sourcing decisions under sensor uncertainty to minimize the total expected cost that consists of printing, ordering, holding and downtime costs.

This work connects two strands of literature. The first relates to advance demand information (ADI) for spare parts inventory management. Several studies use sensor-driven condition monitoring to forecast equipment failure and adjust maintenance and inventory decisions accordingly (e.g., Chen et al. 2021, De Pater and Mitici 2021, Feng et al. 2023). These approaches use predictive analytics to improve availability and reduce costs. The second strand focuses on dual sourcing in spare parts supply chains. Previous work (e.g., Akkerman et al. 2024, Knofius et al. 2021) examines combining conventional manufacturing and AM, showing that this can reduce lead times and increase flexibility. Yee et al. (2024) develop a dual sourcing inventory model under non-stationary and partially observable demand, using a POMDP framework to manage orders from suppliers with varying lead times and costs. Their dual base-stock policy adjusts orders based on belief states inferred from demand observations. While their focus is on stochastic demand management, our work addresses predictive spare parts provisioning using AM, where decisions to print, order, or take no action are based on real-time sensor data and system degradation. Song and Zhang (2024) study predictive spare parts printing using the internet of things, but focus on printing for stock rather than on-demand or dual sourcing. Our model incorporates both printing and ordering options under uncertainty, enabled by a POMDP framework, and supports real-time supply decisions driven by evolving system conditions.

We model the component degradation with three states: good, defective, and failed. The good and defective states are operative and unobservable, while the failed state is self-announcing and non-operative. The sensor provides an imperfect signal on the component's condition due to the false positive and false negatives. Using Bayes' theorem and the sensor's confusion matrix, we infer our belief on the degradation state in each time period. At every time period, the decision-maker decides whether to print, order, or take no action.

We formulate the decision problem as a finite-horizon POMDP and solve it using backward induction. We illustrate the optimal policy using heatmaps that show sourcing decisions across belief states and inventory levels. We observe a clear threshold behavior: if no part is on hand, printing is optimal regardless of the belief. If one

part is in operation and no part is in stock, printing is optimal below a certain belief threshold, beyond which ordering becomes optimal. If there is enough inventory available, taking no action is optimal.

We conduct a numerical analysis over a range of sensor true positive (α) and true negative (β) rates. We find that higher α and β values, indicating more accurate sensor data, lead to lower total expected cost. However, the impact of α and β is higher when the transition probability from defective state to failed state is higher and when printing is the only supply option. We evaluate several performance measures derived from the confusion matrix and calculate their correlation with the total expected cost. For our example case, we find that the false omission rate (FOR) and the negative predictive value (NPV) are the most informative measures for sensor selection. Sensors with low FOR or high NPV consistently lead to low expected total cost.

References

Akkerman F, Knofius N, van der Heijden M, Mes M (2024) Solving dual sourcing problems with supply mode dependent failure rates. *arXiv preprint arXiv:2410.03887*.

Chen C, Zhu ZH, Shi J, Lu N, Jiang B (2021) Dynamic predictive maintenance scheduling using deep learning ensemble for system health prognostics. *IEEE Sensors Journal* 21(23):26878–26891.

De Pater I, Mitici M (2021) Predictive maintenance for multi-component systems of repairables with remaining-useful-life prognostics and a limited stock of spare components. *Reliability Engineering & System Safety* 214:107761.

Feng H, Zhang SH, Zhang Y (2023) Managing production-inventory-maintenance systems with condition monitoring. *European Journal of Operational Research* 310(2):698–711.

Knofius N, van der Heijden MC, Sleptchenko A, Zijm WH (2021) Improving effectiveness of spare parts supply by additive manufacturing as dual sourcing option. *OR spectrum* 43:189–221.

Song, J. S., & Zhang, Y. (2025). Predictive Three-Dimensional Printing of Spare Parts with Internet of Things. *Management Science*, *71*(3), 1925-1943.

Yee, H., van Staden, H. E., & Boute, R. N. (2024). Dual sourcing under non-stationary demand and partial observability. *European Journal of Operational Research*, *314*(1), 94-110.

ReLife: an open-source Python library for data-driven decision-making in asset management based on reliability theory

A. Frémond* T. Guillon** W. Grison***

* RTE, La Défense, 7 C Place du DÔME, immeuble Window, 92073 Puteaux. *alexis.fremond@rte-france.com*
** RTE, La Défense, 7 C Place du DÔME, immeuble Window, 92073 Puteaux. *thomas.guillon@rte-france.com*
*** *Artelys, 81 Rue Saint-Lazare, 75009 Paris. william.grison@artelys.com*

Abstract. In the context of aging infrastructures and climate change, asset managers face critical investment decisions. ReLife is an open-source Python library that aims to provide various statistical methods, including survival analysis and advanced reliability models like condition-based maintenance and optimal asset replacement. The library also provides a renewal process sampling, includes a renewal equation solver, and supports Lebesgue-Stieltjes integration. ReLife helps asset managers select optimal maintenance policies that minimize expected equivalent annual costs and socio-economic impacts while serving as a collaborative platform for reliability professionals and researchers.

Keywords: reliability, survival analysis, renewal process, gamma process, non-homogeneous Poisson process, optimal maintenance policy, open-source software

ReLife is an open-source Python library designed to help asset managers perform quantitative risk analysis to evaluate and compare risk control options based on probability models along with socio-economic criteria When applicable, quantitative risk analysis is highly desirable for justifying investment decisions in accordance with the regulator, maximizing asset value and anticipating stocks and budgets. ReLife is the only library that offers a complete workflow, beginning with failure data and lifetime distribution estimation, followed by the comparison of various maintenance policies using reliability, engineering economy and renewal theory, for predicting the residual consequences of the optimal policy.

Due to intrinsic wear and tear, industrial assets tend to deteriorate over time and eventually fail, necessitating replacement. A common example in the electric transmission network is the breakage of insulator strings. As corrosion builds up over time, the strings deteriorate and ultimately break. ReLife enables the calibration of a variety of reliability models such as asset lifetime distribution or Gamma deterioration processes.

In contrast to non-repairable failures, ReLife also handles recurrent events. These are characterized by repairable failures that do not necessitate the replacement of the asset. A common example in the electric network is sulfur hexafluoride (SF_6) leaks in metal-enclosed substations. When gas pressure falls below a critical level, the compartment is shut down for a minimal repair, which involves refilling the gas. Over time, as the substation deteriorates, SF_6 leaks increase, leading to more frequent minimal repairs. ReLife relies on non-homogeneous Poisson processes for modeling the dynamics of recurrent events with minimal repairs.

These reliability models are then used to compare renewal policies, such as age-based preventive replacement or condition-based replacement. Building on the renewal reward processes, the expected equivalent annual costs of each policy are computed, optimized and compared.

Reliability models in ReLife are provided with failure data and encapsulated in Python objects that handle both simulation and estimation. These encapsulated models are then passed, along with specific cost structures, to maintenance policy objects. These policy objects are responsible for parameter optimization (optimal age of replacement for an age-based replacement policy, optimal inspection frequency and deterioration threshold for preventive replacement upon inspection policy, etc.) and the computation of the associated expected equivalent annuals costs, as shown in Figure 1.

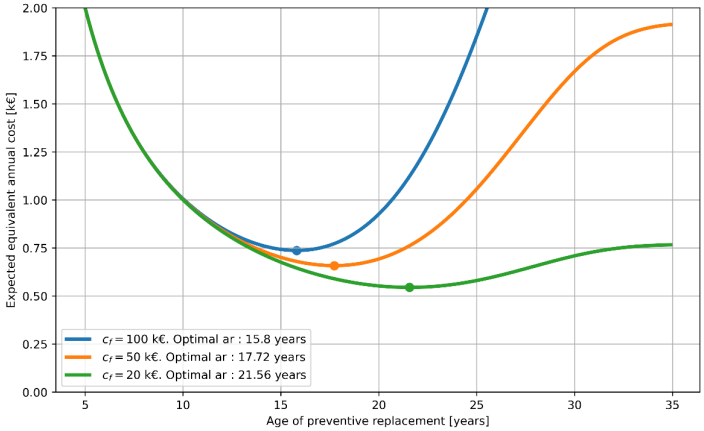


Figure 1. Evolution of the expected equivalent annual cost for different costs of replacement on failure.

Moreover, ReLife supports Lebesgue-Stieltjes integration and implements a fast solver to solve for the renewal equation

$$m(t) = F(t) + \int_0^t m(t-x)dF(x)$$

where m is the renewal function and F is the cumulative distribution for the cycle length. This equation is central to reliability in asset management, as it is used to project the number of replacements and annual costs induced by the replacement policy to guide decision-making, as shown in Figure 2.

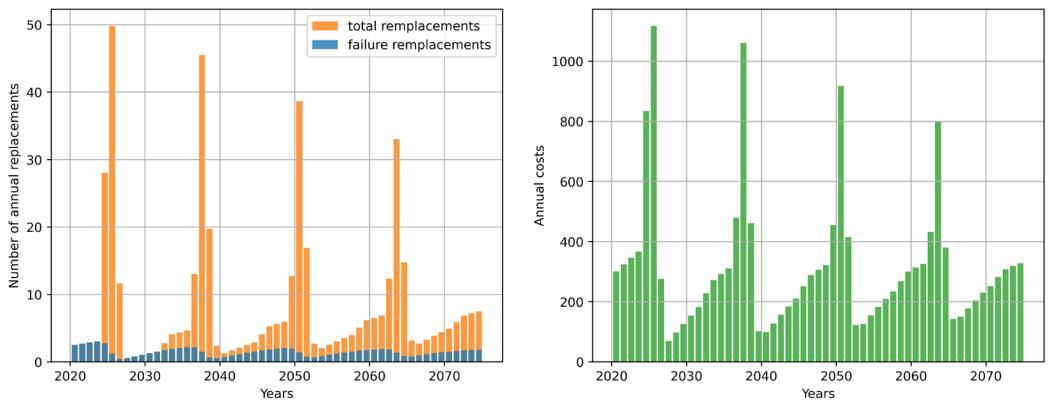


Figure 2. Expected annual replacements (left) and costs (right) for an age-based replacement policy.

ReLife is an open-source initiative designed to foster collaboration among reliability professionals, researchers, and practitioners in asset management. The library has recently undergone significant refactoring to streamline contributions, testing, and code maintenance. Future developments aim to expand the library's capabilities with new reliability models and maintenance policies, such as virtual age models for imperfect repairs, Cox models or causal inference techniques.

References

Barlow R. E., Hunter L.C. (1960) Optimum preventive maintenance policies.
Operations Research **8**, 90-100.

Barlow, R. E., Proschan, F. (1996). Mathematical theory of reliability.
Society for Industrial and Applied Mathematics.

Gertsbakh, I. B. (2000). Reliability theory: with applications to preventive maintenance.
Springer Science & Business Media.

Nakagawa, T. (2006). Maintenance theory of reliability.
Springer Science & Business Media.

Van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance.
Reliability Engineering & System Safety, 94(1), 2-21.

Lawless, J. F. (2011). Statistical models and methods for lifetime data.
John Wiley & Sons.

From prognostics-based predictive maintenance to end-to-end predictive maintenance for complex systems

Mihaela Mitici

Faculty of Science, Utrecht University, Heidelberglaan 8, Utrecht, the Netherlands, m.a.mitici@uu.nl

Abstract. Several studies on predictive maintenance planning make use of condition monitoring data and machine learning models to estimate the remaining useful life of assets. Further, these estimates become the basis for maintenance planning. We discuss the benefits of considering an end-to-end approach for maintenance planning, where data is directly integrated with the specification of the maintenance planning model. We consider a case study for electric aircraft. We show that an end-to-end approach leads to lower maintenance costs and fewer failures, compared with existing maintenance planning approaches that rely on data-driven RUL estimators.

Keywords: predictive maintenance, prognostics, decision-making, end-to-end maintenance

Modern assets such as wind turbines, batteries, aircraft engines are continuously monitored by sensors. These sensors generate over time multi-variate time-series of measurements, which are used to anticipate faults and estimate the remaining useful life (RUL) of the assets, i.e., RUL prognostics. For this, advanced data-driven models and machine learning approaches are considered (Jamshidi, A. et. al., 2018; Mitici, M. et. al. 2023). In turn, such prognostics are used to optimize the maintenance planning of the assets, i.e., predictive maintenance.

Several existing studies on predictive maintenance develop data-driven machine learning models for RUL prognostics, where the focus is on achieving a high accuracy for the RUL estimation (Liu, X. et. al., 2021; Mitici, M. et. al., 2023). However, when the obtained RUL prognostics are integrated into optimization models for maintenance planning, the maintenance decisions and costs may not necessarily improve. This is because the RUL prognostics are obtained in isolation from the maintenance planning model.

In this paper we discuss the performance gap between i) a predictive maintenance approach where RUL prognostics are obtained for complex assets (Lithium-ion batteries) using time-series of measurements, while these prognostics are further integrated into a linear programming model to optimize the replacement time of the assets, and ii) a predictive maintenance approach where the time-series of measurements are directly considered in a linear programming model for end-to-end predictive maintenance. Below we introduce a general model structure for each of the two planning approaches.

We consider a multi-variate time series of measurements $\boldsymbol{x}_a$ for an asset a.

i) Predictive maintenance planning that integrates that integrates RUL prognostics (RUL-M)

Let $g(x_a, \theta)$ be a machine learning model (regressor) that processes the measurements and outputs an estimate of the RUL, $\hat{r}_a$, of the asset such that the loss function L is minimized, i.e., the accuracy of the regressor is maximized, with $L = \left|\left|r - \hat{r}\right|\right|_2$.

Next, the RUL prognostics are integrated in a linear programming model that minimizes the cost with asset replacement by aiming to identify optimal replacement times for the asset. Let the replacement be schedule any day $d \in \{1, 2, \dots, k\}$ from a current day d_0. Consider the decision variables:

$$y_{ad} = \begin{cases} 1, & if\ replacement\ \ for\ asset\ a\ is\ performed\ on\ day\ d\ away\ from\ a\ current\ day\ d_0 \\ 0, & otherwise \end{cases}$$

$$y_a^{postpone} = \begin{cases} 1, & if\ replacement\ for\ asset\ a\ is\ not\ performed\ on\ any\ k\ days\ from\ current\ day\ d_0 \\ 0, & otherwise \end{cases}$$

Then, the cost of maintenance $\hat{c}_a$ when considering the RUL prognostic $\hat{r}_a$ is:

$$\hat{c}_a = \left(c(\widehat{r_a}, 1), c(\widehat{r_a}, 2), \dots, c(\widehat{r_a}, k), \hat{c}^{postpone}(\widehat{r_a})\right)^T.$$

Let $\hat{\boldsymbol{c}} = (\widehat{\boldsymbol{c}_a})_{a \in A}$ denote all maintenance cost for a set of assets $a \in A$.

The overall optimization model for multiple assets is now:

$$\min_{\mathrm{y}} \hat{\boldsymbol{c}}^{\boldsymbol{T}} \boldsymbol{y}$$

$$1^T \boldsymbol{y}_a = 1,\ \forall a \in A$$
$$\sum_{a \in A} y_{ad} = 1,\ \forall d \in \{1, 2, \dots, k\}$$
$$\boldsymbol{y}_a \in \{0,1,\}^{k+1},\ \forall a \in A$$

ii) End-to-end predictive maintenance planning (E2E-M)

For the end-to-end predictive maintenance, we consider the same cost structure as in the approach in i). However, in this case, the cost coefficients are estimated directly using a machine learning regression $f(\boldsymbol{x}_a, \theta')$ with the loss function $L' = \boldsymbol{c}_a{}^{\boldsymbol{T}} \boldsymbol{y}_a^*(\hat{\boldsymbol{c}}) - \boldsymbol{c}_a{}^{\boldsymbol{T}} \boldsymbol{y}_a^*(\boldsymbol{c})$, i.e., the cost gap between the true optimal solution and the one acquired using $\hat{\boldsymbol{c}}$.

In this approach, the maintenance optimization model receives feedback on the decision loss from the maintenance planning problem. Based on this feedback, the parameters $\theta'$of the regressor f are updated.

As case study, we apply the two approaches for predictive maintenance planning of Lithium-ion batteries of electric Vertical Take-off and landing (EVTOL) aircraft. A description of the dataset and feature selection is available in Mitici, M. et. al. 2023b. Some features with high importance are the Voltage measured during take-off, Discharge and Charge capacity, as well as the Temperature recorded (Mitici, M. et. al. 2023b). As regressors $g(.)$ and $f(.)$, we consider a Long-Short Term Memory (LSTM) model. We consider a set of 25 EVTOLs, each performing 10 trips per day for a period of 10 years, where a replacement can be scheduled up to $k = 10$ days. We also consider a cost of replacing a battery of 100, while the cost of an unscheduled replacement due to unforeseen failure is 1000. Table 1 shows the numerical results obtained with the two approaches.

Table 1. Numerical results RUL-M versus E2E-M

	Total cost	Battery failured	Batteries replaced (not failed)
RUL-M	70.712	594	1.125
E2E-M	26.834	63	1.692

The results show that the E2E-M approach leads to the fewest replacements due to unforeseen battery failures and a lower maintenance cost. This shows the benefit considering the optimization objectives of the maintenance planning already in the data processing stage, rather than processing the data in isolation. The main characteristic of the E2E-M approach that explains the performance achieved is the fact that, by informing the data processing step of the maintenance planning objectives, the regressor $f(.)$ is able to better avoid large costs due to unexpected failures.

Overall, this paper proposes a integrative perspective for maintenance planning, taking into account the data directly into the maintenance optimisation model. This is different from many existing studies on predictive maintenance, which make use of an intermediate estimator, the RUL prognostics, to decide on optimal maintenance actions.

References

Jamshidi, A., Hajizadeh, S., Su, Z., Naeimi, M., Núñez, A., Dollevoet, R., & Li, Z. (2018). A decision support approach for condition-based maintenance of rails based on big data analysis.
Transportation Research Part C: Emerging Technologies, 95, 185-206.

Mitici, M., de Pater, I., Barros, A. and Zeng, Z., 2023. Dynamic predictive maintenance for multiple components using data-driven probabilistic RUL prognostics: The case of turbofan engines.
Reliability Engineering & System Safety, *234*, p.109199.

Liu, X., Sun, Q., Ye, Z. S., & Yildirim, M. (2021). Optimal multi-type inspection policy for systems with imperfect online monitoring.
Reliability Engineering & System Safety, *207*, 107335.

Mitici, M., Hennink, B., Pavel, M., & Dong, J. (2023b). Prognostics for Lithium-ion batteries for electric Vertical Take-off and Landing aircraft using data-driven machine learning.
Energy and AI, *12*, 100233.

Reliability Prediction for a Basic Hardware-Software System Using Survival Signature

W. Alahmadi*, F.P.A. Coolen** and T. Coolen-Maturi***

* Department of Mathematics, Taibah University, Saudi Arabia. Email: wahamade@taibahu.edu.sa

* Department of Mathematical Sciences, Durham University, UK. Email: drbq21@durham.ac.uk

** Department of Mathematical Sciences, Durham University, UK. Email: frank.coolen@durham.ac.uk

*** Department of Mathematical Sciences, Durham University, UK. Email: tahani.maturi@durham.ac.uk

Abstract. This study uses the survival signature approach to develop a unified framework for predicting the reliability of a basic hardware–software system. Additionally, failure diagnosis is conducted based on the derived survival signature under specific assumptions. The results show that the reliability function and failure diagnosis measures, incorporating the derived survival signature, accurately capture the distinct failure characteristics of both hardware and software components. This confirms that the survival signature approach effectively captures such systems' behaviour, supporting its future enhancement with advanced software reliability models and practical maintenance improvements.

Keywords: reliability prediction, survival signature, hardware-software systems, system failure diagnosis.

1. Introduction

Modern safety-critical systems—such as those used in aerospace, healthcare, and industrial applications—require accurate reliability assessments to prevent catastrophic failures (Sinha et al., 2019). This study applies the survival signature approach presented by Coolen and Coolen-Maturi (2012) to a basic k-out-of-n hardware–software system. A key feature of this approach is its ability to apply to large systems and networks by separating the structure function from component failure time distributions, making it suited for representing real-world systems that combine hardware and software components. In the system considered, each component represents a subsystem composed of one hardware and one software component connected in series, so both must function for the subsystem to function. All hardware components are assumed exchangeable; similarly, all software components are exchangeable within their type. In addition to reliability prediction, the survival signature also supports system failure diagnosis. For this purpose, several diagnostic measures proposed by Alharshan (2024) are adopted, which are based on the survival signature method. Compared to traditional methods that rely solely on the structure function, these approaches offer a flexible and scalable framework capable of addressing the complexity of modern integrated systems.

2. Methodology

The reliability of the system considered in this study is evaluated using the survival signature framework, which assumes that failure times of components of the same type (hardware or software) are independent and identically distributed (i.i.d.) or exchangeable, while hardware and software failure times are independent. Both hardware and software component failure times are assumed to follow Weibull distributions, with hardware having an increasing failure rate ($\alpha_h > 1$) and software a decreasing failure rate ($\alpha_s < 1$), consistent with typical reliability assumptions and thereby reflecting their distinct degradation behaviours. The survival signature represents the probability that the system functions given that a certain number of each component is functioning. For the proposed system, the probability that the entire system functions, given that l_h hardware and l_s software components are functioning, is expressed as follows:

$$\Phi(l_h, l_s) = \sum_{j=k}^{n} \binom{l_h}{j} \binom{n - l_h}{l_s - j} \Big/ \binom{n}{l_s}$$

By substituting this survival signature into the reliability model—along with the Weibull cumulative distribution functions for hardware and software components—the overall system reliability function is obtained as follows:

$$R(t) = \sum_{l_h=0}^{n_h} \sum_{l_s=0}^{n_s} \left[\sum_{j=k}^{n} \left[\binom{l_h}{j} \binom{n-l_h}{l_s-j} \Big/ \binom{n}{l_s} \right] \binom{n_h}{l_h} \binom{n_s}{l_s} \times \left(1 - e^{-\left(\frac{t}{\lambda_h}\right)^{\alpha_h}}\right)^{n_h - l_h} \left(e^{-\left(\frac{t}{\lambda_h}\right)^{\alpha_h}}\right)^{l_h} \times \left(1 - e^{-\left(\frac{t}{\lambda_s}\right)^{\alpha_s}}\right)^{n_s - l_s} \left(e^{-\left(\frac{t}{\lambda_s}\right)^{\alpha_s}}\right)^{l_s} \right], t \geq 0$$

where $F(t_h) = 1 - e^{-\left(\frac{t}{\lambda_h}\right)^{\alpha_h}}$ and $F(t_s) = 1 - e^{-\left(\frac{t}{\lambda_s}\right)^{\alpha_s}}$ are the cumulative distribution functions for hardware and software failures, respectively.

For system failure diagnosis, we adopt the methodology proposed by Alharshan (2024), which extends the survival signature framework with diagnostic measures. In this setting, the derived survival signature is integrated to evaluate conditional probabilities and expectations related to system failure, including the probability that a system failure is caused by a hardware failure, the number of failed components given preceding failures, the number of failed components at system failure conditioned on system age, and the number of failed components when the system is expected to fail within a future time interval. This allows the model to capture all possible failure scenarios that could lead to system failure at a given time.

3. Conclusion

This study presents a unified framework for analysing the reliability of a basic k-out-of-n hardware–software system using the survival signature approach. By integrating Weibull-based reliability modelling with a derived survival signature function, the methodology quantifies the distinct contributions of hardware and software failures under various operating conditions. The proposed system reliability and failure diagnosis models offer a clear and flexible formulation suitable for complex integrated systems while also enabling targeted maintenance strategies through the identification of critical failure patterns. Future work could extend this framework by incorporating advanced software reliability models to further enhance its predictive capabilities.

Acknowledgement

Wejdan Alahmadi gratefully acknowledges the support provided by Taibah University and the Saudi Arabian Cultural Bureau in the UK. She also extends her appreciation to Durham University for providing a valuable academic environment.

References

Alharshan, A. (2024). *Systems Failure Diagnosis and Repair Kit using Survival Signature*. PhD thesis, Durham University.

Coolen, F. P. A., & Coolen-Maturi, T. (2012). Generalizing the signature to systems with multiple types of components. In *Complex Systems and Dependability* (pp. 115–130). Springer.

Sinha, S., Goyal, N. K., & Mall, R. (2019). Survey of combined hardware–software reliability prediction approaches from architectural and system failure viewpoint. *International Journal of System Assurance Engineering and Management, 10*, 453–474.

Determination of supercapacitors' reliability based on accelerated life tests

J. Sobral*,**, Arcílio B.S. Semente***, R.S. Sampaio***, T.M. Silva*,*** and M.F. Montemor***

** Mechanical Engineering Department, ISEL - Instituto Superior de Engenharia de Lisboa, Instituto Politécnico de Lisboa, Lisboa, Portugal: jsobral@dem.isel.pt*

*** Centre for Marine Technology and Ocean Engineering (CENTEC), Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal*

**** CQE, Institute of Molecular Sciences, Departamento de Engenharia Química, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1049-001 Lisboa, Portugal*

Abstract. Accelerated life tests are experimental methods used to estimate the lifespan or reliability of a product, component, or material by subjecting it to conditions more severe than normal operational environments. The main objective of the present study is to define the lifetime of supercapacitors through accelerated life testing. The experimental data collected in the study was obtained by carrying out floating tests that were fitted to an appropriate degradation model and analyzed. Results showed that this model can be applied to supercapacitors' accelerated life tests, allowing the comparison of different materials and obtaining the best solutions.

Keywords: Accelerated life tests, reliability, supercapacitors

Supercapacitors and batteries can store and use electrical energy whenever necessary. Although batteries are technologically well developed and used in a wide range of applications, they have some limitations, such as relatively long charging times and limited cycle life. To address these issues, supercapacitors have emerged as complementary energy storage technology. Supercapacitors can operate at extremely high electrical currents, feature excellent power density, long cycle lives, and fast charging times. Thus, supercapacitors fill the gap between electrical energy storage devices, such as batteries and common capacitors.

The goal of accelerated life tests is to induce failures more quickly so that data can be collected in a shorter time, allowing engineers and designers to predict the product's performance under normal conditions over a longer period. Accelerated life tests were applied to innovative energy storage technologies. To validate and implement the concepts, the referred tests were conducted on supercapacitor prototypes, subjected to various voltages exceeding normal operating levels, to accelerate their degradation and induce failure, allowing prediction of the lifespan of such devices under normal operating conditions. The objective of test methodologies, together with a statistical model for the effect of the acceleration factor, is to allow the estimation or prediction of the useful life of a component from a much shorter time interval (Collins et al., 2013). Several models (mathematical equations and respective parameters) can be applied in accelerated life tests, such as models for a single load, various load levels, and accumulated risk and damage models. The Arrhenius Life-Stress relationship or model is probably the most commonly used Stress-Life relationship when the topic is accelerated life tests. It has been widely used when the stimulus or accelerating variable is thermal, that is, temperature. The Eyring relationship was formulated from quantum mechanical principles and is most commonly used when a thermal load is the acceleration factor. However, this relationship is used when working with other loads, such as humidity. The inverse power law model is typically used for accelerated tests that do not involve the thermal variable (temperature). The Temperature-Humidity relationship is a variation of the Eyring relationship and was proposed to predict life under operating conditions when temperature and humidity are the accelerating loads in a test. When the loads in an accelerated test are temperature and another non-thermal variable (e.g. voltage), the Arrhenius and Inverse Power relations can be combined to obtain the relation (T-NT). Some other accelerated models can also be used.

In this work, the main objective is the development of a new generation of manganese oxide (MnOx) electrodes, capable of self-repair, for high-performance hybrid supercapacitors. In this context, evaluating the durability of these new materials is a key factor for the success of the devices. The experimental data were obtained by carrying out floating tests, fitted to an appropriate degradation model and analyzed using specialized software. This analysis yielded critical information about the lifespan of these devices, enabling the prediction of their useful life under normal operating conditions. Based on the tests and the data analysis, it was concluded that there is a gradual reduction in time until the device’s failure as the applied voltage

increases. It was also noted that there is a range of voltage values where the degradation of the devices evolves exponentially until failure is reached, thus indicating a possible phenomenon of accelerated wear in that range of values. In these tests, failure is considered to occur when the device's specific capacitance decreases to 80% of its initial value. Figure 1 shows the capacitance retention and float time for a specific set of applied voltages.

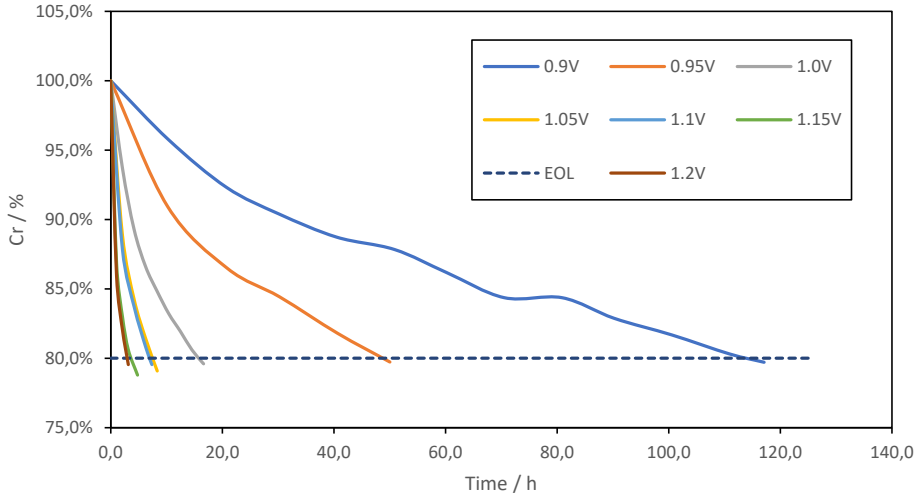


Figure 1. Capacitance retention for accelerated tests

The results achieved correspond to coefficients of the Reverse Power Law and the parameters of the assumed distribution. In this case, the shape parameter (β=1.466), the model parameters (K=0.000673 and n=5.911), and the scale parameter or characteristic life (η=7362.7 min) were determined. Based on the results, for example, when the test voltage is 1.2 V, the acceleration factor (A_F) for a voltage of 0.75 V is 16.08. Figure 2 shows the probability of failure based on the Weibull distribution for each voltage level and for normal usage voltage, as well as the corresponding probability density function for a voltage of 0.75 V.

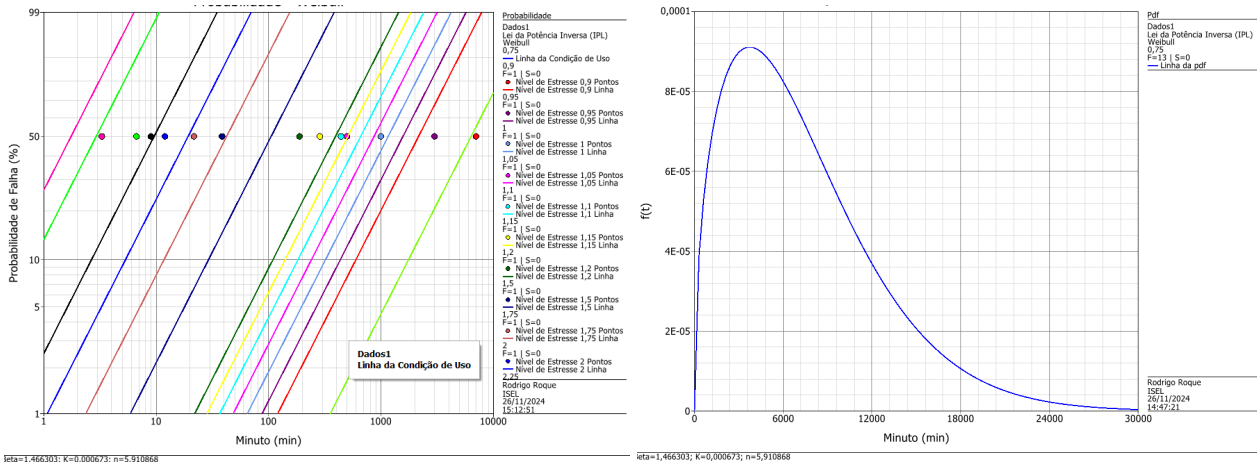


Figure 2. Probability of failure for accelerated tests (left) and probability density function (right)

Results show that at regular use for a voltage of 0.75 V, the characteristic life corresponds to 7362.7 min, meaning that in 100 supercapacitors, it is expected that 36 or 37 present a capacitance over 80% at that time. The results are aligned with other investigations that were carried out to determine the life of supercapacitors.

The following steps involve new supercapacitors' accelerated life tests, allowing the comparison of different materials and obtaining the best solutions.

References

Collins, DH, Freels, JK, Huzurbazar, AV, Warr, RL, Weaver, BP (2013) Accelerated test methods for reliability prediction. *Journal of Quality Technology* **45**(3), 244-259.

A Data-Driven Robust Approach to a Problem of Optimal Replacement in Maintenance

Sina Shahri Majarshin*, Ahmadreza Marandi*, Claudia Fecarotti*, Geert-Jan van Houtum*

**Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands, {s.shahri.maharshin, a.marandi, c.fecarotti, g.j.v.houtum}@tue.nl*

Abstract

Maintenance strategies are pivotal in ensuring the reliability and performance of critical components within industrial machines and systems. However, accurately determining the optimal replacement time for such components under stress and deterioration remains a complex task due to inherent uncertainties and variability in operating conditions. In this paper, we propose a comprehensive approach based on Robust Markov Decision Processes (RMDP) to optimize component replacement decisions in machines with one critical component while addressing uncertainty in a structured manner. RMDP offers a robust framework for decision-making under uncertainty, allowing for the modeling of component degradation and variability in operating conditions. Our methodology uses data-driven ambiguity sets, including likelihood-based and Kullback-Leibler (KL)-based ambiguity sets, to capture and quantify uncertainty in the degradation process. We show the mathematical relationship between the KL-based and Likelihood-based ambiguity sets and provide statistical guarantees for the optimal cost. Through computational experiments, we demonstrate the effectiveness of our RMDP approach in identifying the optimal replacement time that minimizes the total maintenance cost while exhibiting greater stability compared to traditional methods.

Keywords: Robust Markov Decision Processes, Data-Driven Uncertainty, Optimal Replacement.

In many industrial settings, a single deteriorating component can dictate the performance and reliability of an entire system. Determining the optimal time to replace such components is essential for minimizing total maintenance costs, which include both operational inefficiencies due to degradation and direct replacement expenses. While classical dynamic programming methods (e.g., Ross, 1983) can yield optimal strategies when system dynamics are precisely known, in real-world applications, these dynamics must often be inferred from limited or biased data, making resulting policies fragile and overly optimistic.

To address this challenge, we model the replacement problem as a finite-horizon Markov Decision Process (MDP) with unknown transition probabilities. The component's condition is periodically observed, and the decision-maker must decide whether to replace the component or continue operation. The cost structure is twofold: a replacement cost and a state-dependent operational cost that increases with deterioration.

We introduce robustness to this framework via data-driven ambiguity sets that capture uncertainty

in the transition probabilities. Specifically, we investigate two families of ambiguity sets: likelihood-based sets, which constrain candidate transition matrices based on their distance from the Maximum Likelihood Estimator (MLE), and Kullback-Leibler (KL)-based sets, which bound the relative entropy from the empirical distribution. Both constructions allow for tuning robustness via a scalar parameter $\alpha \in (0,1)$, which determines the size of the ambiguity set. A smaller α induces a more conservative policy, hedging against rare but costly outcomes.

We formally establish a connection between these two types of ambiguity sets and show how the resulting robust Bellman recursions can be solved via efficient convex programs. To complement this, we derive finite-sample statistical bounds: we prove that, with high probability, the cost of the robust policy upper-bounds the true expected cost incurred under the unknown transition law. This offers practical assurance to decision-makers that the chosen policy will not significantly underperform in deployment, even when data is scarce or partially biased (Casella & Berger, 2002).

Our computational study involves synthetic data across three component classes—fragile, moderately durable, and sturdy—each characterized by different degradation dynamics. We simulate both unbiased data (reflecting the true degradation behavior) and biased data (mimicking misestimation). Results show that robust policies, particularly KL-based ones, yield substantially lower cost variability in small data regimes. Interestingly, robustness also correlates with replacement behavior: it leads to fewer replacements under unbiased data but increases replacement frequency under biased scenarios that underestimate failure risk. As data availability increases, the robust and empirical policies converge, confirming theoretical expectations.

This research contributes to the field of Condition-Based Maintenance (CBM), extending foundational work (Derman, 1963; Nilim & El Ghaoui, 2005; Arts et al., 2012; Quatrini et al., 2020) by integrating modern statistical insights and tractable optimization into a practical and interpretable framework. The proposed approach enables maintenance engineers and managers to make informed, risk-aware replacement decisions under real-world data constraints.

References

1. Derman, C. (1963). On optimal replacement rules when changes of state are markovian. In Mathematical Optimization Techniques, pp. 201–210.
2. Ross, S. M. (1983). Stochastic Processes. Wiley.
3. Nilim, A., & El Ghaoui, L. (2005). Robust control of Markov decision processes with uncertain transition matrices. Operations Research, 53(5), 780–798.
4. Arts, J., Gopalakrishnan, M., & van Houtum, G. J. (2012). Fifty years of maintenance modeling. Journal of Quality in Maintenance Engineering, 18(4), 338–353.
5. Casella, G., & Berger, R. L. (2002). Statistical Inference. Duxbury Press.
6. Quatrini, E., Costantino, F., Di Gravio, G., et al. (2020). Condition-Based Maintenance—An Extensive Literature Review. Machines, 8(2), 31.

An Actuarial Approach to Data-Driven Warranty Pricing

Anouck Vercammen[a,b], Robert Boute[a,b,c], Brammert Termont[b,d]

[a]*Research Center for Operations Management, KU Leuven, Anouck.vercammen@kuleuven.be*

[b]*Flanders Make@KU Leuven, Leuven*

[c]*Technology and Operations Management Area, Vlerick Business School, Leuven*

[d]*Department of Marketing, Ghent University, Ghent*

Abstract.
The Right to Repair legislation requires original equipment manufacturers to make repairs more accessible. While this legislation helps consumers, manufacturers must consider the additional costs of returning and repairing products. Our model predicts the cost of returning a product, allowing manufacturers to determine optimal warranty pricing before the product is sold. The model is developed using actuarial best practices, combining return costs with claims frequency, and trained using gradient-boosting predictive models. We validate our approach on a dataset of 25,000 returns from an automotive supplier over five years. By predicting return costs, our model supports optimal warranty pricing.
Keywords: Warranty, Product Return, Predictive Model, Machine Learning

While the Right-to-Repair initiative has the potential to benefit consumers, literature suggests that manufacturers will be worse off. Stricter regulations, such as a mandatory 12-month warranty extension, force manufacturers to raise warranty fees to cover additional costs. At the same time, the presence of third-party repairers increases competition, forcing manufacturers to lower prices to retain customers. This complicates setting accurate warranty fees, which are premiums added to the base price to cover potential future return costs. The manufacturer incurs return costs when a customer returns a product under warranty, a process known as a warranty claim. These costs represent the total cost of providing the customer with a functional replacement, whether remanufactured or new. To address this issue, we develop a predictive model that forecasts the return cost of a product based on pre-sale product and operational characteristics, such as product family and country of origin. This model assists manufacturers in estimating return costs before sale and establishing more product-tailored warranty fees.

To develop these predictive models, we use principles from the actuarial industry. More specifically, we adopt the frequency-severity approach to define the total return cost per product by predicting the number of times a product is returned within the warranty period (claim frequency) and the average return cost of a product (severity). Most predictive models in the insurance industry that implement this methodology are Generalized Linear Models (GLMs) and Gradient Boosting Models (GBMs) (Henckaerts et al., 2020).

Existing methods to predict warranty costs

Various methods have been developed to predict claim frequency, including lifetime distributions, stochastic processes, artificial neural networks, and time series models (Kalbfleisch et al., 1991; Wu & Akbarov, 2011). While these studies primarily use historical volumes as predictors for the number of claims, our model instead leverages product characteristics as input variables, as this allows us to make predictions before the product is sold. The literature on warranty cost forecasting (severity) relies on extrapolation and time-based effects (Wang et al., 2021). In contrast, our study builds on the work of Becker and Schölzel (2025), who employed machine learning models for predicting accounting warranty reserves, which are funds set aside by firms to cover future warranty claims and extend their approach to an operational context. Our study makes two contributions to the field. First, actuarial principles are applied to combine the claim frequency with the average return costs. Second, our models allow manufacturers to predict return costs in advance, enabling them to set optimal warranty fees.

Predictive models to forecast total return cost

Our industrial partner operates in the business-to-business automotive sector, supplying heavy and commercial vehicle manufacturing parts. The dataset under consideration encompasses 25,000 product returns

received over five years. We develop GLM and XGBoost models to predict claim frequency and average return cost for each product. The model makes predictions based on eight input variables that represent the pre-sale product and operational characteristics of the component. The training set consists of all returns from products sold during the first four years, while data from the final year is used to test the model. This setup allows for annual updates to the predictions, ensuring suitability for the following year.

We use Poisson and Gamma distributions as the underlying models for claim frequency and return cost severity, respectively, given their widespread use in insurance literature (Henckaerts et al., 2020). To integrate these distributions into our predictions, we apply Poisson deviance as the loss function for claim frequency predictions and Gamma deviance for severity predictions. The Poisson (D_i^{POI}) and Gamma deviance (D_i^{GAM}) for product i are calculated as follows:

$$D_i^{POI}(y_{i,freq}, \hat{y}_{i,freq}) = \sum_{i=1}^{n} 2 * \left(y_{i,freq} * \log\left(\frac{y_{i,freq}}{\hat{y}_{i,freq}}\right) - (y_{i,freq} - \hat{y}_{i,freq})\right),$$

$$D_i^{GAM}(y_{i,sev}, \hat{y}_{i,sev}) = \sum_{i=1}^{n} 2 * \left(\frac{y_{i,sev} - \hat{y}_{i,sev}}{y_{i,sev}} - \log\left(\frac{y_{i,sev}}{\hat{y}_{i,sev}}\right)\right).$$

Here, $y_{i,freq}$ denotes the actual number of claims for product i, while $\hat{y}_{i,freq}$ represents the corresponding predicted values. Similarly, $y_{i,sev}$ refers to the actual return cost per product and $\hat{y}_{i,sev}$ denotes the predicted return cost.

Model Performance

Two models are developed to predict claim frequency and severity: a generalized linear model (GLM), and a gradient boosting model (XGBoost). Both GLM and XGBoost models include all eight predictor variables in their analysis. In addition, we include a simple baseline model that always predicts the mean. This serves as an algorithmic benchmark and reflects the current forecasting approach used by our industry partner. Lower scores suggest better performance when using deviance as a performance measure, as they indicate smaller differences between the actual target values and the model's predictions. Thus, the findings indicate that the XGBoost model outperforms the GLM model for both analyses. Moreover, XGBoost significantly outperforms the naïve baseline, with gains of 34% and 48%. This improved performance can be explained by the gradient-boosting model's ability to capture more complex interactions and manage multicollinearity more effectively than the GLM. Our XGBoost models provide more accurate predictions of return costs, enabling manufacturers to set fair warranty fees and avoid over- and underpricing.

Table 1. Gamma and poisson deviance statistics for GLM and XGBoost models.

Analysis	Loss function	Naïve Baseline	GLM	XGBoost
Frequency	D^{POI}	0.29	0.20	0.19
Severity	D^{GAM}	1.24	0.81	0.64

Conclusion

Our study presents an approach for predicting product return costs based on product characteristics, allowing manufacturers to estimate these costs in advance and set fair warranty fees. The gradient-boosting models outperform traditional statistical models in predicting future return costs. While the right-to-repair movement may put manufacturers at a disadvantage, our methodology provides a way to minimize redundant costs, giving them a tool to better manage and plan for return-related costs.

References

Becker, M., & Schölzel, S. (2025). Warranty Provisions: Machine-Learning Versus Human Estimates. *European Accounting Review*, 34(1), 1-30.

Henckaerts, R., Côté, M. P., Antonio, K., & Verbelen, R. (2021). Boosting insights in insurance tariff plans with tree-based machine learning methods. *North American Actuarial Journal*, *25*(2), 255-285.

Kalbfleisch, J. D., Lawless, J. F., & Robinson, J. A. (1991). Methods for the analysis and prediction of warranty claims. *Technometrics*, *33*(3), 273-285.

Wang, J., Zhu, S., & Du, S. (2021). Analysis of a two-dimensional stair-case warranty policy with preventive maintenance. *IMA Journal of Management Mathematics*, *32*(1), 51-67.

Wu, S., & Akbarov, A. (2011). Support vector regression for warranty claim forecasting. *European Journal of Operational Research*, *213*(1), 196-204.

Zero inflated model for rare failure analysis

M. Skarupski* and M. Veenvliet** and M. Wouters*** and S. Kapodistria****
** Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands & Wrocław University of Science and Technology, Poland, m.skarupski@tue.nl*
*** ASML, Veldhoven, The Netherlands, miklas.veenvliet@asml.com*
**** ASML, Veldhoven, The Netherlands, maarten.wouters@asml.com*
***** Department of Mathematics & Computer Science, Eindhoven University of Technology, The Netherlands, email address: s.kapodistria@tue.nl*

Abstract. Highly complicated machines contain many parts with few failures (up to 5 in the whole history). This makes predicting future failures challenging, hindering a cost-effective stocking and maintenance strategy. This article explores the benefits of switching from the traditional Poisson to the Zero Inflated Poisson (ZIP) distribution for the modeling of the occurrence of rare failures. Using real data, we compare the performance of the two modeling approaches and show that the proposed model significantly improves the forecasting accuracy of spare part storage management. ZIP also provides an improved prediction range that more accurately describes the magnitude of predicted failures.

Keywords: Zero Inflated Poisson, spare parts management, rare events, frequency estimation procedure

INTRODUCTION

In the semiconductor industry and other high-tech industries, the downtime of the machines is often very expensive. So, not being able to repair a machine purely due to an insufficient stock of spare parts is extremely costly and should be avoided as much as possible. Intuitively, one would, therefore aim to put sufficient spare parts on stock with a wide safety margin. However, due to the rapid pace of innovation, the parts on stock will most likely become obsolete then before they are ever used, also resulting in a significant loss. Having a cost-effective stocking and maintenance strategy is therefore vital.

The complicated machines in the high-tech industry can often contain thousands of spare parts. The majority of these parts will, however, have at most a hand full ($N \leq 5$) of observed failures over the entire life span of all field machines. Forecasting the expected number of failures in a given time window will become very challenging with such few data points but is, however of crucial importance for a cost-effective stocking and maintenance strategy.

Another complicating factor is the variance in the time it takes to replenish stock levels. The complicated and specialized service parts often take a significant time to manufacture. Therefore, it is important not only to predict the failure rate for each spare part but also to model its variation accurately. For example, if one accurately predicts that 3 parts will be used over a specific period, it still matters if there 3 parts are used at the same time or homogeneously spread over the period. In the latter case one might suffice with putting a single part on stock, which would result in a long downtime for the former case where the service engineer needs to wait for the manufacturing of two more spare parts.

ZIP AND ITS LIMITATIONS

The ZIP model was introduced by Lambert (1992), who showed its usefulness in statistical analysis of soldering defects on printed wiring boards. Since then, the model has been used in a wide range of research, including health insurance (eg. Mouatassim and Ezzahid (2012)), natural calamities' data (Beckett et al.(2014)) or software reliability engineering (Khoshgoftaar et al. (2001)).

Denks et al. (2020) used a ZIP model for the microbubble counts in super-resolution voxels. They obtained the closed formula for Maximum Likelihood Estimation (MLE) method. In our work, we applied this method. However, while calculating the numerical values of the estimators, we noticed that this method has its limitations, which were to the best of our knowledge not reported in the earlier literature. The main limitation is that if one doesn't thoroughly check if the assumptions for the estimation procedure are met, it might result in results such as an estimated negative probability of inflation. These assumptions are very stringent and in order to satisfy them in real data sets one might be forced to discard a significant portion of the available data, which is unwanted as these data sets capture limited occurrences. As a result of the underlying assumptions for the employment of the MLE method, not every data set can be used to calculate the estimates. In extreme cases, this leads to the rejection of up to 75% of the data sets, mainly data sets where several individual failures appear.

PROPOSED SOLUTION AND REAL DATA ANALYSIS

In this work, we overcome these issues by proposing an alternative estimation approach based on the frequency substitution (FS) method.

We show that this approach is consistent with the MLE method. The proposed approach leads to improved estimations that can be viewed as the correction to the MLE estimators. We also demonstrate that the newly proposed method has a broader applicability (due to having fewer assumptions) than the MLE method without any losses in available data due to rejection.

The application of the proposed corrections was evaluated on real data from industry. Thanks to this, in many cases it was possible to determine the spare parts storage policy even for parts that rarely broke down. Based on real data, we compared the forecasting accuracy resulting from the Poisson and ZIP models. For this purpose, quantiles of the conditional distribution $P(X = k \mid X > 0)$ (where X is a random variable modeling the number of failures in a time unit) were determined to find the amount of usage such that a certain level of coverage of the demand will be satisfied. The rate of the failure was comparable in both models, however, the variation is explained more accurately by ZIP.

CONCLUSIONS

In conclusion, the ZIP distribution is a powerful statistical tool for analyzing count data with excess zeros and with rare failures. Its ability to disentangle the sources of zeros provides researchers and practitioners with valuable insights and improved predictive accuracy. It is more helpful in describing both mean and the variation of expected usage of the spare parts. The MLE method of estimation for ZIP has its limitation, which was not described previously. By applying FS with proposed correction we are able to apply ZIP on a wider range of datasets improving its application in real world cases.

Acknowledgements The work has been partially funded by the NWO grant NWA.1160.18.238. (PrimaVera). The work of Stella Kapodistria is supported by NWO through the Gravitation-grant NETWORKS-024.002.003.

References

Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., Pal, N. (2014) Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve: A Journal of Mathematics* **7(6),** 751 – 767

Dencks, S., Piepenbrock, M., Schmitz, G. (2020) Assessing Vessel Reconstruction in Ultrasound Localization Microscopy by Maximum Likelihood Estimation of a Zero-Inflated Poisson Model. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* **67(8)**, 1603–1612

Khoshgoftaar, T., Gao, K., Szabo, R. (2001) An application of zero-inflated Poisson regression for software fault prediction. In: *Proceedings 12th International Symposium on Software Reliability Engineering*, pp. 66–73

Lambert, D. (1992) Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* **34(1)**, 1–14

Mouatassim, Y., Ezzahid, E.H. (2012) Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. *European Actuarial Journal* **2**, 187–204

Suzanne E. Perumean-Chaney Charity Morgan, D.M., Aban, I. (2013) Zero-inflated and overdispersed: what's one to do? *Journal of Statistical Computation and Simulation* **83(9)**, 1671–1683

A Framework for Maintenance Scheduling in Reconfigurable Manufacturing Systems

Huu-Truong Le*, Phuc Do*, Alexandre Voisin*, Chiara Franciosi*

** Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France*

Email adress: {huu-truong.le@univ-lorraine.frphuc.do@univ-lorraine.fr, alexandre.voisin@univ-lorraine.fr, chiara.franciosi@univ-lorraine.fr}

1. Introduction

Reconfigurable Manufacturing Systems (RMSs) are designed to adapt to fluctuating market demands and evolving product configurations. Unlike traditional manufacturing systems, RMSs are characterized by their ability to reconfigure in response to changes in production capacity and product design [1]. However, maintaining these systems poses unique challenges due to their modular, scalable and dynamic nature. Conventional maintenance models do not fully exploit the unique modularity, scalability, customization and convertibility of RMSs, resulting in inefficient maintenance costs and increased downtime. While maintenance plays a crucial role in enhancing the performance and adaptability of RMSs, most existing research has primarily concentrated on their design and operation [2, 3], creating a gap in the literature. To address these limitations, this research proposes a novel maintenance framework that integrates these key RMS characteristics, providing a structured methodology for maintenance modeling and decision-making. This framework allows the designing of maintenance strategies in RMSs to enhance system reliability and responsiveness. A case study was then conducted to validate the proposed framework, demonstrating its practical use in RMS maintenance planning and serving as a robust tool for engineers and practitioners aiming to optimize maintenance in RMS environments.

2. Proposed maintenance framework

The framework (Figure 1) is structured into two main phases: Phase 1: Maintenance Modeling and Phase 2: Maintenance Decision and Optimization. By leveraging the unique characteristics of RMSs, it aims to maintain production capacity and functionality in dynamic environments.

- Phase 1 - Maintenance Modeling: Maintenance modeling forms the foundation of the framework by characterizing the system's structure, operational behaviors, and dependencies. This phase provides the necessary insights into the RMS, allowing for the development of tailored maintenance strategies. The modeling begins with the characterization of the target system and support system, a critical step to understanding their behavior under different configurations and operational scenarios.
- Phase 2 - Maintenance Decision and Optimization: This phase focuses on defining and optimizing maintenance policies, balancing preventive and reactive strategies. The framework differentiates between short-term and conventional policies to ensure adaptability to changing production conditions. Maintenance optimization aims to enhance efficiency, reduce costs, minimize downtime, and improve overall performance. This is achieved by balancing multiple objectives using advanced optimization techniques, including heuristic, metaheuristic, and data-driven algorithms.

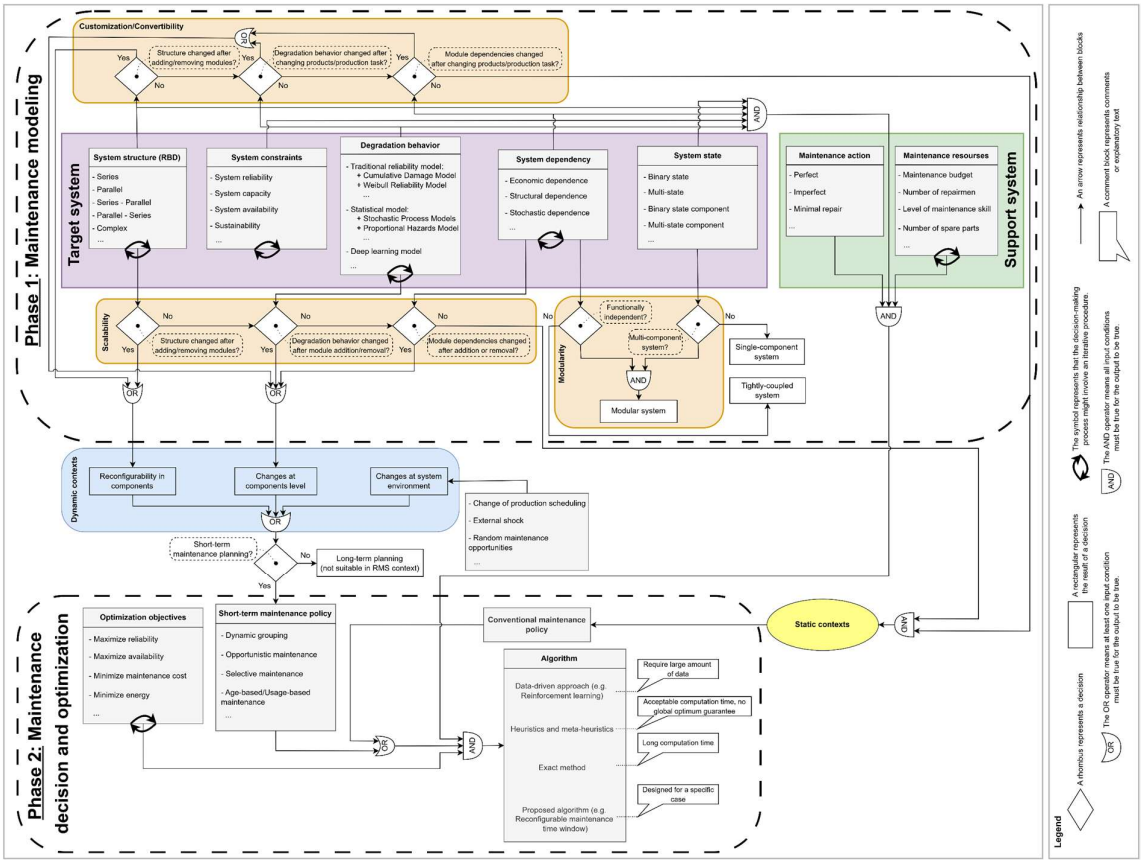


FIG. 1 – A framework for maintenance in RMSs

3. Case study and validation

To validate the proposed maintenance framework for RMS, a case study within the dynamic grouping maintenance approach [4] was investigated. In addition, different RMS contexts were incorporated.

- Dynamic production demand and limited repairman resources: This scenario demonstrates how the modularity and scalability of RMSs influence maintenance decisions under resource constraints, optimizing human resource utilization while maintaining operational efficiency.
- Degradation behavior in a scalable RMS: Focusing on maintenance planning under varying operational loads and degradation rates, this case study showcases the framework's ability to dynamically adjust maintenance schedules to align with capacity changes.
- Opportunistic maintenance for product line reconfiguration: Highlighting the integration of maintenance tasks during system reconfiguration for new product variants, this scenario illustrates how opportunistic maintenance can minimize downtime and enhance system readiness during production transitions.

The results obtained from the case study demonstrated the practical value and adaptability of the proposed framework in managing maintenance within RMS environments while integrating unique RMS characteristics.

In the first case, where the system structure changed due to dynamic production demand and there were not enough repairmen, the framework successfully balanced system capacity and worker availability by considering modularity and scalability. This reduced downtime and kept production running smoothly. In the second case, the framework was able to adjust maintenance schedules in response to system load variations, thus preventing unexpected breakdowns and optimizing maintenance time under abnormal operational modes. The third case, which highlighted the customization and convertibility of the system through reconfiguration for new products, showed that the framework could take advantage of reconfiguration time to perform opportunistic maintenance actions. This helped to reduce transition delays and improved production readiness for new product variants.

However, these results are context-specific. Future research should focus on applying the framework to larger and more varied RMS settings to explore its scalability and applicability.

ACKNOWLEDGMENT

This work was supported by the MODAPTO project (MODULAR MANUFACTURING AND DISTRIBUTED CONTROL VIA INTEROPERABLE DIGITAL TWINS) funded by the European Union's Horizon 2022 (grant agreement No 101091996).

REFERENCES

[1] Bortolini, M., Galizia, F. G., & Mora, C. (2018). Reconfigurable manufacturing systems: Literature review and research trend. Journal of Manufacturing Systems, 49, 93–106.

[2] Koren, Y., U. Heisel, F. Jovane, T. Moriwaki, G. Pritschow, G. Ulsoy, & H. Van Brussel. (1999). Reconfigurable Manufacturing Systems. CIRP Annals, 48(2), 527-540.

[3] Le, H., Franciosi, C., Do, P., & Voisin, A. (2024). Contribution of maintenance to reconfigurable manufacturing Systems: State of the art and challenges. IFAC-PapersOnLine, 58(8), 121–126.

[4] Vu, H. C., Do, P., Barros, A., & Bérenguer, C. (2014). Maintenance grouping strategy for multi-component systems with dynamic contexts. Reliability Engineering & System Safety, 132, 233–249.

Failure Detection in Photovoltaic Inverters

Jorge Ruiz Amantegui[1,3], Phuc Do[1], Hai-Canh Vu[2], Nicolas Favrot[3] and Marko Pavlov[3]
[1]*Université de Lorraine CNRS, CRAN, F-54000, Nancy, France.* .
[2]*Quantom, Feedgy, 75009 Paris,*
[3]*Universit´e de Technologie de Compi`egne, CS 60319, CEDEX, 60203 Compi`egne, France*

Abstract. Inverters are the key component in a photovoltaic operation, as their failures can lead to the disconnection of hundreds of photovoltaic modules. If the inverter fails, several hundred kW of power can be completely lost. Furthermore, this component is the most prone to failure in photovoltaic operations. Out of the all the possible failure modes for the inverter, the capacitor and power component failures are the most prominent. For these reasons, this paper presents a detection and prognostic method based on the creation of a feature using domain knowledge to understand the physics behind the failures.
Keywords: failure detection, Inverter, Photovoltaic

1. Introduction

Photovoltaic (PV) modules generate direct current (DC) power which must be transformed into alternate current (AC) power ~~energy~~ before being supplied into the grid. This task is handled by the inverter. It is a crucial component because its failure involves the loss of power of all the PV modules connected to it. This can generate significant monetary losses and problems for the Operation and Maintenance (O&M) responsible entity since uptime of the PV system is an often used metric.

Despite its crucial role, the inverter is one of the most prone to failure components in the PV systems (Nagarajan *et al.*, 2019). A study of 37,238 O&M tickets (tickets from here onward) from over 100 PV systems showed that 62.5% of the tickets were inverter related (Lindig *et al.*, 2022). In another study of 3500 maintenance systems, 43% of tickets, and 36% of energy losses were due to inverter issues (Gunda *et al.*, 2020).

Out of the many components in the PV inverter, the power components (or switching devices) and the capacitors (Abuelnaga, Narimani and Bahman, 2021; Kurukuru, Khan and Malik, 2021) are the most vulnerable subcomponents in the PV inverter, and a failure of any of these will translate in underperformance or even in a total shutdown of the system.

Large scale PV operation usually employs large central inverters which have a triphasic configuration. These inverters have switching devices in each of their lines. Each of these lines also have their own voltage and current. When these devices are underperforming, they will suffer greater power losses. The main idea in this article is that it is possible to identify cases where one of these devices is failing by the means of domain knowledge and feature engineering.

In the dataset section, the different inverter models, the type of data employed, and the tickets employed to perform this study will be presented.

Following, the methodology section will show how and why the feature indicator was created and how different cases of O&M records were employed to validate the new failure indicator.

The next section, results, will show the difference in alarm composition and frequency during the periods where the indicator is present vs. periods where it is not. This section will also discuss how this method can be used to filter the identified shutdowns in the dataset to only keep those that are related to power component or capacitor issues.

2. Dataset

Two datasets are employed in this study, one for is composed of data coming from 14 different PV plants and 168 inverters. All the inverters in this dataset are Conergy IPG 280K inverters. Second dataset is made up of data from 12 PV plants and 210 Huawei inverters of the series KTL.

The data employed for this analysis includes meteorological data (irradiance), AC power, voltage and current per line of the inverter and O&M records.

3. Methodology

The methodology employed here takes several steps. First step is to use the O&M records to identify cases where maintenance crew was required on site and there is a distinct description of the failure mode.

In second instance a feature is created to identify cases where the power component is malfunctioning. A triphasic inverter has three distinct phases, each of which has two power component (or switches) connected to them, if one of this power components fail the power supplied to its line will be affected, therefore we can expect the voltage and current in that line to be diminished. For this reason the feature craeted is equal to the difference between the maximum line voltage (or current) and the minimum line voltage (current). It represents that one line of the inverter is not working like it should.

Finally, we develop rules to automatically identify the presence of the feature. First, normal behaviour of the voltage and current deviation is defined, this is done by using the data of the first month of operation of the inverter (which we assume to be normal operation). From this data we measure the mean deviations, and the standard deviation. We set as threshold of normal deviation the mean plus three times the standard deviation.

Then for each day the KPI is raised if any of the following conditions are met: (1) if the top 3 values that are above the threshold have an RMSE above 40, (2) over 15% of the datapoints are above the threshold and the values that are above the threshold have an RMSE above 2 and the top 3 values above the threshold have an RMSE above 20.

Performance is benchmarked with the Conergy dataset for which there are available O&M tickets, unlike for the Huawei dataset for which there are no available tickets. The tickets are used to verify that power components are the most common failure mode and that the feature created appears more often than not for the power component and capacitor failures. Later the method is applied in the Huawei dataset.

4. Results

The O&M records from our dataset confirm what other have studied, the power component and capacitor failures are the most prominent among the inverter issues. Also, it has been visually confirmed that in the cases where the maintenance crew described a power component or capacitor failure, the KPI appeared more often (53% and 43% respectively) than in all the other cases (23%).

Using the labelled O&M records as having or not the presence of the KPI, we can asses the performance of the automatic detection of the KPI. The results were as follows:

True Positive	False Positive	True Negative	False Negative
80.3%	2.92%	97.08%	19.7%

The composition of alarms in the days where the KPI is raised vs days where it is not are significantly different, especially in the Huawei Dataset. Days where the KPI is not raised, more than 99% of the alarms are 0 (no failure). While for days where the KPI is raised the error code 301 composes 87% of the alarms. This does not necessarily mean that error code 301 only appears in this case, but it is an indicator that this feature is representative of an inverter that is malfunctioning.

References

Abuelnaga, A., Narimani, M. and Bahman, A.S. (2021) 'A Review on IGBT Module Failure Modes and Lifetime Testing', *IEEE Access*, 9, pp. 9643–9663.

Gunda, T. *et al.* (2020) 'A Machine Learning Evaluation of Maintenance Records for Common Failure Modes in PV Inverters', *IEEE Access*, 8, pp. 211610–211620.

Kurukuru, V.S.B., Khan, M.A. and Malik, A. (2021) *Failure mode classification for grid-connected photovoltaic converters*. Edited by A. Haque et al. (RELIABILITY OF POWER ELECTRONICS CONVERTERS FOR SOLAR PHOTOVOLTAIC APPLICATIONS), p. 249.

Lindig, S. *et al.* (2022) 'Towards the development of an optimized Decision Support System for the PV industry: A comprehensive statistical and economical assessment of over 35,000 O&M tickets', *Progress in Photovoltaics: Research and Applications* [Preprint].

Nagarajan, A. *et al.* (2019) *Photovoltaic Inverter Reliability Assessment*. NREL/TP-5D00-74462, 1573462, p. NREL/TP-5D00-74462, 1573462.

Mitigating Contamination Risks: The Crucial Role of Breathers in Rotating Equipment Lubrication Tanks

Alsuthaywani, Nawaf and Aburoyan, Abdulrahman
**Saudi Arabia, Dhahran + email address: nawaf.alshuthaywani@aramco.com*
*** Saudi Arabia, Dhahran + email address: abdulrahman.aburoyan@aramco.com*

Abstract. Rotating equipment, such as pumps, gearboxes, and turbines, rely heavily on proper lubrication to ensure efficient operation and longevity. However, lubrication tanks can be susceptible to contamination from airborne particles, moisture, and other environmental factors, which can compromise the integrity of the lubricant and ultimately lead to equipment failure. Breathers play a vital role in preventing these contaminants from entering the tank, thereby safeguarding the quality of the lubricant and protecting the equipment. By allowing air to enter and exit the tank while filtering out impurities, breathers help maintain a stable atmosphere within the tank, reducing the risk of oxidation, corrosion, and microbial growth. This paper highlights the significance of breathers in rotating equipment lubrication tanks, discussing their design, functionality, and benefits, as well as best practices for selection, installation, and maintenance. By emphasizing the importance of breathers, this research aims to raise awareness among industry professionals about the critical need for effective contaminant control measures to optimize equipment performance, reduce downtime, and extend lifespan.

1. Technical Aspects

Breathers are devices that allow air to enter and exit the lubrication tank while filtering out impurities. They are typically installed on the top of the tank and are designed to prevent contaminants from entering the tank. Breathers are usually equipped with a filter element that traps particles, moisture, and other impurities, allowing only clean air to enter the tank.

1.1 Why are Breathers Important?

Breathers are important for several reasons:

(1). Prevent Contamination: Breathers prevent airborne particles, such as dust, dirt, and debris, from entering the tank and contaminating the lubricant. Contaminated lubricant can cause equipment failure, downtime, and costly repairs.

(2). Reduce Moisture: Breathers help to reduce moisture in the tank, which can cause lubricant degradation, corrosion, and microbial growth. Moisture can enter the tank through the air, and if not controlled, can lead to serious problems.

(3). Maintain Lubricant Quality: By preventing contamination and reducing moisture, breathers help to maintain the quality of the lubricant. This ensures that the lubricant remains effective and efficient, reducing the risk of equipment failure.

(4). Extend Equipment Life: By protecting the lubricant and preventing contamination, breathers help to extend the life of the equipment. This reduces downtime, maintenance costs, and the need for premature replacement.

(5). Improve Equipment Performance: Breathers help to improve equipment performance by ensuring that the lubricant is clean and free from contaminants. This results in smoother operation, reduced vibration, and increased efficiency.

(6). Reduce Maintenance Costs: By preventing contamination and extending equipment life, breathers help to reduce maintenance costs. This includes the cost of repairs, replacement parts, and labor.

1.2 Types of Breathers

There are several types of breathers available, including:

(1). Desiccant Breathers: These breathers use a desiccant material, such as silica gel, to absorb moisture from the air.

(2). Filter Breathers: These breathers use a filter element to trap particles and impurities from the air.
(3). Combination Breathers: These breathers combine desiccant and filter elements to provide comprehensive protection against contamination and moisture.
(4). Automatic Breathers: These breathers are designed to automatically adjust to changes in temperature and humidity, ensuring optimal performance.
(5). Manual Breathers: These breathers require manual adjustment and maintenance, but are often less expensive than automatic breathers.

1.3 Selection and Installation of Breathers

When selecting a breather, it is essential to consider the following factors:
(1). Tank Size: The breather should be sized to match the tank capacity.
(2). Lubricant Type: The breather should be compatible with the type of lubricant used.
(3). Environmental Conditions: The breather should be designed to withstand the environmental conditions, such as temperature, humidity, and exposure to chemicals.
(4). Maintenance Requirements: The breather should be easy to maintain and replace.
(5). Budget: The breather should fit within the budget and provide good value for the cost.

When installing a breather, it is essential to follow the manufacturer's instructions and ensure that the breather is properly secured to the tank.

1.4 Maintenance and Replacement of Breathers

Breathers should be regularly inspected and maintained to ensure they are functioning correctly. The filter element should be replaced according to the manufacturer's recommendations, and the breather should be cleaned and inspected for damage.

1.5 Common Problems with Breathers

(1). Clogged Filters: Clogged filters can prevent the breather from functioning correctly, allowing contaminants to enter the tank.
(2). Damaged Breathers: Damaged breathers can allow contaminants to enter the tank, causing equipment failure and downtime.
(3). Incorrect Installation: Incorrect installation of the breather can prevent it from functioning correctly, allowing contaminants to enter the tank.
(4). Lack of Maintenance: Failure to maintain the breather can lead to clogged filters, damaged breathers, and equipment failure.

1.6 Best Practices for Breather Maintenance

(1). Regular Inspection: Regularly inspect the breather to ensure it is functioning correctly.
(2). Filter Replacement: Replace the filter element according to the manufacturer's recommendations.
(3). Cleaning: Clean the breather regularly to prevent damage and ensure optimal performance.
(4). Record Keeping: Keep records of breather maintenance and replacement to ensure that the breather is properly maintained.
(5). Training: Provide training to maintenance personnel on the proper installation, maintenance, and replacement of breathers.

Sub-sub sections should be avoided, but if necessary use same title format as subsections. A blank 11pt single space line must follow all section and subsection text.

2. Finally

Breathers are a critical component in rotating equipment lubrication tanks, playing a vital role in preventing contamination and ensuring the longevity of the equipment. By understanding the importance of breathers, selecting the right type, and following proper installation and maintenance procedures, industry professionals can help to optimize equipment performance, reduce downtime, and extend lifespan. Remember, a well-maintained breather is essential for a well-maintained lubrication system, and a well-maintained lubrication system is essential for a well-maintained machine.

Using a Breather with a High-Quality Filter: A high-quality filter will help to prevent contaminants from entering the tank.

Choose a Breather that is Compatible with Your Lubricant: Ensure that the breather is compatible with the type of lubricant used in your equipment.

Consider the Environmental Conditions: Choose a breather that is designed to withstand the environmental conditions in which it will be used.

Follow the Manufacturer's Instructions: Follow the manufacturer's instructions for installation, maintenance, and replacement of the breather.

Keep Records of Breather Maintenance: Keep records of breather maintenance and replacement to ensure that the breather is properly maintained.

By following these tips and best practices, you can help to ensure that your breathers are functioning correctly and providing optimal protection for your equipment.

A flexible maintenance policy for systems with CBM and FBM components: leveraging opportunistic maintenance

Ipek Kivanc*, Claudia Fecarotti*, Néomie Raassens*, Geert-Jan van Houtum*

Eindhoven University of Technology, Eindhoven, the Netherlands

Abstract. Effective maintenance strategies are essential for sustainable and cost-efficient operations in complex systems composed of heterogeneous components. This study proposes a flexible policy that integrates condition-based maintenance (CBM) and failure-based maintenance (FBM), leveraging opportunistic maintenance to reduce both urgent corrective interventions and unnecessary preventive replacements. A Markov decision process (MDP) framework is developed to determine optimal maintenance actions, supported by two heuristic methods. Heuristic 0 relies solely on corrective replacements, leading to higher costs, while Heuristic 1 incorporates semi-urgent visits and opportunistic maintenance to enhance performance and sustainability. Numerical results demonstrate significant reductions in cost, intervention frequency, and material waste. The proposed approach offers a scalable solution that balances economic, social, and environmental objectives, enabling more intelligent and sustainable maintenance planning.

Keywords: sustainable maintenance, opportunistic maintenance, condition-based maintenance, semi-urgent maintenance

The increasing complexity of industrial systems necessitates maintenance strategies that go beyond cost minimization and address sustainability, flexibility, and system heterogeneity. Traditional time-based policies often fail to accommodate differences in component degradation patterns, leading to either premature preventive replacements or costly reactive interventions. Recent research in condition-based maintenance (CBM) and opportunistic maintenance highlights the benefits of aligning maintenance decisions with real-time degradation data and system-level disruptions (e.g., Yang et al., 2018; Dinh et al., 2022; Castro et al., 2020).

This study addresses a finite-horizon maintenance optimization problem for a series-configured system of n heterogeneous components. Components are either subject to CBM or handled based on failures. The system is evaluated over T discrete time periods. At each time step t, the system is in state $\mathbf{x} = (x_1, x_2, ..., x_n)$, where x_i represents the degradation level or binary failure state of component i. Maintenance actions are defined by the vector $\mathbf{a} = (a_0, a_1, ..., a_n)$, with $a_0 \in \{0, 1, 2\}$ representing the type of visit: 0 for no visit, 1 for semi-urgent preventive, and 2 for urgent corrective. For each component, $a_i \in \{0, 1\}$ indicates whether it is replaced. An urgent visit is triggered when any CBM component reaches its failure threshold or any FBM component fails, in which case replacement is mandatory.

The total maintenance cost associated with action $\mathbf{a}$ is given by:

$$C(a) = c^{ssd} \cdot 1_{\{a_0=1\}} + c^{usd} \cdot 1_{\{a_0=2\}} + \sum_{i=1}^{n} c_i^{rep} \cdot 1_{\{a_i = 1\}}$$

where, c^{ssd} and c^{usd} represent the setup costs of semi-urgent and urgent visits, respectively, and c_i^{rep} denotes the replacement cost of component i. Indicator functions are used to activate the appropriate costs depending on the action taken.

Component degradation evolves stochastically over time. Each CBM component follows a discrete-time Markov chain with stationary increments. If not replaced, a component transitions from x_i to y_i with probability $P_{(x_i,y_i)}$; if replaced, it resets to a new state and transitions with probability $P_{(0,y_i)}$. The system-wide transition probability from state $\mathbf{x}$ to $\mathbf{y}$ under action $\mathbf{a}$ is the product of individual component transition probabilities.

The objective is to minimize the total expected maintenance cost across the planning horizon. The

Bellman recursion for the optimal cost function $V_{t(x)}$ is:

$$V_{t(x)} = \min_{a \in A(x)} \{ C(a) + \sum_{y} P(x, a, y) \cdot V_{t+1}(y) \}$$

At the final time T, if any component has failed, an urgent replacement is executed; otherwise, no further cost is incurred.

Due to the combinatorial explosion in state and action spaces, solving the full MDP becomes computationally intractable for large systems. To address this, we develop two heuristics. Heuristic 0 considers only corrective maintenance and serves as a baseline. Heuristic 1 introduces two thresholds per CBM component: one for triggering semi-urgent visits and another for enabling opportunistic replacements during any visit. Components exceeding their opportunistic threshold are also replaced proactively, enabling grouped interventions that reduce setup costs and future unexpected failures.

We analyze relatively small systems composed of both CBM and FBM components and demonstrate that the optimal policy adheres to a threshold-based structure. Specifically, semi-urgent and opportunistic maintenance decisions follow control-limit rules, with actions triggered when degradation levels exceed dynamic, time-dependent thresholds. The results confirm the optimality of threshold-based policies, while the proposed heuristics achieve near-optimal performance with significantly reduced computational complexity. This study contributes to the field of sustainable maintenance by illustrating how real-time data and probabilistic modeling can enable smarter, more adaptive maintenance planning in heterogeneous systems.

References

Castro, I. T., Cavalcante, C. A. V., & Scarf, P. A. (2020). Maintenance optimization models for systems with components subject to degradation and random failure. *European Journal of Operational Research*, *280*(2), 601–613. https://doi.org/10.1016/j.ejor.2019.07.042

Dinh, Q. T., van Houtum, G. J., & Adan, I. J. B. F. (2022). A maintenance optimization model for systems with economic dependencies and partially observable deterioration. *Reliability Engineering & System Safety*, *221*, 108340. https://doi.org/10.1016/j.ress.2022.108340

Peng, H., Zhang, Z., & Chen, X. (2017). Approximate solution for condition-based maintenance with opportunistic maintenance. *Computers & Industrial Engineering*, *112*, 765–777. https://doi.org/10.1016/j.cie.2017.08.041

Poppe, N., Boute, R. N., Lambrecht, M. R., & van Houdt, B. (2018). Joint condition-based and corrective maintenance optimization for a multi-component system with degradation-dependent costs. *IISE Transactions*, *50*(12), 1076–1092. https://doi.org/10.1080/24725854.2018.1474316

Yang, B., Zeng, Y., & Liu, Y. (2018). Opportunistic maintenance for multi-component systems with production stoppages. *Reliability Engineering & System Safety*, *180*, 139–148. https://doi.org/10.1016/j.ress.2018.07.006

Zhu, J., Modarres, M., & Jin, T. (2015). A condition-based opportunistic maintenance strategy for multi-component systems with economic dependence. *Reliability Engineering & System Safety*, *142*, 205–215. https://doi.org/10.1016/j.ress.2015.05.016

Pressure-Driven Resilience Assessment of Water Distribution Networks: An Integrated Probabilistic Approach for Critical Infrastructure Management

Gloria MARUCHU[1], Rasa REMENYTE-PRESCOTT[1], Silvia TOLO[1]
[1] Resilience Engineering Group, University of Nottingham, University Park, UK
email: gloria.maruchu@nottingham.ac.uk
email: r.remenyte-prescott@nottingham.ac.uk
email: silvia.tolo@nottingham.ac.uk

Abstract. Water distribution networks face challenges from urbanization, climate change, and aging infrastructure. This study presents a probabilistic framework for enhancing network resilience under multiple failure scenarios. Traditional pipe criticality assessments overlook dynamic hydraulic conditions; we integrate real-time pressure indicators to improve vulnerability analysis. Using EPANET hydraulic simulations and custom Python code, we model network performance during failures. Our probabilistic failure and recovery model captures uncertainties in response times and operational constraints. Case studies show pressure-based rankings improve maintenance prioritization, shorten recovery times, and enhance resilience. The framework helps utilities optimize resources and maintain service during disruptions.
Keywords: Resilience, Hydraulic Simulations, Probabilistic Modelling, Water Distribution Network

Water distribution networks (WDNs) are facing significant challenges due to rapid urbanisation, population growth, and climate change, compounded by aging infrastructure and increasing demand for reliable and efficient service. Traditionally, a method of identifying pipe criticality ranking has relied on static parameters, such as but not limited to pipe size and population density. While these factors are important, they often fail to capture the dynamic nature of pressure variations within the network, which can provide more detailed insights into the resilience of WDNs. Inadequate water service, as characterised by insufficient pressure and failure to meet customer demand, is a common issue in WDN performance (Moriarty et al., 2011).

In addition, efficient prioritisation of pipe replacements during multiple failure events is crucial for minimising downtime and restoring service quickly, thereby improving overall network resilience. Timely mobilisation of personnel and equipment for replacement is essential for reducing service disruptions and accelerating recovery.

This paper addresses two key gaps in literature. Firstly, it introduces a pressure-based metric for pipe criticality ranking, which is crucial for prioritising replacements in multiple failure scenarios. Current methods do not adequately account for dynamic pressure variations, which are essential for assessing network resilience. Secondly, the study develops a probabilistic framework for modelling network recovery, considering uncertainties, such as variable replacement and response times, as well as operational constraints. By integrating these two approaches, the research aims to provide a more objective and practical method for improving decision-making in the management of pipe failures and enhancing the resilience of WDNs.

Data on pipe properties, nodal demands, and network topology were used from two case studies, one of a small water supply project in Tanzania and the second one Net 3 example network taken from the EPANET, a publicly available software tool for hydraulic modelling. The Darcy-Weisbach equation was employed to simulate head losses across the two example networks. A Python-based script was developed to take in the diverse pipe failure rates based on pipe parameters, such as, but not limited to, location, size and material, that are assumed to simulate failures following a Poisson distribution. Response times were modelled using a beta distribution, and our study considered different response scenarios to understand their impact on network recovery, a minimum of 4 hours and maximum of 24 hours (standard in the UK for authorities to respond after a disruption is reported). The pipe failure rate and water authority response time to failures were obtained from expert consultation, and the Poisson distribution has been adopted due to its common usage in modelling the number of failures over a time period. A recovery model, Eq. (1) was developed to probabilistically simulate the network's return to

normal operation after a failure, which included evaluating the effect of different repair strategies on the speed of recovery:

$$Recovery_i = \frac{\left(\sum_{j=1}^{j=N} Q_{OS,j} - \sum_{j=1}^{j=N} Q_{DS,j}\right)}{\sum_{m=1}^{M} A_{mT}} \quad (1)$$

Where A_{mT} is the total increment of the water age for the disturbed node m in hours between two time steps (t-1, T) before the disruption, T is the time duration in hours of the analysis during the disruption and before the end of the replacement works, N is the number of nodal junctions, M is the number of nodal junctions that are affected by the disruption, Q_{OS} and Q_{DS} are the nodal water supply in l/s before (normal state) and after (disrupted state) the failure of pipe i respectively.

Figure 1 provides a schematic representation of failure occurrences, disrupted states, mean response times, and the patterns of varying recovery rates, used in this study.

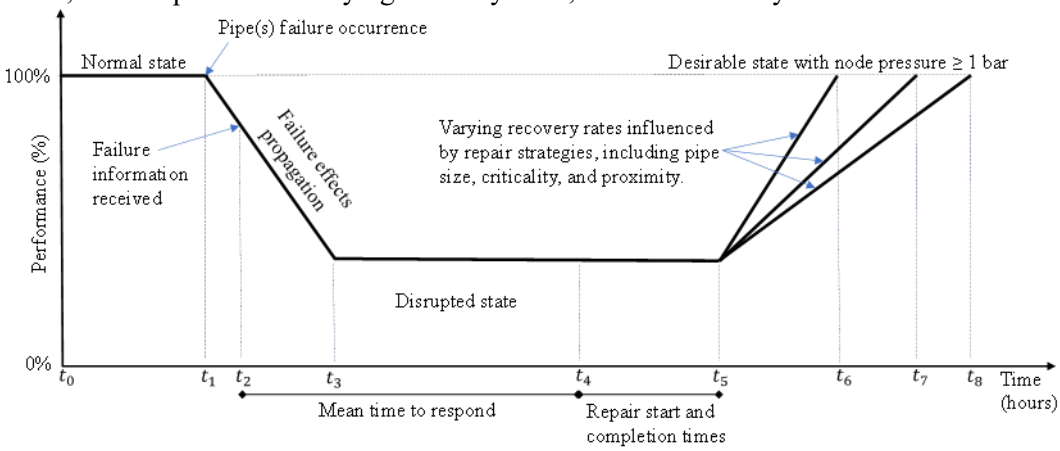


Figure 1. A schematic representation of the performance curve for the simulation of WDN resilience.

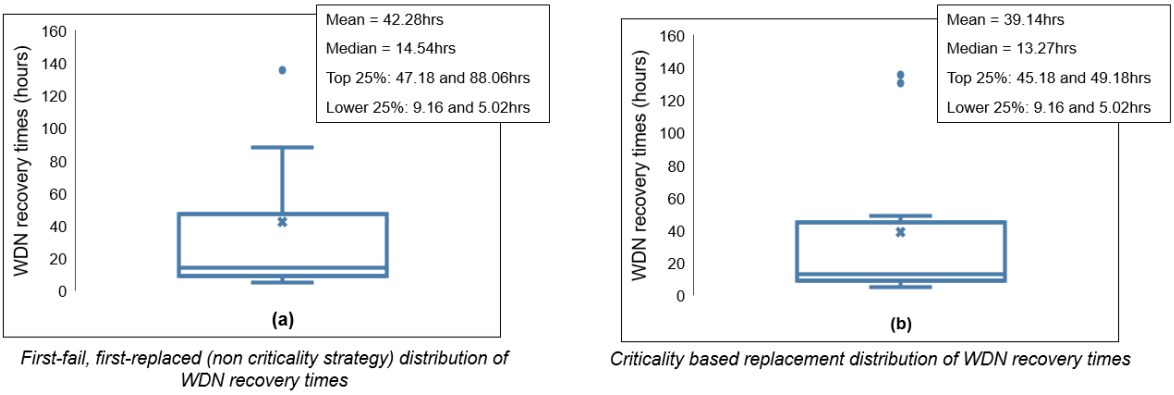


Figure 2. Net 3 Results of responses to multiple failure scenarios (*Data source*: EPANET)

Results in Figure 2 show that the criticality-based approach yields faster recovery times, indicating a more resilient response to pipe failures. This strategy enhances recovery performance, reduces downtime, minimizes customer complaints, and improves regulatory compliance. Utilities can leverage criticality rankings to prioritize preventive replacements, optimize resources, and improve contingency planning for more efficient network management. By conducting comprehensive simulations of pipe replacement scenarios, organizations gain deeper insight into network vulnerabilities and resilience, which supports evidence-based decisions for infrastructure investments, emergency response protocols, and overall asset management strategies. Future work aims to validate the results from Net 3 with comparable metrics found in literature, also perform the sensitivity analysis of the existing parameters.

References

Moriarty, P., Batchelor, C., Fonseca, C., Klutse, A., Naafs, A., Nyarko, K., Pezon, C., Potter, A., Reddy, V. R., & Snehalatha, M. (2011). Ladders for assessing and costing water service delivery. *IRC - International Water and Sanitation Center, April*, 24. http://www.washcost.info/page/753

A Concept of knowledge-based Recommendation System for Sustainable Maintenance in Flexible Manufacturing Systems

G. Princz*, K.Spiess*, S.Erol* and F. Ansari**

*University of Applied Sciences Wiener Neustadt, *Johannes-Gutenberg-Straße 3, A-2700 Wiener Neustadt (gabor.princz@fhwn.ac.at)*

*University of Applied Sciences Wiener Neustadt, *Johannes-Gutenberg-Straße 3, A-2700 Wiener Neustadt (122336@fhwn.ac.at)*

*University of Applied Sciences Wiener Neustadt, *Johannes-Gutenberg-Straße 3, A-2700 Wiener Neustadt (selim.erol@fhwn.ac.at)*

**TU-Vienna, *Theresianumgasse 27, 1040 Wien (fazel.ansari@tuwien.ac.at)*

Abstract. Flexible manufacturing systems (FMS) face escalating challenges due to the vast number of components and potential failure causes, intensified by diverse product portfolios and fluctuating production volumes. This paper introduces an agent-based adaptive recommendation system that signifies a paradigm shift in data-driven maintenance management. By integrating the expertise of maintenance engineers and personnel, the system employs a Natural Language Processing (NLP)-based model to classify symptoms, causes, and solutions with high accuracy (achieving 100% for symptoms, 90% for causes, and 85% for solutions). Leveraging Bayesian networks to model cause-and-effect relationships, the system enables probabilistic inferences and provides resource-efficient recommendations that minimize energy use, resource consumption, and waste. Continuous feedback mechanisms enhance the system's performance over time, improving diagnostic and repair processes while promoting sustainable maintenance practices.

Keywords: Health monitoring and diagnosis, Artificial intelligence, Knowledge modelling and knowledge based systems, Flexible and reconfigurable manufacturing systems

Flexible manufacturing systems (FMS) are widely employed in automotive, semiconductor, and shipbuilding industries, relying on modular production lines that comprise processing units, material handling systems, storage systems, and control modules (Cronin, Conway, and Walsh 2019; Chen et al. 2020). Due to the complexity introduced by diverse product portfolios and fluctuating production volumes, these components endure varying degrees of wear and tear that complicate failure diagnosis (Javaid et al. 2022). The detection of early symptoms often depends on human expertise, as maintenance personnel are typically the first to notice anomalies such as a squeaky bearing, passing this information along for more thorough troubleshooting (Xia et al. 2022). This reliance on direct observation can lead to shutting down entire lines or replacing multiple components at once, resulting in wasted resources and higher energy consumption (Sipsas et al. 2016). To mitigate these inefficiencies, an agent based adaptive recommendation system integrates historical logs, operator feedback, and Bayesian Networks (BNs) to provide refined, data driven recommendations. This includes precise actions such as component lubrication or replacement, thereby reducing downtime and optimizing resource usage (Yi et al. 2023). Feedback from maintenance personnel continuously refines the system's predictive capabilities, ensuring robust performance in rapidly changing production environments and fostering more sustainable maintenance practices through lower resource usage, reduced troubleshooting time, and energy efficient solutions.

Some research has been conducted on integrating advanced maintenance frameworks into FMS, but a crucial need persists to leverage maintenance personnel's cognitive insights for issues not easily captured by sensors. Yetis and Karakose (2018) show how Industry 4.0 enhances flexibility but also increases production error risks, prompting various Decision Support Systems (DSS) for early failure detection and resolution. Data driven approaches (Grosso et al. 2023), knowledge based frameworks (Ansari, Glawar, and Sihn 2020), and agent based systems (Afsar, Crump, and Far 2022) have been examined, typically using sensor data, expert knowledge, and fault logs. Digital twins provide virtual representations that detect anomalies and recommend resource saving strategies (Tuli et al. 2021; Onile et al. 2021), while agent based recommendation systems adopted from e commerce (Guo and Liu 2010) are used to propose industrial solutions, often employing Naive Bayes algorithms for fault detection (He et al. 2021). Although FMECA

data and BNs enable adaptive, tailored maintenance recommendations, the gap remains in systematically capturing and applying operator expertise to address sensor limited symptoms.

To address this, an agent-based recommendation system is developed in the Festo FMS50 environment, guided by a Failure Modes, Effects, and Criticality Analysis (FMECA) that classifies failure modes with a hierarchical data structure of symptom labels, root causes, and solutions. Data preparation integrates text augmentation methods such as lemmatization, paraphrasing, synonym replacement, and back translation to increase textual diversity without compromising semantic quality. In the pre-processing phase, tokenization, punctuation removal, and lemmatization standardize the text, while newly encountered labels are dynamically encoded. TF-IDF feature extraction then transforms the text into weighted term vectors, highlighting domain specific patterns. Three machine learning classifiers predict symptoms, root causes, and solutions, trained via Stochastic Gradient Descent (SGD) with a logistic regression loss function, where inverse frequency weighting handles class imbalances. A Bayesian Network (BN) captures probabilistic relationships among symptoms, root causes, and solutions, using a Dirichlet prior for parameter estimation and variable elimination for inference. Finally, a sustainability impact metric that accounts for energy usage, resource consumption, waste generation, and implementation time guides the system toward low impact solutions when posterior probabilities are similar.

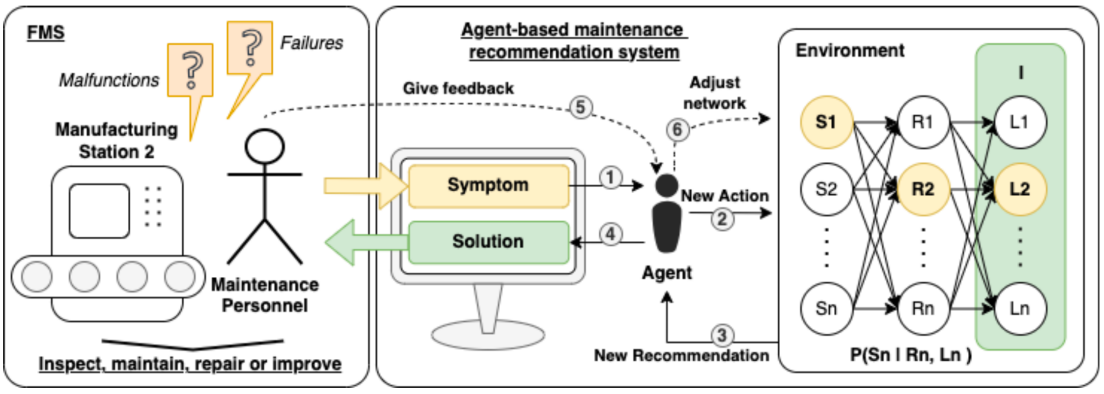


Figure 1: Framework of the agent-based maintenance recommendation system

As shown in Figure 1, malfunctions in the FMS (Flexible Manufacturing System) are initially identified by maintenance personnel, who report specific Symptoms (S) that the agent-based system evaluates using a Bayesian Network (BN) to determine the most likely Root Causes (R) and recommend corresponding Solutions (L). Sustainability metrics (I) are also integrated, enabling the framework to prioritize energy-efficient and resource-saving actions while continuously adapting through operator feedback and new data. Implemented in the FMS50 environment, this approach combines human expertise with probabilistic reasoning and sustainability considerations, relying on three classifiers—Symptom, Root Cause, and Solution—that were trained on augmented and preprocessed maintenance data. The Symptom Classifier demonstrated an accuracy of 100 percent, highlighting its robustness in identifying reported symptoms. In comparison, the Root Cause Classifier achieved a 90 percent accuracy, indicating a need for further representative data in certain classes. The Solution Classifier attained 85 percent accuracy, suggesting that additional training samples could improve feature differentiation. The adaptive model updating mechanism, which integrates new data and sustainability considerations, ensures that the system remains responsive to evolving operational conditions.

Future research aims to enhance text classification through alternative methods or large language models (LLMs), incorporate real-time sensor data for improved residual use (RUL) predictions, and quantitatively assess the framework's influence on energy consumption, resource usage, and waste in dynamic manufacturing settings.

References

Afsar, M. Mehdi, Trafford Crump, and Behrouz Far. 2022. 'Reinforcement Learning Based Recommender Systems: A Survey'. *ACM Comput. Surv.* 55 (7): 145:1-145:38. https://doi.org/10.1145/3543846.

Ansari, Fazel, Robert Glawar, and Wilfried Sihn. 2020. 'Prescriptive Maintenance of CPPS by Integrating Multimodal Data with Dynamic Bayesian Networks'. *Machine Learning for Cyber Physical Systems*, 1–8. https://doi.org/10.1007/978-3-662-59084-3_1.

Chen, Yu-Qiang, Biao Zhou, Mingming Zhang, and Chien-Ming Chen. 2020. 'Using IoT Technology for Computer-Integrated Manufacturing Systems in the Semiconductor Industry'. *Applied Soft Computing* 89 (April):106065. https://doi.org/10.1016/j.asoc.2020.106065.

Cronin, Con, Andrew Conway, and Joseph Walsh. 2019. 'Flexible Manufacturing Systems Using IIoT in the Automotive Sector'. *Procedia Manufacturing* 38:1652–59. https://doi.org/10.1016/j.promfg.2020.01.119.

Grosso, Luca Actis, Andrea De Martin, Giovanni Jacazio, and Massimo Sorli. 2023. 'Development of Data-Driven PHM Solutions for Robot Hemming in Automotive Production Lines'. *International Journal of Prognostics and Health Management* 11 (1). https://doi.org/10.36001/ijphm.2020.v11i1.2591.

He, Sen, An Chang, Hourong Zhang, Yunhai Song, Jianing Shang, Gang Zhao, and Weiming Li. 2021. 'Fault Causes Identification of Transmission Lines Based on Weighted Naive Bayes Classification Algorithm Combined with Complex Algorithm'. In *2021 International Conference on Power System Technology (POWERCON)*, 2312–16. https://doi.org/10.1109/POWERCON53785.2021.9697835.

Javaid, Mohd, Abid Haleem, Ravi Pratap Singh, and Rajiv Suman. 2022. 'Enabling Flexible Manufacturing System (FMS) through the Applications of Industry 4.0 Technologies'. *Internet of Things and Cyber-Physical Systems* 2:49–62. https://doi.org/10.1016/j.iotcps.2022.05.005.

Onile, Abiodun E., Ram Machlev, Eduard Petlenkov, Yoash Levron, and Juri Belikov. 2021. 'Uses of the Digital Twins Concept for Energy Services, Intelligent Recommendation Systems, and Demand Side Management: A Review'. *Energy Reports* 7 (November):997–1015. https://doi.org/10.1016/j.egyr.2021.01.090.

Sipsas, Konstantinos, Kosmas Alexopoulos, Vangelis Xanthakis, and George Chryssolouris. 2016. 'Collaborative Maintenance in Flow-Line Manufacturing Environments: An Industry 4.0 Approach'. *Procedia CIRP* 55:236–41. https://doi.org/10.1016/j.procir.2016.09.013.

Tuli, Tadele Belay, Linus Kohl, Sisay Adugna Chala, Martin Manns, and Fazel Ansari. 2021. 'Knowledge-Based Digital Twin for Predicting Interactions in Human-Robot Collaboration'. In *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–8. Vasteras, Sweden: IEEE. https://doi.org/10.1109/ETFA45728.2021.9613342.

Xia, Liqiao, Pai Zheng, Xinyu Li, Robert.X. Gao, and Lihui Wang. 2022. 'Toward Cognitive Predictive Maintenance: A Survey of Graph-Based Approaches'. *Journal of Manufacturing Systems* 64 (July):107–20. https://doi.org/10.1016/j.jmsy.2022.06.002.

YETİŞ, Hasan, and Mehmet KARAKÖSE. 2018. 'A Data-Driven Method for Decision Support Systems in Mass Production and Mass Customization'. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1–4. https://doi.org/10.1109/IDAP.2018.8620794.

Yi, Zhengyao, Siyao Mi, Tianqi Tong, Haoming Li, Yan Lin, Wenbiao Wang, and Jiangbo Li. 2023. 'Intelligent Initial Model and Case Design Analysis of Smart Factory for Shipyard in China'. *Engineering Applications of Artificial Intelligence* 123 (August):106426. https://doi.org/10.1016/j.engappai.2023.106426.

Joint environmental and economic assessment of cutting tool replacement policies in Ti6Al4V milling under uncertain condition estimate

L. Colantonio*, G. Moroncini*, O. Sénéchal**, P. Dehombreux*, F. Ducobu* and L. Equeter*

**Machine Design and Production Engineering Unit, University of Mons, Place du Parc, 20, 7000 Mons, Belgium [lorenzo.colantonio, gerome.moroncini, pierre.dehombreux, francois.ducobu, lucas.equeter]@umons.ac.be*
*** Université Polytechnique Hauts-de-France, CNRS UMR 8201, LAMIH, Valenciennes, France olivier.senechal@uphf.fr*

Abstract. An optimal cutting tool replacement policy balances the risk of producing out-of-tolerance pieces, and limits both excessive consumption of new tools and production interruption. In this paper, based on advances in cutting tool monitoring, we assume that it is possible to monitor the tool wear. We use a Monte-Carlo simulation method to show, through an example in Ti6Al4V milling, that a condition-based replacement allows improved economic and ecological performance, considering the single indicator of Global Warming Potential, expressed in kg_{eq} CO_2, of the tool replacement policy, even under uncertainty on the condition estimate.

Keywords: Sustainability, machining, condition monitoring, environmental assessment, replacement policy

Untimely management of cutting tools represents 30% of the machining costs (Liu, Li, Hua, Lu & Mou, 2018), either due to out-of-tolerance production or waste of a portion of the tool's useful life. Further, the improper tool replacement also induces an environmental impact, even though the reduction of cutting tool use is underrepresented in the literature (Elnourani, Johansen, & Öhrwall Rönnbäck, 2024).

Current efforts mainly concern tuning the machining parameters (including the choice of tool) to optimize the tool life duration or sustainability (Pimenov et al., 2022), but further, it is also possible to estimate the remaining useful life of a tool in an assessment that includes the economic, social and environmental impact of the tool use (Sun et al., 2020). However, using such a scheme to compare competing replacement policies has not been attempted.

In this contribution, we compare three strategies for replacing cutting tools:

1. Fixed periodicity systematic replacement that is a multiple of the machining duration
2. At the end of machining each workpiece, based on an estimate of the tool wear based on condition monitoring
3. At more frequent opportunities that would interrupt machining, based on the same estimate of the tool wear based on condition monitoring.

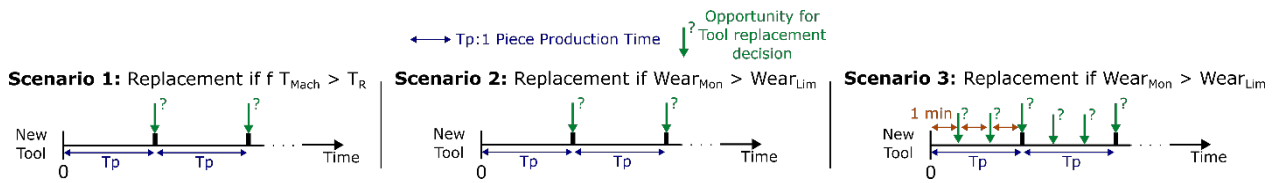


Figure 1. Replacement strategies that are compared in this contribution, with: T_{Mach} the elapsed duration of machining of the current tool, T_p the machining duration for one workpiece, T_R the replacement periodicity of scenario 1, $Wear_{Lim}$ the wear replacement criterion, and $Wear_{Mon}$ the wear evaluation produced by the condition monitoring system.

In order to do so, we study a face milling operation. The economic model is built on the cost of new tools and of workpiece material. The environmental impact is limited to the category of Climate Change, in kg_{eq} CO_2. It includes tool manufacturing, according to the values from Furberg et al., 2019, tool coating through a Physical Vapor Deposition process from Fairoz & Shokrani, 2025, Ti6Al4V workpiece production from Lyons et al., 2021, and energy production from the current Belgian grid. The milling operation can produce three items: discarded workpieces (scrap), good workpieces, and worn tools. The scrap recycling, the use and end of life of good workpieces, as well as the tool repair or recycling, are out of the system boundaries.

In scenarios 2 and 3, the condition (flank wear) estimate is assumed to be affected by a random error following a normal distribution and corresponding to a 15% Mean Absolute Percentage Error, matching

actual performances in similar applications (Colantonio, Equeter, Dehombreux & Ducobu, 2024). The tool is to be replaced whenever the flank wear is greater than 300 µm.

The production is simulated through the Monte-Carlo method, simulating a month's production before assessing the impacts. In each scenario, the impacts are then normalized by the number of good workpieces produced in order to compare the policies in terms of impact (cost or environmental footprint) per produced good workpiece.

All three scenarios are optimized for cost only, with respect to different parameters. In scenario 1, the parameter is the replacement periodicity (in minutes), whereas in scenarios 2 and 3, the parameter is the value of the tool wear estimate that triggers a tool replacement. In all cases, the optimal cost-wise parameter value is retained for comparing the scenarios.

The results (shown in Table 1) show that in all aspects, scenario 3 provides the best results, with a cost reduction of up to 10% in comparison to scenario 1, both limiting the number of used tools and the number of scraps thanks to the possibility of interrupting the machining operation, and despite the random error affecting the tool wear prediction.

Table 1. Comparison of the production for the different scenarios (1 month's production, CO_2 equivalent in metric tons)

	Good workpieces	Scrap	Number of tools	Total Cost [k€]	Total CO_2 [Teq CO_2]
Scenario 1	9030	569	1600	159 (100%)	519 (100%)
Scenario 2	8983	616	1484	153 (96.3%)	492 (94.7%)
Scenario 3	9214	386	1397	149 (93.6%)	472 (90.9%)

Further, scenario 3 has also the greatest number of good workpieces produced, thanks to a strongly limited scrap rate. Despite scenario 2 inducing more scraps than scenario 1, it still represents a cost and environmental gain, because it limits the number of used tools.

References

Colantonio L, Equeter L, Dehombreux P, Ducobu F (2024) Confidence Interval Estimation for Cutting Tool Wear Prediction in Turning Using Bootstrap-Based Artificial Neural Networks. *Sensors* **24**, 3432.

Elnourani M, Johansen K, Öhrwall Rönnbäck A (2024) Enabling Factors for Circularity in the metal cutting Industry – With Focus on High-Value Circular Tools. In *Advances in Transdisciplinary Engineering*. IOS Press.

Fairoz I, Shokrani A (2025) Life Cycle Assessment of Cutting Tool Coatings. *Procedia CIRP* (in Press).

Furberg A, Arvidsson R, Molander S (2019) Environmental Life Cycle Assessment of Cemented Carbide (WC-Co) Production. *Journal of Cleaner Production* **209**, 1126-38.

Liu C, Li Y, Hua J, Lu N, et Mou W (2018) Real-Time Cutting Tool State Recognition Approach Based on Machining Features in NC Machining Process of Complex Structural Parts. *The International Journal of Advanced Manufacturing Technology* **97**, 229-41.

Lyons R, Newell A, Ghadimi P, Papakostas N (2021) Environmental Impacts of Conventional and Additive Manufacturing for the Production of Ti-6Al-4V Knee Implant: A Life Cycle Approach. *The International Journal of Advanced Manufacturing Technology* **112**, 787-801.

Pimenov DY, Mia M, Gupta MK, Machado ÁR, Pintaude G, Unune DR, Khanna N, et al. (2022) Resource Saving by Optimization and Machining Environments for Sustainable Manufacturing: A Review and Future Prospects. *Renewable and Sustainable Energy Reviews* **166**, 112660.

Sun H, Liu Y, Pan J, Zhang J, Ji W (2020) Enhancing Cutting Tool Sustainability Based on Remaining Useful Life Prediction. *Journal of Cleaner Production* **244,** 118794.

Dynamic Vapour Sorption Experiments: analysis of results, advantages and future development in supporting systems modelling.

C.Wells

AWE Aldermaston, Reading, *cathie.wells@awe.co.uk*

Abstract. Predicting how materials will age is of vital importance in establishing component lifetimes in larger systems models. Dynamic Vapour Sorption experiments can help to find properties of materials, such as diffusivity and solubility, through modelling one dimensional flow. The parameters obtained can then be applied to more complex systems. Here we consider how best to analyse raw results from these experiments, showing that unique parameters for an analytical solution and for Henry's mode absorption can be found from a simple non-linear fit to the experimental data, but that including Langmuir terms challenges the assumption of uniqueness.

Keywords: DVS, diffusion, Henry's mode, Langmuir mode, solubility, parameter fitting, PDEs

The Dynamic Vapour Sorption (DVS) process involves small samples being placed in a vessel attached to a balance inside a sealed unit, which is kept at a fixed temperature. The mass of the sample is recorded at intervals, as water vapour is allowed to travel over the vessel, providing a changing relative humidity (RH). The mass of water absorbed by the sample can be used to determine transport properties describing the material.

Estimating water transport properties from DVS in materials has been considered in a number of previous studies. Approaches vary from approximating the governing diffusion equation as a linear relationship (Iyer, et al. 2023), to including Langmuir mode, where adsorption and desorption of molecules around micro-voids are considered (Roy, et al. 2022) and even triple-mode sorption simulations, which add in the complexity of pooling (Sharma, et al. 2017, Sun, et al. 2017). In more recent research, Foley et al use time dependence to estimate an effective diffusivity for each RH step and then find transport properties by solving systems of equilibrium equations chosen by categorising the observed mass gain as one of five material dependent cases (Foley, et al. 2024).

Analytical solutions for a variety of geometries are available, including infinite planes and spheres (Crank 1975) and cylinders (Jones 1973). These are derived under the assumption that relative humidity (RH) will be constant, so although appropriate for the longer equilibrium sections of the data, these are not appropriate for materials where the rate of absorption is fast comparable to the RH ramp up. Fitting these solutions to experimental data allows an initial estimate for the diffusion coefficient to be sought, whilst the solubility can be estimated from the ratio of internal to external concentration at all time steps for which mass is stable.

In the experiments described here a fixed temperature of 35°C was chosen. RH was increased from 0 to 50%, in steps of 10%, allowing the mass increase of the sample to equilibrate before increasing to the next RH level. At 50% RH and below, it is not necessary to consider pooling mechanisms (Roy, et al. 2022), so here we consider two different sorption mechanisms:

- Henry's mode sorption, when we consider just the absorption of free water molecules from the external concentration:

$$\frac{\partial C}{\partial t} = D\nabla^2 C \qquad C(x=0) = \gamma c\,, C(t=0, x \neq 0) = 0, \frac{\partial C}{\partial t}(x=h) = 0$$

  where C is internal concentration in mg cm^{-3}, D is diffusion coefficient in cm^2 min^{-1}, c is external concentration in mg cm^{-3}, γ is dimensionless solubility and h is depth of sample in cm.
- a dual mode combining Henry's mode and the Langmuir mode, to also describe the adsorption of previously free molecules and the desorption of previously bound molecules at sites within the material:

$$\frac{\partial C}{\partial t} = D\nabla^2 C - \frac{\partial n}{\partial t} \qquad C(x=0) = \gamma c\,, C(t=0, x \neq 0) = 0, \frac{\partial C}{\partial t}(x=h) = 0$$

$$\frac{\partial n}{\partial t} = b' k_{de} C_H' C - (b'C + 1) k_{de} n \qquad n(x=0) = 0\,, n(t=0) = 0, \frac{\partial n}{\partial t}(x=h) = 0$$

  where n is bound concentration in mg cm^{-3}, b' is Langmuir affinity in cm^3 mg^{-1}, k_{de} is the desorption

coefficient in min^{-1} and C'_H is Langmuir capacity in mg cm^{-3}.

Software was developed to find parameter sets to allow optimal solution for each of these sets of partial differential equations (PDE), in the first case using MATLAB's nonlinear regression tool and in the second using the Dakota Single Objective Genetic Algorithm (SOGA).

For datasets where the absorption mechanism is dominated by Henry's mode, parameters to give the optimal solution to the diffusion equation, when compared with data simulated using a finite volumes method with added normally distributed noise, can be found in under 3 hours. Results for diffusion and solubility from this single mode method are improved when the time taken to increase to the next RH level is minimised. Initial values are provided by the analytical solution for diffusion and the estimated solubility from equilibrium values.

When datasets exhibiting a significant contribution from the Langmuir terms are considered, the dual mode method is used. This process again starts with the analytical solution to give starting values for the single mode method, which in turn provides boundaries for the genetic algorithm. One key advantage of providing parameter values for each mode, is that no decision on the applicable mechanism needs to be taken a priori. The dual mode method involves three different fits to allow the five parameters to be found more efficiently. An example of the final fit for a sample of Sylgard-184 is given in Figure 1. However, it is clear when experimental data is tested, that not all parameters appear to have unique values for the optimal fit. Figure 2 shows the range of parameter values that provide a fit within a 95% Wilks confidence interval of the optimal fit. From this the fit is shown to be largely insensitive to P1, P3 and P4, the Langmuir capacity, affinity and desorption respectively.

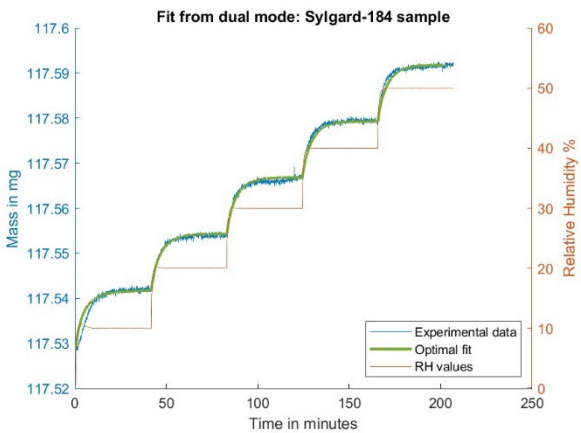


Figure 1. Optimal fit from dual mode.

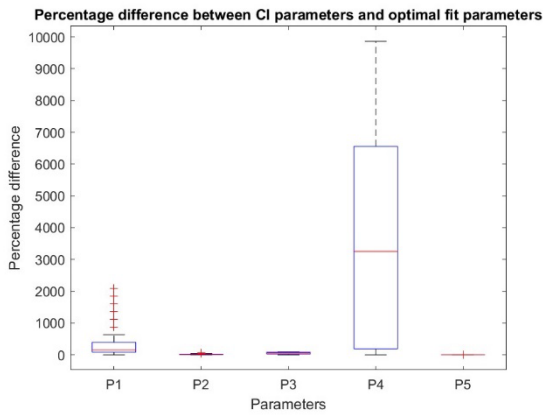


Figure 2. Percentage difference between optimal parameters and those providing a similar fit.

More research is needed to explore these patterns further, but parameter values for solubility and diffusion coefficient obtained to date compare well with those found using the method in Foley et al applied to the same experimental datasets (Foley, et al. 2024), whilst Langmuir parameters are more varied.

This publication is **UK Ministry of Defence © Crown Owned Copyright 2025/AWE**.

References

Crank J, (1975) *The Mathematics of Diffusion.* Clarendon Press.

Foley B, Aubry S, Murialdo M and Glascoe E (2024) Rigorous models for vapor diffusion in polymers with immobilization and chemical potential gradients. *J Phys Chem B* **128** (38) 9274-9283.

Iyer, J, Foley B, Qian F, and Aubry S (2023) Moisture Sorption and Swelling of Sylgard-184 *(LLNL-TR-853364)* Lawrence Livermore National Laboratory.

Jones W (1973) Diffusion controlled exchange of a finite cylinder with a finite amount of well mixed gas. *Journal of Applied Physics* **44**, 2946-2950.

Roy P, Castonguay S, Knipe J, Sun E, Glascoe E and Sharma H (2022) Multi-material modelling of sorption-desorption processes with experimental validation. *Chemical Engineering Science* **253** no. 117542.

Sharma H, Harley S, Sun Y, and Glascoe E (2017) Dynamic Triple-Mode Sorption and Outgassing in Materials. *Scientific Reports* **7** no. 2942.

Sun Y, Tong C, Harley S, and Glascoe E (2017) Expeditious Modeling of Vapor Transport and Reactions in Polymeric Materials. *AIChE Journal* **63** (9) 4079-4089.

Quality indicator prediction utilizing historical production data: A case study in steel manufacturing

T. Kontogiannis[*,**], Wanda Melfo[***], N. Eleftheroglou[*], D. Zarouchas[**]

[*] *Intelligent Sustainable Prognostics Group, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands*

[**] *Center of Excellence in AI for Structures, Aerospace Structures and Materials Department, Faculty of Aerospace Engineering, Delft University of Technology, The Netherlands*

[***] *Tata Steel Nederland Technology B.V., The Netherlands*

Abstract. This study presents a data-driven approach to predicting coiling temperature (CT) in the Hot Strip Mill process, addressing the limitations of traditional physics-based models. A neural network regressor with a custom loss function is proposed to enhance CT prediction and capture fluctuations linked to surface quality defects. The custom loss emphasizes rare deviations in CT profiles, improving both regression and anomaly detection performance. Evaluation against existing models, including a physics-based baseline, demonstrates the superiority of the proposed method, particularly in identifying sudden CT drops. These findings highlight the potential for improved quality control and proactive process monitoring in steel manufacturing.

Keywords: Quality prediction, historical production data, Coiling Temperature, Hot Strip Mill, data-driven, Artificial Neural Networks

Accurate prediction of quality indicators is critical in industrial manufacturing for both product assurance and process control. Traditional models based on physical laws—such as heat transfer and phase transformation—offer interpretability but often rely on simplifying assumptions, making them less effective under real-world conditions where non-ideal operations introduce variability (Yin et al., 2014).

This study targets the coiling temperature (CT) prediction in the Hot Strip Mill (HSM) process, a key quality indicator for final material properties (Min et al., 2012; Oktay et al., 2022; Sun et al., 2020). CT is measured via infrared pyrometers and can fluctuate due to surface conditions. A chipped oxide layer, for example, alters surface emissivity and produces abrupt CT drops (Min et al., 2012). Thereby, by observing the steel strip's coiling temperature (CT) profile, significant deviations from the goal temperature and, more importantly, rapid drops, can indicate a chipped-off oxide layer. These deviations often correlate with poor surface quality – information overlooked by traditional models due to their rigid assumptions.

To overcome these limitations, we propose a data-driven approach using historical production data and artificial neural network (ANN) regressors. Unlike analytical models, ANNs can learn complex relationships without predefined assumptions, improving prediction accuracy and robustness to variability. In a recent study (Panjiar & Murugananth, 2025), an Artificial Neural Network (ANN) has been utilized to predict the CT in the HSM process. However, this work presents two significant flaws that harm its real-world applicability. Firstly, a min-max normalization technique is applied to the entire dataset prior to the train-test split. While normalization techniques are standard practice to improve the accuracy and efficiency of ANN models (Kaufinann & 2006, 2007), only the training data should be used when calculating the transformation. Utilizing both the training and test data violates the requirement that the test set remains unseen during training, thereby constituting data leakage (Al Shalabi & Shaaban, 2006). Adding to that, by reviewing the results, the proposed model fails to capture the fluctuation observed in the CT profiles. Thus, it does not offer a significant improvement from the existing analytical models for CT prediction.

This study proposes ANN regressors with a custom loss function to improve the prediction accuracy of the CT profiles of the HSM process of Tata Steel Europe for heavy-gauge steel and the detection of fluctuations. The custom loss function tackles the challenge posed by data distributions that are centred on normal values, where deviations rarely occur. A sigmoid relevance function (Torgo & Ribeiro, 2007) is utilized to assign higher weights in the regression error of the data containing temperature drops following this equation:

$$L(\hat{y}, y) = MSE(\hat{y}, y) * e^{\Phi(\hat{y}, y)}, \qquad where\ \Phi(\hat{y}, y) = (1 - m) * \Phi(\hat{y}) + m * \Phi(y),$$

where $m \epsilon [0, 1]$ is a hyperparameter that balances the sensitivity to false positives and negatives, $\hat{y}$ and y the predicted and true CT values, respectively.

We benchmark four models: Tata Steel's numerical heat transfer–phase transformation model (HT-PT), gradient-boosted decision trees (GBDT), a standard ANN with MSE loss (ANN-MSE), and the proposed ANN

with custom loss (ANN-CL). Scatter plots of predicted vs. true CT values (Figure 1 -Figure 2) and performance metrics (Table 1) show that ANN-CL achieves the best results. To assess its ability to detect temperature anomalies, we set a -5% deviation threshold to reframe the regression task as anomaly detection. ANN-CL outperforms other models in RMSE, recall, and F1-score, effectively capturing rapid CT drops.
These findings demonstrate the value of data-driven modelling and domain-informed loss functions in enhancing quality prediction and surface defect detection in steel manufacturing.

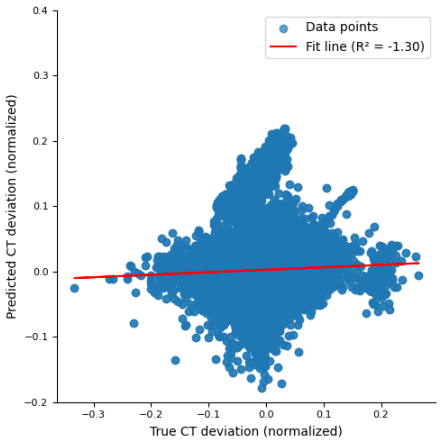


Figure 1 Scatter plot of HT-PT model predictions (CT is presented as the normalized deviation from the goal temperature)

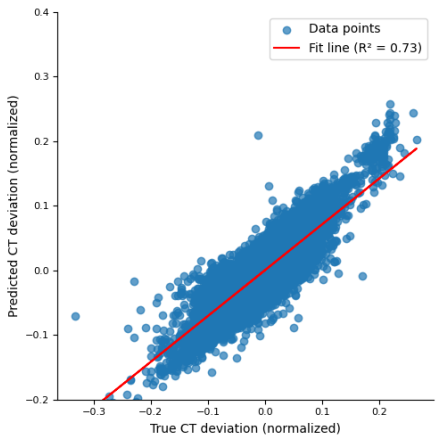


Figure 2 Scatter plot of ANN-CL model predictions (CT is presented as the normalized deviation from the goal temperature)

Table 1 Performance metrics of the four models

	RMSE	Recall	F1-score
HT-PT	0.0296	0.03	0.03
GBDT	0.0139	0.03	0.07
ANN-MSE	0.0126	0.38	0.50
ANN-CL	**0.0111**	**0.43**	**0.54**

Acknowledgements
This research is part of the DEPMAT project (with project number P20-22 / N21022) of the research programme Perspectief which is partly financed by the Dutch Research Council (NWO). It is also part of the Partnership Program of the Materials innovation institute M2i (www.m2i.nl).

References

Al Shalabi, L., & Shaaban, Z. (2006). Normalization as a Preprocessing Engine for Data Mining and the Approach of Preference Matrix. *Proceedings of International Conference on Dependability of Computer Systems, DepCoS-RELCOMEX 2006*, 207–214. https://doi.org/10.1109/DEPCOS-RELCOMEX.2006.38

Kaufinann, W. M.-M., & 2006, undefined. (2007). Data mining: Concepts and techniques. *Liacs.Leidenuniv.Nl*. http://liacs.leidenuniv.nl/~bakkerem2/dbdm2007/05_dbdm2007_Data%20Mining.pdf

Min, K., Kim, K., Kim, S. K., & Lee, D.-J. (2012). Effects of oxide layers on surface defects during hot rolling processes. *Metals and Materials International*, *18*(2),341–348.https://doi.org/10.1007/s12540-012-2020-8

Oktay, S., Nunzio, P. E. Di, Cesile, M. C., Davut, K., & Şeşen, M. K. (2022). Effect of coiling temperature on the structure and properties of thermo-mechanically rolled S700MC steel. *Journal of Mining and Metallurgy, Section B: Metallurgy*, *58*(3), 475–489. https://doi.org/10.2298/JMMB220304028O

Panjiar, H., & Murugananth, M. (2025). Artificial Neural Network Modeling of Coiling Temperature for Dual-Phase Steel. *Lecture Notes in Mechanical Engineering*, 345–358. https://doi.org/10.1007/978-981-97-6667-3_28/TABLES/2

Sun, M., Xu, Y., & Du, W. (2020). Influence of Coiling Temperature on Microstructure, Precipitation Behaviors and Mechanical Properties of a Low Carbon Ti Micro-Alloyed Steel. *Metals 2020, Vol. 10, Page 1173*, *10*(9), 1173. https://doi.org/10.3390/MET10091173

Torgo, L., & Ribeiro, R. (2007). Utility-Based Regression. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *4702 LNAI*, 597–604. https://doi.org/10.1007/978-3-540-74976-9_63

Yin, S., Ding, S. X., Xie, X., & Luo, H. (2014). A review on basic data-driven approaches for industrial process monitoring. *IEEE Transactions on Industrial Electronics*, *61*(11), 6418–6428. https://doi.org/10.1109/TIE.2014.2301773

Preliminary Data Exploration for Bayesian Stochastic Petri Net-Based Degradation Modelling in Hydropower Assets

J. Lally*, S. Gigot**, J.-M. Tacnet*** and C. Berenguer****

* *Compagnie Nationale du Rhone, Univ. Grenoble Alpes, France. E-mail: j.lally@cnr.tm.fr*
** *Compagnie Nationale du Rhone, France. E-mail: s.gigot@cnr.tm.fr*
*** *Univ. Grenoble Alpes, CNRS, INRAE, IRD, Grenoble INP, IGE, France. E-mail: jean-marc.tacnet@inrae.fr*
**** *Univ. Grenoble Alpes, CNRS, Grenoble-INP, GIPSA-lab, France. E-mail: christophe.berenguer@grenoble-inp.fr*

Abstract. In the hydropower industry, traditional reliability models are limited by failure data availability and quality. This research prepares for a Bayesian Stochastic Petri Net (BSPN) model for two assets: a spillway gate and an alternator. Maintenance records, inspection results, and operational data were compiled and analysed to assess the possibility of determining the parameters necessary for the BSPN development, such as the state transition probabilities and the relevant input variables for Bayesian updates. It was determined that the data was insufficient to make strong statistical inferences from. An alternative approach is proposed using simulated data to create the model.

Keywords: Hydropower, maintenance decision-making, asset management, degradation modelling.

The hydropower industry is at a turning point. As its role in the global energy mix grows due to the shift toward renewables and energy security concerns, asset management faces increasing challenges. Aging infrastructure, climate change, demanding operational regimes and inflationary pressures further constraining budgets complicate maintenance decision-making. Traditional reliability analysis methods struggle in this domain due to long asset lifespans resulting in limited, poor quality failure data, while the cost and complexity of condition monitoring have historically constrained degradation modelling (Welte, 2008). Together, these factors have motivated the industry to develop new methods of modelling degradation to inform condition-based preventative maintenance asset management practices. These methods must be adaptable to changes in the degradation drivers, with the capacity to integrate new data, as well as be able to better leverage different imperfect data sources, such as condition monitoring data and expert opinion, in conjunction with the existing failure data and condition assessments.

Previous research identified a Bayesian Stochastic Petri Net (BSPN) model would theoretically have the capacity to leverage the existing data in conjunction with expert knowledge, while maintaining the ability to adapt dynamically to new data as conditions change and monitoring improves (Lally et al., 2025). This graphical formalism, proposed by Taleb-Berrouane et al, combines the capabilities of Stochastic Petri nets (SPN), with Bayesian Networks. This empowers the SPN the ability to adapt dynamically to new data derived from changes in the degradation drivers and the available monitoring data, using Bayesian updating on the state transition functions at the SPN nodes. This characteristic makes the methodology particularly well adapted to modelling dynamic systems, where the drivers of degradation are changing over time, and refine its prediction capabilities in light of new information, such as in the case of many critical hydropower assets.

A review of the existing data in CNR, a French hydropower operator, was carried out across two of its asset classes (radial spillway gates and bulb turbine alternators) to determine if it were possible to directly identify some of the parameters necessary for the BSPN development, such as the state transition probabilities and the critical input variables for Bayesian updates. Several limitations were encountered. Although records existed for maintenance, failure incidents, and a recently commenced condition inspection regime, the data lacked the necessary quality and quantity. The failure data consisted of brief incident logs requiring substantial processing to distinguish actual failures, which was not feasible within the project timeframe. Maintenance data, while extensive, reflected interventions rather than true failure events and, when disaggregated by component, was too sparse to model time-to-intervention robustly (Rausand and Høyland, 2004). The inspection regime offered only one or two condition data points per component, making it impossible to track degradation over time or derive state transition probability laws (Si et al., 2011). Additionally, inconsistencies in how components were defined

across data sources hindered integration and comparative analysis. These issues collectively made it infeasible to directly infer the parameters needed for the BSPN model from the available data.

Due to limitations in the quality and quantity of the available data, an alternative, complementary approach is proposed for use until sufficient new data become available. The aim of this would be to empower the enterprise to put in place the framework necessary for future adoption of the BSPN maintenance model, by illustrating its potential benefits. This would also aid in studying through simulation which data the enterprise should focus on collecting for future model development, and what would be the optimal resolution, considering the collection costs and associated risk costs. This approach would involve the simulation of continuous degradation data and corresponding failure data, using existing degradation models for a series of critical components in an equipment. This data could then be used to determine the parameters needed to develop the BSPN (Taleb-Berrouane et al., 2020). Although this would not use true data, it would allow CNR to put the framework in place and enable us to study the value of information of the degradation level on the maintenance decision performance, by altering the resolution of available information on the degradation data set, using different censoring, time sampling or discretization techniques for the degradation state.

In this way, we could help to optimize the real inspection scheduling and asset monitoring and determine the ideal level of detail required to be gathered during inspections, when the cost-benefit of downtime losses and inspection and monitoring costs are balanced with the benefits of reduced uncertainty regarding asset health state and the prediction of failures. This approach would also then allow for a true equipment specific BSPN maintenance model to be developed over time as more inspections yielded real degradation data, taking advantage of Bayesian updates to refine the models' predictive capabilities, in conjunction with the imperfect existing data and data from other sources such as condition monitoring sensors (Nikulin et al., 2010). It is intended that this model will be used to inform predictive condition-based maintenance strategies, optimising resource allocation and improving maintenance decision-making in hydropower asset management, by more clearly defining the degradation trends and associated uncertainties that assets are exposed to.

Acknowledgements

This work is supported by La Compagnie Nationale du Rhône (CNR).

References

Lally J, Gigot S, Tacnet J.-M., and Berenguer C. (2025), Characterization, dynamic modeling, and monitoring of the degradation of hydroelectric production infrastructures.
Proc. of the 33rd European Safety and Reliability & 33th Society for Risk Analysis Europe Conference - ESREL2025 & SRA-E 2025, E. Bjorheim Abrahamsen, T. Aven, F. Bouder, R. Flage, M. Ylönen

Nikulin M, Limnios N, Balakrishnan N, Kahle W, and Huber-Carol C. (2010).
Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finance. Birkhauser Boston.

Rausand, M. and A. Høyland (2004).
System Reliability Theory: Models, Statistical Methods, and Applications (2nd ed ed.). Wiley Series in Probability and Statistics. Wiley-Interscience.

Si X.-S, Wang W, Hu C.-H, and Zhou D.-H. (2011). Remaining useful life estimation – A review on the statistical data driven approaches.
European Journal of Operational Research 213(1), 1–14.

Taleb-Berrouane M, Khan F, and Amyotte P. (2020). Bayesian Stochastic Petri Nets (BSPN) - A new modelling tool for dynamic safety and reliability analysis.
Reliability Engineering & System Safety 193, 106587.

Welte T. (2008). Deterioration and maintenance models for components in hydropower plants. Ph. D. thesis, Norwegian University of Science and Technology.

Evaluation of the health state of oil-immersed transformer based on dissolved gas analysis: An empirical study

Huy Vu Tran*, Kim Anh Nguyen*, Duc Hanh Dinh**, Dinh Duong Le*, Huu-Truong Le***, Phuc Do***

**Faculty of Electrical Engineering, The University of Danang - University of Science and Technology, Da Nang, Vietnam: thvu@ncs.dut.udn.vn, nkanh@dut.udn.vn (Corresponding author: K.A. Nguyen), ldduong@dut.udn.vn*
*** Faculty of Mechanical Engineering, The University of Danang - University of Science and Technology, Da Nang, Vietnam: ddhanh@dut.udn.vn*
**** Université de Lorraine, CRAN, CNRS, UMR-7039, Nancy 54000, France: huu-truong.le@univ-lorraine.fr, phuc.do@univ-lorraine.fr*

Abstract. Power transformers are vital for grid reliability, but insulation degradation remains a major failure risk. Dissolved gas analysis (DGA) is widely used for diagnostics, yet traditional methods lack sensitivity to early faults and face field implementation challenges. This study proposes the Dissolved Gas Analysis Factor (DGAF), a novel health index integrating concentrations of seven key gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, CO_2) with optimized weights to improve fault detection. Using 871 DGA samples from 240 transformers, optimization boosted sensitivity to 89.36% (97.59% accuracy, 98.06% specificity) and identified 16 latent faults among 824 normal samples, validated by Duval methods showing T2 thermal trends. A logistic regression model links DGAF to failure probability, aiding risk-based maintenance. By enabling early fault detection, DGAF supports proactive interventions to minimize outages and costs.

1. Introduction

Power transformers are vital for electrical grids, managing voltage and power distribution. Failures from insulation degradation, arcing, or thermal faults cause costly outages, environmental risks, and grid instability. Dissolved Gas Analysis (DGA) monitors health by analyzing gases like H_2, CH_4, C_2H_4, and C_2H_2, indicating faults such as partial discharge (PD), discharges (D1, D2), and thermal faults (T2, T3). However, traditional methods like Duval's Triangle and Rogers' Ratio miss early faults and struggle with mixed classifications, limiting proactive maintenance.

Recent studies enhance diagnostics with health indices. Jahromi *et al.*, (2009) combined DGA with oil and electrical tests, but it needs extensive data, impractical for routine use. Wattakapaiboon and Pattanadech (2016) used gas trends, though subjective weighting reduced applicability. Machine learning (Li *et al.*, 2023) improves fault detection but requires large datasets and lacks field-friendly interpretability, highlighting the need for a simple, sensitive DGA-based index.

This study introduces the DGAF, a new health index from 871 DGA samples of 240 transformers (ages 5–35 years, 110 kV). DGAF enhances fault detection with optimized weights and modeling, raising sensitivity from 72.34% to 89.36%, and detects 16 latent T2 faults in 824 normal samples, confirmed by traditional methods. A logistic regression model estimates failure probability for risk-based maintenance. Goals include: *(i)* validating DGAF for accuracy, *(ii)* enhancing early fault detection, and *(iii)* prioritizing high-risk transformers. Sections 2–5 cover methods, methodology, results, and conclusions.

2. Materials and Methods

This study evaluates the health state of oil-immersed transformers using DGA through an empirical approach. The methodology encompasses data collection, gas analysis, statistical processing, and health index formulation based on real-world transformer data. This section details the materials, procedures, and analytical techniques employed, ensuring the study's reproducibility and robustness.

2.1. Dataset description and performance evaluation

This research employed a dataset comprising 871 DGA samples derived from 240 oil-immersed power transformers managed by electrical utilities. The transformers exhibited a diverse range of characteristics, with ages spanning from 5 to 35 years (averaging 22 years), voltage ratings between 110 kV and 220 kV, and capacities from 25 MVA to 63 MVA. Each transformer utilized mineral oil as the insulating medium, adhering

to IEC 60296 standards. Seven key gases were analyzed and reported in parts per million (ppm): hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2). Within the dataset, 824 samples were classified as Normal (N), while 47 samples were identified as faulted, categorized as follows: 6 samples exhibited Partial Discharge (PD), 2 showed Low-Energy Discharge (D1), 4 indicated High-Energy Discharge (D2), 4 displayed Discharge-Thermal (DT) faults, 7 experienced Medium-Temperature Overheating (T2), and 24 demonstrated High-Temperature Overheating (T3). Analysis of 871 DGA samples showed clear gas patterns in Figure 1. These align with IEEE C57.104-2019, supporting DGAF diagnostics.

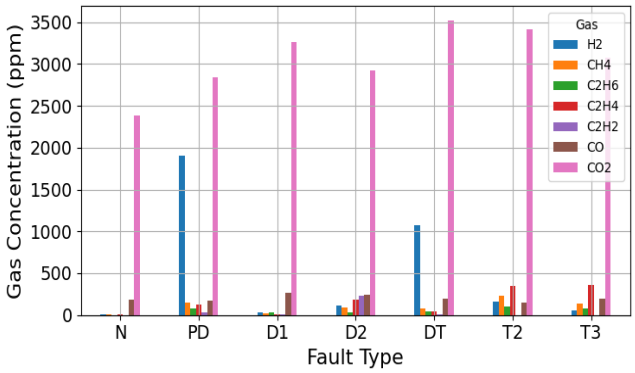


Figure 1. Mean gas concentrations by fault type.

The study evaluates the proposed method using key metrics from Zhu *et al.*, (2010). These metrics ensure a robust assessment of the method's performance in transformer health diagnostics

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}}; \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TN}}{\text{TP+FN}}; \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN+FP}}. \quad (3)$$

Where, TP, TN, FP, and FN are the counts of true positives, true negatives, false positives, and false negatives, respectively.

2.2. *DGAF formulation*

To enhance the accuracy of transformer condition assessment, a health index termed the DGAF was developed based on a weighted scoring system, assigning scoring factors (S_i) and weighting coefficients (W_i) to individual gases. The condition score for each gas is determined based on IEC 60599-2007/ IEEE C57.104-2019 standards (Prasojo *et al.*, 2020; Tamma *et al.*, 2021; *Li et al.*, 2023). These parameters were optimized to improve DGAF's sensitivity in distinguishing between minor and severe fault conditions (Haema *et al.*, 2013; Murugan *et al.*, 2019). DGAF was calculated as follows:

$$\text{DGAF} = \frac{\sum_{i=1}^{7} S_i.W_i}{\sum_{i=1}^{7} W_i}. \quad (4)$$

Where S_i is the score assigned to each gas based on its concentration (1 to 6) and W_i is the weighting factor reflecting its diagnostic importance (1 to 5).

These materials and methods provide a robust foundation for DGA-based transformer health assessment, leveraging a real-world dataset, standardized sampling, and rigorous statistical analysis.

3. Proposed method

This study introduces DGAF, an empirical health index from DGA data, using gas trends, weights, and modeling for reliable transformer assessment. It details DGAF's development, optimization, and validation with 871 samples from 240 transformers.

3.1. *DGAF framework*

The DGAF quantifies transformer health by combining the concentrations of seven key gases, namely hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2) into a single index (Eq. (4)). The DGAF value is then optimized and mapped to a condition rating (Table 1 and 2).

3.2. *Optimization of weighting factors*

Analysis of the 871-sample dataset revealed that the initial W_i values did not fully capture fault-specific gas trends. For instance, H_2 showed extreme elevation in partial discharge (PD) cases (e.g., 10,055.22 ppm), while

C_2H_4 dominated in high-temperature overheating (T3) samples (e.g., 2809.364 ppm). To improve sensitivity, the weighting factors were optimized as follows (Table 1).

Table 1. Optimized scoring and weighting actors.

Gas	Score S_i						Init. W_i	Opt. W_i	Noted
	1	2	3	4	5	6			
H_2	≤100	100-200	200-300	300-500	500-700	>700	2	3	Increased for PD, D1, D2
CH_4	≤75	75-125	125-200	200-400	400-600	>600	3	2	Reduced, less impact in severe faults
C_2H_6	≤65	65-80	80-100	100-120	120-150	>150	3	2	Reduced, similar to CH_4
C_2H_4	≤50	50-80	80-100	100-150	150-200	>200	3	4	Increased for T2, T3, D2
C_2H_2	≤3	3-7	7-35	35-50	50-80	>80	5	6	Increased for arcing
CO	≤350	350-700	700-900	900-1100	1100-1400	>1400	1	3	Increased for insulation aging
CO_2	≤2500	≤3000	≤4000	≤5000	≤7000	>7000	1	2	Increased, less diagnostic alone

The optimization was qualitative, guided by fault distributions (e.g., 824 normal vs. 47 faulted samples) and gas concentration patterns. For example, in a T3 sample (DGAF = 3.090909091), C_2H_4 reached 2809.364 ppm, justifying its higher weight. The total weight denominator increased from 18 to 22, normalizing the DGAF scale. The DGAF value is then mapped to a proposed condition rating (Table 2)

Table 2. Transformer condition ratings based on optimized DGAF.

Rating Code	Condition	Initial W_i	Optimized W_i
A	Good	DGAF < 1.2	DGAF < 1.2
B	Acceptable	1.2 ≤ DGAF < 1.5	1.2 ≤ DGAF < 1.5
C	Need	1.5 ≤ DGAF < 2	1.5 ≤ DGAF < 2
D	Poor	2 ≤ DGAF < 3	2 ≤ DGAF < 2.5
E	Very	DGAF ≥ 3	DGAF ≥ 2.5

3.3. *Failure probability estimation*

To enhance risk-based maintenance, a logistic regression model estimates the failure probability *P(Fail)* of transformers using DGAF values. The model uses the failure probability and is defined as follows:

$$P(Fail) = \frac{e^{(-\beta_0+\beta_1.\mathrm{DGAF})}}{1+e^{(-\beta_0+\beta_1.\mathrm{DGAF})}} \tag{5}$$

Where, β_0 and β_1 are regression coefficients and determined through statistical fitting using the 871-sample dataset (824 normal, 47 faulted). Using python, the logistic regression algorithm iteratively minimized the log-loss function to optimize β_0 and β_1, yielding β_0= -13.665 and $\beta_1 = 7.8342$. The negative β_0 reflects a low baseline failure probability for normal transformers (low DGAF), while the positive β_1 indicates that higher DGAF values exponentially increase *P(Fail)*, aligning with fault severity trends (e.g., PD, T3). This probabilistic approach empowers utilities to prioritize maintenance effectively, reducing outage risks and operational costs.

3.4. *Application procedure*

The proposed method for evaluating transformer health follows a structured application procedure. Initially, gas concentration inputs are obtained by measuring seven key gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, and CO_2) via gas chromatography, reported in parts per million (ppm). Next, scores S_i are assigned to each gas based on predefined concentration ranges outlined. These scores are then used to compute the DGAF by applying optimized weighting factors W_i specified in Table 1, resulting in a single health index. The transformer's condition is subsequently classified into one of five categories (A to E) according to Table 2, reflecting its health status from "Good" to "Very Poor".

Following this, the failure probability $P(Fail)$ is estimated using a logistic regression model, providing a quantitative basis for risk assessment. Based on the DGAF value, appropriate maintenance decisions are

recommended. Specifically, inspection activities are advised for DGAF ≥ 1.5, while immediate maintenance actions are required for DGAF ≥ 2.5.

4. Results

This section presents the outcomes of applying the DGAF to assess the health of oil-immersed power transformers using a dataset of 871 DGA samples from 240 transformers. The results compare the performance of initial and optimized DGAF parameters, evaluate fault detection accuracy, analyze failure probability predictions, and explore practical implications. Limitations and future research directions are also discussed to contextualize the findings.

4.1. *DGAF performance*

The performance of the DGAF method was evaluated by comparing its initial and optimized parameters across various fault types. The initial and optimized DGAF results for 871 transformer samples are compared in Table 3. Initially, 84 samples were "Acceptable", 753 "Good", 7 "Need caution", 10 "Poor", and 17 "Very poor". 824 Normal samples included 75 "Acceptable" and 749 "Good" with 824 true negatives (TN) and 0 false positives (FP). Faulted samples showed: PD (6) with 3 "Good", 1 "Need caution", 2 "Very poor" (3 true positives (TP), 3 false negatives (FN)); T3 (24) with 4 "Acceptable", 4 "Good", 3 "Need caution", 6 "Poor", 7 "Very poor" (16 TP and 8 FN). The total initial number of TP was 34, and the number of FN was 13.

Table 3. Optimized DGAF results by fault type.

Fault Type	Optimized DGAF										Initial DGAF			
	Acceptable	Good	Need caution	Poor	Very poor	Total	TP	FN	TN	FP	TP	FN	TN	FP
D1	1		1			2	1	1			1	1		
D2			1	2	1	4	4				4			
DT			1	1	2	4	4				3	1		
N	102	706	16			824			808	16			824	0
PD			4	1	1	6	6				3	3		
T2			1	3	3	7	7				7			
T3		4	7	8	5	24	20	4			16	8		
Total	**103**	**710**	**31**	**15**	**12**	**871**	**42**	**5**	**808**	**16**	**34**	**13**	**824**	**0**

After optimization, classifications shifted to 103 "Acceptable", 710 "Good", 31 "Need caution", 15 "Poor", and 12 "Very poor". Normal samples had 102 "Acceptable", 706 "Good", and 16 "Need caution" (808 TN and 16 FP). PD improved to 4 "Need caution", 1 "Poor", 1 "Very poor" (6 TP and 0 FN); T3 had 4 "Good", 7 "Need caution", 8 "Poor", 5 "Very poor" (20 TP and 4 FN). Total optimized TP rose to 42, FN dropped to 5, but FP increased to 16.

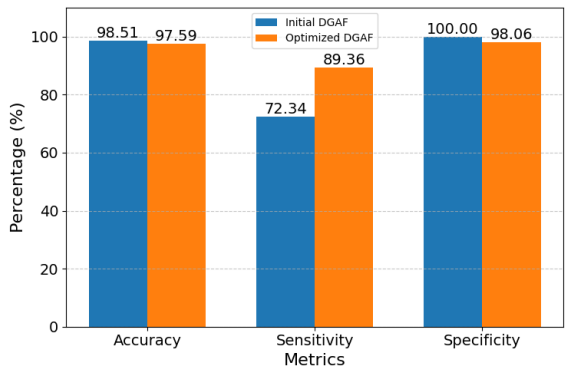


Figure 2. Performance comparison between initial dgaf and optimized DGAF methods.

Figure 2 compares the performance metrics of the initial and optimized DGAF methods. The initial DGAF achieved an accuracy of 98.51%, a sensitivity of 72.34%, and a specificity of 100.00%. In contrast, the optimized DGAF had a slightly lower accuracy of 97.59%, a higher sensitivity of 89.36%, and a reduced specificity of 98.06%. The decrease in accuracy and specificity indicates that the optimized method flagged more N samples as needing caution (16 FP), but the significant increase in sensitivity highlights its improved ability to detect latent faults, particularly in samples initially classified as N.

Table 4 details the 16 Normal samples flagged as "Need caution" by the optimized DGAF, revealing elevated gas levels that suggest latent issues. For instance, one sample showed C_2H_4 at 36.131 ppm and CH_4 at 18.666 ppm, levels indicative of potential thermal faults. Validation using traditional methods, such as the Duval Triangle and Pentagon, confirmed T2 fault tendencies in all 16 samples, aligning with the DGAF's assessment. The optimized DGAF's ability to identify these 16 potential issues within 824 Normal samples underscores its enhanced reliability for maintenance decisions. By detecting latent faults early, it enables proactive interventions, reducing outage risks and associated costs.

Table 4. Details of 16 "Need caution" cases.

H2	CH4	C2H6	C2H4	C2H2	CO	CO2	Actual Fault	DGAF	Duval Triangle 1	Pentagon 1
5.744	4.535	1.416	8.099	0	446.836	7841.675	N	1.590909	N	T2
0	2.619	1.192	1.562	0	406.105	7623.438	N	1.590909	N	T2
7.92	7.744	5.026	28.465	0	614.962	12110.64	N	1.590909	N	T2
…	…	…	…	…	…	…	…	…	…	…
10.321	5.217	2.524	3.825	0	634.073	7042.789	N	1.590909	N	T2
12.37	10.148	4.275	14.241	0	587.658	7488.928	N	1.590909	N	T2
7.777	15.123	5.787	36.131	0	441.52	8053.905	N	1.590909	N	T2

The optimized DGAF method demonstrates a notable improvement in sensitivity (89.36% vs. 72.34%), allowing it to better identify incipient faults in transformers initially diagnosed as Normal by traditional methods. However, this comes at the cost of slightly lower accuracy (97.59% vs. 98.51%) and specificity (98.06% vs. 100.00%), as it introduces some false positives. The ability to flag 16 out of 824 Normal samples for further inspection, as detailed in Table 4, highlights the optimized DGAF's potential to enhance transformer health monitoring and support proactive maintenance strategies. By detecting latent faults early, it enables proactive interventions, reducing outage risks and associated costs.

4.2. *Failure probability analysis*

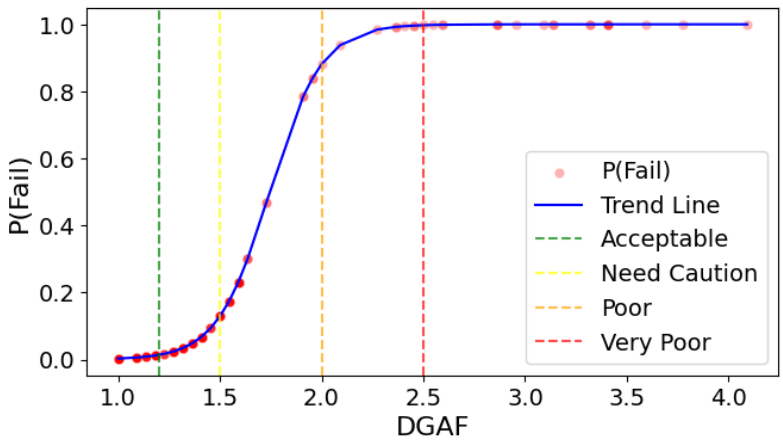


Figure 3. *P(Fail)* probability distribution chart according to DGAF on 871 samples.

The logistic regression model for the optimized DGAF, as depicted in Figure 3, effectively estimates the failure probability *P(Fail)* based on 871 samples, plotting the probability curve against DGAF values alongside data points for normal and faulted samples. Normal samples exhibited a low *P(Fail)* of 0.0029 at DGAF = 1.07, indicating negligible risk. In contrast, faulted samples showed significantly higher probabilities: a PD sample with DGAF = 3.11 had *P(Fail)* = 0.9987, reflecting severe risk. A threshold analysis revealed that a *P(Fail)* cutoff of 0.5 (corresponding to DGAF ≈ 1.8) correctly separated 95% of faulted samples (44/47) from normal ones, supporting robust classification. Table 5 details *P(Fail)* for representative samples. This method enhances maintenance prioritization by quantifying risk, enabling proactive interventions to mitigate transformer failures.

5. Conclusion

This study introduces the DGAF, a novel index for transformer health assessment, leveraging 871 DGA samples from 240 oil-immersed transformers. By synthesizing seven gas concentrations into a single metric, DGAF simplifies diagnostics while enhancing early fault detection. The optimized DGAF achieves 89.36% sensitivity, identifying 16 latent T2 thermal faults among 824 normal samples–outperforming conventional methods like the Duval Triangle. Initial accuracy (98.51%) and specificity (100%) slightly declined post-

Table 5. Example *P(Fail)* values for selected samples.

Fault Type	DGAF (Optimized)	*P(Fail)*	Maintenance Action Recommended
N	1.07	0.0029	Routine monitoring
PD	3.11	0.9987	Immediate inspection
D1	1.95	0.8385	Urgent inspection
D2	2.56	0.9946	Priority inspection
DT	3.59	0.9999	Immediate inspection
T2	3.41	0.9999	Immediate inspection
T3	2.71	0.9973	Priority inspection

optimization (97.59%, 98.06%) but balanced improved fault detection with minimal false positives. A logistic regression model translates DGAF values into failure probabilities, distinguishing low-risk (e.g., 0.0029 probability at DGAF 1.07) from critical cases (0.9946 at DGAF 2.56) to prioritize maintenance. Designed for field use, DGAF supports proactive, risk-based strategies to reduce outages. Future integration with real-time sensors and machine learning could automate monitoring, addressing current limitations in dataset diversity and weighting methods. DGAF advances transformer diagnostics, offering a practical, scalable tool for grid resilience.

References

Jahromi A, Piercy R, Cress S, Service J, Fan W (2009) An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine* **25**, 20-34.

Wattakapaiboon W, Pattanadech N (2016) The new developed health index for transformer condition assessment. In *Proc. 2016 International Conference on Condition Monitoring and Diagnosis*, pp. 32–35.

Li S, Li X, Cui Y, Li H (2023) Review of transformer health index from the perspective of survivability and condition assessment. *Electronics* **12**, 2407.

Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In *Proc. NESUG: Health Care and Life Sciences*, Baltimore, Maryland, **19**:67.

Prasojo RA, Gumilang H, Suwarno, Maulidevi NU, Soedjarno BA (2020) A fuzzy logic model for power transformer faults' severity determination based on gas level, gas rate, and dissolved gas analysis interpretation. *Energies* **13**, 1009.

Haema J, Phadungthin R (2013) Development of condition evaluation for power transformer maintenance. In *Proc. Fourth International Conference on Power Engineering, Energy and Electrical Drives (POWERENG)*, Istanbul, Turkey, 13–17 May 2013, pp. 620–623.

Murugan R, Ramasamy R (2019) Understanding the power transformer component failures for health index-based maintenance planning in electric utilities. *Engineering Failure Analysis* **96**, 274–288.

Brandtzæg G (2015) Health indexing of Norwegian power transformers. Master's thesis, Norwegian University of Science and Technology (NTNU).

Tamma WR, Prasojo RA, Suwarno (2021) High voltage power transformer condition assessment considering the health index value and its decreasing rate. *High Voltage* **6**(2), 314–327.

Guo H, Guo L (2022) Health index for power transformer condition assessment based on operation history and test data. *Energy Reports* **8**, 9038–9045.

---------- REVIEW 1 ----------
SUBMISSION: 80
TITLE: Evaluation of the health state of oil-immersed transformer based on dissolved gas analysis: An empirical study
AUTHORS: Huy Vu Tran, Kim Anh Nguyen, Duc Hanh Dinh, Dinh Duong Le, Huu-Truong Le and Phuc Do

---------- Overall evaluation ----------
SCORE: 3 (strong accept)
----- TEXT:
An interesting contribution on an interesting topic that aligns well with the conference themes

---------- REVIEW 2 ----------
SUBMISSION: 80
TITLE: Evaluation of the health state of oil-immersed transformer based on dissolved gas analysis: An empirical study
AUTHORS: Huy Vu Tran, Kim Anh Nguyen, Duc Hanh Dinh, Dinh Duong Le, Huu-Truong Le and Phuc Do

---------- Overall evaluation ----------
SCORE: 2 (accept)
----- TEXT:
The paper is well written, and the novelty is explained well.
It would be good if you comment on the difference between putting all scores together in one health index, and analysing the different scores separately?
Could the latter approach be more appropriate, as sometimes the collective indexes can mask severe problems with some scores (which are very low) if other scores are very high? Please comment. And justify why you went for a collective index approach.

Dear Conference Organizing Committee MIMAR2025 and Reviewers,

#REVIEW 1

We sincerely thank Reviewer 1 for their positive evaluation and encouraging remarks. We truly appreciate your recognition of our contribution and are pleased that our work aligns well with the conference themes. Your support is a great motivation for us to continue advancing research in this important area.

#REVIEW 2

R2-1: It would be good if you comment on the difference between putting all scores together in one health index, and analysing the different scores separately?

<u>Answer:</u>

We sincerely thank the Reviewer for their valuable and insightful comments, which we greatly appreciate. The motivation for developing a health index that integrates all relevant information related to the degradation of liquid insulation in power transformers and incipient faults based on DGA is grounded in the following rationale: Power transformer faults typically arise from complex interactions among multiple dissolved gases, rather than from the concentration of any individual gas. For instance, thermal faults (T2, T3) may involve elevated levels of CH4, C2H4, and CO, while arcing faults are strongly associated with C2H2. This multivariate nature of fault manifestation is the basis for many established diagnostic methods such as gas ratio techniques (e.g., IEC 60599, Rogers, Dornenburg) and geometric approaches (e.g., Duval's triangle and pentagon). These methods attempt to capture, as comprehensively as possible, the relationships among gas concentrations in order to identify specific fault types related to partial discharge, arcing, or thermal faults. However, these traditional methods often suffer from limited accuracy due to overlapping and inherent nonlinearities, which cannot be effectively represented by simple ratio-based or geometric rules. As a result,

various alternative approaches, particularly those based on AI, have been developed in recent years. Most of these efforts focus on improving fault classification accuracy once a fault has already occurred.

Nevertheless, a key challenge in the context of Condition-Based Maintenance (CBM) and Predictive Maintenance (PdM) strategies is the need to estimate the health state of a power transformer before faults visibly occur. Approaches that rely on individual gas concentrations to propose an optimal maintenance decision-making model pose significant challenges, and in some cases, may even prove infeasible. For this reason, in this study, we intentionally adopted a collective health index by putting all scores together into one integrated indicator, referred to as the DGAF. We believe that this unified index provides an effective basis for evaluating power transformer health in its early stages and can support more proactive and optimized maintenance strategies under CBM and PdM frameworks.

R2-2: Could the latter approach be more appropriate, as sometimes the collective indexes can mask severe problems with some scores (which are very low) if other scores are very high? Please comment. And justify why you went for a collective index approach.

Answer:

We acknowledge the Reviewers' insightful comment regarding the potential risk that a collective index may obscure severe issues when individual scores are critically low but are compensated by higher values from other components. This concern is valid in general health index design, and we have carefully considered it in the development of our Dissolved Gas Analysis Factor (DGAF).

To address this, our DGAF formulation is constructed such that it not only aggregates individual gas scores but also reflects the impact of the most critical gas indicators. In particular, the index is sensitive to very low individual scores (Si), which significantly reduce the overall DGAF value, thereby triggering early warnings even if other gases appear normal.

To illustrate the intended behavior of the DGAF in good transformer conditions, we clarify that when the DGA gas concentrations are low—specifically, H_2 < 100 ppm, CH_4 < 75 ppm, C_2H_6 < 65 ppm, C_2H_4 < 50 ppm, C_2H_2 < 3 ppm, CO < 350 ppm, and CO_2 < 250 ppm—all individual scores Si are equal to 1. Consequently, the DGAF equals 1, which clearly indicates that the transformer is in a good operating condition with no signs of degradation or fault.

We chose a collective index approach because it provides a holistic and continuous representation of transformer health, enabling early detection of incipient faults that may not yet be classifiable by traditional methods. Moreover, the DGAF facilitates straightforward health tracking over time and can be directly incorporated into predictive maintenance frameworks. Nonetheless, to mitigate the masking effect mentioned, the DGAF can be complemented by visualizing the individual scores in parallel, allowing maintenance personnel to identify which gas or gases are deteriorating even if the overall index has not yet crossed the warning threshold.

Please let us know if further revisions are needed. Thank you for your time and consideration.

Best regards,

Kim Anh Nguyen
On behalf of the authors,
The University of Danang – University of Science and Technology,
54 Nguyen Luong Bang, Danang 550000, Vietnam
Tel: +84 (0)962 369 525
Email: nkanh@dut.udn.vn

Innovative Modeling of Age Replacement Policy for Infrastructure Renewal Decisions

Y. Debbagh*, T.Guillon**, J-Y.Dauxois***

* RTE, 7C place du dôme, *immeuble Window, Puteaux France* + yasmine.debbagh@rte-france.com

** *RTE, 7C place du dôme, immeuble Window, Puteaux France* + thomas.guillon@rte-france.com

*** *Institut de Mathématiques de Toulouse, UMR5219, Université de Toulouse, INSA, F-31077 Toulouse, France* + jean-yves.dauxois@insa-toulouse.fr

Abstract. Transmission system operators (TSOs) must balance the renewal of aging infrastructure with integrating renewable energy sources. Traditional age replacement models assume identical replacements and often overlook financial factors like discounting. This work introduces a model based on the delayed renewal reward process, showing that the optimal replacement age of the first asset depends on both existing and replacement asset characteristics. It also highlights that such an optimal age exists only if discounting is considered. Moving beyond traditional assumptions provides TSOs with more realistic, cost-effective asset management strategies in a changing energy landscape.

Keywords: age replacement policy, preventive maintenance, delayed renewal reward process, optimization models, expected equivalent annual cost, discounting, infinite horizon, infrastructure renewal with different asset, investment decision-making, asset management.

Europe's electricity transmission networks are reaching a critical turning point. Primarily built after World War II, these infrastructures are significantly aging, while the ongoing energy transition imposes ever-increasing demands for modernization and adaptation. Transmission System Operators (TSOs) must, therefore, balance the renewal of aging assets with the expansion of the network, particularly to integrate renewable energy sources. In this context, optimizing investment strategies for asset renewal is becoming crucial, especially under growing constraints related to financial resources, workforce availability, equipment supply and planned outages.

A widely adopted strategy for asset renewal is the age replacement policy (cf. Barlow & Hunter, 1960, Barlow & Proschan, 1996, Nakagawa 2005, Mazzuchi & al., 2007). Under this approach, an asset is preventively replaced at a predetermined age a, incurring a cost c_p, or immediately upon failure at a higher cost c_f, whichever occurs first. This strategy relies on renewal reward processes to compute the expected cost per unit of time over an infinite horizon (cf. Nakagawa, 2005, Rausaud & Hoyland, 2003, Resnick, 1992, Nakagawa, 2011). The optimal replacement age a^* minimizes this long-term cost. To refine decision-making and incorporate the time value of money, some extensions introduce continuous discounting functions (cf. Fox, 1966, Ran & Rosenlund, 1976, Van der Weide & al., 2008).

However, the classical age replacement model assumes like-for-like replacement, an assumption that may not hold in today's evolving technological landscape. Environmental standards, sustainability goals, and technological obsolescence often require replacing assets with non-identical alternatives. This reality calls for more flexible models that can accommodate changes in both cost structures and lifetime distribution. While previous works (cf. Schrijner & Coolen, 2006, Schrijner & Coolen 2007) have proposed strategies based on a finite number of cycles, they may not adequately reflect the long-term planning horizons of infrastructure managers aiming to minimize asymptotic costs.

Our paper introduces an extended age replacement framework based on a delayed renewal reward process that allows the initial asset to be replaced by one with different characteristics. The objective is to identify the optimal pair of replacement ages (a_1^*, a^*) that minimizes the expected equivalent annual cost. The cost function integrates exponential discounting with rate δ, ensuring temporal consistency in decision-making, meaning that preferences remain stable over time.

One of our main results is that the optimal replacement age a^* for the second asset remains unchanged compared to the classical model as if it had been deployed from the beginning. In contrast, the optimal age a_1^* for the initial asset is adjusted to account for both its own characteristics and those of its successor.

We show that considering the initial and replacement assets heterogeneity can lead to substantial cost reductions, exceeding *30%* in some scenarios, particularly when the first asset has a significantly longer mean

lifetime than the second. Furthermore, we highlight the decisive impact of discounting on the existence and value of a_1^*. In the specific case where the initial asset follows an exponential lifetime distribution, the optimal strategy becomes binary: either replace immediately or operate until failure.

Overall, this enhanced model provides TSOs with a more realistic and economically sound tool for optimizing infrastructure investments. It supports long-term grid sustainability while improving cost-effectiveness by capturing the evolution of technological options and integrating financial principles such as discounting.

References

Barlow, Richard and Hunter, Larry. (1960). Optimum preventive maintenance policies. *Operations research*, 90-100.

Barlow, Richard E and Proschan, Frank (1996). *Mathematical theory of reliability.* SIAM.

Coolen-Schrijner, P and Coolen, FPA. (2006). On optimality criteria for age replacement. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability* 220, no. 1 : 21--29.

Coolen-Schrijner, Pauline and Coolen, Frank PA. (2007). Nonparametric adaptive age replacement with a one-cycle criterion. *Reliability Engineering & System Safety* 92, no. 1 : 74-84.

Fox, Bennett. (1966). Age replacement with discounting. *Operations Research* 14, no. 3 : 533-537.

Mazzuchi, Thomas A and Van Noortwijk, Jan M and Kallen, Maarten-Jan. (2007). Maintenance optimization. *Encyclopedia of Statistics in Quality and Reliability* : 1000-1008.

Nakagawa, Toshio. (2005). Age replacement. *Maintenance theory of Reliability*, 2005: 69-94.

Nakagawa, Toshio. (2011). *Stochastic processes: With applications to reliability theory.* Springer Science & Business Media, 2011.

Ran, Arne and Rosenlund, Stig I. (1976) Age replacement with discounting for a continuous maintenance cost model. *Technometrics* 18, no. 4 : 459-465.

Rausand, Marvin and Hoyland, Arnljot. (2003) *System reliability theory: models, statistical methods, and applications.* John Wiley & Sons.

Resnick, Sidney I. (1992) *Adventures in stochastic processes.* Springer Science & Business Media.

Van der Weide, JAM and Van Noortwijk, JM and others. (2008) Renewal theory with exponential and hyperbolic discounting. *Probability in the Engineering and Informational Sciences* 22, no. 1 : 53-74.

Sewer pipe inspection, maintenance, and replacement optimization under incomplete information

Ragnar Eggertsson*, Rob Basten*, Geert-Jan van Houtum* and Hajo Molegraaf**

**Department of Industrial Engineering and Innovation Sciences, Eindhoven University of Technology, PO Box 513, 5600 MB Eindhoven, The Netherlands, {r.h.eggertsson, r.j.i.basten, g.j.v.houtum}@tue.nl*

***Rolsch Assetmanagement, Hengelosestraat 581, 7521AG Enschede, The Netherlands, hajo.molegraaf@rolsch.nl*

Abstract. We formulate the inspection, maintenance, and replacement problem for sewer pipes and model this problem as a partially observable Markov decision process. We perform a case study based on real-world data and parameters established together with an expert. Using an open-source solver, we obtain near-optimal policies. For a location with low and high failure costs, the policies inspired by practice perform at least 34% and 251%, respectively, worse than the near-optimal policy. Our results imply that large cost savings are possible by making the planning dependent on the failure costs of a sewer pipe.

Keywords: Maintenance optimization, Condition-based maintenance, Opportunistic maintenance, Inspections, Partially observable Markov decision process

The Netherlands has roughly 126,000 kilometers of sewer pipes, together valued at around 100 billion euros. Sewer pipes are used to dispose of waste and excess water. A sewer pipe failure can lead to environmental contamination and sinkholes. The latter pose a large risk to traffic at night as holes in the road can go unnoticed by cars until it is too late (see van Nuland, 2023). To avoid failures, sewer pipes can be maintained or replaced preventively. A challenge when making maintenance and replacement decisions is that condition information can only be obtained via costly inspections. To avoid inspecting too often, decisions are often based on condition data obtained via previous inspections. As this data does not accurately reflect the current system state, the decision is based on *incomplete information*. This leads to the problem of optimizing inspection, maintenance, and replacement decisions for sewer pipes with incomplete information.

All municipalities face the problem of optimizing inspection, maintenance, and replacement decisions for underground assets. Our partner company, Rolsch Assetmanagement, aids municipalities by providing inspection, maintenance, and replacement optimization software. The company addresses the following questions: “How to plan inspections smartly?” and “Should replacement opportunities be utilized, and if so, when?” Replacement *opportunities* arise when another underground asset or an overlaying road is replaced. By combining activities, costs, such as those associated with road removal and bringing expensive tooling to the location, only need to be paid once, and can be shared among multiple parties. This leads to lower replacement costs for sewer pipes. In addition to both questions, Rolsch Assetmanagement is interested in the features of a well-performing inspection, maintenance, and replacement policy for scenarios with different failure costs. The failure costs are e.g. very high for a sewer pipe under a busy road intersection and low for a sewer pipe in an easily accessible field.

Maintenance and replacement optimization relies on recorded damages. The international standards EN13508 and EN13508-2 advise on how to classify sewer pipe damages during inspections. These standards distinguish *multiple failure modes*. Each failure mode is assigned several severity levels corresponding to damage observations. These severity levels are translated by municipalities to a damage class from 1 to 5, where 1 corresponds to no observed damage, and 5 means the worst score for that damage type. For most failure modes, in particular those we consider, pipes can still operate if they are in damage class 5; however, the probability of failure is high in this damage class. During an inspection, all damages for all failure modes are recorded. Maintenance and replacement interventions are performed to keep sewer pipes in good condition and thus to prevent failures. We formulate the company’s inspection, maintenance, and replacement problem as a partially observable Markov decision process (POMDP).

Based on the practice for sewer pipes, we formulate an advanced model for the inspection, maintenance, and replacement decisions for a single sewer pipe. The model includes inspection, maintenance, and replacement decisions, multiple competing failure modes, the possibility of maintenance per failure mode, and opportunistic to execute replacements against a low setup cost. These opportunities arise in practice when a piece of a road is opened, for example. The closest related papers are by Byon and Ding (2010) and Byon (2013), which both consider multiple competing failure modes and include low cost replacement opportunities but do not allow separate maintenance actions per failure mode. Karabağ et al. (2020) and Eggertsson et al. (2025) study maintenance models with multiple competing failure modes and separate maintenance actions per failure mode, but their models have no inspections and low-cost replacement opportunities.

Our main contributions are as follows. First, we study a new, advanced maintenance model. Second, we perform a case study on sewer pipes in the Dutch city of Breda. The degradation processes are derived from real-world data by Jimenez-Roa et al. (2022), and the remaining parameters were established with our partner company. Hence, this work adds to the limited literature on maintenance optimization based on real-world data. Third, we show that a near-optimal policy outperforms two policies used in practice for the described setting. The policies used in practice perform at least 34% worse, and in the most severe case 251% worse. Fourth, we draw the following managerial insights from the case study:

- A sewer pipe's inspection frequency should depend on the replacement and failure costs of the sewer pipe, i.e., as the replacement and failure costs increase, the inspection intervals should become shorter.
- Inspection planning should be condition-based. This means that, in some scenarios, inspections should be performed more often if the condition worsens. Otherwise, a periodic inspection policy currently used in practice is expected to perform well.
- Replacement planning should be condition-based. Replacements should not be performed while the condition is still good. The latter occurs under the age-based replacement policies as used currently in practice (Dekkers, 2019).
- If replacement and failure costs are high, then replacement opportunities are very important for cost savings and should be utilized more often than for low replacement and failure costs.
- Maintenance is preferable to replacements when possible.

The single-sewer pipe model as developed in this paper, may constitute a building block for the optimization of inspections, maintenance, and replacements of *sewer pipe lines*. A direct modeling of that problem would lead to problem sizes that are computationally intractable. Instead, one can consider to decompose the sewer pipe line problem into single-sewer pipe problems where maintenance/replacement actions for one pipe form low setup cost opportunities for other sewer pipes. This is similar to the decomposition used by e.g. Kivanç et al. (2023) for a multi-component system.

References

- Byon E (2013) Wind turbine operations and maintenance: A tractable approximation of dynamic decision making. *IIE Transactions* **45**, 1188–1201.
- Byon E, Ding Y (2010) Season-dependent condition-based maintenance for a wind turbine using a partially observed Markov decision process. *IEEE Transactions on Power Systems* **25**, 1823–1834.
- Dekkers M (2019) Verbreed Gemeentelijk Rioleringsplan, Broks-Messelaar Consultancy.
- Eggertsson R, Eruguz AS, Basten RJI, Maillart L (2025) Maintenance optimization for multi-component systems with a single sensor. *European Journal of Operational Research* **320**, 559-569.
- Jimenez-Roa LA, Heskes T, Tinga T, Molegraaf HJA, Stoelinga M (2022) Deterioration modeling of sewer pipes via discrete-time Markov chains: A large-scale case study in the Netherlands. In: *32nd European Safety and Reliability Conference, ESREL 2022: Understanding and Managing Risk and Reliability for a Sustainable Future*, pages 1299–1306.
- Karabağ O, Eruguz AS, Basten RJI (2020) Integrated optimization of maintenance interventions and spare part selection for a partially observable multi-component system. *Reliability Engineering & System Safety* **200**, article 106955.
- Kivanç I, Fecarotti C, Raassens N, van Houtum GJ (2023) A scalable multi-objective maintenance optimization model for systems with multiple heterogeneous components and a finite lifespan. *European Journal of Operational Research* **315**, 567-579.
- van Nuland, M (2023) Sewer science. https://www.nwo.nl/en/cases/sewer-science. Accessed: 2023-09-01.

Sustainable Condition-based Maintenance: Balancing Circularity and Spare Parts Availability Within Finite Maintenance Horizons

Ghofrane Boujemaa* Christophe Bérenguer** and Elodie Robert***

** Volvo Group Trucks technology, 69802 Saint Priest, France and GIPSA-LAB 38000 Grenoble: ghofrane.boujemaa@volvo.com*
*** Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, 38000 Grenoble, France : Christophe.Berenguer@grenoble-inp.fr*
**** Volvo Group Trucks technology, 69802 Saint Priest, France : Elodie.Robert@volvo.com*

Abstract. This research work proposes a condition-based maintenance model integrating the circularity constraint by offering the possibility of replacement with remanufactured parts while considering the availability of parts in the decision-making process. Maintenance optimization under these constraints is carried out over a finite time horizon corresponding to the duration of the maintenance contract in the context of industrial vehicles.

Keywords: Dynamic programming, remanufacturing, Spare parts availability, condition-based maintenance, Sustainable maintenance

The optimization of condition-based maintenance models is a crucial issue for the industrial vehicle sector [1], in particular when taking into account operational constraints such as resource management, circularity and customer expectations. This research focuses on the modeling and optimization of maintenance policies, an area that has become essential for automakers, who have become aware of the importance of after-sales services in their revenues. Customers' expectations have also evolved [7,9,11]: they are no longer content with simple breakdown alerts; they are now looking for comprehensive, optimized maintenance solutions, aimed at limiting workshop visits, avoiding breakdowns and unscheduled downtime, while ensuring that their missions are fulfilled.
The aim of this research work is to develop an approach for optimizing maintenance decisions on a finite horizon, using a condition-based maintenance model that explicitly integrates operational constraints. These constraints include circularity [12,13], which allows for the replacement of parts with remanufactured or reconditioned parts in addition to new parts. Furthermore, availability constraints are also considered, referring to the stock availability of parts, whether they are new or remanufactured. Based on this model, the aim is to optimize maintenance decisions while maximizing operational efficiency and minimizing associated costs, all while meeting customer expectations regarding reliability.
Maintenance optimization is carried out over a finite horizon [8,10], corresponding to the duration of the maintenance contract with the customer. First, we define what is a remanufactured component and analyze its impact on the level of system degradation [2,6]. We then integrate into the same maintenance model the consideration of parts availability in the replacement decision-making process. The proposed maintenance policy takes account of available stocks and provides a maintenance schedule based on this availability. In addition, we propose to consider the recovery of the vehicle after the term of the contract. Optimization is therefore carried out on the total cost of maintenance throughout the life of the contract, taking into account whether the vehicle will be recovered after the end of the contract. These optimizations are carried out using simulations, and the results obtained are used to assess the effectiveness of circularity in a maintenance model. We also examine how the addition of the availability constraint allows us to opt for a more complete model, offering additional gains and better-planned workshop interventions.

In conclusion, this study shows how condition-based maintenance can be optimized when considering both circularity and spare parts availability within a finite maintenance horizon. Within dynamic programming [5], different scenarios are considered to evaluate the performance of the system over time and optimize the maintenance policy, and the best one is retained. This approach ensures that maintenance decisions are cost-effective, by balancing the compromise between operational efficiency and resource availability. Ultimately, this method provides a robust framework for sustainable condition-based maintenance that meets both operational needs and sustainability goals [3,4].

References

[1] Bouvard, K. (2010). *Planification de la maintenance basée sur la dégradation de composants. Application à un véhicule industriel.* PhD Thesis, Univ. Grenoble Alpes.

[2] Boujemaa, G., Bérenguer, C., & Robert, E. (2024). Remanufacturing in vehicle maintenance: Logistic challenge and constraint or opportunity for increasing sustainability. *In Proceedings of the 34th European Safety and Reliability Conference (ESREL).*

[3] Burggräf, P., Wagner, J., Steinberg, F., Heinbach, B., Wigger, M., & Saßmannshausen, T. (2022). Life cycle assessment for adaptive remanufacturing: Incorporating ecological considerations into the planning of maintenance activities – A case study in the German heavy machinery industry. *Procedia CIRP*, 105, 320–325.

[4] Burggräf, P., Adlon, T., Dackweiler, J., Bröhl, F., & Fölling, C. (2022). Adaptive remanufacturing – Decision model for the intelligent maintenance of production resources. *Procedia CIRP*, 105, 219–224.

[5] Kivanc¸, Ipek, Claudia Fecarotti, Neomie Raassens, and Geert-Jan van ´ Houtum (2024), 'A scalable multi-objective maintenance optimization model for systems with multiple heterogeneous components and a finite lifespan,' *European Journal of Operational Research*, 315 (2), 567- 579

[6] Leroy, M., Bérenguer, C., Doyen, L., & Gaudoin, O. (2022). Statistical inference for a Wiener-based degradation model with imperfect maintenance actions under different observation schemes. *Statistical Modelling*, 39, 352–371.

[7] Lesobre, R. (2015). *Modélisation et optimisation de la maintenance et de la surveillance des systèmes multi-composants – Applications à la maintenance et à la conception de véhicules industriels*. PhD Thesis, Univ. Grenoble Alpes.

[8] Liu, B., Pandey, M. D., Wang, X., & Zhao, X. (2021). A finite-horizon condition-based maintenance policy for a two-unit system with dependent degradation processes. *European Journal of Operational Research,* 295(2), 705–717.

[9] Longhitano, P. D. (2024). *Prescriptive maintenance for optimal management of systems subject to degradation – Application to the joint prescription of usage modes and maintenance actions for heavy vehicles.* PhD Thesis, Univ. Grenoble Alpes.

[10] Nakagawa, T., & Mizutani, S. (2009). A summary of maintenance policies for a finite interval. *Reliability Engineering & System Safety*, 94(1), 89–96.

[11] Robert, E. (2019). *Health management of industrial vehicles and fleet maintenance optimization: Taking into account operation constraints and mission planning.* PhD Thesis, Univ. Grenoble Alpes.

[12] Santos, A. C. J., Cavalcante, C. A. V., & Wu, S. (2023). Maintenance policies and models: A bibliometric and literature review of strategies for reuse and remanufacturing. *Reliability Engineering & System Safety,* 231, 108983.

[13] Santos, A. C. de J., Cavalcante, C. A. V., Ren, J., & Wu, S. (2023). *A novel delay time modelling method for incorporating reuse actions in three-state single-component systems. Reliability Engineering & System Safety*, 234, 109129.

Interpretable Maintenance Impact Quantification Using JEPA-KAN with Self-Supervised Contrastive Learning: an Aircraft case study

Weikun DENG*, Khanh T.P. NGUYEN**, Phuc DO***, and Kamal MEDJAHER****

*47 Av. d'Azereix, 65000 Tarbes, UTTOP, **weikun.deng@uttop.fr**

47 Av. d'Azereix, 65000 Tarbes, UTTOP, **thi-phuong-khanh.nguyen@uttop.fr

***Campus sciences BP 70239, 54506 Vandoeuvre Cedex, University of Lorraine, **phuc.do@univ-lorraine.fr**

****47 Av. d'Azereix, 65000 Tarbes, UTTOP, **kamal.medjaher@uttop.fr**

Abstract.

Optimizing maintenance strategies is critical for ensuring the safety, reliability, and cost-effectiveness of aircraft operations. A fundamental challenge in predictive maintenance lies in quantifying the impact of maintenance interventions. Existing approaches for evaluating this impact generally fall into two categories: model-based methods, which rely on predefined mathematical formulations but often fail to capture real-world complexities, and data-driven methods, which leverage real-time monitoring data but function as black-box models without explicitly characterizing maintenance effects. To address these challenges, this paper introduces a novel self-supervised learning approach for interpretable maintenance impact quantification. In particular, by employing a Siamese architecture with interpretable Kolmogorov-Arnold Networks (KANs), the proposed framework derives an analytical expression that quantitatively describes maintenance effects. This allows for a transparent and physically meaningful assessment of maintenance efficiency, bridging the gap between data-driven adaptability and model-based interpretability. Experimental evaluations on real-world aircraft maintenance datasets demonstrate that the proposed approach outperforms conventional methods in predicting post-maintenance system health while providing clear, explainable rationale for its predictions, ultimately supporting more informed decision-making and improving aviation safety and efficiency.

Keywords: General aviation aircraft maintenance, Self-supervised learning, JEPA, Interpretable maintenance impact identification.

1. Introduction

The aviation industry depends on rigorous maintenance strategies to ensure safety and reliability while managing costs. Predictive maintenance has thus become essential Stanton et al. (2023). However, its implementation is hindered by the difficulty of quantifying the true maintenance impact due to the lack of clear cause-and-effect baselines in real-world Lundgren et al. (2018).

To address the above challenge, different approaches are developed for quantification, each tailored to their specific objectives. In structural health contexts, maintenance impact is modeled by integrating sensor data into Health Indicator (HI) models (like physics-based digital twin Jeon et al. (2024)), allowing for the tracking of key performance indicators (KPIs) that reflect structural degradation. These health-sensitive metrics are derived through sensitivity analysis to identify measurable responses—such as strain, deflection, or stress—that are both degradation-sensitive and reliably captured by sensors. On assessment of the capacity to fulfill the mission, the impact of maintenance is quantified using service-level KPIs that combine dynamic asset criticality and asset degradation levels, such as the influence on the passenger flow disruption risk Consilvio et al. (2024). In addition, the analysis includes economic costs and a comprehensive set of assessments from a mixed perspective, similar to the approach in Dinh et al. (2024). The impact of maintenance is quantified through changes in component hazard rates and reliability functions, modeled using disassembly-sensitive indicators that capture both physical degradation and the economic consequences of maintenance decisions.

The above approaches are generally classified into two categories: data-driven and physics-based methods. A key limitation of many data-driven methods is their black-box nature. For example, while deep networks can accurately predict impending failures, they often fail to explain how much life was

restored by a repair or why a maintenance action was effective Serradilla and Diaz (2021). In contrast, physics-based models explicitly link maintenance interventions to measurable improvements in health indices but typically rely on simplifying assumptions and require extensive calibration. Recent literature highlights a growing interest in explainable AI for predictive maintenance Serradilla and Diaz (2021); however, most deep learning (DL) models still lack analytical descriptions of maintenance effects. Maintenance events are often modeled as abrupt resets without estimating the magnitude of the benefit, a limitation further emphasized by research on benchmark datasets such as NASA's C-MAPSS, which often overlook detailed labeling of maintenance events B. Saxena and Eklund (2008).

Therefore, assessing maintenance impact requires models that not only handle complex real-world data but also provide transparent, interpretable reasoning—particularly in the absence of ground truth or historical labels. Such interpretability, though seldom achieved, is critical for trust, insight, and actionable predictions. To address this need, inspired by the research on Kolmogorov-Arnold Networks (KAN) Liu et al. (2024) and Joint Embedding Predictive Architecture (JEPA) Assran et al. (2023), in Section 2, we propose a novel dual-branch Siamese self-supervised framework that provides the solution for quantifying the impact of maintenance operations on health indicators (HI) without the ground truth labels. Section 3 presents a case study on general aviation aircraft to evaluate the framework's performance, while Section 4 concludes with key findings and future research directions.

2. Proposed Methodology

Our framework enables interpretable health assessment by adopting a JEPA-inspired architecture that models system behavior in a semantically meaningful latent space, rather than relying on explicit numerical labels. Prior knowledge of degradation dynamics is embedded as structural constraints (Fig.1). The model jointly processes pre- and post-maintenance time-series sensor data through parallel feature extraction branches, yielding health indicators (HIs) that characterize degradation and recovery patterns (see Section2.1). These branches are optimized via a multi-objective loss that balances structural constraints and supervised targets, enhancing the interpretability (see Section 2.2).

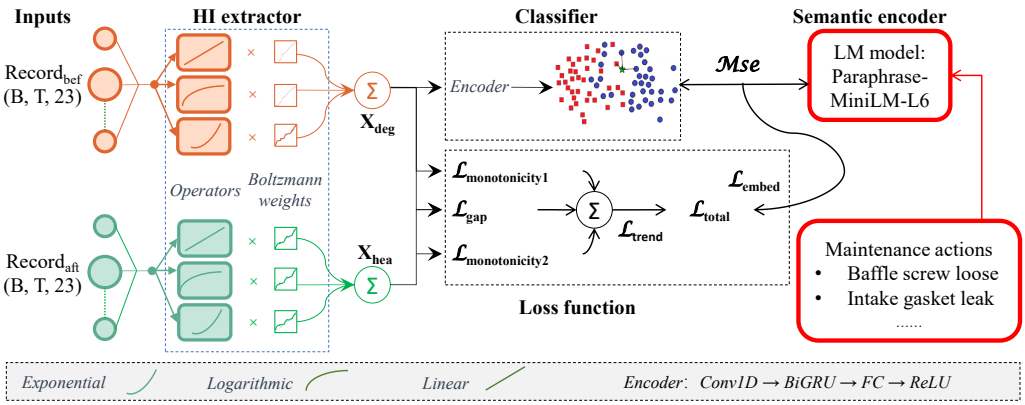


Figure 1: Siamese dual-branch KAN structure.

Specifically, the maintenance impact is quantified by analyzing the shift in HI trajectories before and after intervention, where interpretable operator weights and trend parameters (e.g., slope, scale, growth rate) learned by the KAN modules reveal how degradation patterns transition toward recovery, providing a transparent, analytical description of maintenance effects.

2.1. Siamese-KAN for dual-branch representation learning

The proposed framework (Fig. 1) is an end-to-end architecture for learning interpretable HIs directly from raw multivariate time-series data recorded before ($Record_{\text{deg}}$) and after ($Record_{\text{hel}}$) maintenance. Each input has shape (B, T, C), denoting batch size, sequence length, and number of sensor channels.

HI extractor module first extracts temporal features per channel and estimates sensor-wise importance via a learnable fully connected layer, producing normalized weights that guide feature aggregation.

Table 1: KAN candidate operator formulas, parameter roles, and their respective modeling intents. $\tau = \frac{t}{T-1}$; $\tilde{\tau} \in [1, 2]$; $\varepsilon = 10^{-6}$; $x(t) = [X_{\text{deg}}$ or $X_{\text{hel}}]$.

Operator	**Formula** $y(t)$	**Parameters**	**Design Purpose**
Linear	$y(t) = x(t) + s_{\text{lin}} \cdot (\text{ReLU}(w) \cdot \tau + b)$	s_{lin}: scale amplitude; w: slope; b: offset	Model steady, linear degradation or recovery
Exponential	$y(t) = x(t) + s_{\text{exp}} \cdot (e^{\alpha\tau} - 1)$	s_{exp}: scale; α: growth rate	Capture accelerating trends (e.g., rapid degradation)
Logarithmic	$y(t) = x(t) + s_{\text{log}} \cdot \log(\tilde{\tau} \cdot \alpha_+ + \varepsilon)$	s_{log}: scale; $\alpha_+ = \text{softplus}(\alpha)$	Model early fast change and influence that gradually saturates

HI extractor comprises two parallel branches: the upper (orange) branch models degradation from $Record_{\text{deg}}$, while the lower (green) branch captures recovery from $Record_{\text{hel}}$. Each branch incorporates a set of interpretable operators—**Linear**, **Exponential**, and **Logarithmic** (Table 1)—to characterize diverse temporal trends. Their outputs are combined via a Boltzmann-weighted softmax mechanism, where learnable trend confidences are normalized using a global temperature parameter, enabling adaptive and probabilistic trend modeling.

The degradation representation X_{deg} is passed to a temporal classifier (Conv1D $\rightarrow$ BiGRU $\rightarrow$ FC $\rightarrow$ ReLU) to predict a semantic embedding $\text{Pred}_{\text{semantic}}$, which is aligned with the ground-truth label embedding $\text{Real}_{\text{semantic}}$ via an embedding loss $\text{Loss}_{\text{embed}}$. This promotes semantic coherence across instances sharing maintenance actions. To regularize learning, two additional constraints are applied: a gap loss (Loss_{gap}) to ensure a separation between X_{deg} and X_{hel}, and monotonicity losses ($\text{Loss}_{\text{monotonicity}}$) to enforce consistent degradation and recovery trajectories across time.

2.2. Priori degradation knowledge based learning constraint

To promote interpretable health trajectories, we define a multi-objective loss that integrates structural priors with task supervision. As outlined in Table 2, the unsupervised term $\mathcal{L}_{\text{trend}} = \mathcal{L}_{\text{monotonicity1}} + \mathcal{L}_{\text{monotonicity2}} + \mathcal{L}_{\text{gap}}$ serves as a global regularizer shaping the temporal dynamics. Specifically, monotonicity terms enforce directional progression, the gap constraint ensures separation between degraded and healthy states, and semantic alignment promotes consistency with label embeddings. These constraints jointly guide the model toward physically coherent representations of maintenance impact.

Table 2: Constraints and loss functions summary

Constraint name	**Formula and purpose**
$X_{\textbf{deg}}$ **monotonicity**	$\mathcal{L}_{\text{monotonicity1}} = \frac{\sum \text{ReLU}(-\Delta X_{\text{deg}})}{\sum \text{valid_len}_{\text{deg}}}$. Encourages monotonicity for HI before maintenance.
$X_{\textbf{hel}}$ **monotonicity**	$\mathcal{L}_{\text{monotonicity2}} = \frac{\sum \text{ReLU}(-\Delta X_{\text{hel}})}{\sum \text{valid_len}_{\text{hel}}}$. Encourages monotonicity for HI after maintenance.
Gap constraint	$\mathcal{L}_{\text{gap}} = \text{ReLU}(\text{gap}_{\text{margin}} - \min(X_{\text{deg}}) + \max(X_{\text{hel}}))$. Prevents overlap between degraded and healthy sequences.

Continued on next page

Table 2 – continued from previous page

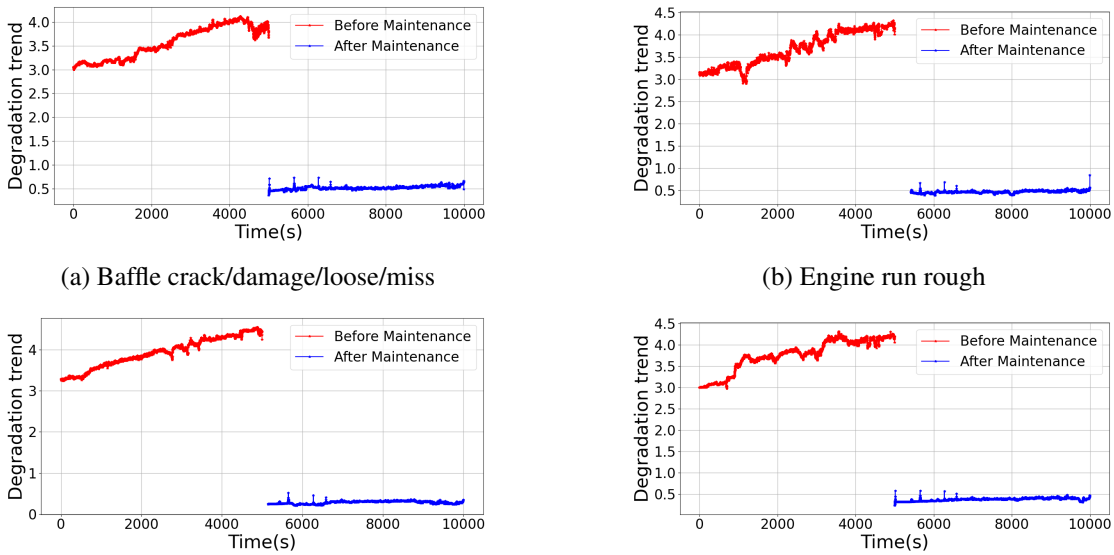
Constraint name	Formula and purpose
Semantic alignment loss	$\mathcal{L}_{\text{embed}} = \frac{1}{N}\sum_{i=1}^{N} \lVert \mathbf{s}_i - \mathbf{e}_i \rVert_2^2$. Matches predicted maintenance types $\mathbf{s}_i$ with true maintenance labels $\mathbf{e}_i$.
Total loss	$\text{loss}_{\text{final}} = w_{\text{unsup}} \cdot \mathcal{L}_{\text{trend}} + w_{\text{cls}} \cdot \mathcal{L}_{\text{embed}}$.

3. Case study

To evaluate the performance of the proposed methodology, we use the NGAFID Aviation Maintenance Dataset introduced by Yang and Desell (2022), which contains high-resolution flight data recorded from 23 sensors installed on 172 aircraft. These sensors monitor various subsystems, including the electrical, fuel, oil, and piston engine modules, as well as environmental parameters. The dataset spans over 31,177 flight hours across 28,935 flights, with each flight labeled as occurring either before or after an unplanned maintenance event. To mitigate class imbalance, we grouped sequences based on their labels and intervention status (i.e., **before** vs. **after** maintenance) and employed balanced sampling to ensure equal representation for each class. Here, the "before" sequences capture the progression of degradation leading up to the intervention, while the "after" sequences reflect system recovery or post-maintenance stabilization. Then, the data is partitioned into training (60%), validation (20%), and test (20%) sets using stratified sampling to preserve label distributions across splits. The detailed layer-wise architecture and hyperparameters in this study are provided in the supplementary reference Deng et al. (2025).

3.1. Qualitative impact of maintenance operations

Fig. 2 illustrates several important insights into the degradation and recovery dynamics before and after maintenance interventions. Most notably, the **sharp contrast between pre- and post-maintenance**



(a) Baffle crack/damage/loose/miss

(b) Engine run rough

(c) Intake gasket leak/damage

(d) Baffle seal loose/damage

Figure 2: Predicted degradation trends for different before-post-maintenance fault types.

trajectories confirms that the proposed framework effectively captures the **impact of maintenance operations** by modeling the transition from degradation to recovery. First, the **post-maintenance health trends** (blue curves) are remarkably consistent across different fault types, suggesting that maintenance actions generally succeed in restoring the system to a stable, low-degradation state regardless of the specific failure. Second, although all **pre-maintenance trajectories** (red curves) exhibit increasing

trends indicative of degradation, they differ in **shape, slope, and fluctuation patterns**, even for faults affecting the same component. For instance, both "baffle crack/damage/loose/miss" and "baffle seal loose/damage" involve baffle-related issues; however, the former shows a smoother, more gradual progression, while the latter exhibits sharper increases and intermittent plateaus. Finally, the degradation pattern for "engine run rough" is characterized by steeper and more volatile behavior compared to other cases, likely reflecting the rapid deterioration associated with combustion-related anomalies. These qualitative distinctions demonstrate the model's ability to capture nuanced failure progression and the effectiveness of maintenance in halting or reversing degradation.

3.2. Quantitative impact of maintenance operations

To derive a transparent and quantitative representation of maintenance impact, we formulate the predicted health trajectory as a weighted combination of candidate trend operators applied to the time-varying sensor signal $x(t)$ after a linear NN layer projection. Each operator captures a distinct degradation behavior: linear growth, exponential rise, and logarithmic saturation. The aggregated channel representation $x(t)$ is computed based on the raw input signal $x_{\text{raw}}(t)$ using the formula:

$$u = \frac{1}{T}\sum_{t=1}^{T} x_{\text{raw}}(t), \quad x(t) = \mathbf{W}\cdot u + \mathbf{b} \tag{1}$$

$$\begin{aligned}\text{Output}(t) = {} & w_{\text{linear}}\left[x(t) + \text{scale}_{\text{linear}}\cdot(at + \text{bias}_{\text{linear}})\right] \\ & + w_{\text{exp}}\left[x(t) + \text{scale}_{\text{exp}}\cdot\left(e^{\alpha_{\text{exp}}t} - 1\right)\right] \\ & + w_{\log}\left[x(t) + \text{scale}_{\log}\cdot\log\left(t\cdot\text{softplus}(\alpha_{\log})\right)\right]\end{aligned} \tag{2}$$

The primary difference before and after maintenance lies in the operator parameters shown in Table **??**. As illustrated in Fig. 2, maintenance shifts the degradation trend from linear to logarithmic. Before maintenance, the linear operator dominates (weight $= 0.9882$, $a = 0.8400$), reflecting steady degradation with large variations caused by the exponential degradation pattens. After maintenance, its influence vanishes (weight $= 0.0050$, $a = 0$), while the logarithmic operator becomes dominant (weight $= 0.9930$), producing flattened trends. The exponential operator remains marginal in post-maintenance. This shift offers a transparent explanation of how maintenance alters system dynamics from active degradation to post-maintenance stabilization.

Table 3: Candidate operator parameters before vs. after maintenance

Candidate	**Before maintenance**					**After maintenance**				
	weights	a	bias	scale	alpha	*weights*	a	bias	scale	alpha
Linear	0.9882	0.8400	1.3585	1.5358	–	0.0050	0.0000	0.2907	0.5016	–
Exp	0.0099	–	–	0.7975	1.8713	0.0019	–	–	0.1296	1.0355
Log	0.0018	–	–	0.7495	1.9035	0.9930	–	–	0.5518	1.6093

4. Conclusion

This paper proposes a self-supervised framework that combines the interpretability of JEPA-KAN with the discriminative power of contrastive learning to quantify maintenance impact in a transparent, data-efficient, and label-free way. Physics-informed constraints (monotonicity, temporal ordering, semantic separability) guide the learning of health indicators (HIs) aligned with realistic degradation and recovery. Without manual labels, the model distinguishes pre/post-maintenance states and assigns interpretable operator weights capturing the magnitude, shape, and trend of interventions. Applied to

real aircraft maintenance data, it delivers strong performance and visual interpretability, revealing clear fault-specific trends and transforming linear degradation into stabilized logarithmic recovery.

Future work will extend the framework to multi-objective maintenance planning with risk prioritization, enhanced fault separability, and integration with attention-based diagnostics and interactive visualization for real-time, explainable decision support in safety-critical domains.

5. Acknowledgements

The authors acknowledge the support of the French Agence Nationale de la Recherche (ANR), under grant ANR-22-CE10-0011-01 (project X-IMS).

References

Assran, M., Duval, Q., Misra, I., Bojanowski, P., Vincent, P., Rabbat, M., LeCun, Y., and Ballas, N. (2023). Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 15619–15629.

B. Saxena, S. Goebel, D. S. and Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. *Proceedings of the Prognostics and Health Management Conference*, pages 31–39.

Consilvio, A., Vignola, G., López Arévalo, P., Gallo, F., Borinato, M., and Crovetto, C. (2024). A data-driven prioritisation framework to mitigate maintenance impact on passengers during metro line operation. *European Transport Research Review*, 16(1):6.

Deng, W., Nguyen, K. T., Do, P., and Medjaher, K. (2025). https://docs.google.com/document/d/139w$_w arx - fjqfbvrggngmhqmigcmh_4 i/edit?usp = drive_l inkouid = 101270498230458171003 rtpof = truesd = true.$

Dinh, D.-H., Do, P., Iung, B., and Nguyen, P.-T.-N. (2024). Reliability modeling and opportunistic maintenance optimization for a multicomponent system with structural dependence. *Reliability Engineering & System Safety*, 241:109708.

Jeon, C.-H., Shim, C.-S., Lee, Y.-H., and Schooling, J. (2024). Prescriptive maintenance of prestressed concrete bridges considering digital twin and key performance indicator. *Engineering Structures*, 302:117383.

Liu, Z., Wang, Y., Vaidya, S., Ruehle, F., Halverson, J., Soljačić, M., Hou, T. Y., and Tegmark, M. (2024). Kan: Kolmogorov-arnold networks. *arXiv preprint arXiv:2404.19756*.

Lundgren, C., Skoogh, A., and Bokrantz, J. (2018). Quantifying the effects of maintenance–a literature review of maintenance models. *Procedia CIRP*, 72:1305–1310.

Serradilla, S. and Diaz, M. (2021). Explainable ai in predictive maintenance: A survey. *IEEE Internet of Things Journal*, 8:55–63.

Stanton, I., Munir, K., Ikram, A., and El-Bakry, M. (2023). Predictive maintenance analytics and implementation for aircraft: Challenges and opportunities. *Systems Engineering*, 26(2):216–237.

Yang, H. and Desell, T. (2022). A large-scale annotated multivariate time series aviation maintenance dataset from the ngafid. *arXiv preprint arXiv:2210.07317*.

Optimal maintenance decision-making for continuously degrading systems with imperfect repairs using Markov decision processes

Q.K. Tran, K.T. Huynh, A. Grall, Y. Langeron

University of Technology of Troyes, 10004 Troyes CEDEX, France, Email: {quang_khai.tran, tuan.huynh, antoine.grall, yves.langeron}@utt.fr

Abstract Repair serves as a cost-effective strategy to extend a system's useful life, yet its efficiency is significantly influenced by stochastic factors such as maintenance engineer expertise and part quality. To optimize maintenance decisions, it is critical to model these uncertainties and evaluate their impact on resulting policies. Motivated by this challenge, we address in this paper two important problems in elaborating the degradation-based maintenance policy for repairable systems: (1) developing a framework to identify optimal maintenance policies for a single-unit system experiencing continuous degradation, periodic inspections, perfect replacements, and imperfect repairs with stochastic repair coefficients; and (2) systematically evaluating how variability in repair efficiency impacts economic performance. For such a maintenance problem, we assume that the system degradation follows an inverse Gaussian process with general shape functions, while the stochastic repair efficiency is modeled using a beta-distributed random variable. Unlike conventional parametric policy approaches, we formulate the problem as a Markov decision process to directly learn state-to-action mappings under the long-run average cost criterion. To overcome computational challenges posed by the mixed discrete-continuous state space, the policy is optimized through dynamic programming combined with function approximation. Regarding the evaluation of the framework's performance, we conduct comparative studies against alternative maintenance policies, such as replacement-only policies, demonstrating the effectiveness of the proposed approach.

1. Introduction

Maintenance plays a crucial role in industrial systems by ensuring efficient operation, enhancing safety, and minimizing downtime. According to Pham and Wang (1996), maintenance activities can be categorized based on their impact on the system degradation state. In particular, perfect maintenance actions, such as corrective or preventive replacement, restore the system to an as-good-as-new (AGAN) state. In contrast, repairs typically return the system to a degradation level between AGAN and as-bad-as-old (ABAO). In the literature, replacement-only policies have been extensively studied (Elwany, Gebraeel, and Maillart 2011; Zhang and Zeng 2017) due to their simplicity and effectiveness, particularly for non-repairable or critical systems where repair is limited by technological and economic constraints. However, their practical application presents notable challenges. Determining an appropriate replacement threshold is essential: replacing the entire system too early leads to excessive costs due to both the high replacement expense and the increased frequency of interventions, while delaying replacement increases the risk of unexpected failures. To address this trade-off, incorporating repair actions provides a more flexible and cost-efficient alternative. As a mitigating intervention, repair reduces the system's degradation at a moderate cost, thereby extending its service life and enhancing the balance between reliability and maintenance expenditure (Cheng and Zhao 2023; Mosayebi Omshi and Grall 2021). Despite its advantages, incorporating repair actions into maintenance policies introduces additional challenges. A key issue is the inherent stochasticity of repair operations, influenced by factors such as maintenance engineer skill levels and the quality of spare parts. Accurately accounting for these stochastic factors is a crucial step in optimizing maintenance policies and understanding their long-term effects.

To develop and optimize repair-replacement policies, one can adopt either a parametric or a non-parametric approach. The parametric approach imposes a structured framework parameterized by decision variables, leveraging Markov renewal theory to determine optimal values (Huynh, Grall, and Berenguer 2019). Alternatively, the maintenance problem can be formulated as a Markov decision process (MDP) or its extensions, such as semi-MDPs or partially observed MDPs, allowing for the derivation of an optimal, non-parametric state-to-action mapping (Puterman 2005). While the parametric approach has been widely applied to continuous degra-

dation states, e.g., see Huynh (2021), the non-parametric counterpart is typically restricted to discrete states or discretized state spaces, e.g., see Cheng and Zhao (2023) and Yu et al. (2018).

To bridge this gap, we propose a framework for non-parametric repair-replacement policies in a single-unit system subject to continuous degradation, monitored via periodic inspections at intervals of δ time units. The system degradation process is modeled by an inverse Gaussian process $\{X(t), t \geq 0\}$ with a degradation rate parameter $\mu > 0$, a monotonic increasing shape function $\Lambda(t)$, and a scale parameter $\lambda > 0$. The degradation increment $\Delta X_{t_j,t_i} := X(t_j) - X(t_i), t_i < t_j$ is an inverse Gaussian-distributed random variable, i.e., $\Delta X_{t_j,t_i} \sim \mathrm{IG}(\mu \cdot \Delta\Lambda_{t_j,t_i}, \lambda \cdot \Delta\Lambda^2_{t_j,t_i})$ with the associated probability density function (PDF)

$$f_{\Delta X_{t_j,t_i}}(x; \mu \cdot \Delta\Lambda_{t_j,t_i}, \lambda \cdot \Delta\Lambda^2_{t_j,t_i}) = \sqrt{\frac{\lambda \cdot \Delta\Lambda^2_{t_j,t_i}}{2\pi \cdot x}} \cdot \exp\left(-\frac{\lambda}{2\mu^2} \cdot \frac{(x - \mu \cdot \Delta\Lambda_{t_j,t_i})^2}{x}\right), \tag{1}$$

where $\Delta\Lambda_{t_j,t_i} := \Lambda(t_j) - \Lambda(t_i)$. We refer to the inspection time as a *decision epoch* since after assessing the system degradation, the maintenance manager (also referred to as the decision maker) must decide which action to take on the system. Specifically, for operation safety, whenever the system's degradation exceeds a fixed critical threshold $L > 0$, it is considered failed and must undergo corrective replacement (CR) at a cost c_{CR} to bring it back to the AGAN condition. Otherwise, the decision maker can choose one of the following actions: doing nothing (DN), performing an imperfect repair (IR) to reduce some portion of the degradation, or conducting a preventive replacement (PR) to restore the system to the AGAN state. DN will only cost the inspection cost c_{I}, while IR and PR will respectively induce costs c_{IR} and c_{PR}, with the assumption that $c_{\mathrm{I}} < c_{\mathrm{IR}} < c_{\mathrm{PR}} < c_{\mathrm{CR}}$. Additionally, we take into account the benefit loss, which is proportional to the system downtime D with a downtime cost rate c_{D}. The aim of this problem is to seek a non-parametric, optimal maintenance policy $\pi^\star$ that optimizes the long-run maintenance cost rate. To achieve this, we formalize the problem within a Markov decision process (MDP) framework and propose a dynamic programming-with-function-approximation method for policy optimization.

The remainder of this paper proceeds as follows. Section 2 establishes the MDP formulation and details the policy optimization methodology. Following this, Section 3 presents numerical experiments and analyses. Finally, Section 4 concludes the paper with key insights and future research directions.

2. MDP formulation and policy optimization

We model the degradation-based maintenance problem in Section 1 as an MDP, denoted by $\mathscr{M} : (\mathscr{S}, \mathscr{A}, r, P)$, where:

- $\mathscr{S} = \{\boldsymbol{s} = (x, y) \mid x \in \mathbb{R}^+, y \in \mathscr{Y} = \{n\Delta \mid n \in \mathbb{N} \cup \{0\}\}\}$ is the state space, with x representing the system degradation and y denoting its age.
- $\mathscr{A} = \{\mathrm{DN}, \mathrm{IR}, \mathrm{PR}, \mathrm{CR}\}$ is the action space, where $\mathscr{A}(\boldsymbol{s})$ denotes the feasible actions in state $\boldsymbol{s}$.
- $r : \mathscr{S} \times \mathscr{A} \times \mathscr{S} \to \mathbb{R}^+$ is the immediate cost function, where $r(\boldsymbol{s}, a, \boldsymbol{s}')$ represents the cost of taking action a in state $\boldsymbol{s}$ and transitioning to state $\boldsymbol{s}'$.
- $P_{S_{i+1}|S_i,A_i} : \mathscr{S} \times \mathscr{A} \to \mathscr{P}(\mathscr{S})$ is the transition model, with $\mathscr{P}(\mathscr{S})$ as the space of probability distributions over $\mathscr{S}$. Specifically,

$$P_{S_{i+1}|S_i,A_i}(\mathscr{H} \mid \boldsymbol{s}, a) = \Pr(S_{i+1} \in \mathscr{H} \mid S_i = \boldsymbol{s}, A_i = a) \tag{2}$$

for any subset $\mathscr{H} \subset \mathscr{S}$.

This MDP is evaluated using the long-run cost rate metric

$$J(\pi) = \lim_{N \to \infty} \frac{1}{T \cdot \delta} \mathbb{E}\left[\sum_{i=0}^{T-1} r(S_i, A_i, S_{i+1}) \mid S_0 = \boldsymbol{s}\right], \tag{3}$$

where π is the current policy. In this paper, we focus on stationary deterministic policies, such that $\pi(S_u = \boldsymbol{s}) = \pi(S_v = \boldsymbol{s}) = \pi(\boldsymbol{s}) = a$ for any decision epochs u and v.

The state space $\mathscr{S}$ is a mixed continuous-discrete space because, under the periodic inspection assumption and the system starts its mission from AGAN state, the system age y is restricted to discrete values $n\Delta$, while the degradation level x can take any non-negative real value. In $\mathscr{M}$, the system state at decision epoch i is

denoted $S_i = (X_{Y_i}, Y_i)$, where Y_i is the system age and X_{Y_i} is the degradation level at that age. Note that Y_i does not necessarily equal the elapsed time t_i, as actions like PR or CR can reset it to 0 at some previous decision epochs. We assume all actions take negligible time, and, except for DN, all other actions instantaneously update the current state. In particular, if the system's current state is $S_i = (X_{Y_i}, Y_i) = \boldsymbol{s} = (x, y)$, then the its state immediately after action A_i, denoted S_{i^+}, is given by

$$S_{i^+} = (X_{Y_{i^+}}, Y_{i^+}) = \boldsymbol{s}^+ = (x^+, y^+) = \begin{cases} (x, y) & \text{if } A_i = \text{DN}, \\ (0, 0) & \text{if } A_i \in \{\text{PR}, \text{CR}\}, \\ (Kx, y) & \text{if } A_i = \text{IR}, \end{cases} \tag{4}$$

where $K \in [0, 1]$ is the repair efficiency coefficient. To model repair uncertainty, we assume that $K \sim \text{Beta}(a, b)$, with a and b as shape parameters. Fig. 1 presents the probability distributions (PDFs) and cumulative distribution functions (CDFs) of $\text{Beta}(a, b)$ for selected values of (a, b). For the transition model, it is easy to verify

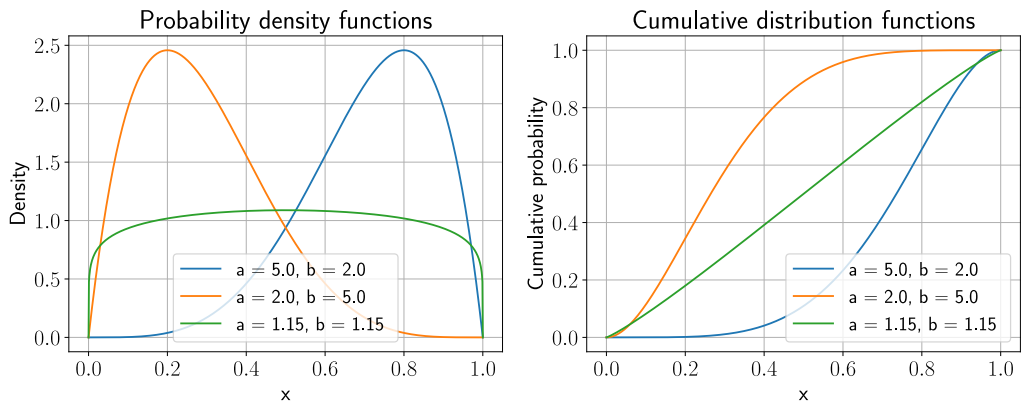


Figure 1: Some shapes of the Beta distribution's PDFs and CDFs.

that the next system age is deterministically updated as $Y_{t_{i+1}} = Y_{t_i^+} + \delta$. Considering a subset $\mathscr{H} = \mathscr{H}_X \times \mathscr{H}_Y$ of the state space, where $\mathscr{H}_X \subset \mathbb{R}^+$ and $\mathscr{H}_Y \subset \mathscr{Y}$, then it can be shown that

$$P_{S_{i+1}|S_i,A_i}(\mathscr{H} \mid \boldsymbol{s}, a) = \Pr(X_{i+1} \in \mathscr{H}_X \mid S_i = \boldsymbol{s}, A_i = a, Y_{i+1} = Y_{i^+} + \delta). \tag{5}$$

This reduces the problem of computing $P_{S_{i+1}|S_i,A_i}(\mathscr{H} \mid \boldsymbol{s}, a)$ to determining the probability distribution over $\mathscr{H}_X$. Since the next degradation level is $X_{Y_{i+1}} = X_{Y_{i^+}} + \Delta X_{Y_{i^+}, Y_{i^+}+\delta}$, we can leverage the properties of the IGP and Beta distribution to derive this distribution explicitly.

For the cost function r, its formula takes the form

$$r(S_i = \boldsymbol{s}, A_i = a, S_{i+1} = \boldsymbol{s}') = c(a) + c_{\text{D}} \cdot \mathbb{E}[D \mid S_{i^+} = \boldsymbol{s}^+, S_{i+1} = \boldsymbol{s}'], \tag{6}$$

where $c(a)$ is the cost of maintenance action $A_i = a$, and $\mathbb{E}[D \mid S_{i^+} = \boldsymbol{s}^+, S_{i+1} = \boldsymbol{s}']$ represents the conditional expected downtime given $S_{i^+} = \boldsymbol{s}^+$ and $S_{i+1} = \boldsymbol{s}'$. For $A_i \in \{\text{DN}, \text{PR}, \text{CR}\}$, this expected downtime is

$$\mathbb{E}[D \mid S_{i^+} = \boldsymbol{s}^+, S_{i+1} = \boldsymbol{s}'] = \int_{y^+}^{y^+ + \delta} \left(1 - F_{X_u|X_{Y_{i^+}}, X_{Y_{i+1}}}(L \mid x^+, x')\right) \mathrm{d}u, \tag{7}$$

where $F_{X_u|X_{Y_{i^+}}, X_{Y_{i+1}}}(L \mid x^+, x')$ is the conditional CDF of the degradation level X_u at age u satisfying $y^+ \leq u \leq y^+ + \delta$, given $X_{Y_{i^+}} = x^+$ and $X_{Y_{i+1}} = x'$. For $A_i = \text{IR}$, where the repair efficiency introduces uncertainty, it follows that

$$\mathbb{E}[D \mid S_{i^+}, S_{i+1}] = \int_{y^+}^{y^+ + \delta} \int_0^x \left(1 - F_{X_u|X_{Y_{i^+}}, X_{Y_{i+1}}}(L \mid x^+, x')\right) f_{X_{Y_{i^+}}|X_{Y_i}, A_i, X_{Y_{i+1}}}(x^+ \mid x, \text{IR}, x') \, \mathrm{d}x^+ \, \mathrm{d}u, \tag{8}$$

where $f_{X_{Y_{i^+}}|X_{Y_i}, A_i, X_{Y_{i+1}}}(x^+ \mid x, \text{IR}, x')$ is the conditional PDF of the degradation level immediately after the action IR, given the system's degradation level $X_{Y_i} = x$ before taking action $A_i = \text{IR}$ and the next degradation level $X_{Y_{i+1}} = x'$.

If the transition model $P_{S_{i+1}|S_i,A_i}$ and the immediate cost function r are well-defined, one can rely on dynamic

programming methods such as value iteration or policy iteration (Puterman 2005) to derive the optimal policy $\pi^\star$ of the given MDP. Value iteration indirectly derives $\pi^\star$ by firstly computing the optimal value fuction $V^\star(\boldsymbol{s})$ that satisfies Bellman optimality equation (Mahadevan 1996)

$$V^\star(\boldsymbol{s}) = (\mathscr{B}V^\star)(\boldsymbol{s}) - J(\pi^\star), \tag{9}$$

where $\mathscr{B}$ is the Bellman operator defined as $(\mathscr{B}V^\star)(\boldsymbol{s}) = \max_a \left\{ \int_{\mathscr{S}} \left(r(\boldsymbol{s}, a, \boldsymbol{s}') + V^\star(\boldsymbol{s}') \right) P_{S_{i+1}|S_i,A_i}(\mathrm{d}\boldsymbol{s}' \mid \boldsymbol{s}, a) \right\}$. Subsequently, the optimal policy $\pi^\star$ then can be extracted as

$$\pi^\star(\boldsymbol{s}) = \arg\max_a \left\{ \int_{\mathscr{S}} \left(r(\boldsymbol{s}, a, \boldsymbol{s}') + V^\star(\boldsymbol{s}') \right) P_{S_{i+1}|S_i,A_i}(\mathrm{d}\boldsymbol{s}' \mid \boldsymbol{s}, a) \right\}. \tag{10}$$

In the average cost setting, it can be proven that, if $V^\star(\boldsymbol{s})$ is an optimal value function, then $V_R(\boldsymbol{s}) = V^\star(\boldsymbol{s}) + C$ is also an optimal value function. Exploiting this property together with $\mathscr{B}$ is a span-contraction operator, with an initial guess $V^{(0)}$, one can iteratively apply $\mathscr{B}$ to approximate the value of $V^\star(\boldsymbol{s})$, i.e., for large K, $V^{(K)}(\boldsymbol{s}) = (\mathscr{B}V^{(K-1)})(\boldsymbol{s}) \approx V^\star(\boldsymbol{s})$. However, $V^{(K)}(\boldsymbol{s})$ can grow at a linear rate, leading to unstable approximation. Knowing that $V_R(\boldsymbol{s}) = V^\star(\boldsymbol{s}) - V^\star(\boldsymbol{s}_R)$, where $\boldsymbol{s}_R$ is an arbitrary state, is also an optimal value function. We can modify the update rule of $V^{(K)}$ to $V^{(K)}(\boldsymbol{s}) = (\mathscr{B}V^{(K-1)})(\boldsymbol{s}) - V^{(K-1)}(\boldsymbol{s}_R)$ to get a more stable approximation. This kind of update is called the *relative value iteration* (RVI). However, if $\mathscr{S}$ has an infinite number of elements, it is impossible to iterate over all states. To tackle this problem, we can approximate the values of a finite number of states and then apply some kinds of function approximation to compute the values of other points. Since the state space $\mathscr{S}$ of $\mathscr{M}$ is mixed discrete-continuous, we combine the relative value iteration method with k-nearest neighbors (KNN) regression (Tran et al. 2024) to obtain the near-optimal value function and then derive the associated policy using Eq. (10).

3. Numerical results

The standard parameter configurations for our experimental framework are summarized in Table 1. For the IGP shape function $\Lambda(t)$, we analyze three distinct cases: linear ($3t$), convex ($1.75t^{1.25}$), and concave ($5.5t^{0.75}$). Correspondingly, we investigate three parameterizations (a,b) for the beta-distributed repair coefficient K: right-skewed $(2,5)$, left-skewed $(5,2)$, and quasi-uniform $(1.15,1.15)$. The derived maintenance policies using RVI-KNN algorithm under these configurations (Figs. 2, 3, 4) consistently exhibit, at each system age Y_i, a control-limit structure governed by two degradation thresholds (R, M). When system degradation X_{Y_i} remains below R, DN constitutes the optimal action. For degradation levels between R and M, IR becomes preferable, while PR is recommended between M and L. Beyond L, CR becomes mandatory according to the initial maintenance rule assumption. In linear degradation scenarios (Fig. 2), thresholds R and M demonstrate remarkable

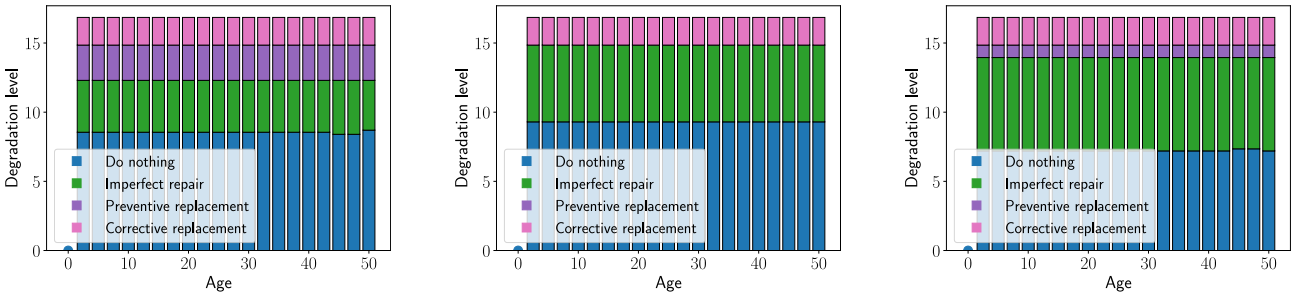
Table 1: Default settings for the maintenance problem

μ	λ	δ	c_{I}	c_{IR}	c_{PR}	c_{CR}	c_{D}
0.5	1	2.5	1	10	50	100	75

age-invariance, reflecting the constant degradation rate characteristic of this case. Conversely, convex degradation patterns (Fig. 3) exhibit progressively decreasing R and M values with system aging, necessitating earlier interventions to counter accelerating degradation rates. The concave case (Fig. 4) presents an inverse relationship where aging systems develop higher R and M thresholds, enabling delayed maintenance actions due to decelerating degradation rates.

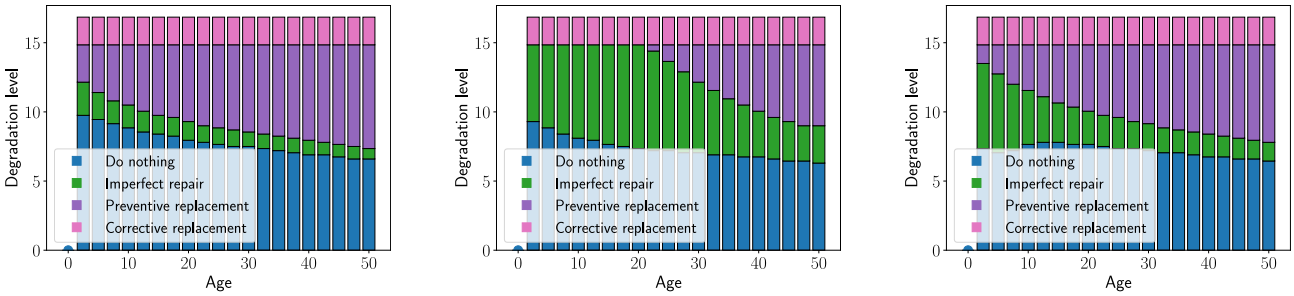
Imperfect repair effectiveness shows distinct patterns across repair coefficient distributions. Left-skewed K distributions produce narrow IR intervals ($[R,M)$), particularly under linear and convex degradation cases. Right-skewed K distributions conversely yield substantially wider IR ranges, occasionally eliminating PR requirements through cost-effective degradation reduction. quasi-uniform K distributions generate intermediate

IR intervals, though with more frequent PR occurrences compared to left-skewed cases. Notably, convex degradation scenarios exhibit age-dependent PR zone ($[M,L)$) expansion due to IR's inability to mitigate degradation rate acceleration, while concave degradation shows progressive contraction of both IR and PR zones with system aging, ultimately favoring DN actions as degradation rates diminish.



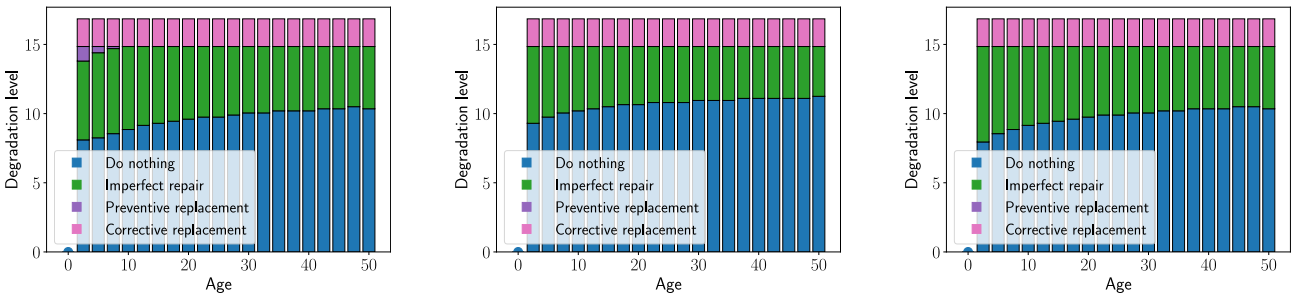
(a) $K \sim \text{Beta}(5,2)$ (b) $K \sim \text{Beta}(2,5)$ (c) $K \sim \text{Beta}(1.15,1.15)$

Figure 2: Near-optimal policy for linear degradation paths



(a) $K \sim \text{Beta}(5,2)$ (b) $K \sim \text{Beta}(2,5)$ (c) $K \sim \text{Beta}(1.15,1.15)$

Figure 3: Near-optimal policy for convex degradation paths



(a) $K \sim \text{Beta}(5,2)$ (b) $K \sim \text{Beta}(2,5)$ (c) $K \sim \text{Beta}(1.15,1.15)$

Figure 4: Near-optimal policy for concave degradation paths

We further present in Table 2 comparative average cost rates between our repair-replacement policies and replacement-only benchmarks. Notably, repair-replacement policies consistently outperform replacement-only approaches, even under low-efficiency K conditions.

4. Conclusion and future perspective

This paper develops a framework for deriving optimal maintenance policies in continuously degrading repairable systems. The proposed approach adapts effectively to diverse repair coefficient distributions (left-skewed, right-skewed, and quasi-uniform) while consistently outperforming replacement-only strategies, even under low-efficiency repair conditions. The age-responsive two-threshold structure (R, M) dynamically adjusts to degradation patterns: decreasing thresholds for convex (accelerating) degradation, increasing thresholds for

Table 2: Average cost rates

	Repair-replacement			
Degradation path	$K \sim \text{Beta}(5,2)$	$K \sim \text{Beta}(2,5)$	$K \sim \text{Beta}(1.15,1.15)$	Only replacement
Convex	4.54721	6.74160	6.13037	6.89711
Concave	1.30829	2.89608	1.82131	4.94837
Linear	2.36979	6.19485	4.17993	6.72737

concave (decelerating) cases, and stable thresholds for linear progression. Future works can consider non-periodic inspection scheme, allowing a more flexible maintenance planning.

Acknowledgement

This work is funded by the Grand Est region and the French Ministry of Higher Education and Research.

References

Pham, Hoang and Hongzhou Wang (Nov. 1996). "Imperfect Maintenance". In: *European Journal of Operational Research* 94.3, pp. 425–438.

Elwany, Alaa H., Nagi Z. Gebraeel, and Lisa M. Maillart (June 2011). "Structured Replacement Policies for Components with Complex Degradation Processes and Dedicated Sensors". In: *Operations Research* 59.3, pp. 684–695.

Zhang, Xiaohong and Jianchao Zeng (Oct. 2017). "Joint Optimization of Condition-Based Opportunistic Maintenance and Spare Parts Provisioning Policy in Multiunit Systems". In: *European Journal of Operational Research* 262.2, pp. 479–498.

Cheng, Wanqing and Xiujie Zhao (Dec. 2023). "Maintenance Optimization for Dependent Two-Component Degrading Systems Subject to Imperfect Repair". In: *Reliability Engineering & System Safety* 240, p. 109581.

Mosayebi Omshi, Elham and Antoine Grall (Nov. 2021). "Replacement and Imperfect Repair of Deteriorating System: Study of a CBM Policy and Impact of Repair Efficiency". In: *Reliability Engineering & System Safety* 215, p. 107905.

Huynh, Khac Tuan, Antoine Grall, and Christophe Berenguer (Mar. 2019). "A Parametric Predictive Maintenance Decision-Making Framework Considering Improved System Health Prognosis Precision". In: *IEEE Transactions on Reliability* 68.1, pp. 375–396.

Puterman, Martin L. (2005). *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley Series in Probability and Statistics. Hoboken, NJ: Wiley-Interscience.

Huynh, Khac Tuan (Sept. 2021). "An Adaptive Predictive Maintenance Model for Repairable Deteriorating Systems Using Inverse Gaussian Degradation Process". In: *Reliability Engineering & System Safety* 213, p. 107695.

Yu, Jiawei et al. (Apr. 2018). "Reliability Modeling of Multi-state Degraded Repairable Systems and Its Applications to Automotive Systems". In: *Quality and Reliability Engineering International* 34.3, pp. 459–474.

Mahadevan, Sridhar (1996). "Average Reward Reinforcement Learning: Foundations, Algorithms, and Empirical Results". In: *Machine Learning* 22.1/2/3, pp. 159–195.

Tran, Quang Khai et al. (2024). "Average Reward Reinforcement Learning for Optimizing Condition-Based Maintenance Policies of Continuously Deteriorating Systems". In: *ESREL 2024 Monograph Book Series, Advances in Reliability, Safety and Security, Part 4*, pp. 213–222.

Uncertainty Quantification as a Complementary Latent Health Indicator for Remaining Useful Life Prediction on Turbofan Engines

L. Thil*, J. Read**, R. Kaddah***and G. Doquet****

*LIX Ecole Polytechnique & IRT SystemX: thil@lix.polytechnique.fr
** LIX Ecole Polytechnique: jesse.read@polytechnique.edu
*** IRT SystemX: rim.kaddah@irt-systemx.fr
**** Safran Tech: guillaume.doquet@safrangroup.fr

Abstract. Health Indicators (HIs) are essential for predicting system failures in predictive maintenance. While methods like RaPP (Reconstruction along Projected Pathways) improve traditional HI approaches by leveraging autoencoder latent spaces, their performance can be hindered by both aleatoric and epistemic uncertainties. In this paper, we propose a novel framework that integrates uncertainty quantification into autoencoder-based latent spaces, enhancing RaPP-generated HIs. We demonstrate that separating aleatoric uncertainty from epistemic uncertainty and cross combining HI information is the driver of accuracy improvements in Remaining Useful Life (RUL) prediction. Our method employs both standard and variational autoencoders to construct these HIs, which are then used to train a machine learning model for RUL prediction. Benchmarked on the NASA C-MAPSS turbofan dataset, our approach outperforms traditional HI-based methods and end-to-end RUL prediction models and is competitive with RUL estimation methods. These results underscore the importance of uncertainty quantification in health assessment and showcase its significant impact on predictive performance when incorporated into the HI construction process.

1. Introduction

Accurate estimation of the RUL of complex systems, such as industrial machinery or aerospace components, is critical for enabling predictive maintenance, reducing operational costs, and mitigating catastrophic failures. By translating raw sensor data into HIs that reflect system degradation, RUL prognostics provide actionable insights for real-time decision-making. Traditional approaches to prognostics have relied on statistical methods, including regression models, survival analysis, and stochastic techniques, which often require strong assumptions about degradation patterns or failure modes. However, the advent of deep learning has allowed the field to leverage data-driven models to autonomously learn complex degradation trends directly from high-dimensional sensor data, eliminating the need for handcrafted feature engineering. Among these data-driven methods, autoencoder (AE)-based frameworks have gained prominence for their ability to learn healthy system representations. By training on nominal operational data, AEs reconstruct input signals, and deviations in reconstruction error serve as a measure of novelty detection indicating a degradation from this nominal healthy state. Recent advances, such as RaPP Ki Hyun et al. (2020) and González-Muñiz et al. (2022), refine this approach by iteratively projecting reconstruction errors into the latent space of the encoder, generating more accurate HIs to model degradation trends. While promising, such methods often overlook the inherent uncertainty in model predictions. This is a critical gap for safety-critical applications where unreliable HIs can led to costly overhauls or unexpected downtime.

In this work, we propose to augment the RaPP HIs with uncertainty quantification (UQ) through Monte Carlo dropout, enabling probabilistic interpretation of degradation signals. Our contributions are in the fusion of these indicators and benchmarking the HIs directly on RUL estimation for prognosability clarity.

2. Literature Review

There exist various methods to estimate degradations in systems using data-driven approaches, such as measuring a degradation trend, doing linear regression. But a drawback from these is that often they are not interpretable, or do not offer the possibility to assess in real time the current system's status for a domain expert to comprehend. Therefore, the notion of health indicators (HIs) is crucial to produce a set of quantifiable metrics providing a real time description of the health of the system. Common HIs could be the

remaining capacity of a lithium-ion battery, the State of Health (SOH) or any relevant metric that provides an assessment of a system or its inner key components.

The State of Health (SOH) of unmaintained engineering systems is modeled as a continuous monotonous function, reflecting the expectation that such systems degrade progressively until failure without self-recovery or maintenance. Health Indicators (HIs) must therefore provide actionable insights into this degradation trajectory. Common evaluation criteria for HIs include monotonicity (consistent degradation trend), trendability (detectable degradation patterns), and prognosability (predictive utility for remaining useful life, or RUL) (Coble, 2010). Traditional data-driven HIs rely on statistical features of sensor signals, such as skewness, kurtosis, or variance over time, whose evolution correlates with degradation states (Fink et al., 2020). However, emerging approaches leverage machine learning to design more robust HIs. Among these, autoencoders (AEs) have gained prominence: trained on healthy, non-degraded data (e.g., sensor readings from intact turbofan engines), AEs learn to reconstruct normal operating patterns. When exposed to degraded inputs (e.g., signals from aging components), the AE's reconstruction error—measured as the difference between the input (original data) and output (reconstructed data)—increases, serving as a direct HI (Huang et al., 2023). This error, often quantified via metrics like mean squared error (MSE), acts as a novelty score, where higher deviations signal greater degradation severity. For instance, in turbofan engine monitoring, the rising reconstruction error mirrors progressive mechanical wear, enabling RUL estimation. While most methods focus on input-output reconstruction, recent work exploits latent space distances in AEs, leveraging their disentangled representations to capture degradation patterns more robustly.

Current metrics focus on the entire input/output space, and these individual parameters do not necessarily reflect the hierarchical nature of deep learning models in the inner latent spaces. Comes into play the RaPP by Ki Hyun et al. (2020) which uses the projections of the activations of the model's parameters to compute the errors. It is based on two families: one is a Euclidean distance of the activations of the model's parameters computed by each layer, and the second is based on the Mahalanobis distance as it takes into consideration the variance and correlation of the different components. These RaPP metrics were initially proposed in the context of an encoder hidden layer space, but González-Muñiz et al. (2022) highlighted better results when it was used only in the latent space of the AE architecture.

In UQ, noise is categorized as aleatoric when arising from inherent data variability that cannot be reduced with additional samples, or epistemic originating from model uncertainty, which diminishes with more training data. Building on this distinction, Akbari and Jafari (2020) leveraged a Variational Auto-Encoder (VAE) with Monte Carlo dropout (Gal & Ghahramani, 2015) between its latent space and decoder layers to disentangle these uncertainties. While prior work highlights methodological advances, the interplay of HIs for degradation modeling remains understudied. In this study, we investigate how integrating such HIs enhances data-driven degradation modeling, with a focus on their combined efficacy for accurate remaining useful life (RUL) estimation.

3. Methods

3.1 *Reconstruction and Latent Distance Measures*

In recent years, autoencoders architectures have been used as HI by using the reconstruction error of its output $\hat{x}$ from an input x. The underlying idea is to train an autoencoder architecture over non-degraded inputs such that when presented with degraded inputs, it will fail to perform an accurate reconstruction. This error can be measured to identify degraded samples, while quantifying their degree with a distance measure based on the reconstruction error ϵ_{REC} defined as:

$$\epsilon_{REC} = |x - \hat{x}|. \tag{1}$$

Current metrics focus on the entire input/output space, and these individual parameters do not necessarily reflect the hierarchical nature of deep learning models in the inner latent spaces Ki Hyun et al. (2020). Using a larger manifold with the hidden representation of an encoder shows superior capabilities in novel detection than measuring only ϵ_{REC} (1).

The RaPP methods use the projections of the activations of the model's parameters to compute the errors. It is based on two families: ϵ_{SAP} which is a Euclidean distance of the activations of the model's parameters computed by each layer, and ϵ_{NAP} which is based on the Mahalanobis distance and and differs

from ϵ_{SAP} as it takes into consideration the variance and correlation of the different parameters. ϵ_{NAP} attributes less weight to sparser components. In the case of signal analysis with the assumption that each component be a random Gaussian variable, it leads to minimizing the influence of the noisiest components in the computation of the distance metric.

$$\epsilon_{SAP}(x) = |h_i(x) - h_i(\hat{x})|. \tag{2}$$

$$\epsilon_{NAP}(x) = |(d(x) - \mu_X)^T V \Sigma^{-1}|. \tag{3}$$

Where $h_i(x)$ are the activations of the ith hidden layer h from input x, $d(x) = h(x) - h(\hat{x})$, μ_X is the column wise mean of the matrix D with the rows collecting $d(x)$, and V containing the right singular vectors from the SVD of D, and Σ the diagonal of $\bar{D}$ denoting the column-wise centered version of D. These RaPP metrics were initially proposed in the context of an Encoder space, but González-Muñiz et al. (2022) highlighted better results when it was used only in the latent space of the AE architecture. Similarly, we denote the RaPP HIs from the encoder as ϵ_{SAP} and ϵ_{NAP}, and the latent those computed on z only noted $\epsilon_{SAP_{LS}}$ and $\epsilon_{SAP_{LS}}$.

3.2 *Adding Uncertainty Quantification: Aleatoric versus Epistemic*

Aleatoric uncertainty noted σ_a comes from noise in the data and cannot be reduced by adding more data, whereas epistemic uncertainty σ_e originates from the lack of knowledge of the model and can be mitigated with more samples. Uncertainty in an autoencoder (AE) model can be quantified using Monte Carlo dropout applied to the decoder. However, since aleatoric noise cannot be explicitly modeled in the latent space of an AE, disentangling aleatoric and epistemic uncertainties remains challenging. In contrast, variational autoencoders (VAEs) offer a more nuanced approach. By modeling the latent space as a Gaussian distribution, VAEs enable the separation of these two types of uncertainties, providing a clearer distinction between inherent data noise and model-related uncertainties.

In Bayesian deep learning, encoder model parameters W_E and latent variables z are treated probabilistically. For regression tasks, the likelihood is modeled by a Gaussian:

$$p(y \mid W, x) = \mathcal{N}(y; \mu^{W_E}(x), \sigma^{W_E}(x)^2). \tag{4}$$

Where $\mu^{W_E}(x)$ and σ^{W_E} are encoder-derived mean and variance. MC dropout approximates Bayesian inference by sampling from the posterior via stochastic dropout mask during inference. The predictive distribution integrates over both W and latent variables z:

$$p(y \mid x) = \int N(y; \mu^{W_E}(x, z), \sigma^{W_E}(x, z)^2) p(W \mid X) p(z \mid x) d W_E dz. \tag{5}$$

where X is the training data. This intractable integral is approximated via MC sampling:

$$p(y \mid x) \approx \frac{1}{n} \sum p(y \mid \hat{W}_E, x, \hat{z}). \tag{6}$$

With $\hat{W}_E \sim p(W_E)$ (via dropout) and $\hat{z} \sim p(z \mid x)$. VAE explicitly models aleatoric uncertainty through a stochastic latent space $\hat{z} \sim q(z \mid x)$ is the encoder's approximate posterior. The training objective combines reconstruction loss and KL divergence regularization:

$$\mathcal{L}_{VAE} = -E_{q(z|x)}[\log p(x \mid z)] + \beta . D_{KL}(q(z \mid x) \mid p(z)). \tag{7}$$

With $p(z)$ as a standard Gaussian prior. During inference: 1) Aleatoric uncertainty is sampled via $\hat{z} \sim q(z \mid x)$, 2) Epistemic uncertainty is introduced through dropout on decoder weights W_E. In the case of vanilla AEs, aleatoric uncertainty σ_a cannot be isolated as z is deterministic, rendering it undefined. This highlights the advantage of VAEs for joint uncertainty estimation.

The randomness of the latent space z can be used to quantify the aleatoric uncertainties, and where W_h being the weights of a classification head can be used to quantify the epistemic uncertainties by applying dropout over multiple passes. To perform Monte Carlo sampling, from an input x we infer the model over n passes while applying dropout over the weights and averaging the outputs as the Bayesian inference. Critically, because this is a reconstruction regression task, we perform UQ on the full output rather than a specific label, but we need a reduction operation on the dimensions to have the uncertainty value to 1. As such, UQ is applied to the reconstruction error ϵ_{REC} instead of the direct full output y.

Finally, our AE model will model a global noise that is likely to contain mostly epistemic noise, whereas in the VAE we will be able to separate the two. We will use the uncertainties quantifications over the data as standalone HIs that we can combine with the RaPP ones.

3.3 *Creating and Benchmarking Health Indicators*

Compared to previous HI approaches using RaPP, we will use these series to predict a RUL. We train a random forest (RF) regressor over those and perform an ablation study to assess their impact on predicting a RUL. Furthermore, this benchmark will allow us to directly compare the predictive capabilities of these indicators compared to more traditional models trained directly to predict the RUL from direct sensor data. Because these latter approaches do not produce HIs they do not have clear evaluation metrics over the values we produce. In this work, we bridge this gap by benchmarking our created HIs with existing methods.

There are various benchmarks performed over C-MAPSS Saxena et al. (2008) but they often carry different protocols. Indeed, the underlying degradation mechanism in C-MAPSS is rather linear until it enters an exponential growth in the latter stages. Thus, many approaches defined a cutoff threshold called $\boldsymbol{R}_{early}$ that was commonly used at RUL=125, sometimes they would clip values larger than 125 as they assumed that the degradation was not sufficiently present.

3.4 *Datasets: 4 subsets from the C-MAPSS repository with a mix of flight conditions and degradation mechanisms*

The train trajectories are complete until a point of failure indicated by RUL=0, while the test set cuts off before failure. In some of the protocols over the test set, we only predict up to that point. The datasets contain an increasing number of conditions and fault modes, and it can be assumed that the difficulty increases for each set, FD001 being the simplest one and FD004 the hardest.

	FD001	**FD002**	**FD003**	**FD004**
# Train trajectories	100	260	100	248
# Test trajectories	100	259	100	249
# Conditions	1	6	1	6
# Fault modes	1	1	2	2

Table 1. Description of the C-MAPSS Dataset

3.5 *Evaluating the Health Indicators*

Classical HIs are assessed through monotonicity, trendability, and prognosability. However, monotonicity and trendability lack direct relevance to SOH estimation, offering limited interpretability and inconsistent performance insights. Prognosability often yields near-perfect scores (e.g., 0.99), misleadingly implying precise SOH knowledge despite poor RUL prediction accuracy. To overcome these limitations, we propose benchmarking HIs directly on RUL estimation using a RF regressor, grounding evaluation in empirical performance. An ablation study further examines HI combinations' impact on prediction efficacy. Our work evaluates the complementary role of UQ with RaPP-generated HIs in RUL estimation, contrasting results against conventional HI metrics.

4. Results

We show that both AE in Fig. 1 and VAE in Fig. 2 models demonstrate consistent reductions in RMSE when combining HIs derived from RaPP with UQ. For instance, on dataset FD001, the combined use of latent RaPP features (NAP and SAP) and UQ achieves near state-of-the-art (SOTA) performance (RMSE = 11.23 vs. SOTA = 11.17). Notably, our approach surpasses existing benchmarks on more complex datasets (FD003 and FD004), where the inclusion of diverse HIs leads to superior RUL estimation accuracy. We attribute this improvement to the complementary behavior of combined HIs: while individual indicators exhibit oscillatory patterns, their integration stabilizes the health status representation, enabling clearer degradation trends.

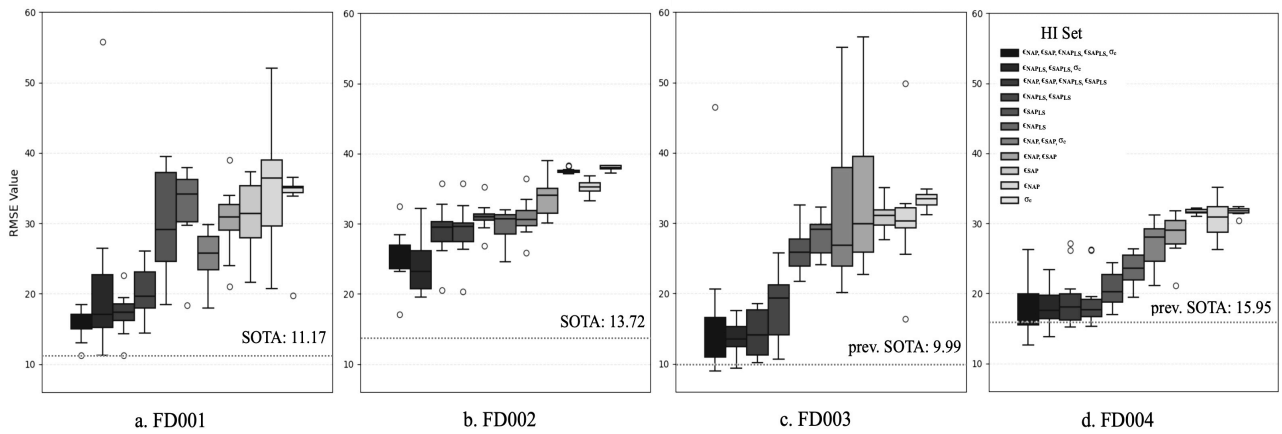


Figure 1. Autoencoder (AE) benchmark results over C-MAPSS dataset for different HI groups. Lower RMSE scores means better accuracy in RUL estimation. We observe that the groups combining the different HIs together achieve SOTA it for FD003 and FD004 thought as the most complex groups despite using a simple RF regressor.

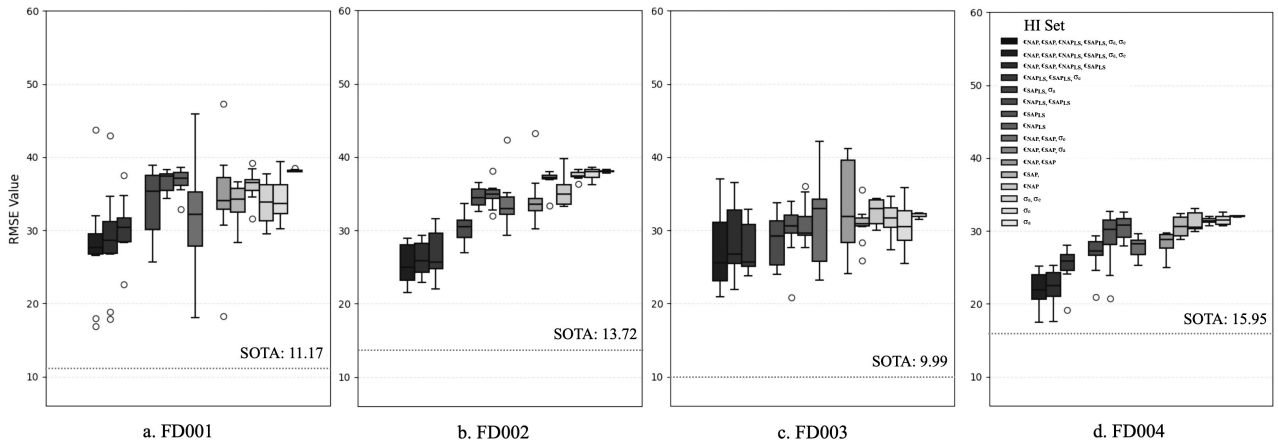


Figure 2. Variational Autoencoder (VAE) benchmark results over C-MAPSS dataset for different HI groups. We do not find better results than the AE counterpart shown in fig. 1 contrary to the expectations that the VAE architecture outperforms it. This is likely explained by the high noise in the data, but on FD004 the results are converging near SOTA performance suggesting potential over more complex datasets.

	FD001				FD003			
HI	Mono	Trend	Progno	RUL RMSE	Mono	Trend	Progno	RUL RMSE
σ_e	0,601	5,900E-05	0,748	19,74	0,513	4,000E-05	0,617	31,21
ϵ_{SAP}	0,691	3,320E-04	**0,850**	**21,64**	0,821	1,430E-04	**0,774**	**27,63**
ϵ_{NAP}	0,330	3,700E-05	0,617	20,67	0,421	1,400E-05	**0,148**	**16,4**
$\epsilon_{SAP_{LS}}$	0,728	5,830E-04	0,883	18,42	0,711	2,700E-05	0,325	21,66
$\epsilon_{NAP_{LS}}$	0,395	3,200E-05	**0,545**	**18,32**	0,306	1,900E-05	0,630	24,05

Table 2: Example of HIs evaluation metrics in respect to RUL RMSE (lower is better). The best RUL estimation metrics over C-MAPSS are not necessarily those showing the best prognosability (in bold), or with monotonicity and trendability.

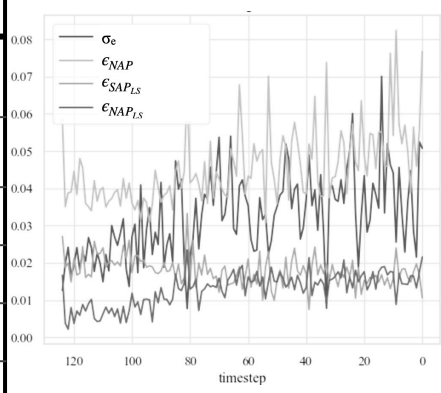


Figure 3: HI Evolution of engine 1 from FD001 test. Encoder_SAP not displayed because it has different scale values.

To further evaluate HI quality, Fig. 3 compares traditional metrics (monotonicity, trendability, prognosability) against RUL prediction RMSE. Results reveal weak links between these classical metrics and actual prognostic performance. For example, the feature achieves the lowest prognosability score but delivers the best RMSE for RUL prediction. This discrepancy highlights the limitations of relying solely on conventional HI evaluation criteria and underscores the need for task-specific metric alignment.

5. Discussion and Conclusions

We showed that coupling different HIs computed in the latent space of deep learning models is an effective approach in modeling degradation dynamics. We've shown that adding an uncertainty quantification component made the predictions more robust. To adapt the evaluation of our HIs, we propose to diverge from the traditional metrics of monotonicity, trendability and prognosability because they poorly reflect degradation mechanisms, and we instead propose to directly benchmark on RUL prediction to assess their quality. This also allows for an ablation study between different HI combinations to assess their performance. In short, the RaPP metric combined with uncertainty quantification isolating aleatoric and epistemic noise is a solid approach to model degradation dynamics from HIs in the latent space of deep learning models.

For future works, we aim at testing with architectures combining different window sizes, better regressor to evaluate the HIs or propose an architecture allowing to isolate sub-system HIs. The next area of focus would be over robustness and interpretability of HIs generated from the latent space of deep learning models.

Acknowledgements

This work has been supported by the French government under the "France 2030" program, as part of the SystemX Technological Research Institute within the JNI3 project.

References

Gal, Y., Ghahramani, Z., 2015. Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. https://doi.org/10.48550/ARXIV.1506.02142

Akbari, A., Jafari, R., 2020. Personalizing Activity Recognition Models Through Quantifying Different Types of Uncertainty Using Wearable Sensors. *IEEE Trans. Biomed. Eng*. 67, 2530–2541. https://doi.org/10.1109/TBME.2019.2963816

Fink, O., Wang, Q., Svensén, M., Dersin, P., Lee, W.-J., Ducoffe, M., 2020. Potential, challenges and future directions for deep learning in prognostics and health management applications. *Engineering Applications of Artificial Intelligence* 92, 103678. https://doi.org/10.1016/j.engappai.2020.103678

Saxena, A., Goebel, K., Simon, D., Eklund, N., 2008. Damage propagation modeling for aircraft engine run-to-failure simulation, in: *2008 International Conference on Prognostics and Health Management. Presented at the 2008 International Conference on Prognostics and Health Management (PHM)*, IEEE, Denver, CO, USA, pp. 1–9. https://doi.org/10.1109/PHM.2008.4711414

Huang, L., Pan, X., Liu, Y., Gong, L., 2023. An Unsupervised Machine Learning Approach for Monitoring Data Fusion and Health Indicator Construction. Sensors 23, 7239. https://doi.org/10.3390/s23167239

RaPP: Novelty Detection with Reconstruction along Projection Pathway, Ki Hyun Kim, Sangwoo Shim, Yongsub Lim, Jongseob Jeon, Jeongwoo Choi, Byungchan Kim and Andre S. Yoon, ICLR, 2020

González-Muñiz, A., Díaz, I., Cuadrado, A.A., García-Pérez, D., 2022. Health indicator for machine condition monitoring built in the latent space of a deep autoencoder. *Reliability Engineering & System Safety* 224, 108482. https://doi.org/10.1016/j.ress.2022.108482

Coble, J.B., 2010. Merging data sources to predict remaining useful life–an automated method to identify prognostic parameters.

Task Grouping Optimization for Gas Insulated Substations Predictive Maintenance: Trade-off between cost and environmental impact

Wenbo Wu*, Rim Kaddah*, Miguel F. Anjos**, Paul-Henri Langlois ***, Sihem Tebbani****

* Technological Research Institute SystemX, 2 Bd Thomas Gobert, 91120 France {wu,wenbo, rim.kaddah}@irt-systemx.fr)

** School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh, Edinburgh EH9 3FD United Kingdom and GERAD – HEC Montréal, 3000, ch. de la Côte-Sainte-Catherine, Montréal (Québec) H3T 2A7 Canada (anjos@stanfordalumni.org)

*** Réseau de Transport d'Electricité, 7C Place du DÔME, Puteaux, France (paul-henri.langlois@rte-france.com)

**** Université Paris-Saclay, CentraleSupélec, CNRS, Laboratoire des Signaux et Systèmes (L2S). 3 rue Joliot-Curie, 91192 Gif-Sur-Yvette, France (sihem.tebbani@centralesupelec.fr)

Abstract: An effective predictive maintenance plan ensures operational reliability while achieving key optimization goals. In the context of Gas-Insulated Substations (GIS), our core concerns lie in minimizing maintenance costs while mitigating the environmental impact of sulfur hexafluoride (SF_6) gas leakage, a potent greenhouse gas. Building on our previous work optimizing maintenance schedules for individual GIS components, this study extends the approach to a second-stage task-grouping optimization consolidating maintenance tasks across multiple GIS components to address operational inefficiencies such as increased costs, underutilized resources, and prolonged downtime. Specifically, our approach combines maintenance tasks that are temporally close and operationally compatible, thereby reducing travel costs and improving resource utilization. The proposed method formulates the grouping problem as an Integer Linear Programming model, incorporating key constraints such as task time windows, resource availability, and operational priorities. Furthermore, the trade-off between maintenance costs and SF_6 amount leaked is analyzed to account for its environmental impact.

1. Introduction

In modern power systems, Gas-Insulated Substations (GIS) play a critical role in ensuring the reliability and efficiency of electricity transmission and distribution. A GIS is a multi-component system which comprises many Pressurized Metal-Enclosed Compartments (PMEC), which house SF_6 gas-insulated equipment (CIGRE, 2014). As these systems age and their operational complexity increases, effective maintenance planning for PMECs —particularly gas refilling (GR) interventions—is essential to ensure system reliability (Burns, 2024). This study adopts an optimization framework for coordinating maintenance tasks across multiple components, similar to the scalable bi-objective maintenance models proposed by (Kivanç, 2025).

Currently, maintenance tasks have been planned independently for each PMEC based on predefined schedules or reactive responses to alarms. The scheduling of maintenance activities is made possible by the integration of real-time monitoring systems with GIS, as the presence of sensors enables continuous tracking of pressure parameters within each PMEC. Utilizing these data, power utilities have developed alarm prediction algorithms and, based on these forecasts, have established long-term proactive maintenance plans for each PMEC. However, strictly adhering to individual maintenance schedules and addressing each intervention separately is not a globally optimal strategy, as it often results in redundant resource allocation, thus increased operational costs (including transportation, labor, and equipment mobilization), and potential disruptions to power system operations. For instance, dispatching a crew to address a single PMEC's GR incurs certain costs, and recurring visits for temporally proximate tasks increase the total costs. To address this challenge, dynamic update of maintenance planning that groups tasks across multiple PMECs emerges as a strategy to minimize maintenance costs while reducing the frequency of interventions.

This work extends our previous research on predictive maintenance optimization for GIS under operational and environmental constraints. In that study, we developed a framework that provides a decision-supporting Mixed Integer Linear Programming (MILP) optimization model for maintenance scheduling (Wu, et al., 2025). In this paper, we introduced a novel sliding window framework to dynamically group GR tasks for multiple

PMECs. The method identifies tasks scheduled in close temporal proximity, clusters them into intervention batches, and updates subsequent schedules iteratively. By optimizing task grouping within the sliding window, the approach minimizes redundant site visits, thereby reducing transportation costs and labor costs associated with frequent interventions. The sliding window approach proposed in this paper addresses this gap by introducing a flexible and scalable framework for multi-unit system task grouping.

This paper makes the following contributions: (i) The problem of multi-unit maintenance task grouping is formulated as an Integer Linear Programming (ILP) optimization problem with temporal and operational constraints. (ii) A sliding window framework is developed to dynamically identify and group tasks based on their temporal proximity. (iii) The proposed approach is evaluated in the context of GIS maintenance planning using real-world-inspired scenarios, demonstrating its effectiveness in improving efficiency and reducing costs.

2. Problem description

Each PMEC houses electrical components insulated with SF_6 gas, which is a potent greenhouse gas (IPCC, 2018), maintained at pressures between 3.0 and 3.5 bar. While SF_6 is essential for reliable operation, leaks are inevitable due to design and operational factors, posing environmental risks and increasing maintenance demands (Phillip Widger, 2018). Maintenance tasks for PMECs fall into two categories: Gas Refilling (GR): A routine, low-cost intervention triggered when pressure drops to 3.2 bar and typically takes only a few hours to carry out. Repair (RE): A comprehensive repair required after several GR interventions to reduce leakage rates or restore PMECs to near-new condition. RE operations are costly and require extensive planning across multiple departments. Given the impracticality of combining RE operations due to their complexity and long execution times, this study focuses exclusively on optimizing GR task grouping.

The operational challenge arises from the scale and distribution of GIS networks, which often lead to fragmented GR schedules across large numbers of PMECs at different substations. Reactive maintenance practices frequently lead to the mobilization of teams for isolated GR tasks even when multiple maintenance needs occur at the same location within a short time frame. This inefficiency increases transportation costs and personnel deployment frequency while failing to capitalize on opportunities for strategic clustering. By strategically rescheduling GR tasks within acceptable temporal ranges, multiple interventions can be consolidated into single maintenance visits—reducing operational costs without compromising system reliability or safety thresholds. In the present study, we focus on the optimization of task grouping within each site, while the allocation of personnel is regarded as a separate planning process.

3. Methodology

3.1 Sliding window framework

In this section, we present the methodology used to optimize maintenance task grouping across one GIS where all the PMECs are supposed to be in the same location. The sliding window approach serves as the backbone of the proposed methodology, enabling dynamic task grouping over time. The framework divides the entire planning horizon into overlapping time intervals, referred to as "windows." Within each window, maintenance tasks of different PMECs are considered for optimization. Figure 1 illustrates the algorithmic workflow designed to systematically decompose the maintenance grouping problem into manageable temporal segments. It should be noted that the sliding period of the window in this context is flexible, rather than fixed at each iteration. Two scenarios are considered: if there remain ungrouped tasks within the current window, the window slides forward to the position of the next such task; if no such tasks exist, the window advances by the full length of the window. This sliding mechanism ensures that the potential for grouping every task is considered.

The proposed sliding window approach decomposes the global maintenance clustering problem into a series of subproblems. This decomposition stems from two important observations: (i) it is computationally difficult to optimize all maintenance tasks simultaneously in real-world problems due to the long-time intervals (usually measured in months) between actual GR interventions, which makes it impossible to have a clear understanding of the future for a long period of time. (ii) each clustering decision impacts the timing of future tasks. The algorithm maintains the continuity of the solution through a window switching mechanism.

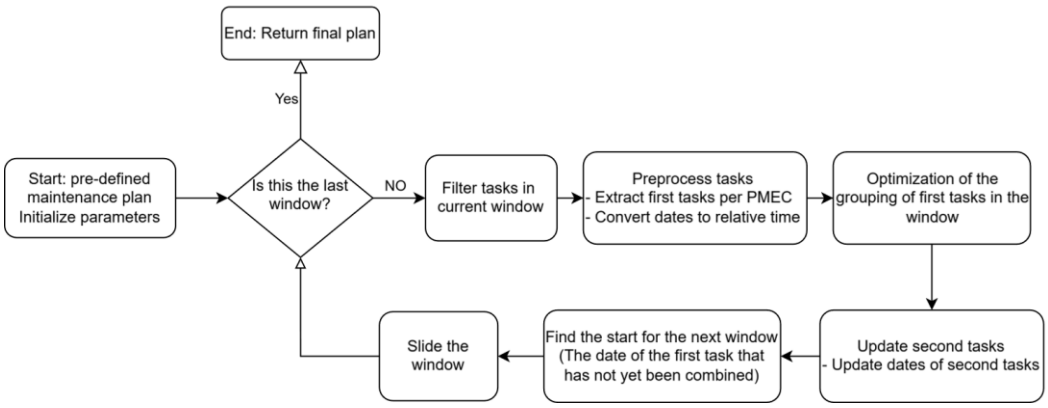


Figure 1. Task grouping in sliding window optimization framework

3.2 Grouping ILP optimization model

In this section, we formulate the grouping optimization problem of one GIS (all PMECs in the same location) as an ILP model that minimizes the number of clusters, i.e., groups as many tasks as possible together while respecting the operational constraints in each window. Table 1 gives the definitions of the parameters and variables.

Problem sets	
$\mathbf{N}$	Set of all the maintenance tasks.
$\mathbf{P}$	Set of PMECs.
Parameters	
$\mathbf{T}$	Planning horizon. The sampling time is set to 1 day.
A	Maximum number of days each task can be advanced.
$\mathbf{t}_{\boldsymbol{n}}$	Original planned time of task *n* before grouping.
$\boldsymbol{\alpha}_{\boldsymbol{i(n)}}$	Degradation rate of PMEC i associated with task n.
$\mathbf{C}_{\mathbf{max}}^{\mathbf{GR}}$	Maximum number of GR in one cluster.
L	Length of the sliding window.
Integer variables	
$\mathbf{a}_{\mathbf{n}}$	Earliest time that task n can be brought forward to.
$\mathbf{b}_{\boldsymbol{n}}$	The latest time that task n can be delayed to.
Binary variables	
$\boldsymbol{\theta}_{\boldsymbol{nt}}$	= 1 if a GR task *n* is assigned to a cluster at time t. Otherwise, 0.
$\mathbf{cluster}_{\boldsymbol{t}}$	= 1 if cluster at time t is activated. Otherwise, 0.

Table 1. Notations and definitions

The model treats each day in the planning horizon as a potential cluster, i.e., we assume that only one cluster is possible for each day. It is then possible to adopt a time-cluster equivalence principle and unify temporal and cluster indices: a cluster at day t is activated if any task n is assigned to it (θ_{nt}). This is also the decision variable in the objective function which directly links task-to-cluster.

The proposed ILP model is formulated as follows: the objective function (1) minimizes the total number of activated maintenance clusters across the sliding window, promoting resource consolidation. Constraint (2) enforces cluster activation when containing at least one task, eliminating empty clusters. Each maintenance task receives a unique assignment through constraint (3) to prevent task duplication or omission. The cluster capacity constraint is addressed through constraint (4), which limits the number of tasks assigned to each cluster (i.e., each day), a key operational parameter that reflects resource availability. Constraints (5) to (8) define the time boundaries within which each task can be brought forward or delayed, denoted respectively by a_n and b_n. Specifically, a_n represents the earliest time a task can be brought forward within the scheduling

window. Here, we use the parameter A representing a maximum allowable advancement of 3 days for any task. This value is selected based on the operational needs and restrictions. Considering the length of the window, (5) and (6) effectively define $a_n = \max(0, t_n - A)$, where t_n is the initially scheduled time of task n. Similarly, b_n indicates the latest time task n can be postponed. This value depends on the time remaining before the pressure reaches the critical low-pressure alarm threshold of 3.0 bar. To ensure system safety and avoid reaching this threshold, a conservative safety margin of 2 days is introduced. Constraints (7) and (8) yield b_n as the minimum between this value and the window length L. Temporal feasibility is governed by constraints (9) and (10): if certain tasks are grouped into a cluster scheduled for day t, the day t must fall within the dynamic interval [a_n, b_n] for each task. Otherwise, the constraints are automatically satisfied.

$$\min \ Obj = \sum_{t=1}^{L+1} cluster_t \tag{1}$$

$$\theta_{nt} \leq cluster_t \quad \forall n \in N \quad \forall t \in [1, L+1] \tag{2}$$

$$\sum_{t=1}^{L+1} \theta_{nt} = 1 \quad \forall n \in N \tag{3}$$

$$\sum_{t=1}^{|N|} \theta_{nt} \leq C_{max}^{GR} \quad \forall t \in [1, L+1] \tag{4}$$

$$a_n \geq t_n - A \quad \forall n \in N \tag{5}$$

$$a_n \geq 0 \quad \forall n \in N \tag{6}$$

$$b_n \leq t_n + (3.2 - 3.0)/\alpha_{i(n)} - 2 \quad \forall n \in \mathrm{N} \tag{7}$$

$$b_n \leq L \quad \forall n \in N \tag{8}$$

$$t\,\theta_{nt} + a_n\,(1 - \theta_{nt}) \geq a_n \quad \forall n \in N \ \forall t \in [1, L+1] \tag{9}$$

$$t\,\theta_{nt} + b_n\,(1 - \theta_{nt}) \leq b_n \quad \forall n \in N \ \forall t \in [1, L+1] \tag{10}$$

4. Experiments

4.1 Maintenance plan generation

The maintenance schedule generation simulates degradation patterns and maintenance events for GIS, serving as input data for subsequent optimization. Each PMEC is initialized with randomly generated parameters (drawn from uniform distributions): initial pressure ((3.2–3.5] bar), daily degradation rate ([0.01–0.05] bar/day), and a number between 3 and 5 for the allowable GRs before RE is required. Pressure decreases daily under normal operation until reaching a critical threshold of 3.2 bar, triggering a GR intervention. GR restores pressure linearly to 3.5 bar within one day. Maintenance events are recorded in the form of a data frame. The parameter values used in this study (shown in Table *2*) are hypothetical and designed to suit the small-scale case. However, they are based on experience drawn from similar real-world situations. These values were carefully chosen to match the experiment size and to clearly demonstrate the model's functionality in a simplified yet representative way.

Parameter	Value (unit)	Parameter	Value (unit)
C_{max}^{CR}	5 (task)	T	30 (day)
L	10 (day)	$\alpha_{i(n)}$	[0.01, 0.05] (bar/day)
A	3 (day)	|P|	10 (unit)
|N|	22 (task)		

Table 2. Table of parameters with units

4.2 Grouping results

Figure 2 illustrates the progressive clustering of maintenance tasks across the planning horizon, with each subfigure representing a successive window execution. The first panel of Figure 2 presents the initial

distribution of tasks across 10 PMECs, showing a temporally dispersed pattern that would traditionally necessitate 22 separate maintenance deployments. This fragmentation represents the baseline scenario where each task would be executed at its originally scheduled time without any strategic grouping. The subsequent panels reveal the sequential operation of the sliding window, with each red-bordered rectangle representing the active window under consideration. The green circles indicate the positions to which the tasks have been adjusted through the clustering process, with the arrows visualizing the direction of these adjustments. The algorithm systematically processes consecutive windows, identifying possible grouping opportunities while respecting the maximum allowable time-shift constraints.

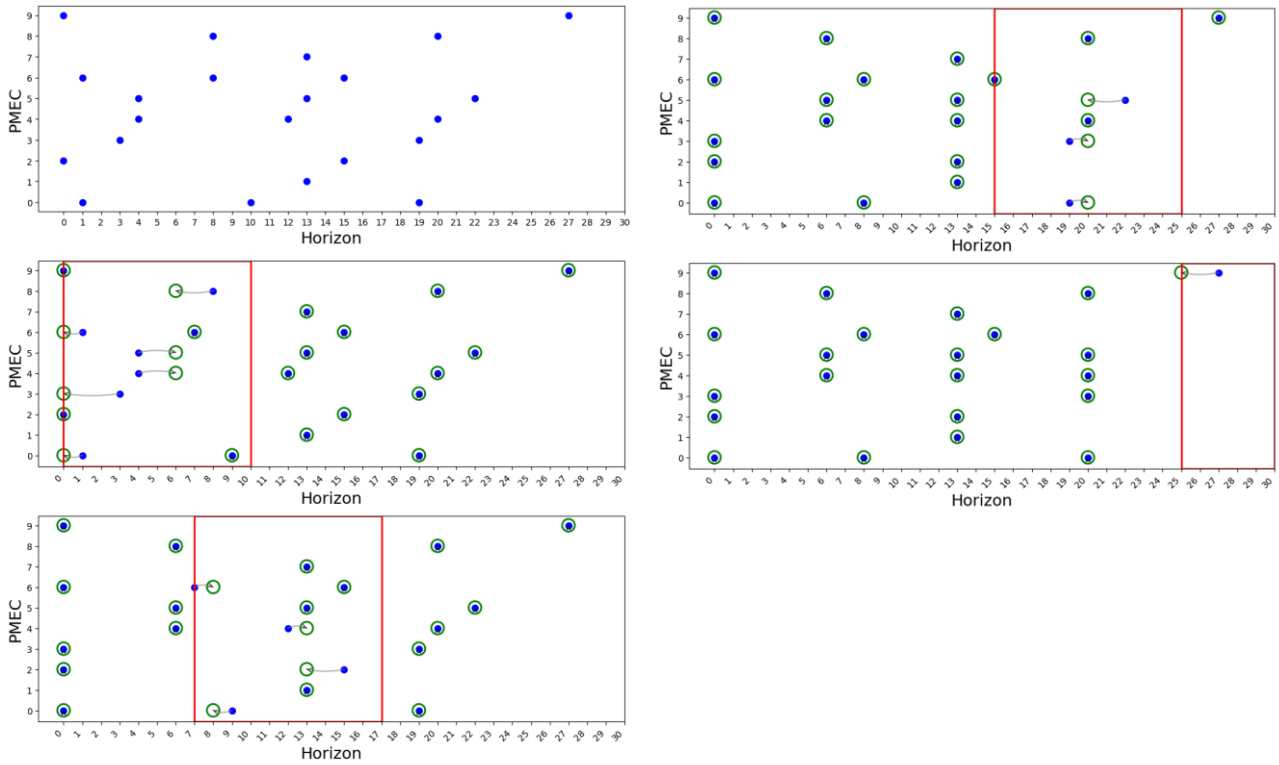


Figure 2. Sliding window grouping results

Figure *3* illustrates the temporal adjustments during clustering, comparing original task schedules to their assigned cluster centers. The algorithm optimizes task grouping by advancing or postponing tasks within a restrained range, adhering to operational constraints. A slight preference for advancing tasks is observed, reflecting a priority for prompt maintenance to ensure system reliability.

Economically, the clustering approach achieves significant cost savings. By consolidating 22 individual GR tasks (each costing 10 units) into 7 maintenance clusters, total expenditure is reduced from 220 units to 70 units—a 68.2% reduction. This efficiency stems from eliminating redundant transportation and setup operations while keeping interventions within acceptable time frames. These results validate the method's practicality for reducing costs in large-scale GIS networks without compromising reliability. The cost reduction discussed here primarily refers to costs associated with transportation and personnel, while the execution cost of the tasks themselves remains unaffected. It should also be noted that the potential increase in costs resulting from a greater number of tasks being executed on the same day due to grouping is currently beyond the scope of our study.

Environmentally, SF_6 gas leakage increased slightly from 42.114 kg to 42.696 kg post-clustering (1.38%). Our analysis across multiple scenarios indicates this is not a consistent trend, suggesting the environmental impact of maintenance clustering is context-dependent rather than systematic. The observed variation likely stems from two key limitations in our current modeling approach: (i) constant leakage rate assumption: our model presumes fixed leakage rates for each PMEC throughout the planning horizon, which may not accurately reflect dynamic degradation patterns in real-world systems. (ii) maintenance interdependency: the current formulation does not account for potential interactions between maintenance types, particularly how the timing of GR operations might influence the scheduling of RE activities, which directly affect leakage rates.

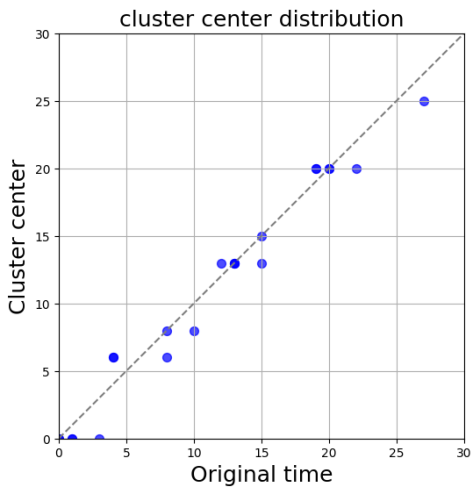


Figure 3. Cluster center vs. original time

5. Conclusion and discussion

This study demonstrates that clustering GIS maintenance tasks using a sliding window algorithm significantly improves operational efficiency, reducing deployment costs by 68.2% while maintaining system reliability. The method consolidates temporally dispersed GR tasks into coordinated interventions, minimizing redundant site visits. Although minor variations in SF_6 leakage were observed post-clustering, these were context-dependent, highlighting the need for dynamic leakage models in future research. The framework offers a scalable solution for optimizing maintenance logistics in large-scale power networks, balancing economic and operational constraints.

The current solution framework may yield multiple equivalent solutions, as it lacks explicit preferences for advancing or delaying maintenance tasks. In practice, certain scheduling behaviors may be preferred, so future work could introduce behavior-specific penalties or preferences to better align with operational priorities. The model's daily maximum capacity can also lead to uneven cluster distributions, with some clusters overloaded and others containing only isolated tasks. Incorporating soft balancing constraints or fairness considerations could improve efficiency and practicality. Furthermore, the framework does not yet account for logistical constraints such as the geographic dispersion of substations and associated travel times. Factoring in these real-world logistical considerations would make the model more realistic and enhance its applicability to large-scale GIS maintenance planning.

Acknowledgements

This work was supported by the French government under the "France 2030" program, as of the SystemX Technological Research Institute. This work was carried out within the JNI3 project at SystemX.

References

Burns, M. &., 2024. Why Gas-Insulated Switchgear (GIS) Substations Are Key to Meeting Accelerating U.S. Power Demand. URL: https://www.tdworld.com/substations/article/55138424/burns-mcdonnell-why-gas-insulated-switchgear-gis-substations-are-key-to-meeting-accelerating-us-power-demand

CIGRE, 2014. CIGRE Position Paper on the Application of SF6 in Transmission and Distribution Networks. IPCC, 2018.

Phillip Widger, A. H., 2018. Evaluation of SF6 Leakage from Gas Insulated Equipment on Electricity Networks in Great Britain. Energies. 2018, p. 11(8):2037.

Wu, W. et al., 2025. Environmental impact-aware and cost-efficient optimal predictive maintenance scheduling for Gas Insulated Substations. s.l., In: *Proceedings of the joint IFAC SSSC 2025, TDS 2025*, and COSY 2025.

Kivanç, İ. (2025). Sustainable maintenance strategies for multi-component systems. Phd Thesis, Eindhoven University of Technology.

Toward Automating xGSPN-mBSPN Model Generation for Scalable Fault Diagnosis in Dynamic Systems

Taofeeq Alabi Badmus*, Darren Prescott** and Rasa Remenyte-Prescott***
**Resilience Engineering Research Group, University of Nottingham, Nottingham, UK & Department of Computer Engineering, Federal University Oye-Ekiti, Ekiti State, Nigeria; email address: taofeeq.badmus@nottingham.ac.uk & taofeeq.badmus@fuoye.edu.ng*
***Resilience Engineering Research Group, University of Nottingham, Nottingham, UK; email address: darren.prescott@nottingham.ac.uk*
****Resilience Engineering Research Group, University of Nottingham, Nottingham, UK; email address: r.remenyte-prescott@nottingham.ac.uk*

Abstract. Efficient fault diagnosis is fundamental to industrial reliability and diagnostic applications, particularly for ensuring system safety and operational performance. While traditional model-based fault detection approaches have proven to be effective, they face scalability challenges due to the manual effort required in model construction. This paper introduces an automated framework, integrating extended Generalised Stochastic Petri Nets (xGSPNs) for system modelling with modified Bayesian Stochastic Petri Nets (mBSPNs) for fault diagnosis. A set of novel algorithms is proposed to enable the automatic generation of the xGSPN model from system specifications and derivation of the mBSPN diagnostic module from the xGSPN representation, including the automated construction of input Conditional Probability Tables (iCPTs), required for diagnostic reasoning. Such automation processes reduce manual effort, improve model accuracy, and enhance adaptability for large-scale and time-varying systems. The effectiveness of the proposed approach is validated using a water tank level control system, demonstrating its capability in detecting and diagnosing single and multiple faults. The findings contribute to advancing hybrid fault detection and diagnostic methodologies, making them more practical for industrial reliability assessment and diagnostics applications.

1. Introduction and Background

Timely and accurate fault detection and diagnosis processes are critical for ensuring the operational reliability and safety of dynamic systems, such as water networks, chemical plants, and aerospace systems. Traditional fault analysis methods, such as Failure Modes and Effects Analysis (FMEA), Fault Trees (FTA), and Reliability Block Diagrams (RBD), are widely used but they can be limited in modelling concurrency, temporal behaviour, and fault propagation under feedback control (Kabir, 2017; Shafiee et al., 2019).

To address these limitations, Petri Nets (PNs) have been adopted for modelling dynamic system behaviour due to their ability to represent discrete events, time-dependent transitions, and component interactions (Kabir & Papadopoulos, 2019). Further enhancements, such as Generalised Stochastic Petri Nets (GSPNs), enable the inclusion of timed transitions and stochastic behaviour (Balbo G., 2007). However, while PNs capture system operations effectively, they lack mechanisms for probabilistic reasoning under uncertainty (Codetta-Raiteri & Portinale, 2014; Weber et al., 2012).

Bayesian Networks (BNs), on the other hand, offer powerful probabilistic inference capabilities and have been integrated with PNs in hybrid approaches, such as Bayesian Petri Nets (BP-Net), Plausible Petri Net (PPN), and Bayesian Stochastic Petri Net (BSPNs) (Andrews & Fecarotti, 2017; Chiachío et al., 2019; Taleb-Berrouane et al., 2020). However, a persistent challenge lies in the manual effort required to construct some of these hybrid diagnostic models. Processes of defining the network structure, specifying dependencies, and parameterising conditional probability tables (CPTs) become prohibitively complex and error-prone for large or evolving systems, hindering scalability and practical adoption (Forner et al., 2013).

Recent advancements have sought to address these limitations via a proposed hybrid PN modelling approach, xGSPN-mBSPN, that combines an extended GSPN (xGSPN) for system modelling with a modified Bayesian Stochastic Petri Net (mBSPN) for fault diagnosis (Badmus, et al., 2023a). The xGSPN formalism introduced richer modelling elements, such as Conditional Reset Transitions (CRTs), Conditional Output Arcs (COAs), and Conditional Output Reset Places (CORPs), to better capture complex control logic, sensor behaviour, and fault propagation in dynamic feedback systems. Concurrently, the mBSPN framework refined diagnostic reasoning by defining specific transition types, independent (ICPT), dependent (DCPT), and

observable (DCPT*) Conditional Probabilistic Transitions (CPTs), to represent prior fault probabilities, causal dependencies, and sensor evidence, respectively, enabling inference via sampling algorithms (Melani et al., 2021; Russell & Norvig, 2010).

Previous work partially automated this process by developing an algorithm to generate input Conditional Probability Tables (iCPTs) automatically from xGSPN sub-module specifications to significantly reduce diagnostic model configuration time (Badmus et al., 2023b). However, a critical bottleneck remained: the structural generation of the mBSPN diagnostic model itself based on the system's operational model to further reduce model construction time and minimise potential human errors still required manual expertise and encoding. While the complete automation pipeline ideally starts from high-level system specifications, the focus of this paper is specifically on automating the crucial and complex transformation from a pre-defined, detailed xGSPN model representation into a ready-to-use mBSPN diagnostic model.

Hence, this paper presents a unified, automated framework designed to bridge this structural generation gap. Given an xGSPN system model described in structured text files, the presented algorithmic pipeline parses the xGSPN model representation, analyses its structure to identify functionally distinct sub-modules (representing components, propagation paths, sensors), automatically constructs the corresponding mBSPN diagnostic structure by mapping operational elements to appropriate diagnostic transitions (ICPT, DCPT, DCPT*) and places (Conditional Probabilistic and Evidence places), and integrates the automated iCPT generation algorithm to parameterise the diagnostic model. This dual automation, encompassing both the diagnostic structure and its probabilistic parameters, distinguishes the approach from prior efforts, which were often focused on only one aspect.

In this study, the entire workflow is implemented within a custom C++ simulation engine supporting model loading, simulation, and Bayesian inference. This targeted automation significantly reduces modelling effort for the diagnostic layer, ensures consistency between operational and diagnostic representations, and enhances model scalability. The framework is validated using a multi-component water tank control system benchmark, demonstrating its effectiveness and showing alignment with results from established tools like HUGIN (HUGIN, 2016). This work advances hybrid Petri Net-Bayesian approaches to fault detection and diagnosis (Melani et al., 2021) by automating a critical step towards making scalable model generation a practical reality.

The remainder of this paper is structured as follows: Section 2 describes the methodology for xGSPN analysis and automated mBSPN generation. Section 3 presents the application of the proposed method to a case study. Section 4 concludes the paper with future directions.

2. Methodology – The Proposed Automated xGSPN-mBSPN Model Framework

This section presents the proposed automation framework for generating fault diagnostic models by transforming an extended Generalised Stochastic Petri Net (xGSPN) into a Modified Bayesian Stochastic Petri Net (mBSPN). The approach comprises four main stages as depicted in Figure 1: parsing the xGSPN model from text-based files, identifying functional sub-modules, generating the diagnostic mBSPN structure, and assigning probabilistic parameters through automated conditional probability table (iCPT) generation.

2.1 *Explanation of the Automated xGSPN-mBSPN Framework*

The automation workflow, presented in Figure 1, illustrates the complete transformation process from a structured xGSPN system description to a fully parameterised mBSPN diagnostic model. The pipeline begins with three input files: `Places.txt`, `Transitions.txt`, and `Arcs.txt`—which together encode the extended Generalised Stochastic Petri Net (xGSPN) model of the target system. These files can be produced manually using pre-defined templates, or automatically generated from model-based design tools that embed state-transition semantics. Examples include transformation from UML/SysML state machines and activity diagrams, or export from platforms such as Simulink/Stateflow. Specifically: `Places.txt` defines the places and their functional types (e.g., component state, sensor output, propagated variables); `Transitions.txt` records transition labels, timing or probabilistic firing types, and control logic roles; and `Arcs.txt` specifies connections and arc types (e.g., inhibitor, reset). This structured representation serves as a tool-independent interface and supports both manual and model-driven workflows, enabling automated model transformation across various application domains.

A C++ parser ingests these files and reconstructs the full xGSPN model in memory. The parser preserves type semantics and recognises extended Petri Net constructs (e.g. Conditional Output Arcs, Conditional Reset Transitions), embedded in the system description. Once the xGSPN model is instantiated, it is passed to the `Analyse_xGSPN()` module for sub-structure identification. This analysis function systematically traverses the model and classifies each transition and its associated places into one of three sub-module types: component, propagation, or sensor. The result is a set of logically segmented sub-modules, each representing a distinct portion of the system's operational or observational behaviour. These are stored in containers (e.g. vectors) for subsequent diagnostic transformation.

From the classified modules, the automation process proceeds along two parallel paths. The first path involves structural transformation via the `Create_mBSPN()` function, which converts each xGSPN sub-module into its corresponding diagnostic element. Specifically, component sub-modules are transformed into Independent Conditional Probabilistic Transitions (ICPTs), propagation sub-modules into Dependent Conditional Probabilistic Transitions (DCPTs), and sensor sub-modules into observable transitions (DCPT*). The output of this step is a diagnostic model structure consisting of interconnected places, transitions, and specialised arcs, such as Conditional Probabilistic Arcs (CPAs) and Evidence Arcs (EAs).

In parallel, the second path focuses on probabilistic parameterisation. The `Create_iCPTs()` function analyses the input-output dependencies of each diagnostic transition and generates the associated Conditional Probability Tables. Both outputs, the diagnostic structure and its parameterisation, are integrated into a single, coherent mBSPN model that supports inference over component states given sensor evidence from the system xGSPN operational model. The final output is a fault diagnostic Petri Net that is fully synthesised from system specifications, eliminating the need for manual fault diagnostic model construction.

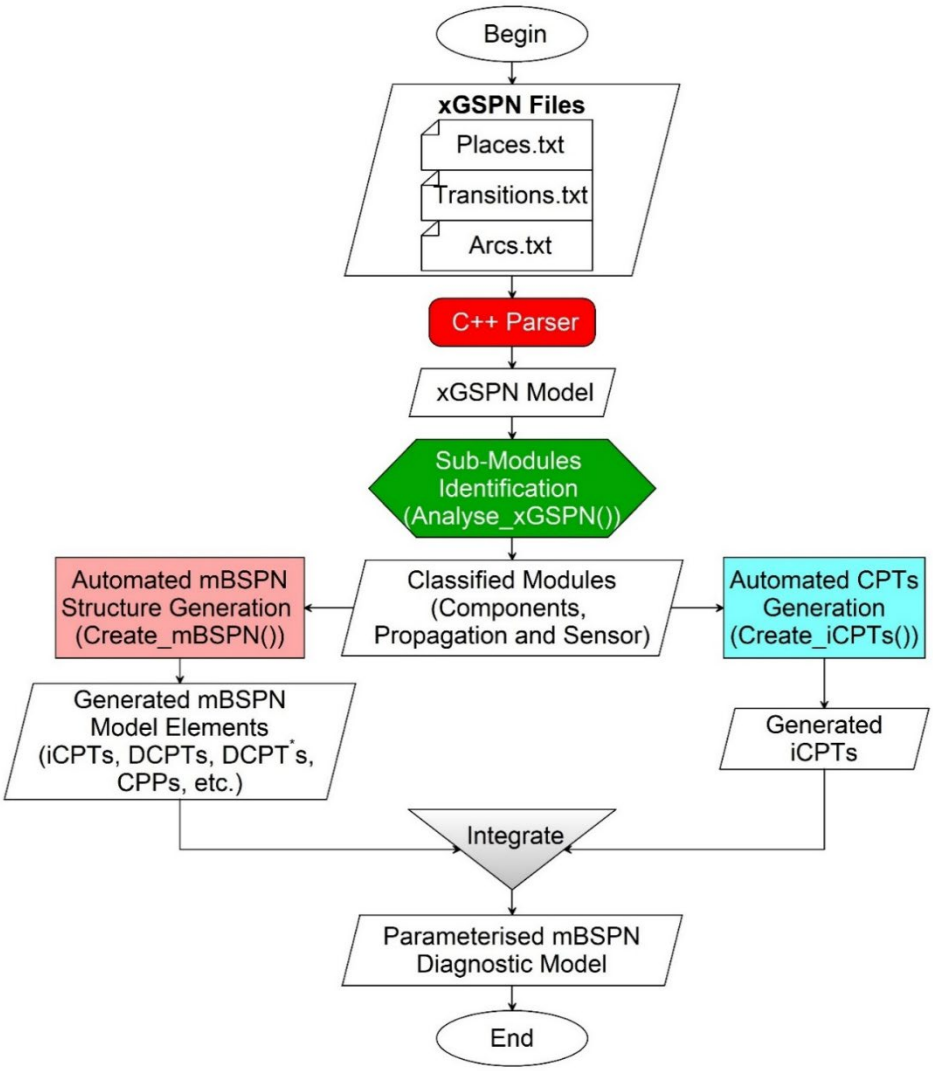


Figure 1: Framework of the automated xGSPN–mBSPN approach

2.2 *The Automated Transformation Algorithms with an Illustrative Mapping Example*

Building on the four-stage framework introduced in Section 2 and illustrated in Figure 1, this section details the core automation algorithms responsible for transforming the structured xGSPN system model into a fully parameterised mBSPN diagnostic model. Each stage is encapsulated in a dedicated algorithm to support scalability, modularity, and probabilistic inference.

To improve the interpretability of the algorithmic logic, a simplified illustrative example is also provided in Figure 2. This example is based on a segment of a tank-filling system comprising three interacting components: a valve (V), a pipe (P), and a flow sensor (FS). These are encoded using structured xGSPN sub-modules, and their diagnostic equivalents are generated via the proposed automation framework. The mapping from system structure to xGSPN and mBSPN representations is summarised in Table 1 and referenced throughout this section to support clarity.

Algorithm 1 (*Generate_mBSPN_Model*) coordinates the entire automation process. It accepts as input a structured xGSPN described by `Places.txt`, `Transitions.txt`, and `Arcs.txt`, alongside a diagnostic sampling mode `InferenceCode ∈ {1, 2, 3}`, denoting forward, rejection, or likelihood weighting sampling, respectively. As shown in **line 2**, the function `Analyse_xGSPN()` decomposes the system model into three functional sub-module vectors: components (C), sensors (S), and process propagation modules (P). While the pseudocode of `Analyse_xGSPN()` is omitted for brevity, its logic is based on parsing Petri Net elements and identifying propagation patterns via arc types and place semantics. In the illustrative case shown in Table 1, the valve and pipe are identified as component modules, while the sensor is classified as an observable monitoring component.

In **line 4** of Algorithm 1, **Algorithm 2 (Create_mBSPN)** is invoked to construct the diagnostic model structure. Each sub-module is translated into a conditional probabilistic transition: components are mapped to ICPTs (e.g., the transition *ti* labelled *V* for valve *V* in Table 1), propagation modules to DCPTs, and sensor modules to observable DCPT*s (e.g., the transition labelled *FS* in Table 1). Each diagnostic transition is connected to places using appropriate arc types defined in the mBSPN formalism (Badmus et al., 2023a), including: input arcs (A^d_{INA}), causal arcs (A^d_{CA}), evidence arcs (A^d_{EA}), and conditional probabilistic arcs (A^d_{CPA}). **Lines 6–32** of **Algorithm 2** implement this logic. For instance, line 7 handles component modules by generating ICPT transitions and linking them to their corresponding probabilistic places (CPPs) via CPA arcs (**line 10**). This is illustrated in the mapping of pipe P in Table 1, row 1. Similarly, propagation and sensor sub-modules are handled in **lines 12–21** and **lines 23–32**, respectively, where causal (CA), conditional probabilistic (CPA) and evidence arcs (EA) are generated to preserve interdependencies within the fault inference structure. As shown in Table 1, the pipe's operational state (row 3) is influenced by the working/failure status of both upstream components, while the sensor (row 4) receives evidence arcs from CORP places and causal arcs from flow state CPPs.

The structural mapping ensures that each transition maintains a minimum of two conditional probabilistic places (CPPs), reflecting the number of component/process/propagated variable states within the corresponding xGSPN sub-module. Moreover, causal arcs of DCPT and DCPT* transitions are restricted to originate from the CPPs of upstream ICPT transitions upon which they depend. Similarly, evidence arcs for DCPT* transitions are derived from CORP places in the xGSPN model (e.g., *pkCORPofFS* in Table 1), thereby maintaining conformance with the diagnostic interconnection rules, defined in the xGSPN-mBSPN formalism (Badmus et al., 2023a).

In **line 6** of Algorithm 1, the function `Create_iCPTs(C, S, P)` is invoked to generate input Conditional Probability Tables (iCPTs) for all diagnostic transitions. The detailed pseudocode and validation of this function are available in prior publication (Badmus et al., 2023b). The function uses the sub-module structure and interconnections to produce consistent probabilistic parameters required by an inference sampling method during fault diagnosis.

In the final stage (**lines 8–10** of Algorithm 1), each diagnostic transition is assigned its iCPT and the inference sampling method. The final output, returned in **line 12**, is a fully parameterised mBSPN diagnostic model that integrates structural and probabilistic components for diagnostic inference.

It is important to note that several low-level helper functions, such as `CreateDiagnosticTransition`, `ExtractComponentPlaces`, and `ConnectCPA`, are abstracted for clarity. These routines are responsible for generating structurally valid interconnections between places and transitions, adhering to the rules of the mBSPN formalism while maintaining algorithmic modularity.

Together, the three algorithms enable the automated and scalable construction of diagnostic Petri Net models from formalised system specifications. This approach enhances model consistency and significantly reduces manual effort in diagnosing faults within complex, dynamic systems.

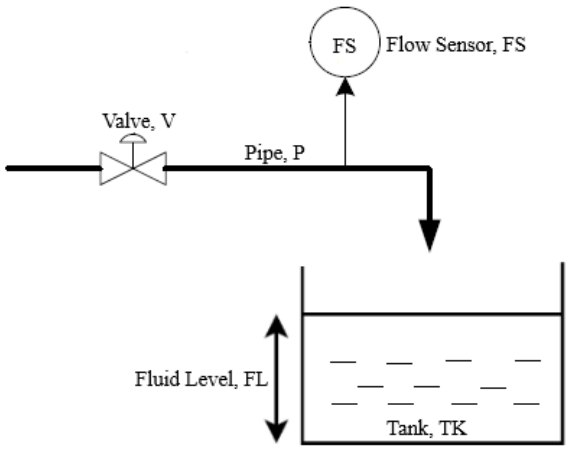


Figure 2. A simplified diagram of a tank-filling system comprising a control valve (V), pipe (P), and flow sensor (FS).

Algorithm 1: Generate_mBSPN_Model: Main Automation Algorithm for Parameterised mBSPN from xGSPN

Input: Structured xGSPN model created from `Places.txt`, `Transitions.txt`, and `Arcs.txt` input files, $InferenceCode \in \{1, 2, 3\}$ `// 1 = forward, 2 = rejection, 3 = likelihood weighting sampling`

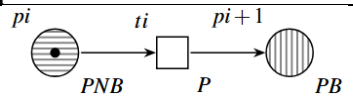
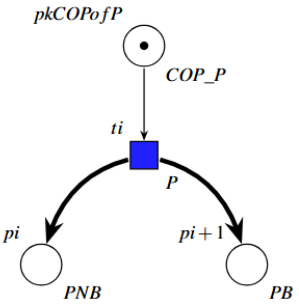
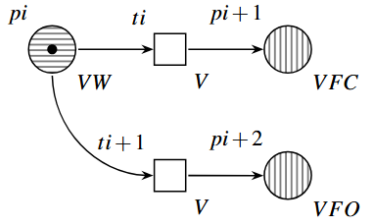
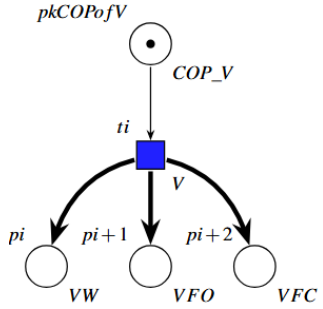
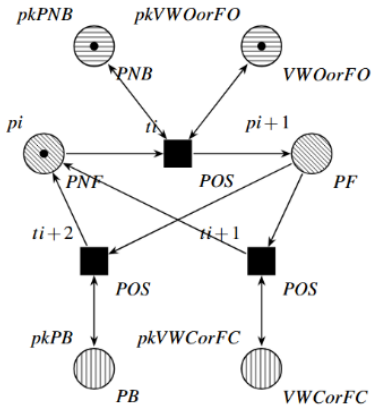
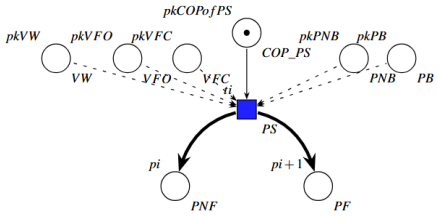
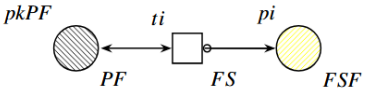
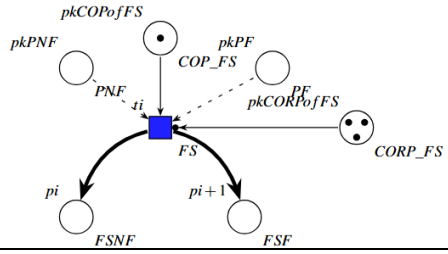
Output: Parameterised mBSPN: $\{P^d, T^d_{CPT}, Type_{t^d_{CPT}}, F_{cpt}, A^d\}$ where: `// ` P^d, T^d_{CPT}, `and` A^d `are places, transitions, and arcs in the diagnostic model.` $Type_{t^d_{CPT}}$ `and` F_{cpt} `are transition types and iCPTs assignment functions`

Step 1: Analyse the xGSPN structure and extract sub-modules:
$\{C, S, P\} \leftarrow$ `Analyse_xGSPN`$(xGSPN)$ `// Vectors of sub-modules: {C, S, P} = {Components, Sensors, Propagation}`
Step 2: Create diagnostic model transitions, places, and arcs:
$\{P^d, T^d_{CPT}, A^d, Type_{t^d_{CPT}}\} \leftarrow$ `Create_mBSPN`(C, S, P)
Step 3: Generate iCPTs for the T^d_{CPT}**:**
$F_{cpt} \leftarrow$ `Create_iCPTs`(C, S, P)
Step 4: Assign CPTs to each diagnostic transition:
foreach $t^d_{CPT} \in T^d_{CPT}$ **do**
  $t^d_{CPT}.assignCPT(F_{cpt}[t^d_{CPT}])$
  $t^d_{CPT}.assignInferenceCode(InferenceCode)$
Step 5: Return the parameterised diagnostic model:
return $\{P^d, T^d_{CPT}, Type_{t^d_{CPT}}, F_{cpt}, A^d\}$ `// ` P^d: `diagnostic places,` T^d_{CPT}: `CPT transitions,` $Type_{t^d_{CPT}}$: `transition types,` F_{cpt}: `CPT assignment,` A^d: `diagnostic arcs`

Algorithm 2: Create_mBSPN: Diagnostic Structure Generation from Sub-module Vectors

Input: Vectors of classified xGSPN sub-modules: C, S, P
Output: $\{P^d, T^d_{CPT}, A^d, Type_{t^d_{CPT}}\}$
`// ` P^d: `diagnostic places,` T^d_{CPT}: `transitions,` A^d: `diagnostic arcs,` $Type_{t^d_{CPT}}$: `transition type assignment map`
Initialise data structures:
$P^d \leftarrow \emptyset$, $T^d_{CPT} \leftarrow \emptyset$, $A^d \leftarrow \emptyset$, $Type_{t^d_{CPT}} \leftarrow \emptyset$
Process component sub-modules (ICPTs):
foreach $c \in C$ **do**
  $t_{ICPT} \leftarrow$ `CreateDiagnosticTransition`(c)
  $T^d_{CPT} \leftarrow T^d_{CPT} \cup \{t_{ICPT}\}$
  $Type_{t^d_{CPT}}[t_{ICPT}] \leftarrow$ `ICPT`
  `// Map component states to conditional probabilistic places`
  $P^d_{CPP} \leftarrow$ `ExtractComponentPlaces`(c)
  $P^d \leftarrow P^d \cup P^d_{CPP}$
  `// Create output CPA arcs from` t_{ICPT}
  $A^d_{CPA} \leftarrow$ `ConnectCPA`(t_{ICPT}, P^d_{CPP})
  $A^d \leftarrow A^d \cup A^d_{CPA}$
Process propagation sub-modules (DCPTs):
foreach $p \in P$ **do**
  $t_{DCPT} \leftarrow$ `CreateDiagnosticTransition`(p)
  $T^d_{CPT} \leftarrow T^d_{CPT} \cup \{t_{DCPT}\}$
  $Type_{t^d_{CPT}}[t_{DCPT}] \leftarrow$ `DCPT`
  `// Extract input causal places and output places`
  $P^d_{CPP}, P^d_{CP} \leftarrow$ `ExtractPropagationPlaces`(p)
  $P^d \leftarrow P^d \cup P^d_{CPP} \cup P^d_{CP}$
  $A^d_{CA} \leftarrow$ `ConnectCA`(P^d_{CPP}, t_{DCPT})
  $A^d_{CPA} \leftarrow$ `ConnectCPA`(t_{DCPT}, P^d_{CP})
  $A^d \leftarrow A^d \cup A^d_{CA} \cup A^d_{CPA}$
Process sensor sub-modules (DCPT*):
foreach $s \in S$ **do**
  $t_{DCPT*} \leftarrow$ `CreateDiagnosticTransition`(s)
  $T^d_{CPT} \leftarrow T^d_{CPT} \cup \{t_{DCPT*}\}$
  $Type_{t^d_{CPT}}[t_{DCPT*}] \leftarrow$ `DCPT*`
  `// Extract causal and evidence (observable) places`
  $P^d_{CPP}, P^d_{EP} \leftarrow$ `ExtractSensorPlaces`(s)
  $P^d \leftarrow P^d \cup P^d_{CPP} \cup P^d_{CP} \cup P^d_{EP}$
  $A^d_{CA} \leftarrow$ `ConnectCA`(P^d_{CPP}, t_{DCPT*})
  $A^d_{EA} \leftarrow$ `ConnectEA`(P^d_{EP}, t_{DCPT*})
  $A^d_{CPA} \leftarrow$ `ConnectCPA`(t_{DCPT*}, P^d_{CPP})
  $A^d \leftarrow A^d \cup A^d_{CA} \cup A^d_{EA} \cup A^d_{CPA}$
return $\{P^d, T^d_{CPT}, A^d, Type_{t^d_{CPT}}\}$

Table 1. Diagnostic transformation of xGSPN sub-modules derived from the system configuration in Figure 2. Each row shows the mapping of an individual component, propagation, or sensor sub-module into its corresponding mBSPN diagnostic representation.

System Configuration		xGSPN Sub-module Representation	mBSPN Sub-module Representation
Component	Pipe P		
	Valve V		
Propagation	Pipe P operational status		
Observable Monitoring System (Sensor)	Flow Sensor FS operational status		

Note:

- *pkPNB* is a **causal place** corresponding to the component state place *PNB* of pipe *P*.
- *pkCOPofP* and *pkCORPofFS* are, respectively, the conditional output place of pipe *P* and the conditional output reset place of sensor *FS*, as used in the fault detection module of the xGSPN system.
- Full definitions of these place types and their roles within the xGSPN–mBSPN formalism are detailed in Badmus et al. (2023a).

3. Application of the Proposed Automated Petri Net Framework

To evaluate the effectiveness of the proposed automation framework, it was applied to a dynamic water tank level control system used in prior research (Badmus et al., 2023a; Badmus et al., 2023b). This system

comprises multiple components, including pipes, level sensors, a controller, and an outlet valve, all interconnected within feedback control loops.

The xGSPN representation of the system was modelled in structured text format (`Places.txt`, `Transitions.txt`, and `Arcs.txt`) and served as input to the automated transformation pipeline implemented in a custom C++ tool. The tool was executed on a Windows 10 (64-bit) machine equipped with an Intel(R) Core(TM) i3 processor (3.60 GHz) and 8 GB RAM.

The transformation process comprised three main steps: parsing and sub-module analysis, automated mBSPN structure generation, and conditional probability table assignment. The input xGSPN comprised 128 places, 124 transitions, and 860 arcs, while the resulting mBSPN model included 85 places, 36 conditional probabilistic transitions, and 232 conditional output arcs, excluding evidence and trigger places and arcs, which serve as interface points to the xGSPN system model (Badmus et al., 2023a). Although the input xGSPN model contains more elements overall, the `Analyse_xGSPN()` function executed in approximately 60 seconds compared to the 100 seconds taken by `Create_mBSPN()` algorithm. This timing difference arises not from the quantity of model elements processed, but from the nature of the operations involved. The analysis step primarily involves structural parsing and pattern-based classification into component, propagation, or sensor sub-modules. In contrast, the creation phase requires instantiating probabilistic transitions, systematically generating interconnected places and arcs, and enforcing formal mapping rules—each of which introduces additional computational overhead. Thus, the runtime of `Create_mBSPN()` reflects the complexity of rule-based transformation and resolution of sub-module dependencies, rather than the total number of elements alone. For parameterisation, the previously published `Create_iCPTs()` algorithm (Badmus et al., 2023b) was integrated, generating 36 iCPTs, each in an average time of 28 ms.

To assess diagnostic performance, three fault scenarios (two single faults and one multiple fault) were simulated using 30,000 runs per test case. Table 2 presents the results, including prior fault probabilities and posterior probabilities obtained from the automated mBSPN model, compared with values from the previously semi-automated implementation (Badmus et al., 2023a).

Table 2. Posterior Probabilities for Single and Multiple Fault Scenarios in the Water Tank Level Control System

Fault Scenarios	Component Name	Failure Mode	Prior Probability	Posterior Probability	
				Previous Work (Badmus et al., 2023a)	This Study
1	Pipe P2	Blocked (B)	0.1	0.1187	0.1206
2	Sensor S1	Failed High (FH)	0.0475	0.0486	0.0484
3	Pipe P3	P3B	0.1	0.3248	0.3056
	Sensor S1	Failed Low (FL)	0.0475	0.0720	0.0490
	Sensor S2	S2FL	0.0475	0.0160	0.0481

The results demonstrate that the diagnostic probabilities generated from the automated mBSPN model align closely with those reported in the previous study. Minor deviations in posterior estimates are attributed to the refined diagnostic structure generated by the automated `Create_mBSPN()` algorithm, which ensures consistent mapping of conditional probabilistic transitions and systematic allocation of conditional probabilistic places (CPPs). Notably, the new framework also resolved previously missed diagnoses. In Scenario 3, the failure of Sensor S2 (S2FL) was not successfully detected in the earlier approach, as its posterior probability (0.0160) is less than its prior (0.0475). In contrast, the automated framework yielded a posterior of 0.0481, exceeding the prior, thereby confirming the presence of the fault. This improvement can be attributed to the enhanced structural consistency introduced by the automation. Although the manually developed diagnostic model was thorough, small inconsistencies in constructing dependencies and probabilistic mappings could inadvertently compromise inference results, particularly for hidden faults. The Create_mBSPN() algorithm systematically enforces the xGSPN–mBSPN formalism, eliminating such inconsistencies and ensuring correct propagation of evidence from system operation to the fault diagnostic module. As a result, the automated structure supports more accurate evidence flow, thereby enhancing inference reliability.

In addition to accuracy enhancements, the proposed framework delivers significant benefits in modelling efficiency and consistency. The reduction in manual effort, coupled with the automated handling of sub-module interdependencies and seamless integration of Bayesian parameterisation, contributes to a scalable and reproducible diagnostic process. These outcomes substantiate the practical applicability of the xGSPN–mBSPN automation pipeline for robust fault diagnosis in dynamic, feedback-driven systems.

4. Conclusion and Future Work

This paper has presented a structured automation framework for transforming extended Generalised Stochastic Petri Net (xGSPN) models into parameterised Modified Bayesian Stochastic Petri Net (mBSPN) diagnostic models. The approach, underpinned by a sequence of modular algorithms: `Analyse_xGSPN()`, `Create_mBSPN()`, and the previously established `Create_iCPTs()`, effectively eliminates the need for manual construction of diagnostic transitions, places, and conditional probability tables. The resulting models are structurally consistent and adaptable to a wide range of system modelling and fault diagnostic scenarios.

Empirical validation using a water tank level control system confirmed the effectiveness of the proposed method in diagnosing both single and multiple faults. The diagnostic outputs were consistent with previously published results, and the new approach successfully resolved cases previously undetected. Model construction time was significantly reduced, with diagnostic structures generated in under two minutes and iCPTs computed within milliseconds per diagnostic transition. These findings highlight the framework's potential for improving efficiency and scalability in model-based fault diagnosis.

The proposed automation has important implications for industrial reliability and diagnostic applications. By reducing the effort and expertise required to build diagnostic models, the framework opens the door to wider adoption of formal model-based diagnostics in complex systems. It supports more proactive maintenance planning and increases system resilience through timely fault detection.

Future work will focus on enhancing the usability and extensibility of the framework. Planned developments include applying the approach to more complex case studies, including multi-unit systems and interdependent components, such as oil and gas systems. Further, adaptive refinement of iCPTs through machine learning techniques will be explored to continuously improve diagnostic precision based on operational feedback.

Automating diagnostic model generation represents a significant step toward more intelligent and scalable maintenance solutions. With continued refinement, the proposed xGSPN–mBSPN pipeline has the potential to play a transformative role in the deployment of fault diagnosis capabilities across a broad range of dynamic and safety-critical systems.

Acknowledgement

The authors would like to thank Petroleum Technology Development Fund (PTDF), Abuja, Nigeria, for funding and supporting Taofeeq Alabi Badmus PhD research through the fund Overseas Scholarship Scheme (OSS) programme [award number P6797054741222445].

References

Andrews, J., & Fecarotti, C. (2017). System design and maintenance modelling for safety in extended life operation. *Reliability Engineering and System Safety*, *163*, 95–108. https://doi.org/10.1016/j.ress.2017.01.024

Badmus, T. A., Prescott, D., & Remenyte-Prescott, R. (2023b). Improving the Modelling Capability of an Integrated Fault Detection and Diagnostic Petri Net Methodology for Dynamic Systems. *IMA International Conference on Modelling in Industrial Maintenance and Reliability*, 1–7. https://doi.org/https://doi.org/10.19124/ima.2023.01.19

Badmus, T. A., Remenyte-Prescott, R., & Prescott, D. (2023a). A novel fault detection and diagnostic Petri net methodology for dynamic systems. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*. https://doi.org/10.1177/1748006X231212539

Balbo G. (2007). Introduction to Generalized Stochastic Petri Nets. In M. Bernardo & J. Hillston (Eds.), *Formal Methods for Performance Evaluation* (In:, Vol. 4486, pp. 83–131). Springer.

Chiachío, M., Chiachío, J., Prescott, D., & Andrews, J. (2019). Plausible Petri nets as self-adaptive expert systems: A tool for infrastructure asset monitoring. *Computer-Aided Civil and Infrastructure Engineering*, *34*(4), 281–298. https://doi.org/10.1111/mice.12427

Codetta-Raiteri, D., & Portinale, L. (2014). Approaching dynamic reliability with predictive and diagnostic purposes by exploiting dynamic Bayesian networks. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, *228*(5), 488–503. https://doi.org/10.1177/1748006X14533958

Forner, L., Kumar, V. S., & Kinshuk. (2013). Assessing design of online courses using bayesian belief networks. *Proceedings - 2013 IEEE 5th International Conference on Technology for Education, T4E 2013*, 36–42. https://doi.org/10.1109/T4E.2013.17

HUGIN. (2016). *HUGIN researcher version 8.0*. https://www.hugin.com/hugin-developerhugin-researcher/

Kabir, S. (2017). An overview of fault tree analysis and its application in model based dependability analysis. *Expert Systems with Applications*, *77*, 114–135. https://doi.org/10.1016/J.ESWA.2017.01.058

Kabir, S., & Papadopoulos, Y. (2019). Applications of Bayesian networks and Petri nets in safety, reliability, and risk assessments: A review. In *Safety Science* (Vol. 115, pp. 154–175). Elsevier B.V. https://doi.org/10.1016/j.ssci.2019.02.009

Melani, A. H. de A., Michalski, M. A. de C., da Silva, R. F., & de Souza, G. F. M. (2021). A framework to automate fault detection and diagnosis based on moving window principal component analysis and Bayesian network. *Reliability Engineering and System Safety*, *215*. https://doi.org/10.1016/j.ress.2021.107837

Russell, S. J., & Norvig, P. (2010). *Artificial Intelligence A Modern Approach* (M. J. Horton, M. Hirsch, T. Dunkelberger, M. Haggerty, A. Michael, M. Lou Nohr, G. Dulles, & D. Sandin, Eds.; Third Edition). Pearson Education, Inc.

Shafiee, M., Enjema, E., & Kolios, A. (2019). An integrated FTA-FMEA model for risk analysis of engineering systems: A case study of subsea blowout preventers. *Applied Sciences (Switzerland)*, *9*(6). https://doi.org/10.3390/app9061192

Taleb-Berrouane, M., Khan, F., & Amyotte, P. (2020). Bayesian Stochastic Petri Nets (BSPN) - A new modelling tool for dynamic safety and reliability analysis. *Reliability Engineering and System Safety*, *193*. https://doi.org/10.1016/j.ress.2019.106587

Weber, P., Medina-Oliva, G., Simon, C., & Iung, B. (2012). Overview on Bayesian networks applications for dependability, risk analysis and maintenance areas. *Engineering Applications of Artificial Intelligence*, *25*(4), 671–682. https://doi.org/10.1016/j.engappai.2010.06.002

Predicting the remaining life of lithium-ion batteries: a frugal data-based approach

A. Gaignard*, F. Cazaurang** and C. Farges***

*Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400, Talence, France: antoine.gaignard@ims-bordeaux.fr

**Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400, Talence, France: franck.cazaurang@ims-bordeaux.fr

***Univ. Bordeaux, CNRS, Bordeaux INP, IMS, UMR 5218, F-33400, Talence, France: christophe.farges@ims-bordeaux.fr

Abstract Predictive maintenance aims to anticipate the end-of-life of components, thereby optimising human intervention and use of parts, but requires constant monitoring. Calculating the Remaining Useful Life (RUL), which estimates the time remaining for a system to operate satisfactorily, is a crucial stage in predictive maintenance. Classically, there are three approaches for calculating RUL: with models, with data or with hybrid methods. Methods using data, whether statistical or using neural networks, often require very large quantities of data. In this paper, a frugal approach in terms of the data required is proposed to calculate the RUL of lithium-ion batteries. This method use a polynomial as an approximation of the capacity over the cycles, whose coefficients are obtained through least mean squares optimization under linear constraints. With an average prediction horizon of 30 % remaining lifetime, this method is relevant in terms of computational complexity and required quantity of data.

1. Introduction

There are three main approaches to maintenance: curative, preventive and predictive (Cheng, Azarian, and Pecht 2010). Curative maintenance is carried out after a failure, but entails the risk of unexpected stoppages. Preventive maintenance is carried out at fixed intervals, reducing the risk of incidents, but can lead to sub-optimal use of parts. Predictive maintenance anticipates the end of life of components, optimising human intervention and use of parts, but requires constant monitoring (Greitzer et al. 2001, Jardine, Lin, and Banjevic 2006). Calculating the RUL, which estimates the time remaining for a system to operate satisfactorily, is a crucial step in predictive maintenance.

Different approaches can be used to obtain the RUL, such as data-based methods, for instance statistical techniques and neural networks (Chakraborty et al. 1992, J. Choi et al. 2010), model-based methods (Christer, W. Wang, and Sharp 1997, Sikorska, Hodkiewicz, and Ma 2011) or hybrid approaches combining both (Li, Wilson Wang, and Ismail 2013, Liao and Köttig 2014).

In this paper, on the calculation of the RUL of a lithium-ion battery is addressed. A model-free approach is proposed. At regular intervals, the capacity value is collected and then used to optimise the coefficients of a polynomial function that predicts the future values of the capacity. The RUL is then the time remaining before the battery capacity reaches a certain threshold, usually 70 % of the initial capacity (Canals Casals et al. 2019).

2. Impact of ageing on battery capacity

2.1. *Presentation of the database*

The experimental data used in this paper comes from the database of the CALCE research team at the A. James Clark School, University of Maryland, USA. Specifically, the cycling tests for the *CS2-33*, *CS2-34*, *CS2-35*, *CS2-36*, *CS2-37* and *CS2-38* batteries were used (Reis et al. 2021, He et al. 2011). They are part of a batch of CS2 type batteries, which are lithium-ion batteries with a cobalt oxide cathode ($LiCoO_2$). They have a nominal capacity of 1.1 A h and are of prismatic geometry. Batteries 33 and 34 were cycled with a constant current of 0.55 A (0.5 C) and batteries 35 to 38 at 1.1 A (1 C).

To compute the RUL of the battery, the usual measure is the battery capacity. It can be calculated using the integral of the current delivered during the battery discharge phase (also known as Coulomb counting). In the

data set considered, each discharge and charge cycle is complete, from 0 to 100 % State Of Charge (SOC). By convention, a SOC of 100 % corresponds to the maximum energy that the battery can store in its current state, even if this maximum is less than the initial capacity.

The curve showing the evolution of the capacity over the cycles is drawn in figure 1 by taking the capacity at each cycle. The curves on the left plot represent the values of the capacity measured at the end of each cycle for the batteries 33, 34 and 35 while the curves on the right plot represent the capacity of the batteries 36, 37 and 38.

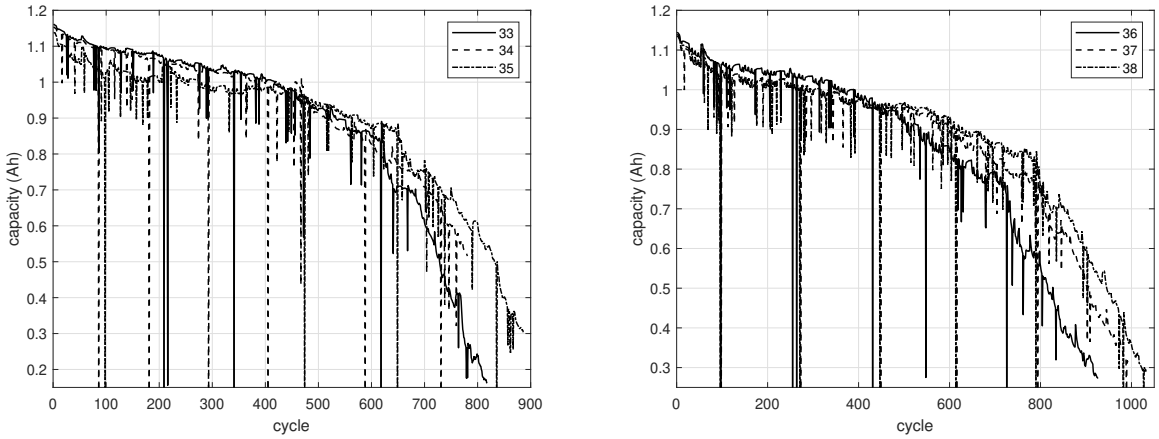


Figure 1. Raw measured capacity of CS2 batteries from CALCE

The raw curves exhibit a lot of outliers, which correspond to large peaks downwards and then upwards stemming from the collection of the datafiles. To remove these outliers, the data is pre-filtered. Observation of the data shows that there are never two points to be filtered in a row. It is therefore proposed to set the criterion that if the capacity decreases by more than a threshold set at 0.1 A in a single cycle, i.e. just under 10 % of the initial capacity, then the data point is an outlier and must be filtered. Removed points are replaced by the average capacity between the cycle before and the cycle after. The heuristic choice of 0.1 A for the threshold allows all false points to be removed without eliminating any true point. Using the average implies that the drop in capacity is locally linear, which is a legitimate approximation since the local variation in capacity is small. By filtering out the outliers and replacing them with consistent values as described above, we obtain the curves of the figure 2. It is from these capacity curves that the RUL is calculated afterwards.

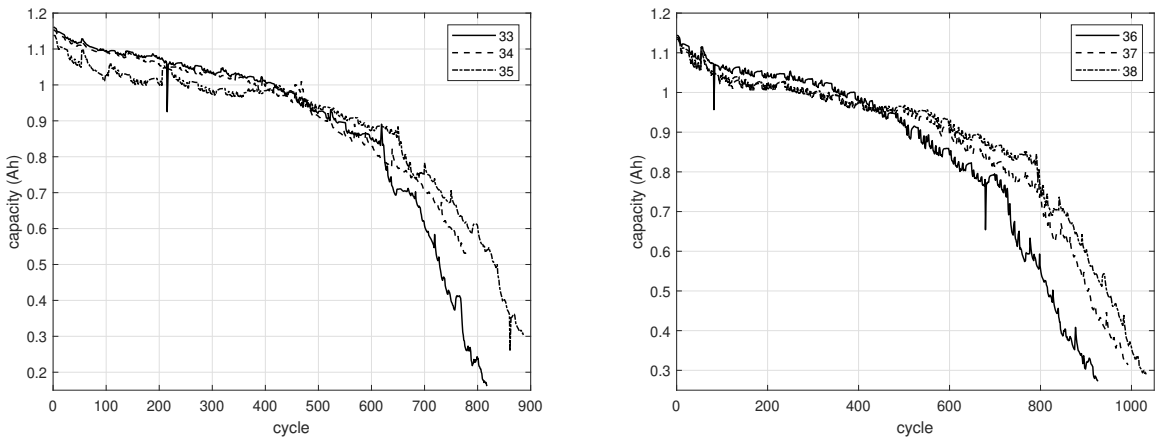


Figure 2. Filtered capacity of CS2 batteries from CALCE

2.2. *Evolution of the capacity over the cycles*

Although the batteries come from the same batch, there is a wide dispersion of the capacity between the batteries in the final cycles, as depicted in the figure 2. Despite the dispersion, the 6 curves exhibit a same behaviour

regarding their evolution. Physically, the formation of an initial layer of Solid Electrolyte Interface (SEI) causes a rapid drop in capacity during the first few cycles. A moderate decline in capacity is then observed. This is due to the increase in this SEI layer (McArthur, Trussler, and Dahn 2011) as well as other phenomena, but of lesser importance in order of magnitude, which are described for example in Atalay et al. 2020. At the end of life, the lithium plating phenomenon leads to a very rapid fall in battery capacity. It is thus proposed in this paper to use a third order polynomial to account for these three phases in the life of the battery.

Figure 3 shows the actual measured capacity of the battery *CS2-33* and a third-order polynomial of the form

$$C_{\text{poly}}(N_{cycle}) = \frac{a}{3} \cdot N_{\text{cycle}}^3 + \frac{b}{2} \cdot N_{\text{cycle}}^2 + c \cdot N_{\text{cycle}} + d \tag{1}$$

which approximates it.

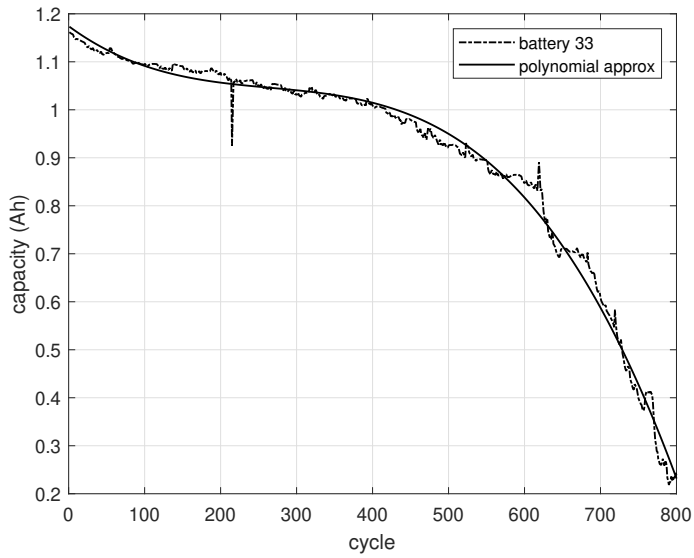


Figure 3. CS2-33 battery capacity and third-order polynomial approximation

Figure 3 demonstrates that the choice of a polynomial is appropriate to describe the evolution of capacity as a function of the number of cycles. A polynomial of order 3 allows the derivative function to be calculated explicitly and simply:

$$\frac{\mathrm{d}C_{\text{poly}}}{\mathrm{d}N_{\text{cycle}}} = a \cdot N_{\text{cycle}}^2 + b \cdot N_{\text{cycle}} + c, \tag{2}$$

as well as the second derivative, which cancels out at the inflection point:

$$N_{\text{I}} = \frac{-b}{2a}. \tag{3}$$

Equations (2) and (3) are used to set linear constraints during the optimisation of the polynomial coefficients in the next section.

3. Prediction of the battery capacity and RUL estimation

3.1. *Optimisation of the polynomial coefficients*

The proposed methodology for calculating the RUL is based on a prediction of battery capacity using a polynomial approximation whose coefficients are optimised at regular intervervals. The coefficients of the polynomial, θ, are the solutions of the following optimisation problem minimising the 2-norm between the measurements and a linearly constrained polynomial:

$$\min_{\theta} \qquad \frac{1}{2}||M \cdot \theta - V_{\text{mes}}||_2^2 \tag{4}$$

subject to:

$$\begin{bmatrix} -2N_{\text{I}}^{\min} & -1 & 0 & 0 \\ 2N_{\text{I}}^{\max} & 1 & 0 & 0 \end{bmatrix} \cdot \theta \leq \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{5}$$

$$\begin{bmatrix} N_1^2 & N_1 & 1 & 0 \\ N_2^2 & N_2 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ N_{\text{total}}^2 & N_{\text{total}} & 1 & 0 \end{bmatrix} \cdot \theta \leq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{6}$$

where

$$M = \begin{bmatrix} 0 & 0 & 0 & 1 \\ \frac{1}{3} & \frac{1}{2} & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots \\ \frac{N^3}{3} & \frac{N^2}{2} & N & 1 \end{bmatrix}, \; V_{\text{mes}} = \begin{bmatrix} C_{\text{mes}}(0) \\ \vdots \\ C_{\text{mes}}(N) \end{bmatrix}, \; \theta = \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} \tag{7}$$

with $C_{\text{mes}}(0)$ corresponding to the measured initial capacity and $C_{\text{mes}}(N)$ corresponding to the capacity measured at cycle N, from which the RUL is estimated.

The inequality (5) avoids the inflexion point to be placed too close or too far from the beginning of the curve:

$$N_{\text{I}}^{\min} \leq N_{\text{I}} \leq N_{\text{I}}^{\max} \tag{8}$$

here $N_{\text{I}}^{\min} = 150$ and $N_{\text{I}}^{\max} = 600$.

The inequality (6) imposes that the capacity is decreasing for all cycles:

$$\frac{\mathrm{d}C_{\text{poly}}}{\mathrm{d}N_{\text{cycle}}} \leq 0. \tag{9}$$

Let N_{final} be the solution to the equation

$$\frac{a}{3}.N_{\text{cycle}}^3 + \frac{b}{2}.N_{\text{cycle}}^2 + c.N_{\text{cycle}} + d = 0.7 \cdot C_{\text{mes}}(0) \tag{10}$$

and N the current cycle. Then the remaining number of life cycles is:

$$RUL(N) = N_{\text{final}} - N. \tag{11}$$

This optimisation problem can be solved using the MATLAB function *lsqlin* based on the interior point algorithm (Forrest and Goldfarb 1992). The optimisation calculations were carried out using Matlab R2024b under Windows 11 with an Intel Core i9-12900H processor and 64 GB of available RAM. The computation times for batteries 33, 34, 35, 36, 37 and 38 are respectively 167 s, 170 s, 164 s, 265 s,177 s and 212 s.

3.2. *RUL graphical representation and evaluation metric*

Figure 4 displays a graphical representation of the estimated RUL for the six batteries presented above. The abscissa represents the current number of battery cycles and the ordinate the number of remaining cycles, calculated at this cycle. In these figures, the oblique lines with slope -1 represent the true RUL. In the figure 5, the dotted lines represent the tube more or less 10 % from the true RUL. The quality of the results can be measured using several metrics (Kim, An, and J.-H. Choi 2017), the main one being the prediction horizon. Let us call N_{tube10} the cycle for which the RUL curve enters the tube at 10 % around the true RUL without leaving it afterwards and $N_{0.7}$ the current end-of-life cycle, then the prediction horizon is:

$$P_H = N_{0.7} - N_{\text{tube10}}. \tag{12}$$

The greater this value, the greater the time available to organize maintenance.

The calculation of the prediction horizons for the six batteries is summarised in table 1. It varies between 150 and 300 cycles, i.e. between 25 % and 50 % of the life of the battery. As with capacity, which exhibits

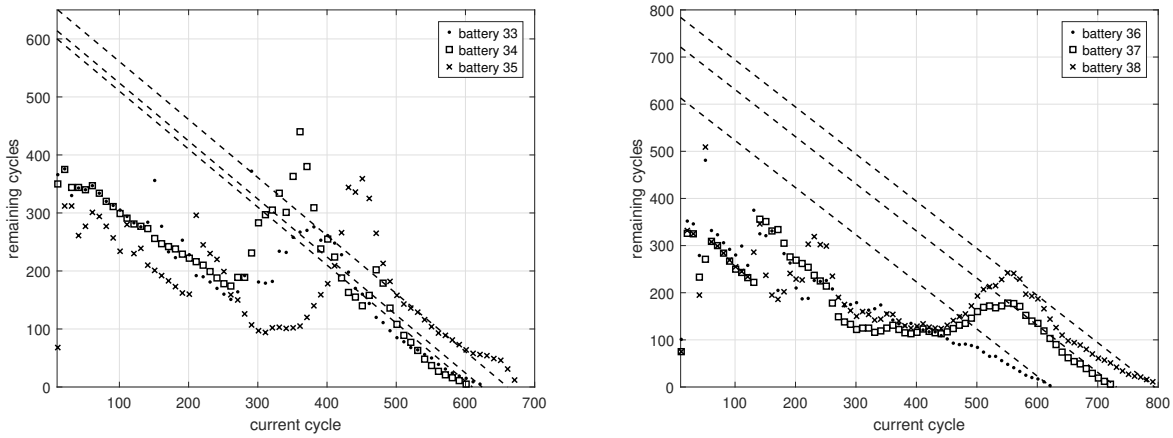


Figure 4. Graphical representation of the RUL over the cycles

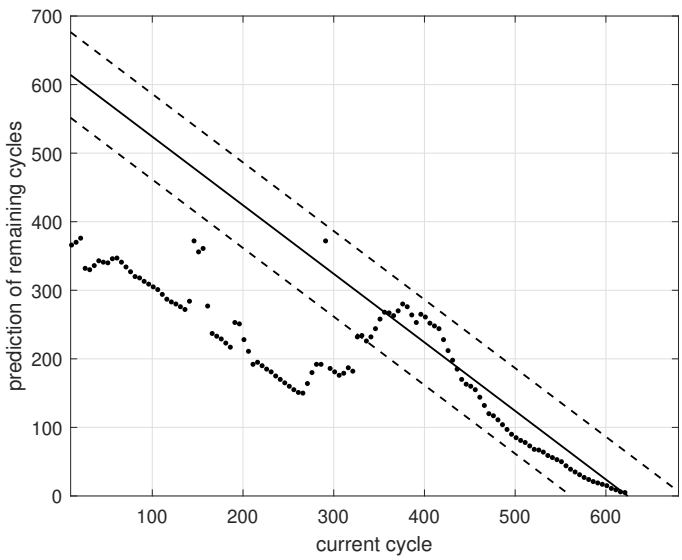


Figure 5. Battery 33 RUL estimation over the cycles

	bat 33	bat 34	bat 35	bat 36	bat 37	bat 38	mean
True life duration ($N_{0.7}$)	624	610	661	623	731	794	673
Cycle N_{tube10}	331	386	478	476	501	511	447
Prediction horizon (P_H)	293	224	183	147	230	283	227
% of remaining life at cycle N_{tube10}	46.9	36.7	27.7	23.6	31.5	35.6	33.6

Table 1. Summary table of RUL calculations for the 6 batteries

a wide dispersion, prediction and RUL also exhibit a wide dispersion. As seen in figure 4 the RUL is largely underestimated in the first cycles because the formation of the initial SEI layer implies a sharp drop in capacity. In the intermediate cycles, the RUL is overestimated because many values of the polynomial coefficients may correspond to the same start of the curve until the inflection has happened. After the inflection, the results stabilize towards the final value of the coefficients.

4. Conclusion and perspectives

The presented method allows for an accurate prediction starting from half of the battery's life, which is comfortable for maintenance planning. With an average of more than 30 % of remaining life when the prediction enters the tube at 10 % of the actual RUL, this method is relevant because it does not require a large database or training time like other statistical or neural network approaches. The authors plan to use these results as a basis for comparison with other methods, in particular, model-based methods.

References

Atalay, Selcuk et al. (2020). "Theory of battery ageing in a lithium-ion battery: Capacity fade, nonlinear ageing and lifetime prediction". In: *Journal of Power Sources* 478, p. 229026.

Canals Casals, Lluc et al. (Oct. 2019). "Evaluation of the End-of-Life of Electric Vehicle Batteries According to the State-of-Health". In: *World Electric Vehicle Journal* 10, p. 63.

Chakraborty, Kanad et al. (1992). "Forecasting the behavior of multivariate time series using neural networks". In: *Neural Networks* 5.6, pp. 961–970.

Cheng, Shunfeng, Michael Azarian, and Michael Pecht (June 2010). "Sensor Systems for Prognostics and Health Management". In: *Sensors (Basel, Switzerland)* 10, pp. 5774–97.

Choi, Jooho et al. (2010). "Bayesian Approach for Parameter Estimation in the Structural Analysis and Prognosis". In: *Annual Conference of the PHM Society*.

Christer, A.H., W. Wang, and J.M. Sharp (1997). "A state space condition monitoring model for furnace erosion prediction and replacement". In: *European Journal of Operational Research* 101.1, pp. 1–14.

Forrest, John J. and Donald Goldfarb (May 1992). "Steepest-edge simplex algorithms for linear programming". In: *Mathematical Programming* 57.1, pp. 341–374.

Greitzer, Frank et al. (Feb. 2001). "Determining how to do prognostics, and then determining what to do with it". In: *IEEE Systems Readiness Technology Conferencce Held in Valley Forge*, pp. 780–792.

He, Wei et al. (2011). "Prognostics of lithium-ion batteries based on Dempster–Shafer theory and the Bayesian Monte Carlo method". In: *Journal of Power Sources* 196.23, pp. 10314–10321.

Jardine, Andrew K.S., Daming Lin, and Dragan Banjevic (2006). "A review on machinery diagnostics and prognostics implementing condition-based maintenance". In: *Mechanical Systems and Signal Processing* 20.7, pp. 1483–1510.

Kim, Nam-Ho, Dawn An, and Joo-Ho Choi (2017). *Prognostics and Health Management of Engineering Systems*. Springer International Publishing.

Li, Dezhi, Wilson Wang, and Fathy Ismail (2013). "Enhanced fuzzy-filtered neural networks for material fatigue prognosis". In: *Applied Soft Computing* 13.1, pp. 283–291.

Liao, Linxia and Felix Köttig (2014). "Review of Hybrid Prognostics Approaches for Remaining Useful Life Prediction of Engineered Systems, and an Application to Battery Life Prediction". In: *IEEE Transactions on Reliability* 63.1, pp. 191–207.

McArthur, M. A., S. Trussler, and J. R. Dahn (Dec. 2011). "In Situ Investigations of SEI Layer Growth on Electrode Materials for Lithium-Ion Batteries Using Spectroscopic Ellipsometry". In: *Journal of The Electrochemical Society* 159.3, A198.

Reis, Gonçalo dos et al. (Sept. 2021). "Lithium-ion battery data and where to find it". In: *Energy and AI* 5, p. 100081.

Sikorska, J.Z., M. Hodkiewicz, and L. Ma (2011). "Prognostic modelling options for remaining useful life estimation by industry". In: *Mechanical Systems and Signal Processing* 25.5, pp. 1803–1836.

Optimizing a maintenance strategy by combining age-based maintenance and an imperfect fault detection model

Ingeborg de Pater*

*Faculty of Aerospace Engineering, Delft University of Technology, HS 2926 Delft, The Netherlands: i.i.depater@tudelft.nl

Abstract In predictive maintenance, fault detection models are often imperfect, with both false positives and false negatives. By solely planning maintenance based on an imperfect fault detection model, the number of failures, maintenance tasks and maintenance costs might increase compared to age-based maintenance. In this paper, we therefore combine age-based maintenance with an imperfect fault detection model. First, we derive the probability of a false positive given the age of the component, using Bayes theorem. We subsequently use the renewal reward theory to optimize the age at which to preventively replace components, if we also replace components based on the alarms of the fault detection model. Here, we may also ignore alarms when the component is just installed. The optimal parameters of this approach indicate whether it is optimal to perform preventive maintenance based on i) only the age (age-based maintenance), ii) only the alarms of the fault detection model, iii) based on the alarms and the age or iv) to have no preventive replacements. We illustrate this approach in a small case study, where we show that even fault detection models with many false positives and false negatives can lower the maintenance costs, if combined with age-based replacements.

1. Introduction

A popular maintenance strategy in industry is age-based, also called time-based, maintenance. In age-based maintenance, components are always preventively replaced after a fixed amount of time. The age-replacement model is often used to find the optimal preventive replacement age with the renewal-reward theorem (Barlow and Proschan (1996)). In recent years, the abundant availability of sensors installed around components has incentivized the development of predictive maintenance instead. In predictive maintenance, the measurements of these sensors are used to develop prognostic models. Components often degrade according to the delay time model, exhibiting three health states: Healthy, degraded and failed. In this case, predictive maintenance models are often divided into two type of models (Lei et al. (2018)). First, the measurements are used to generate an alarm when a component goes from the healthy state to a degraded state. This is called a fault detection model. Once a component is identified as degraded, the Remaining Useful Life (RUL, the time left until failure) is optionally predicted. For instance, neural networks (Fink et al. (2020)) and random forests (Nunes et al. (2023)) are popular methods to develop prognostic models.

In literature, prognostic models are often developed with simulated or lab data, which gives good results. When optimizing the maintenance planning, it is assumed that these good results will also be obtained in practice. For instance, in de Pater et al. (2022) and Wang et al. (2024), accurate RUL predictions with small errors are obtained for a simulated aircraft engine data set. The components are consequently maintained just before the predicted failure time, with a safety margin. With real-life data, however, training a prognostic model is difficult due to noisy data, a lack of failure data, changing operating conditions, multiple failure modes, etcetera. For instance, in Bieber et al. (2023), the authors obtain an F1-score of approximately 60% for anomaly detection with real data from satellites reaction wheels.

Prognostic models trained on real data thus often contian errors, i.e., are imperfect. Specifically, an imperfect fault detection model has false negatives (the model identifies a component as healthy, while it is degraded) and false positives (the model identifies a component as degraded, while it is healthy). When planning maintenance solely based on the outcomes of such an imperfect fault detection model, the number of maintenance tasks might actually increase, compared to age-based maintenance, due to false positives. At the same time, the number of failures might increase as well due to false negatives. Some studies therefore combine health information of the component with age-based replacements. Specifically, in Berrade et al. (2013); Driessen et al. (2017) and Zhang et al. (2021), it is assumed that imperfect inspections reveal whether a component is

degraded or healthy. To handle the errors of the inspections, preventive replacements are performed based on both the estimated degradation and the age of the component.

In this paper, we combine an imperfect fault detection model with age-based replacements. Instead of inspections, we assume that the health state of a component is directly assessed from the sensor measurements gathered during operations. Updates on the health state thus become available frequently. We first derive the probability of a false positive alarm given the age of the component using Bayes theorem. We subsequently optimize a maintenance strategy where we combine preventive replacements based on both the age of the component, and on the alarms of the fault detection model. We formulate the average maintenance costs using the renewal-reward theorem considering two variables: i) from what age of the components onward to act upon alarms of the fault detection model (to handle false positives) and ii) the age to preventively replace the component (to handle false negatives). We illustrate our approach for a component of which the lifetime follows a Weibull distribution and with a fault detection model with varying false positive and false negative rates. Here, the combination of age-based replacements with a fault detection model with a true positive rate of 0.5 (50% of the failures is detected beforehand) and a false positive rate of 0.001 (we expect one false positive per 1000 alarm moments) lowers the maintenance costs with 23.8% compared to only age-based replacements.

In the rest of this paper, we first present the methodology in Section 2, and we illustrate the maintenance optimization strategy in Section 3. Last, we present our conclusions in Section 4.

2. Methodology: Combining preventive replacements based on age and a fault detection model

2.1. *Considered fault detection model*

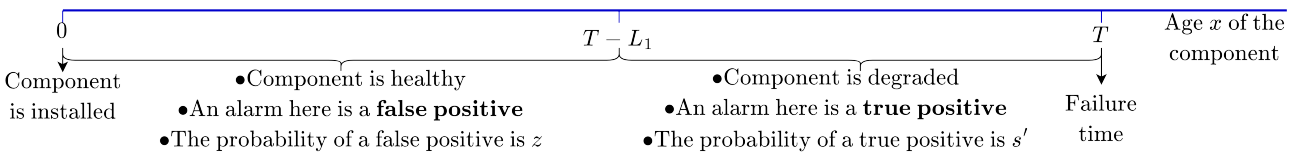


Figure 1. An overview of the fault detection model.

We consider a component that degrades according to the delay-time model, with three health states: Healthy, degraded and failed. The aim of the fault detection model is to raise alarm(s) when the component is degraded. We assume that the fault detection model can only raise an alarm at *discrete* time alarm moments, for instance after each flight. Let $x(i)$ denote the age of a component at alarm moment i, and let T denote the failure time of a component. We have two types of alarms. First, we assume that the component is degraded in the time-window $[T-L_1,T)$, with $L_1>0$ a constant. An alarm raised if the age of the component is in $[T-L_1,T)$ is therefore a true positive. Let $0\leq s'<1$ denote the true positive rate, i.e., probability that we raise a true alarm at any alarm moment i with $x(i)\in[T-L_1,T)$. Second, we assume that the component is healthy before time $T-L_1$. An alarm raised when the age of the component is smaller than $T-L_1$ is therefore a false positive. Let z denote the false positive rate, i.e., the probability that we generate a false positive at an alarm moment i with $x(i)<T-L_1$. We assume that the alarms are independent of each other. An overview of this fault detection model is in Figure 1.

After a single alarm, we employ a RUL prediction model. We assume that this RUL prediction model perfectly predicts the failure time T if the alarm is a true positive. Let $0\leq s<1$ denote the probability that at least one alarm is generated between $[T-L_1,T)$:

$$\begin{aligned} s =& p\left(\text{At least one alarm is raised between } [T-L_1,T)\right)\\ =& 1-p\left(\text{No alarm is raised between } [T-L_1,T)\right)\\ =& 1-\left(1-s'\right)^m. \end{aligned} \tag{1}$$

where m denotes the number of alarm moments in $[T-L_1,T)$ and $p(\cdot)$ denotes the probability of an event. Last, we assume that the lifetime distribution of the component is known. Let $f(\cdot)$ and $F(\cdot)$ denote the PDF and CDF of the lifetime distribution of the component, respectively.

2.2. *Probability of a false positive alarm given the age of a component*

Let A_i denote the event that the fault detection model raises an alarm at alarm moment i, when the age of the component is $x(i)$. Using the law of total probability, we derive the probability of A_i:

$$\begin{aligned} P(A_i) =& P(A_i|T \leq x(i))\,P(T \leq x(i)) + P(A_i|x(i) < T \leq x(i)+L_1)\,P(x(i) < T \leq x(i)+L_1) + \\ & P(A_i|T > x(i)+L_1)\,P(T > x(i)+L_1) \\ =& 0 \cdot F(x(i)) + s'\left[F(x(i)+L_1) - F(x(i))\right] + z\left[1 - F(x(i)+L_1)\right]. \end{aligned} \tag{2}$$

With Bayes theorem, we derive the probability that an alarm A_i at alarm moment i is a false positive:

$$\begin{aligned} P(T > x(i)+L_1|A_i) =& \frac{P(A_i|T > x(i)+L_1)\,P(T > x(i)+L_1)}{P(A_i)} \\ =& \frac{z \cdot \left[1 - F(x(i)+L_1)\right]}{s'\left[F(x(i)+L_1) - F(x(i))\right] + z\left[1 - F(x(i)+L_1)\right]}. \end{aligned} \tag{3}$$

The probability that an alarm is a false positive thus depends on the age $x(i)$ of the component, s and z.

2.3. *Combining preventive replacements based on the fault detection model and on the age*

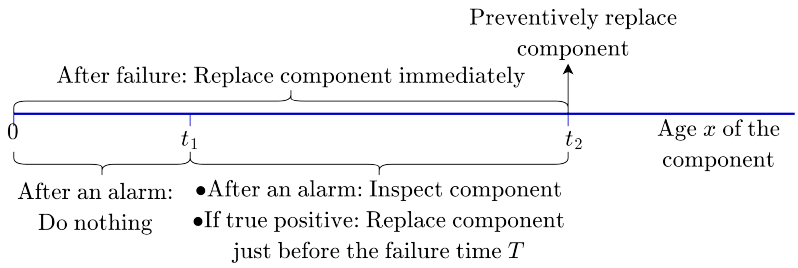


Figure 2. Overview of the considered actions when combining age-based replacements with a fault detection model.

If a fault detection model has a small true positive rate, then only replacing components based on alarms might lead to more failures than with age-based maintenance. If a fault detection model has a high false positive rate, then always acting upon alarms might lead to more maintenance tasks than in the age-based maintenance. To prevent this, we combine preventive replacements based on the age of the component, with replacements based on the alarms of the fault detection model. Specifically, we consider the following four maintenance actions, depending on two variables t_2 and $t_1 \leq t_2$ (see Figure 2):

(a) We replace a component immediately upon failure. A replacement after failure costs c_f.

(b) We replace a component immediately once it reaches the age t_2. Such a preventive replacement costs $c_p < c_f$.

(c) If the fault detection model raises an alarm when the age is smaller than t_1, we do nothing.

(d) If the fault detection model raises an alarm when the age is between t_1 and t_2, we employ the RUL prediction model. If the predicted failure time is smaller than t_2, we immediately inspect the component. This inspection reveals whether the alarm is a false positive or a true positive. If the alarm is a false positive, we do nothing. We only incur the inspection costs c_i. However, if the alarm is a true positive, we preventively replace the component just before the failure time T. We incur the costs $c_i + c_p < c_f$. We assume that all false positives between t_1 and t_2 lead to an inspection.

With these actions, we model four different maintenance strategies at once. Depending on the values of the variables t_1 and t_2, we have preventive replacements based on i) only the age of the component, ii) the age of the component and the alarms, iii) only on the alarms or iv) no preventive replacements. The optimal maintenance strategy for preventive replacements depends on the optimal value of t_1 and t_2.

According the the renewal-reward theorem (see Arts et al. (2024)), the long-term average maintenance costs per time unit equal the expected costs of one replacement, divided between the expected time between two replacements. We use this theorem to derive the long-term average cost per time-unit $g(t_1, t_2)$ with our mainte-

nance strategy:

$$g(t_1,t_2) = \frac{c_f F(t_1) + s(c_p + c_i)\left[F(t_2) - F(t_1)\right] + (1-s)c_f\left[F(t_2) - F(t_1)\right] + c_p\left[1 - F(t_2)\right] + zc_i \sum_{i=m(t_1)+1}^{m(t_2)} \left[1 - F(x(i) + L_1)\right]}{t_2\left[1 - F(t_2)\right] + \int_0^{t_2} uf(u)du}$$

$$= \frac{s\left[c_f - c_p - c_i\right] F(t_1) + c' F(t_2) + c_p + zc_i \sum_{i=m(t_1)+1}^{m(t_2)} \left[1 - F(x(i) + L_1)\right]}{\int_0^{t_2} \left[1 - F(u)\right] du}, \tag{4}$$

where $c' = (1-s)c_f + (s-1)c_p + sc_i$, while $m(t)$ denotes the number of alarm moments up to, but not including, time t. We find the optimal value of t_1 and t_2 with a grid search.

3. Illustration of combining age-replacements with a fault detection model

We illustrate the combination of the age-replacements with a fault detection model with an example. Here, we assume that the lifetime distribution of a component follows a Weibull distribution with PDF $f(x)$:

$$f(x) = \frac{kx^{k-1}}{\lambda^k} exp\left(-\frac{x}{\lambda}^k\right), \tag{5}$$

where k is the shape parameter and λ is the scale parameter. We use $k = 2$ and $\lambda = 1000$ (see Figure 3), which gives a mean lifetime of 886 hours with a standard deviation of 463 hours. The Weibull distribution is often used in literature to model the lifetime of a component. However, our proposed approach works with other lifetime distributions as well. Note that a prognostic fault detection model purely raises alarms based on sensor measurements and/or on a degradation model, and not on the lifetime distribution or the age of the component.

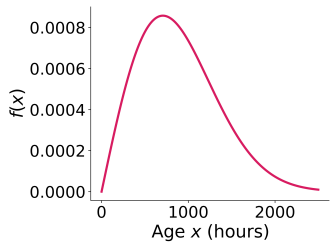


Figure 3. The pdf $f(x)$ of the considered lifetime distribution.

3.1. *Probability of a false positive alarm given the age of a component*

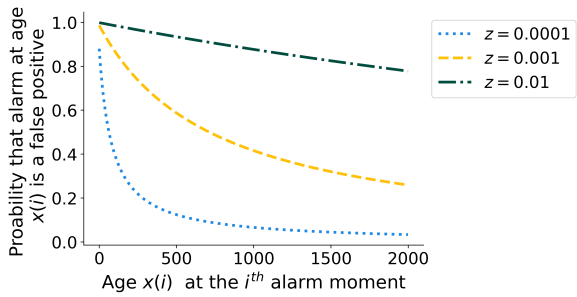


Figure 4. The probability that an alarm at the ith alarm moment, when the age of the component is $x(i)$ hours, is a false positive for $s = 0.5$ and several values of z.

Figure 4 shows the probability that an alarm at age $x(i)$ is a false positive alarm for $s = 0.5$ and several values of z. As expected with this lifetime distribution, the probability that an alarm is a false positive decreases when the component gets older. We might thus indeed save costs by ignoring alarms when the component is just installed, as these alarms have a higher probability of being a false positives.

Table 1. The optimal t_1, t_1^*, and t_2, t_2^* and the corresponding minimum average costs per hour $g(t_1^*, t_2^*)$ for several values of s and z. Strategy "Age only" - Only preventive replacements based on age. Strategy "Age & alarms" - Preventive replacements based on both age and alarms.

z	0	0.0001	0.001	0.01
		$s = 0.1$		
t_1^*	0	14	142	336
t_2^*	354	354	351	336
$g(t_1^*, t_2^*)$	5.76	5.78	5.96	6.06
Strategy	Age & alarms	Age & alarms	Age & alarms	Age only
		$s = 0.5$		
t_1^*	0	3	29	282
t_2^*	474	474	473	382
$g(t_1^*, t_2^*)$	4.38	4.40	4.62	5.99
Strategy	Age & alarms	Age & alarms	Age & alarms	Age & alarms
		$s = 0.9$		
t_1^*	0	0	16	158
t_2^*	1018	1018	1020	921
$g(t_1^*, t_2^*)$	2.29	2.32	2.53	4.48
Strategy	Age & alarms	Age & alarms	Age & alarms	Age & alarms

3.2. *Combining the fault detection model with age-based replacements*

In this case study, a corrective replacement costs $c_f = 10,000$, a preventive replacement costs $c_p = 1000$, an inspection costs $c_i = 250$, $L_1 = 20$ hours, and we have one alarm moment per hour. Without a fault detection model, it is optimal to preventively replace the components at an age of $t^* = 336$ hours, with the average maintenance costs per hour of 6.06.

Table 1 shows the optimal value of t_1, t_1^*, and of t_2, t_2^*, and the corresponding minimum costs per hour $g(t_1^*, t_2^*)$ for several values of z and s. We find these values with a grid search, which takes less than two minutes for all cases. As expected, if $z = 0$, $t_1^* = 0$: If there are no false positives, it is optimal to perform a preventive replacement after each alarm. When z increases, the probability of a false positive increases as well (see Figure 4). As expected, t_1^* therefore increases as well, i.e., we ignore more alarms when z increases. If s increases, we detect more failures before they occur, and the optimal preventive replacement time t_2^* therefore increases.

Moreover, the optimal solution also shows the optimal maintenance strategy. If $s = 0.1$ while $z = 0.01$, it is optimal to only have age-based preventive replacements and to ignore all alarms of the fault detection model. In all other considered cases, it is optimal to mix preventive replacements based on the age and on the alarms. For instance, if $s = 0.5$ (50% of the failures is detected beforehand) while $z = 0.001$ (we expect one false positive per 1000 alarm moments) , the maintenance costs are 23.8% lower with our strategy compared to only age-based replacements.

4. Conclusions

In predictive maintenance, fault detection models are usually not perfect. If we only plan maintenance based on the outcomes of such an imperfect model, the number of maintenance tasks and the number of failures might actually increase, compared to age-based maintenance. In this paper, we therefore combine age-based maintenance with an imperfect fault detection model. We use the renewal-reward theorem to find an expression for the average maintenance costs per time unit. This formula contains two variables, i) from which age onwards to act upon alarms of the imperfect fault detection model and ii) at which age to preventively replace the component. These variables are optimized with a grid search. The optimal solution shows if it is optimal to have preventive replacements based on i) only the age of the component, ii) the age of the component and alarms, iii) only on the alarms or iv) no preventive replacements. We illustrate our approach with an example, where the lifetime distribution of a components follows a Weibull distribution. Here, even combining a fault detection model with a high false positive rate and a low true positive rate with age-based replacements leads to lower maintenance costs, than having only age-based replacements.

This research is a first step in combining fault detection models with the lifetime distribution of a component and/or age-based maintenance strategies. In future research, we will analyse our approach on a real life case study, where we can fit a lifetime distribution and develop a real fault detection model. For this, we will have to make more realistic assumptions, such as having an imperfect or no RUL prognostic model, dependent alarms, no possibility for inspections and a varying value for L_1.

References

Arts, J., Boute, R. N., Loeys, S., and van Staden, H. E. (2024). Fifty years of maintenance optimization: Reflections and perspectives. *European Journal of Operational Research*, **322**:725–739.

Barlow, R. E. and Proschan, F. (1996). *Mathematical theory of reliability*. SIAM.

Berrade, M. D., Scarf, P. A., Cavalcante, C. A. V., and Dwight, R. A. (2013). Imperfect inspection and replacement of a system with a defective state: A cost and reliability analysis. *Reliability Engineering & System Safety*, **120**:80–87.

Bieber, M., Verhagen, W. J., Cosson, F., and Santos, B. F. (2023). Generic diagnostic framework for anomaly detection—application in satellite and spacecraft systems. *Aerospace*, **10**(8):673.

de Pater, I., Reijns, A., and Mitici, M. (2022). Alarm-based predictive maintenance scheduling for aircraft engines with imperfect Remaining Useful Life prognostics. *Reliability Engineering & System Safety*, **221**:108341.

Driessen, J., Peng, H., and Van Houtum, G. (2017). Maintenance optimization under non-constant probabilities of imperfect inspections. *Reliability Engineering & System Safety*, **165**:115–123.

Fink, O., Wang, Q., Svensen, M., Dersin, P., Lee, W.-J., and Ducoffe, M. (2020). Potential, challenges and future directions for deep learning in Prognostics and Health Management applications. *Engineering Applications of Artificial Intelligence*, **92**:103678.

Lei, Y., Li, N., Guo, L., Li, N., Yan, T., and Lin, J. (2018). Machinery health prognostics: A systematic review from data acquisition to RUL prediction. *Mechanical systems and signal processing*, **104**:799–834.

Nunes, P., Santos, J., and Rocha, E. (2023). Challenges in predictive maintenance–a review. *CIRP Journal of Manufacturing Science and Technology*, **40**:53–67.

Wang, L., Zhu, Z., and Zhao, X. (2024). Dynamic predictive maintenance strategy for system Remaining Useful Life prediction via deep learning ensemble method. *Reliability Engineering & System Safety*, **245**:110012.

Zhang, F., Shen, J., Liao, H., and Ma, Y. (2021). Optimal preventive maintenance policy for a system subject to two-phase imperfect inspections. *Reliability Engineering & System Safety*, **205**:107254.

A Probabilistic and Machine Learning Approach to Predictive Maintenance of Railway Tracks

*N G.D. Santos., **B.M. Alkali., ***A.J. Kumar, ****O. Niculita

* 240 Dobbie's Loan, Glasgow G4 0JF, Glasgow nelson.santos@cgcu.ac.uk

**Cowcaddens Rd, Glasgow G4 0BA, Babakalli.Alkali@gcu.ac.uk

Abstract

Railway track infrastructure is subject to continuous degradation, with Rolling Contact Fatigue (RCF) being a major cause of defects such as squats and cracks. Effective maintenance and frequent inspections are essential to ensure operational safety and cost efficiency. This study introduces a hybrid probabilistic and machine learning framework to model and predict railway track defects. The methodology integrates Negative Binomial (NB) and Poisson distributions (PD) for defect count modelling, Random Forest (RF) for predictive accuracy, and Hidden Markov Models (HMM) for degradation state transitions. Real-world railway track data, including operational, environmental, and maintenance parameters, are used to validate the models. The Random Forest model demonstrates superior predictive performance compared to traditional regression methods, achieving a Mean Absolute Error (MAE) of 0.53 and a high regression coefficient ($R^2 = 0.86$). In parallel, the HMM analysis uncovers critical deterioration dynamics, revealing a 76% probability that track sections bypass moderate states and transition directly to severe degradation state within a single inspection cycle. Moreover, there is an 83% likelihood that once a section reaches the severe state, it remains in that condition persistently. These insights highlight the limitations of conventional maintenance approaches and underscore the necessity for predictive maintenance strategies. By enabling early identification of high-risk zones, optimizing inspection intervals, and proactively mitigating irreversible damage, predictive models significantly enhance both safety and cost-efficiency in railway infrastructure management.

1. Introduction

Railway infrastructure degradation is driven by mechanical wear, environmental factors, and operational loads, with Rolling Contact Fatigue (RCF) being a primary contributor to defects like squats and cracks (Esveld, 2001; Zarembski, 2013). Traditional maintenance approaches (corrective and time-based preventive) often fail to address RCF-induced defects efficiently, leading to unplanned downtime and high repair costs.

Recent advances in probabilistic modelling and machine learning offer new opportunities for predictive maintenance. While binomial and Poisson models provide a statistical foundation for defect prediction, ensemble methods like Random Forest and state-transition models like HMM enhance accuracy by capturing nonlinear relationships and degradation dynamics (Breiman, 2001; Rabiner, 1989). This study bridges the gap between statistical and machine learning approaches by:

- Modelling defect occurrences using Negative Binomial and Poisson distributions.
- Predicting defect hotspots with RF, incorporating track curvature, speed, and environmental factors.
- Analysing degradation trends via HMM to identify critical transition states.
- Providing maintenance recommendations based on predictive insights.

2. Literature Review - Railway Infrastructure Degradation Mechanisms

Railway infrastructure degradation is a critical concern for rail operators, primarily driven by mechanical wear, environmental factors, and operational loads. Among these, Rolling Contact Fatigue (RCF) stands out as a major contributor to defects such as squats, head checks, and gauge corner cracks (Esveld, 2001; Zarembski, 2013). RCF results from repeated cyclic stresses at the wheel-rail interface, exacerbated by factors like high axle loads, track misalignment, and insufficient lubrication (Grassie and Kalousek, 1993). Environmental conditions, including temperature variations, moisture, and ballast fouling, further accelerate degradation through mechanisms like corrosion and subgrade weakening (Selig and Waters, 1994). Operational parameters such as train speed, traffic density, and track curvature also significantly influence defect initiation and propagation (Li et al., 2015).

Traditional maintenance strategies primarily rely on corrective (reactive) approaches and time-based preventive measures, both of which have notable limitations. Corrective maintenance leads to unplanned downtime and high emergency repair costs, while scheduled preventive maintenance often results in either over-maintenance (wasting resources) or under-maintenance (missing critical defects) (Jardine et al., 2006). Recent studies advocate for Condition-Based Maintenance (CBM) and Predictive Maintenance (PdM) as more efficient alternatives, enabled by advances in data analytics and machine learning.

Early defect prediction models employed Poisson and Negative Binomial (NB) regression to estimate failure frequencies (Meier-Hirmer et al., 2009). While these statistical models provide a foundation, they struggle to capture nonlinear relationships and temporal degradation patterns. To address these limitations, machine learning techniques have gained prominence:

- Random Forest (RF) excels at identifying defect hotspots by analysing complex interactions between variables like track geometry, load history, and environmental conditions (Breiman, 2001).
- Hidden Markov Models (HMMs) model degradation as a sequence of hidden states, enabling early detection of critical transitions (Rabiner, 1989).

Hybrid approaches combining statistical models with ML techniques show particular promise. For instance, Bayesian Networks integrate domain knowledge with data-driven insights, while ensemble methods like Gradient Boosting improve prediction accuracy (Famurewa et al., 2017). Despite these advances, most existing studies focus either on statistical models or ML techniques, missing opportunities for synergy. This study bridges that gap by modeling defect occurrences using NB and Poisson regression for probabilistic estimates; Predicting high-risk zones with RF, incorporating operational and environmental factors; Analysing degradation trends via HMMs to identify critical transition states; Providing actionable maintenance insights, such as optimized grinding schedules.

The proposed integrated framework aims to minimize unplanned downtime, reduce maintenance costs through data-driven decision-making.

3. Methodology

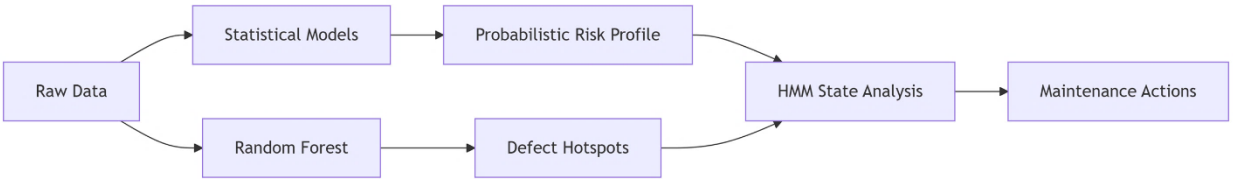


Figure 1: Predictive maintenance model in railway infrastructure

3.1. Data Collection and Preprocessing

Analysis was performed on a five-year dataset. The dataset comprised defect records primarily focused on squat lengths and crack propagation patterns identified through ultrasonic inspections. Additional variables included track geometry (curvature), axle load, train speed (miles/hour), Million Gross Tonnage (MGT), environmental conditions such as rainfall (mm) and temperature (°C), and maintenance records covering squat defect interventions, Mean Time Between Failures (MTBF), and Track Intervention Intervals. Missing data were handled using linear interpolation for temporal gaps and K-Nearest Neighbour (KNN) imputation for spatial inconsistencies.

The primary defect type analysed in this study is Squat. Squats are surface-initiated defects resulting from RCF and are widely recognized as precursors to severe rail degradation. A total of 223 squat incidents are recorded from a 44-mile railway section. Their frequency and impact on ride quality, safety, and rail integrity justify their selection as the central focus of the modelling framework. In addition, Squat on a weld (39 incidents) and Squat at rail end (3 incidents) were recorded. Although less frequent, these variants were included in exploratory analysis to assess their degradation behaviour. However, due to their small sample size and distinct stress environments, they were not included in the Hidden Markov Model calibration to avoid model overfitting.

3.2. Defect Prediction Models

The Negative Binomial (NB) regression model was selected to address overdispersion in defect counts per section, a common issue in infrastructure degradation studies (Lord and Mannering, 2010; (Sresakoolchai and Kaewunruen, 2022; Hilbe, 2011). Equation 1 expresses the model's probability mass function:

$$P(X = k) = \binom{k + r - 1}{k} p^r (1 - p)^k \tag{1}$$

Where:
X: Number of defect incidents.
k: Observed count of defect (non-negative integer)
r: Shape parameter reflects the clustering of defects estimated via maximum likelihood estimation (MLE). Higher r suggests greater variability in clustering data.
p: Probability of defect occurrence in a single inspection interval.

$\binom{k + r - 1}{k}$: Binomial coefficient.

The Poisson Distribution (PD) was used as a baseline, the model in equation 2 assumes defect counts as:

$$P(X = k) = \frac{e^{-\lambda}\lambda^k}{k!} P(X = k). \tag{2}$$

Where:
$P(X = k)$: The probability of observing k events.
λ: The expected number of incident occurrences (the mean rate).
e: is Euler's number.
$k!$: is the factorial of k (i.e., k x (k-1) x …x 1).

- The NB and PD models are useful for establishing baseline failure frequencies and capturing overdispersion in data, for instance more variance than mean (Payne et al., 2017).
- Output: A probabilistic risk profile, indicating areas with statistically high likelihood of failure.

Random Forest (RF) regression model, introduced by Breiman (2001), is a robust ensemble method capable of handling nonlinear and high-dimensional railway degradation data. In this study, we employ 200 decision trees to enhance predictive accuracy while enabling interpretable feature importance analysis. The RF model predicts defect counts (Y) as:

$$Y = f(X) + \epsilon\,. \tag{3}$$

$f(X)$: Aggregated output of 200 decorrelated decision trees (bootstrap-aggregated).
ϵ: Residual error minimized through ensemble averaging.

- The RF identifies important predictors (e.g., curvature, temperature) and forecasts where defects are most likely to appear.
- Output: Defect Hotspots - geographical segments or operational zones that are highly susceptible to degradation.

Probabilistic risk profile synthesizes outputs from statistical models to generate a quantitative risk estimate per track segment. This estimate is based on historical trends and influencing variables. The profile is particularly valuable for prioritizing high-risk areas when inspection resources are constrained.

Defect hotspots are identified by the RF model as areas with elevated risk and are mapped onto route plans to help prioritize decision of maintenance actions.

3.3. The HMM State Analysis

The HMM State Analysis uses defect features, for example, crack length and historical transition data to assess both the current and future degradation states of track segments. The degradation states are defined as follows: State 0 – Good (<10 mm), State 1 – Moderate (10–30 mm), and State 2 – Severe (>30 mm). The model computes transition probabilities to estimate future conditions under prevailing operational loads (Zarembski, 2015; Network Rail, 2022). Table 1 presents predictive maintenance actions aligned with defect state transition probabilities.

Table 1: Predictive Maintenance Actions Based on Transition Probabilities (Network Rail, 2022).

State	Action Plan	Priority
State 0 (Good Condition)	Routine inspections, preventive maintenance.	✅ Low
State 1 (Moderate Degradation)	Scheduled maintenance, tamping, rail grinding.	⚠ High
State 2 (Severe Degradation)	Urgent repairs, possible track replacement.	● Critical

State transitions are modeled via the probability matrix $T_{ij} = P(S_t = j \mid S_{t-1} = i)$ (Rabiner, 1989), where observed studies show a 72% probability of transitioning from State 0 to State 1 when crack growth rates exceed 0.5mm/week (Li et al., 2021). This integration aligns with predictive maintenance frameworks proposed by Kumar et al. (2020), enabling data-driven prioritisation of interventions: grinding for moderate cracks (10-30mm) (Ferreira and López-Pita, 2016) or urgent replacement for severe cases (>30mm) (Network Rail, 2022).

4. Results and Discussion

4.1. Model Performance Comparison

The comparative evaluation of Negative Binomial (NB), Poisson, and Random Forest (RF) models across 253 track sections (755 defects) revealed stark performance differences (Table 2). The RF model significantly outperformed statistical approaches, achieving a MAE of 0.53 (vs. 2.02 for NB and 1.96 for Poisson) and an R^2 of 0.86, demonstrating its capability to capture nonlinear degradation patterns as shown in Figure 2. The NB and Poisson models exhibited overestimation biases in high-traffic segments (25–32 miles), attributed to their fixed variance assumptions (Lord and Mannering, 2010). RF's ensemble structure (200 trees) effectively modeled interactions between variables like curvature-speed-load dynamics, aligning with Breiman's (2001) findings on robustness in high-dimensional data.

Table 2:Model Performance Metrics.

Model	MAE	R^2 Score
Negative Binomial	2.02	0.62
Poisson	1.96	0.65

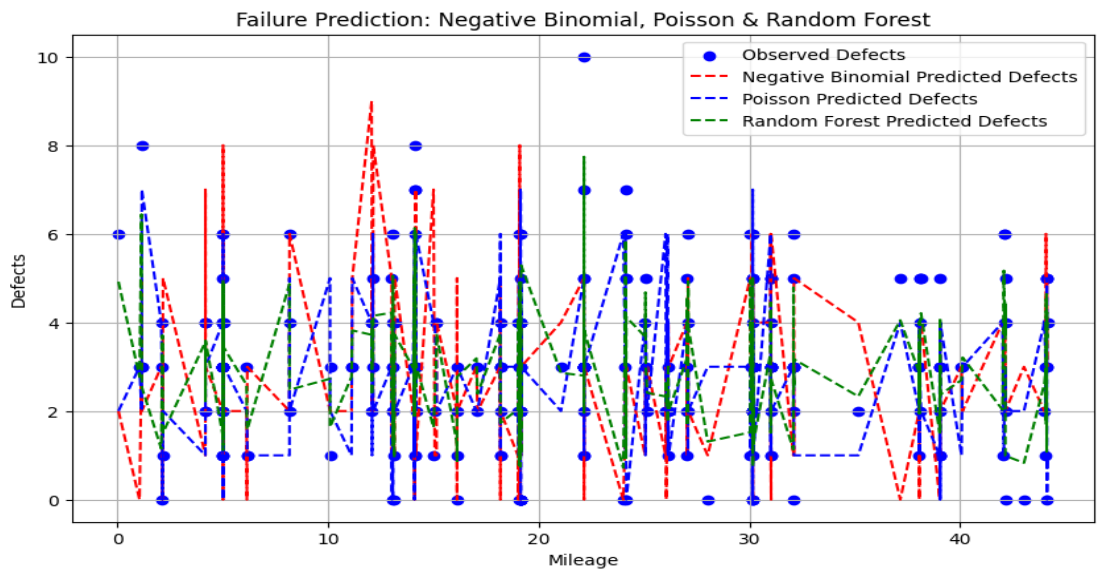


Figure 2: Failure Prediction for railway tracks using NB, PD and RF models.

4.2. Feature Importance and High-Defect Hotspots

The Random Forest model's feature importance analysis, presented in Table 3, quantifies the contribution of each input variable to the model's predictive accuracy. Curvature (0.1779), Speed (0.1473), and Temperature (0.1431) emerged as the most influential predictors of squat defect due to Rolling Contact Fatigue (RCF).

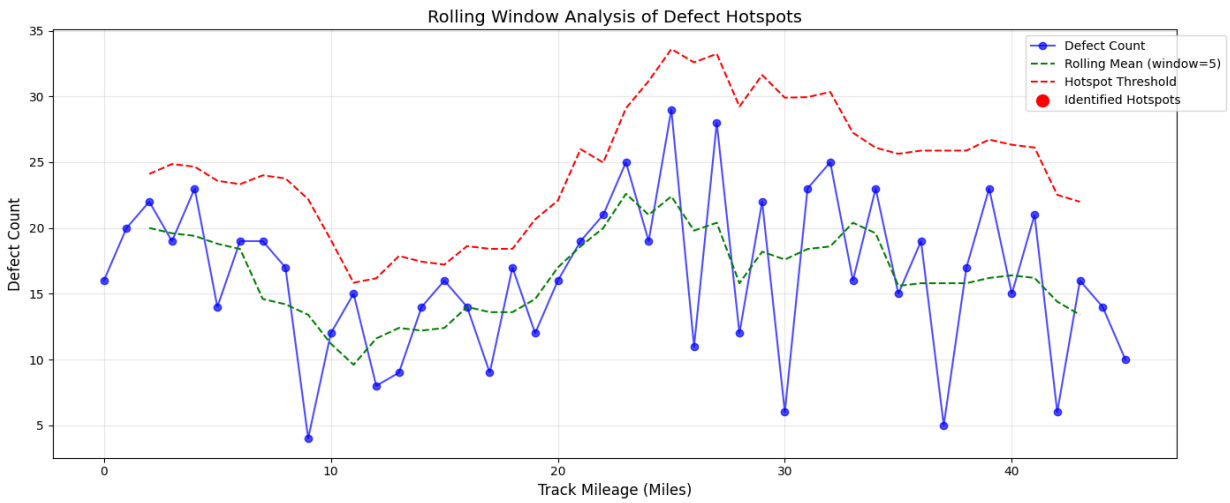
These findings are consistent with empirical studies (Grassie & Kalousek, 1993), which demonstrate that high curvature increases contact stress, speed affects frictional dynamics, and temperature alters rail material properties, factors known to influence squat initiation.

Table 3: Top Influential Features.

Factor	Importance Score
Curvature	0.1779
Speed	0.1473
Temperature	0.1431

4.3. High-Defect Hotspot Identification

The graph in Figure 3 successfully pinpoints specific mileage intervals around 10–15, 25–30, and 35–40 miles.



where defect counts exceed the threshold, enabling targeted maintenance.

Figure 3 High-Defect Hotspot Identification (Spatial Analysis)

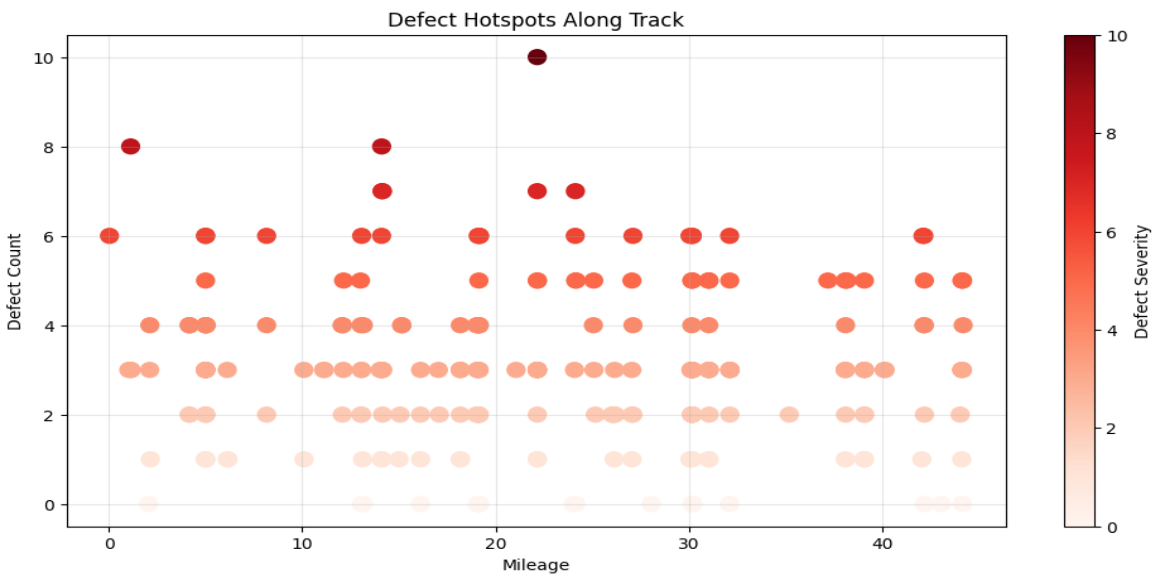


Figure 4: The distribution of defect counts and severities across different mileage segments of a railway track

4.4. HMM Transition Insights

The Hidden Markov Model (HMM) revealed degradation dynamics through transition probabilities as shown in Table 4.

Table 4: HMM State Transition Probabilities

Current State	To State 0 (Good)	To State 1 (Moderate)	To State 2 (Severe)
Good (0)	0%	24%	76%
Moderate (1)	8%	23%	69%
Severe (2)	14%	3%	83%

The HMM analysis reveals two critical findings regarding track deterioration. First, tracks initially classified as Good (State 0) have a 76% probability of rapidly degrading to Severe (State 2) within a single inspection cycle, emphasizing the importance of early intervention to prevent accelerated decline. Second, once tracks reach State 2, they exhibit an 83% likelihood of remaining in this severe condition, demonstrating the self-perpetuating nature of advanced deterioration and highlighting the significant costs associated with delayed maintenance. These findings underscore the need for proactive monitoring and timely repairs to mitigate escalating infrastructure risks.

5. Conclusion and Maintenance Recommendations

This study presents an advanced hybrid framework combining probabilistic modeling and machine learning techniques for railway track maintenance. The integrated approach utilizes Negative Binomial/Poisson regression, Random Forest (RF) algorithms, and Hidden Markov Models (HMM) to significantly improve predictive capabilities. Key findings demonstrate the RF model's superior performance with an R^2 of 0.86 and MAE of 0.53, outperforming traditional statistical methods in capturing complex degradation patterns.

To further improve the R^2 beyond 0.86, future iterations of the model should consider the integration of more granular and context-aware data streams. One approach involves incorporating high-resolution dynamic inputs, such as real-time axle load measurements, vibration and acceleration data from onboard sensors, and wheel-rail contact force metrics. These can provide a more detailed understanding of load-induced stresses and transient fatigue phenomena, enhancing predictive precision.

The HMM analysis reveals critical insights into track deterioration dynamics, particularly highlighting a concerning 76% probability of tracks transitioning directly from Good (State 0) to Severe (State 2) condition within a single inspection cycle. This finding underscores the critical importance of early intervention strategies. Furthermore, the analysis shows an 83% probability of severely degraded tracks remaining in critical condition, emphasizing the substantial cost implications of delayed maintenance actions.

To effectively transition to predictive maintenance, railway operators should first adopt the Random Forest model as their primary defect prediction tool, leveraging its demonstrated accuracy in identifying high-risk sections. Maintenance teams should prioritize the 10-20- and 25-32-mile segments identified as critical hotspots, implementing enhanced monitoring through IoT sensors (strain gauges, accelerometers) inspections for comprehensive coverage.

A three-tiered intervention strategy should be established: routine inspections for State 0 (Good) tracks, preventive grinding within six months for State 1 (Moderate) sections, and immediate repair or replacement for State 2 (Severe) degradation. Organizations must concurrently invest in workforce training programs to build data literacy among maintenance crews and implement standardised digital reporting systems to ensure consistent, high-quality input data for the predictive models. This implementation should be phased, beginning with pilot programs on high-traffic corridors to validate the approach and measure ROI before full network deployment, while continuously refining the models through Bayesian updating as new inspection data becomes available.

A comprehensive cost-benefit analysis framework should be developed to quantify the savings from preventive interventions compared to traditional reactive repairs, helping secure necessary budget allocations and demonstrating the long-term value proposition of this predictive maintenance approach.

6. References

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp.5-32. https://doi.org/10.1023/A:1010933404324

British Standards Institution, 2019. *EN 13231-3: Railway applications - Track - Acceptance of works - Part 3: Acceptance of rail grinding*.

Esveld, C., 2001. *Modern railway track*. 2nd ed. Zaltbommel: MRT-Productions.

Famurewa, S.M., Asplund, M. and Rantatalo, M., 2017. Maintenance decision support for railway infrastructure using condition monitoring. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 231(6), pp.667-679. https://doi.org/10.1177/0954409716656218

Ferrari, A., et al., 2019. Structural crack detection in rails using Bayesian networks. *Mechanical Systems and Signal Processing*, 128, pp.1-15.

Ferreira, P.A. and López-Pita, A., 2016. Optimization of rail grinding cycles. *Transportation Research Part A: Policy and Practice*, 89, pp.1-14.

Grassie, S.L. and Kalousek, J., 1993. Rolling contact fatigue of rails: Characteristics, causes and treatments. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, 207(1), pp.57-68. https://doi.org/10.1243/PIME_PROC_1993_207_227_02

Hilbe, J.M., 2011. *Negative Binomial Regression*. 2nd ed. Cambridge: Cambridge University Press.

Jardine, A.K., Lin, D. and Banjevic, D., 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, 20(7), pp.1483-1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Kumar, A., et al., 2020. Advanced predictive maintenance approaches for railway infrastructure. Transportation Research Part C: Emerging Technologies, 118, 102712. https://doi.org/10.1016/j.trc.2020.102712

Li, X., et al., 2021. Crack growth rate modeling in rails using stochastic methods. *Reliability Engineering & System Safety*, 215, 107812.

Li, Z., et al., 2015. Differential wear and plastic deformation as cause of squat at track local stiffness change. *Wear*, 338-339, pp.79-93. https://doi.org/10.1016/j.wear.2015.05.007

Meier-Hirmer, C., et al., 2009. Maintenance optimization for railway tracks. *IEEE Transactions on Reliability*, 58(2), pp.324-333. https://doi.org/10.1109/TR.2009.2020152

Lord, D. and Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A: Policy and Practice, 44(5), pp.291-305. https://doi.org/10.1016/j.tra.2010.02.001

Network Rail, 2022. *GNGE246 - Defect Response Standards*.

Payne, E.H., et al., 2017. An empirical comparison of methods for analyzing overdispersed count data in public health. Statistics in Medicine, 36(18), pp.2850-2866. https://doi.org/10.1002/sim.7338

Rabiner, L.R., 1989. A tutorial on hidden Markov models. *Proceedings of the IEEE*, 77(2), pp.257-286. https://doi.org/10.1109/5.18626

Selig, E.T. and Waters, J.M., 1994. *Track geotechnology and substructure management*. London: Thomas Telford.

Sresakoolchai, J. and Kaewunruen, S., 2022. Railway track defect detection using machine learning. *Engineering Structures*, 250, 113418.

Zarembski, A.M., 2013. Managing the wheel/rail interface. *AREMA Conference Proceedings*.

Zarembski, A.M., 2015. Managing rail defects. *Procedia Engineering*, 114, pp.3-21.

Evaluating Maintenance Strategies for Locomotive Wheelsets Using Petri Net-Based Modelling Approach

Maksym Ocheretniuk*, Darren Prescott** and Rundong Yan***

*Resilience Engineering Research Group, University of Nottingham, UK, *maksym.ocheretniuk@nottingham.ac.uk*
**Resilience Engineering Research Group, University of Nottingham, UK, *darren.prescott@nottingham.ac.uk*
***Resilience Engineering Research Group, University of Nottingham, UK, *rundong.yan@nottingham.ac.uk*

Abstract. Effective locomotive fleet management is crucial for ensuring the safe, reliable and cost-efficient operation of railway systems. One of the primary maintenance tasks in railway transportation is the condition monitoring and management of rolling stock wheelsets, which directly influence operational safety, ride comfort, and maintenance costs. This paper proposes a Petri net-based approach to model the state transitions of wheelset flange and tread degradation to evaluate different maintenance strategies for locomotive wheelsets. The methodology aims to enhance decision-making regarding maintenance intervals, minimize locomotive downtime, and reduce maintenance costs, while maintaining high safety and reliability standards. A key feature of the proposed approach is the ability to simulate different maintenance strategies for the locomotive wheelsets and evaluate their impact on fleet performance metrics, such as availability, reliability, and life-cycle costs. By varying the inspection and maintenance policies within the simulation, the optimal strategy that balances safety, cost, and operational efficiency can be identified. Another important outcome is that the proposed model allows the safety risks to be evaluated when the locomotive wheelset condition is outside the standard's limits that are established by maintenance policy. Additionally, the model can be used as a decision-support tool that helps to evaluate the trade-offs between different maintenance policies and select the most cost-effective strategy.

1. Introduction

In the railway industry, effective maintenance strategies for rolling stock components are essential to ensure operational safety, minimise costs, and maintain high levels of service availability. One of the most critical components influencing locomotive performance and safety is the wheelset, specifically its flange and tread dimensions. Over time, these dimensions deteriorate due to natural wear from wheel-rail interaction and are further affected by maintenance activities such as corrective and preventive re-profiling. All this maintenance activity reduces wheelset lifespan due to a specific profile restoration process where the wheelset tread must be reduced in order to restore the wheelset flange profile. Thus, the wheelset loses its potential operational lifetime since the wheelset tread reaches its established operational limit sooner than it otherwise would. Therefore, railway companies seek to find optimal maintenance practices to extend the wheelset lifespan alongside saving costs and increasing ride safety.

Several studies were investigated to evaluate existing approaches to solving the challenges mentioned above. Zhang et al. (2022) suggested a joint maintenance optimisation model for the railway wheelset which is based on a compound Poison process and adapted Weibull distribution, where a wheelset joint maintenance strategy is proposed which allows wheelset lifecycle extension up to 12% and maintenance costs reduction up to 27% compared to conventional wheelset maintenance strategies. Additionally, Andrade and Stow (2017) investigated a maintenance approach for the wheelset known as "Economic Tyre Turning" (ETT) for Great Britan Railways using statistical degradation models alongside Monte Carlo simulation and found that ETT maintenance can potentially save up to 4.7% of the maintenance costs. Moreover, Costa et al. (2020) proposed a data-driven model which helps the decision-making process for wheelset reprofiling, and replacement based on survival analysis and Markov process.

Although the studies mentioned above aimed to assess the life cycle of the wheelset and reduce maintenance costs by varying maintenance strategies, there is a gap such not investigating time the wheelset being outside operational limits what may occur in real life due to untimed or delayed inspection and re-profiling. Therefore, the Petri Net framework proposed in this paper describes the entire wheelset degradation process considering flange and tread natural wear, tread wear due to maintenance actions such as re-profiling and presents results which show the wheelset lifespan, number of inspections and re-profiling

activities alongside total time the wheelset operates outside operational limits which can be considered as safety risk. The results presented in section 5 show differences in wheelset lifetime and time the wheelset is out of limits for seven wheelset maintenance strategies.

2. Methodology

Petri Nets (PN) are a graphical and mathematical modelling tool used to describe and model different system behaviour such as the wheelset degradation process. As recent research shows, the PN methodology is widely used as a powerful modeling tool for stochastic and discrete engineering systems which deal with unpredictable system behavior and system condition changes which happen at a specific time, such as railway track (Saleh et al, 2023), wind turbines (Wu et al., 2024) and, nuclear power plants (Yan et al., 2023). Thus, the PN method was chosen for the wheelset degradation process since it is considered to be a stochastic and discrete process at the same time due to the random nature of degradation events and the discrete transitions between different wheelset degradation states over time. Moreover, PN clearly represents all the possible states of the wheelset throughout the entire degradation process, as well as the inspection and maintenance actions applied to it. In addition, the PN method provides a good framework to which a Monte Carlo simulation approach can be applied since the model dynamically works between various system states depending on the system's structure and operations.

This paper shows the applicability of the PN technique and describes its basic features below (Prescott and Andrews, 2013). The PN is a directed graph consisting of two types of nodes: places represented as circles and timed or immediate transitions represented as rectangles shown in Figures 1*a*, 1*b* and 1*c*, respectively. Directed edges are used to connect places and transitions as shown in Figure 1, *d*. Additionally, tokens inside places are used to represent the actual system state: the distribution of tokens within the PN is known as its marking. Therefore, PN's places usually represent system states or conditions, transitions usually represent events or activities that change the system state, and edges play a linking role between places and transitions to allow tokens to input places and added to output places, thereby changing the system condition. Figure 1 (*e*, *f*) shows the basic rules of the PN firing process. Figure 1 (*e*) represents the state of the system before the transition fires and generates the token to the output place. The number of tokens generated in the output place depends on the edge's weight. If the weight is bigger than one it is displayed as shown in Figure 1, (*e* and *f)*. Figure 1 (*f*) represents the state of the system after the transition fired and generated tokes to the output place and back to the input place which also serves as an output place due to the edges directed from the input place to the transition and back. As mentioned above, transitions can be either immediate or timed. An immediate transition implies that an event happens Immediately as soon as all input places satisfy the enabling conditions — meaning that each input place has at least as many tokens as the weight of the corresponding edge. Once these conditions are met, the transition becomes enabled and fires immediately. Timed transitions have an associated delay that must pass before they can fire after they become enabled. Moreover, both an immediate and timed transition can serve as a reset transition, a transition that resets marking of some specified places to particular values when it fires (Prescott and Andrews, 2013). A reset transition is represented as green rectangular unlike an immediate transition.

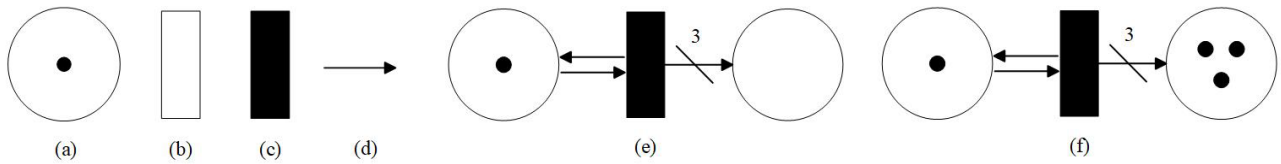


Figure 1. Petri Net features

The main PN firing rule is that all input places connected to a transition must have at least as many tokens as the corresponding edge weight is to make the transition enabled and fire it. Additionally, the PN technique can also make use of a prioritisation rule to determine the transition firing sequence. Transitions may be assigned a priority between x and y, and those with higher priority fire first if several transitions are due fire simultaneously. The PN presented in this paper uses a basic priority value of 500, with a priority range between 495 and 515.

Alongside the PN methodology, Monte Carlo simulation is used to perform repeated simulation runs, with random input time values uniformly sampled from predefined ranges that were defined based on normal distribution analysis for specified PN transitions. Monte Carlo simulation is a mathematical method used to forecast probable outcomes of unpredictable events. Additionally, Monte Carlo simulation refers to stochastic simulations, whether discrete, real-time, or a combination considering random variability in the simulated model (Bonate, 2001).

3. Problem statement

This section shows why such a huge loses of tread thickness occurs when re-profiling wheels and how this affects the wheelset lifespan. The term lifespan used in this paper relates to the wheelset utilisation that starts at its best condition after production until its functional failure when the wheelset tread reaches its lowest allowable limit. The wheelset flange thickness (F_{th}) and tread thickness (T_{th}) are the two main wheelset dimensions which directly influence ride safety and comfort. Both, the flange thickness degradation and tread thickness degradation can be conditionally divided into two parts. The first is natural wear, where the wheelset degrades due to its usual daily operation and is caused by wheel-rail contact. The second is artificial or affected wear, which occurs during re-profiling. The railway industry considers two types of wheelset re-profiling. The first type is corrective reprofiling which is fulfilled to eliminate sharpness, roughness, shelling, wheel flats, etc. on the wheelset surface. The second type is preventive reprofiling, which aims to restore the flange thickness to its required operational dimension and prevent an operational failure. The flange and tread thickness are evenly reduced by certain number of millimetres during corrective reprofiling. Generally, this depends on the damage that the wheelset surface has obtained within operation. On the contrary, the amount of the affected tread wear (ΔT_{th}) that happens during preventive re-profiling mainly depends on the actual flange thickness (F_{th1}) and the required flange thickness after re-profiling (F_{th2}) that a maintenance engineer intends to restore. Below, Figure 2 shows a locomotive wheelset, and its main parts and Figure 3 presents the wheelset profile shape before (solid line) and after reprofiling (dash line).

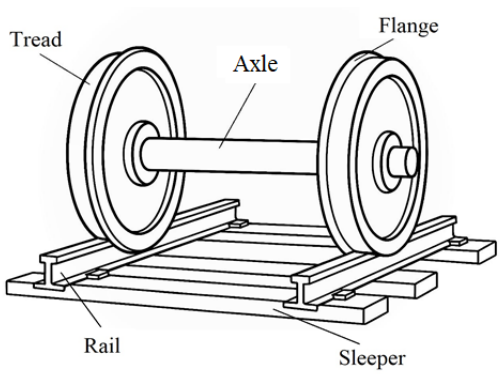


Figure 2. Locomotive wheelset

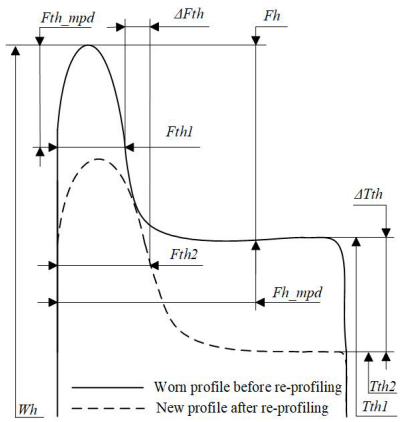


Figure 3. Wheelset profile

When preventive reprofiling takes place the worn tread thickness (T_{th1}) is reduced to the re-profiled tread thickness (T_{th2}), at the same time, the actual flange thickness (F_{th1}) is restored to the re-profiled flange thickness (F_{th2}) by the reduction of the tread thickness and the wheel heigh (W_h) through the reprofiling process. Thus, the increase in flange thickness (ΔF_{th}) represents the difference before and after re-profiling. Additionally, ΔT_{th} represents the loss of tread thickness which could be used for the potential operation of the wheelset.

This study investigates a conventional type of re-profiling and the Economic Tyre Turning (ETT) method mentioned above, where flange is restored only to its initial value and any acceptable value within the highest and lowest established limit, respectively. The results presented in this paper represent seven different maintenance strategies. In the main strategy, the wheelset flange, which initially measures 33.0 mm, is fully restored to its original value after reaching its critical limit. In the other strategies — 32.5 mm, 32.0 mm, 31.5 mm, 31.0 mm, 30.5 mm, and 30.0 mm — the flange is restored to these lower values, following

the ETT concept. This approach means that the flange is only restored to the established value defined by the specific maintenance strategy, rather than to its original thickness. Alongside mentioned data, the model assumes a 30-day inspection interval and considers corrective reprofiling every 60 days, until the flange thickness reaches 27.0 mm. The model accepts that beyond this point, corrective reprofiling is no longer performed, as regular preventive reprofiling will be carried out as soon as the flange thickness reaches 25.0 mm or lower, as identified during an inspection. The results show how the wheelset lifespan and other criteria will be affected using distinct input data.

4. Petri Net modeling

This section describes a PN framework designed for modelling and evaluation a single wheelset lifetime, numbers of inspections and reprofiling actions. It further proposes that the time during which the wheelset exceeds operational limits can be considered as a safety risk parameter, tracking the critical period when the wheelset operates outside allowable dimensions. Three main PN modules such as the wheelset degradation module, the inspection module, and re-profiling module are developed, however, only degradation (Figure 4) and re-profiling (Figure 5) modules are presented in this paper. Also, submodules like flange and tread inspection and failure are developed. Figure 4 (*a, b*) shows the process of the flange and tread thickness degradation. The locomotive wheelset chosen for this research has the highest flange thickness limit allowed for the operation that equals 33.0 mm and the lowest allowed operational limit equals 25.0 mm. However, the PN model considers that the flange can degrade lower than 25.0 mm due to untimely or delayed inspection and re-profiling as can happen in real life. The places from Flange_33.0 to Flange_24.0 represent a range of flange thickness states, with each place corresponding to a 0.1 mm interval, meaning that each discrete state covers a thickness range, for example, Flange_24.0 denotes the thickness range [24.0, 24.1] mm. Transitions from *Tf1* until *Tf90* represent the degradation events (time) after which the flange gets degraded by 0.1 mm. Referring to the PN rules and Monte Carlo simulation, each transition (*Tf1 – Tf90*) represents input data (days of wheelset operation) as following a normal distribution μ = 4.5 days, σ = 0.5 day. Hence, it means that the flange usually decreases by 0.1 mm over 4 -5 days of wheelset operation and this represents the natural flange wear. All further input parameters used in this paper have been chosen based on a comprehensive analysis of the flange and tread measurements of forty wheelsets for passenger locomotives. Additionally, it is worth noting that the standard measure of locomotive wheelset mileage is kilometres, however, due to the inspection intervals that are established as time (days), the locomotive mileage has been converted to days and represents 500 km of wheelset operation as one day.

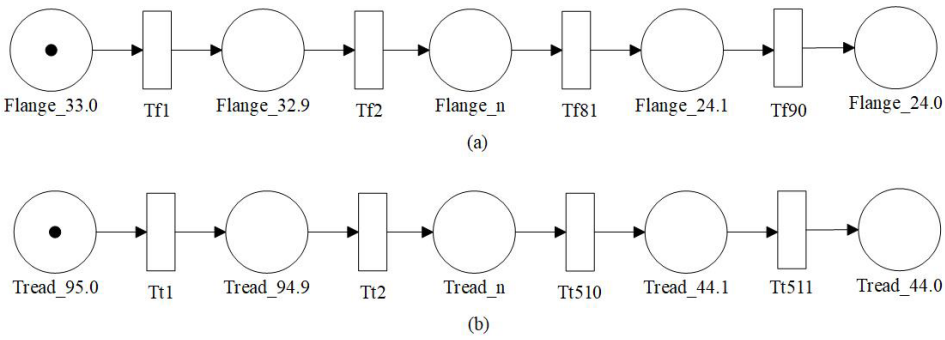


Figure 4. PN wheelset flange and tread degradation module

The same approach is used to describe the tread degradation process, where the highest tread thickness operational limit is 95.0 mm and the lowest is 45.0 mm. However, the transitions (*Tt1 – Tt511*) represent a normal distribution μ = 6.75 days with the standard deviation σ = 0.75 day. It means that the flange degrades faster than the tread. This is due to the lateral forces that act when the wheelset moves along curves, and additional load is placed on the wheel flange (Yang et al, 2025). Figure 5 shows how the reprofiling process is performed. In the reprofiling module, the place *Make_Flange_28.0* represents a maintenance strategy which is chosen for the simulation scenario. For this model, the state *Flange_25.0* and below were chosen as critical values which trigger preventive re-profiling actions. So, if the critical flange state was identified

within the inspection, then the wheelset re-profiling must occur in order to restore the flange thickness to the allowed operational value.

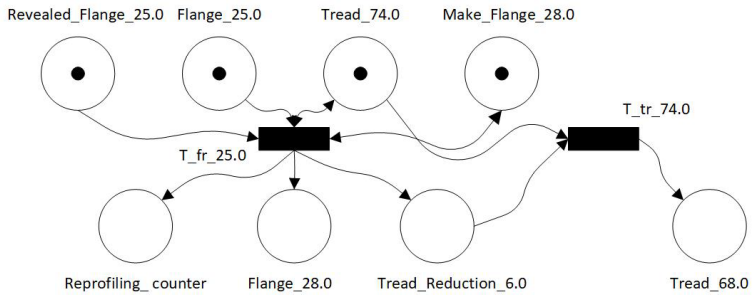


Figure 5. PN reprofiling module

In Figure 5 transitions such as *T_fr_25.0* and *T_tr_74.0* represent re-profiling events for both the flange and the tread respectfully. Transition *T_fr_25.0* fires as soon as the inspection process is finished and the critical flange value that was established for the simulation scenario is identified. Thus, once *Revealed_Flange_25.0* place received a token and inspection process is done, transition *T_fr_25.0* fires moving tokens from its input places to output places as *Reprofiling_counter* to count numbers of reprofiling, *Flange_28.0* which is the actual state of the flange after re-profiling, *Tread_Reduction_6.0* which represents number of centimeters the tread reduced by within reprofiling. Then, *T_tr_74.0* transition fires moving tokens from its input places *Tread_74.0* and *Tread_Reduction_6.0* to its output place *Tread_68.0* which represents the actual tread state after re-profiling. Since re-profiling process for flange and tread occurs simultaneously in real, transitions *T_fr_25.0* and *T_tr_74.0* represent the same re-profiling even, however, due to PN's modeling logic they fire in sequence with distinct priority set up. Once the re-profiling process is completed, inspection takes place to confirm that the wheelset condition meets railway standard requirements after the re-profiling. Following reprofiling, wheelset returns to operation and wheelset degradation process continues, however, after certain time wheelset operational failure occurs and the wheelset cannot be further operated due to its critical tread thickness value.

5. Results

As mentioned in the previous section, this study investigated one conventional — 33.0 mm, and six ETT maintenance strategies - 32.5, 32.0, 31.5, 31.0, 30.5, and 30.0 mm for a locomotive wheelset to evaluate the wheelset lifespan, number of inspections and reprofiling and time the wheelset is operated beyond the flange critical condition. Each simulation scenario was run 5000 time to obtain the results converged. The results shown in Figure 6 (*a*) present how the wheelset lifespan, number of inspections, corrective reprofiling, and preventive reprofiling correlate with each other depending on the applied maintenance strategies.

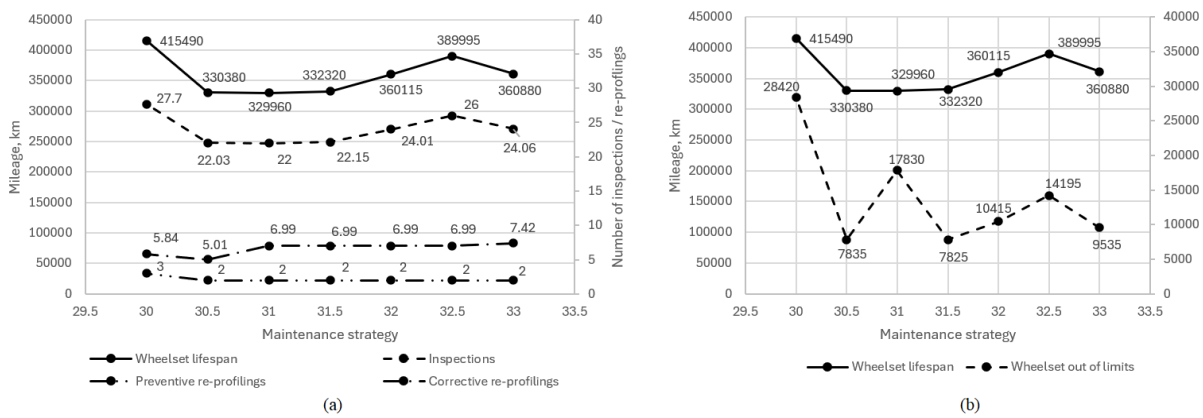


Figure 6. (a) Distribution of wheelset average mileage, number of inspections, corrective and preventive re-profiling depending on the maintenance strategy. (b) Distribution of wheelset average and out of limits mileage depending on the maintenance strategy.

Additionally, Figure 6 (*b*) shows time the wheelset being outside operational limits for each strategy. The result analysis shows that fluctuation of the wheelset lifespan value mainly caused by the time the wheelset spends being out of operational limits, which in its turn caused by established duration between inspections. Thus, for example, if for 33.0 mm conventional maintenance strategy re-profiling usually happens when 24.9 – 24.8 mm flange thickness identified within the inspection, then for 32.5 - ETT strategy this value is around 24.7 – 24.6 mm. Hence, the wheelset flange operating under the 32.5 mm strategy in comparison to the 33.0 mm strategy degrades slightly further before reprofiling occurs. This happens due to the shift in inspection timing when the maintenance strategy is changed while the initial flange thickness remains the same. As a result, after the first corrective reprofiling, when the strategy differs from the initial one 33.0 mm, inspections begin to occur at different points in time compared to the original 33.0 mm maintenance strategy. Consequently, reprofiling also occurs at different flange thickness values.

Conclusion

In this paper, a Petri net simulation model is proposed to evaluate maintenance strategies for a locomotive wheelset, allowing the investigation of their influence on wheelset performance. As results show, the choice of maintenance strategy notably influences the wheelset lifespan, particularly due to shifts in inspection timing and flange thickness restoration values. Strategies with lower flange restoration values lead to increased wheelset degradation before reprofiling, resulting in greater variability in the wheelset's operational lifespan, however, increasing the time the wheelset spends out of operational limits.

Considering future research, the first step is to investigate how other maintenance strategies will affect the wheelset lifespan and the time wheelset out of limits. Furthermore, since the existing model only considers the net operating time of the wheelset, the second step is to expand the model to the number of wheelset and locomotives considering the re-profiling machine and staff availability, locomotive downtime and a locomotive workshop load. Additionally, future work will explore how the time spent out of operational limits can be represented for risk assessment purposes.

References

Andrews, J. D., & Moss, T. R. (2002). *Reliability and risk assessment / by J.D. Andrews and T.R. Moss.* (2nd ed.). Professional Engineering Publishing.

Antonio Ramos Andrade, Julian Stow, Assessing the potential cost savings of introducing the maintenance option of 'Economic Tyre Turning' in Great Britain railway wheelsets, Reliability Engineering & System Safety, Volume 168, 2017, Pages 317-325, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2017.05.033.

Bonate, P. L. (2001). A brief introduction to Monte Carlo simulation. In Clinical Pharmacokinetics (Vol. 40, Issue 1, pp. 15–22). Adis International Ltd. https://doi.org/10.2165/00003088-200140010-00002.

Joaquim A. P. Braga, António R. Andrade, Data-driven decision support system for degrading assets and its application under the perspective of a railway component, Transportation Engineering, Volume 12, 2023, 100180, ISSN 2666-691X, https://doi.org/10.1016/j.treng.2023.100180.

Maozhenning Yang, Youpei Huang, Zhangxing Huang, Huijie Zhang, Yuanjin Ji, Lihui Ren, Optimization of abnormal wheel wear in articulated trams with independent wheelset, Journal of Engineering Research, 2025, ISSN 2307-1877, https://doi.org/10.1016/j.jer.2025.04.004.

Prescott, Darren & Andrews, John. (2013). A track ballast maintenance and inspection model for a rail network. Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability. 227. 251-266. 10.1177/1748006X13482848.

Rundong Yan, Sarah Dunnett, John Andrews, A Petri net model-based resilience analysis of nuclear power plants under the threat of natural hazards, Reliability Engineering & System Safety, Volume 230, 2023, 108979, ISSN 0951-8320, https://doi.org/10.1016/j.ress.2022.108979.

Saleh, Ali. (2023). An optimized asset management petri net model for railway sections. 285-291.

Wen Wu, Darren Prescott, Rasa Remenyte-Prescott, Ali Saleh, Manuel Chiachio Ruano, An asset management modelling framework for wind turbine blades considering monitoring system reliability, Reliability Engineering & System Safety, Volume 252, 2024, 110478, ISSN 0951-8320.

Zhang, Huixian & Wei, Xiukun & Qinlguan, Guan & Zhang, Wei. (2022). Joint Maintenance Strategy Optimization for Railway Bogie Wheelset. Applied Sciences. 12. 6934. 10.3390/app12146934.

Trust is Good, Monitoring is Better: FPGA- & TEE-Based Monitoring for Malware-Detection

Friederike Bruns*, Georg Gläser†, Florian Kögler†, Jonas Lienke†, Nithin R. Nanjundaswamy‡, Gregor Nitsche‡, Behnam R. Perjikolaei§, Jörg Walter§

*Carl von Ossietzky Universität Oldenburg, Germany, email: {first.last}@uol.de; †IMMS Institut für Mikroelektronik- und Mechatronik-Systeme gemeinnützige GmbH (IMMS GmbH), Ilmenau, Germany, email: {first.last}@imms.de; ‡DLR Institut für Systems Engineering, Oldenburg, Germany, email: {first.last}@dlr.de; §OFFIS e.V. Institut für Informatik, Oldenburg, Germany, email: {first.last}@offis.de

Abstract Ensuring trustworthiness in electronic systems is crucial to maintain safety and data integrity. Safety properties of robotic components are rigorously validated during development and, similarly, security requires ongoing monitoring during system operation as well. However, this monitoring must also safeguard its own components from tampering. We propose a novel runtime monitoring approach using application-specific monitors within an FPGA-based Trusted Execution Environment (TEE). To protect these monitors from supply chain attacks during design, fabrication, testing, or packaging, the TEE is programmed as the final step before deployment. The monitors are directly generated from formal constraint specifications established during the design and test phases. Our approach is demonstrated on a RISC-V-based System-on-Chip (SoC) for robotic applications, featuring a force sensor and a CAN-bus interface. We monitor the timing behaviour of hardware and software to detect malicious modifications affecting data transmission to a control unit. In an FPGA prototype, the monitors successfully identified hardware and software tampering. In real ASIC implementations, programming the TEE post-packaging ensures resilience against supply chain attacks.

This work was supported by the German Federal Ministry of Education and Research (BMBF) under funding codes 16ME0243K to 16ME0254.

1. Introduction

Ensuring the trustworthiness of electronic systems is increasingly critical due to risks from hardware faults, supply chain tampering, and environmental changes, which can compromise both functional and non-functional properties, such as timing or energy usage. Runtime verification can be implemented in software, hardware, or a combination of both. Software-based approaches, such as those by Do Tran et al. (2020); Havelund (2008); Tabakov & Vardi (2010), rely on finite-state machines or contract-based methods but introduce performance overhead by running alongside the system software. To address this, hardware-based solutions have emerged. For instance, Tracy et al. (2020) translate LTL formulas into FPGA automata, while Lu & Forin (2008) developed a compiler converting PSL specifications into Verilog. Decker et al. (2018) proposed real-time hardware-based monitoring on MPSoCs, and Nanjundaswamy et al. (2023) extended RISC-V to measure software execution time, though without considering execution time variability. Hybrid approaches, such as Solet et al. (2018), combine lightweight software instrumentation with hardware trace analysis but require variable mapping and incur significant overhead. Laurent et al. (2021) showed that hybrid techniques effectively detect hardware faults and software anomalies, especially in fault injection scenarios. However, most monitoring solutions neglect security threats like supply chain attacks or hardware Trojans, which Pan & Mishra (2022) highlight as critical concerns—especially AI-based Trojans that evade ML detection. Moreover, while runtime monitoring offers a way to detect deviations from nominal behaviour, it faces the same vulnerabilities as the systems it observes such as FPGA-based solutions like Rahman et al. (2024). Attacks may be introduced early in the design phase yet only manifest during operation, making detection challenging (Figure 1). This underscores the need for complementary, software-independent monitoring strategies. Compared to existing methods, our solution uniquely targets both functional faults and malicious modifications (e.g., HW-Trojans or firmware tampering), covers the full system lifecycle, and supports secure updates. The logical and physical isolation of monitors ensures their integrity, even in compromised environments.

We identify three core requirements: (i) monitoring both hardware and software behaviour, achievable only with hardware-based monitoring; (ii) updateability to adapt to changing system software; and (iii) logic isolation to ensure monitors cannot be detected or compromised by the system under observation. Potential threats considered include trojans that manipulate the software or hardware implementation and will lead to unexpected behaviour such as leaking internal data over communication links or executing additional functions, which will inevitably lead to different execution timing. While this work focuses on timing behaviour, approaches that observe the functional behaviour could be integrated in a similar way. We propose a dual-domain monitoring approach in Section 2. Lightweight hardware monitors are automatically generated from timing constraints, while firmware is observed using Statistical Timing Monitors (STMos), which compare execution time against learned timing profiles. To counter supply chain attacks, we embed an on-chip FPGA-based Trusted Execution Environment (TEE) for secure, out-of-band monitoring in Section 3. Our method is validated on a System-on-Chip with a RISC-V CPU and a manipulated CAN subsystem, where our monitors successfully detect malicious modifications and restore system integrity. Moreover, we compare our approach against existing techniques. Lastly, we discuss our results in Section 4.

2. Monitor Definition & Integration

We propose a runtime monitoring approach based on formally specified system behaviour, such as assume-guarantee contracts for timing properties using the MULTIC Timing Specification Language (MTSL) Böde et al. (2019). However, other specification languages or temporal logics could be used in a similar manner to express contracts, for example to adress functional behaviour. These constraints define the expected behaviour and are checked by monitors running inside a TEE, ensuring integrity even in compromised systems (Figure 1). One concern is the corruption of the monitors themselves, because a typical production flow consists of many supply chain participants, which can all be attack vectors. Figure 2 shows how our monitors are deployed post-production. Unlike conventional supply-chain-integrated monitors, ours are deployed post-production, enabling detection of tampered timing behaviour. This resides on two underlying assumptions: the trusted execution environment always executes monitors correctly, and there is a secure deployment/upgrade path for the monitors themselves. This is addressed via secure boot and cryptographic signatures.

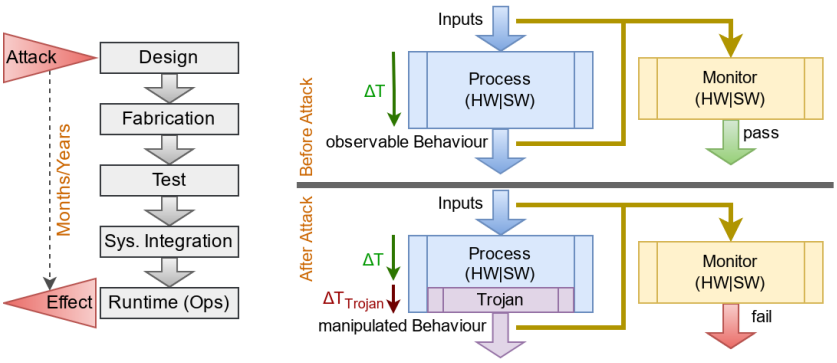


Figure 1. Monitoring (e.g. timing-fingerprints) to identify trojans that could be introduced at any stage of the supply chain.

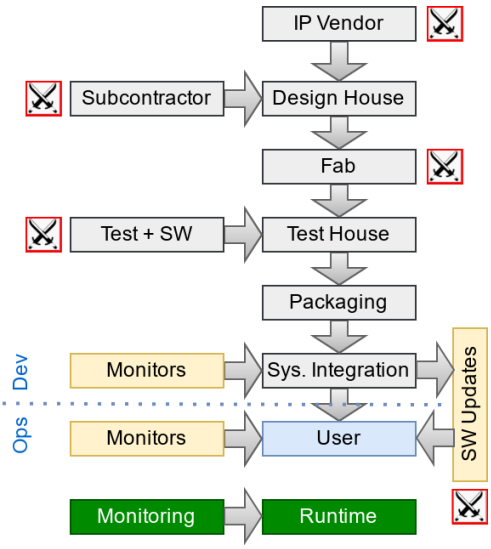


Figure 2. Integration of monitors at the end of development, bypassing multiple attack vectors (crossed swords).

Moreover, we distinguish between hardware and software monitoring, due to their differing timing characteristics:

For hardware, monitors operate at full clock speed to detect even single-cycle violations. Hardware IP-Monitoring (left Figure 3) includes event-specific observers that forward relevant signals to monitors, which verify compliance

with MTSL constraints. Depending on the hardware, observers are combinatorial or sequential (considering different signals) and, thus, it is impossible to give a generic definition of event observers and their logic. They are design specific. The MTSL verification algorithm within each hardware monitor includes a violation detection mechanism. Monitoring starts at event-triggered such as system startup or environmental stimuli at time $t = 0$ (representing physical time in clock cycles). Then, proceeds through Idle and Active states, where each subsequent request in the *Active* state resets the time to indicate the start of a new monitoring cycle. After evaluation of the timing condition and upon a detected violation, the monitor decides how to proceed. The monitors can handle recoverable (e.g., noisy input) and unrecoverable (e.g., critical violations) failures. This high-resolution, out-of-band monitoring enables low-overhead detection of rare timing anomalies, maintaining reliability. Consequently, Deviations from expected behaviour can be addressed promptly.

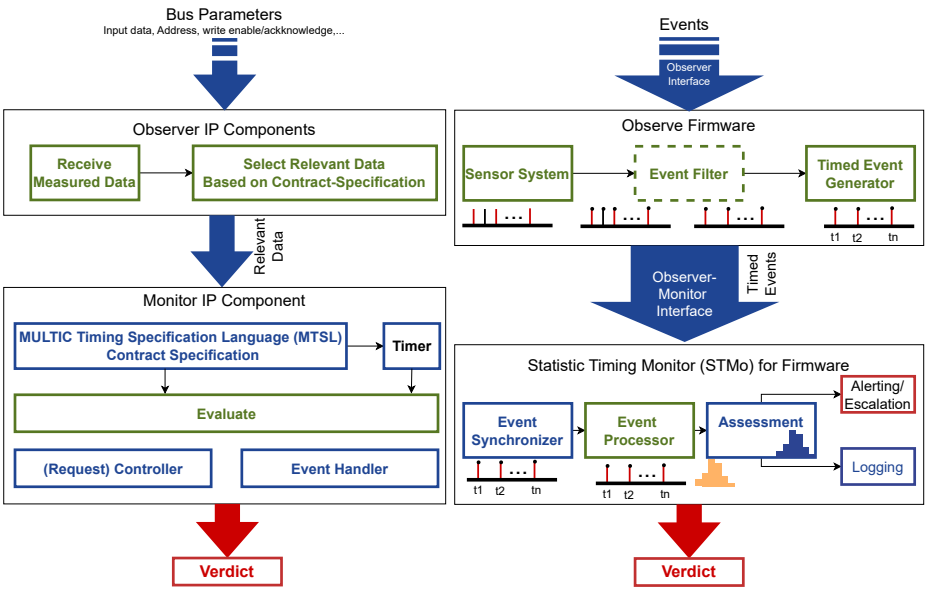


Figure 3. Monitoring hardware (left) and software (right).

On the other hand, timings for software on a modern CPU core are not necessarily fully deterministic, therefore we use a statistical approach. The developed Statistical Timing Monitors (STMo) system (Figure 3) effectively identifies modifications such as tampering that can affect functionality and security. Any modification (e.g., trojan insertion) will invariably change timing, which is captured as a shift Δt from the normal execution time t_{Norm}. Rather than relying solely on each single execution time, which can be influenced by factors like memory hierarchy, branch prediction, and I/O operations, STMo compares run-time timing statistics to a pre-defined histogram specification allowing for precise detection of software modifications. Here, an observer (integrated with a RISC-V core) timestamps software events using a synchronised clock. These timed events represent significant value- or state-changes of the software are filtered and passed to the STMo unit for synchronisation. The STMo builds a runtime histogram of execution timing and compares it to a reference histogram representing the expected behaviour within the assessment block. Deviations signal potential tampering and the assessment block triggers alerts or developer-defined recovery actions. All data is logged for traceability, enabling the back-tracing of any issues that may arise within the STMo system.

3. Case Study: Demonstrator ASIC

We present a RISC-V-based ASIC (Figure 4) designed for accurate force measurement in robotics, enabling collision detection via a Wheatstone bridge frontend and CAN interface. The chip includes a 32-bit RISC-V core, 16 kB SRAM, external QSPI NOR flash for program memory, and an eFPGA-based TEE attached via internal GPIOs for flexible runtime monitoring. The FPGA, built using the OpenFPGA toolchain Tang et al. (2019), features 200 LUTs to monitor processor activity. The FPGA bitstream is stored within the internal memory, allowing for rapid prototyping and testing.

For secure applications, access to the FPGA would need to be restricted —- either through JTAG or ROM-preloaded bitstreams. Fabricated in 180nm CMOS, the chip uses 155 886 standard cells (8.72 mm^2), with the FPGA accounting for 70 816 cells (1.5 mm^2) and 54% chip utilization. Estimated power consumption for the digital logic is 88 mW. A 'dummy Trojan' in the CAN-Controller demonstrates the monitor's capability to detect security threats effectively. We developed a custom firmware that retrieves and pre-processes data from a sensor before send to the robotic demonstrator's control unit via CAN, serving as a testbed for both hardware and software attack vectors.

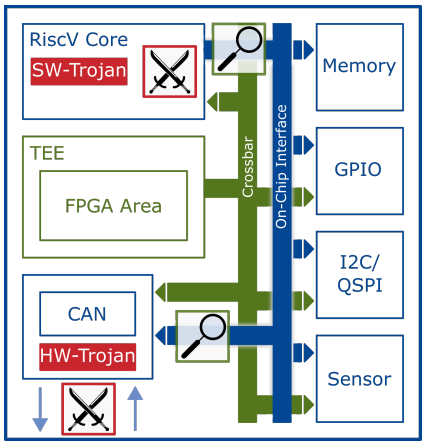


Figure 4. ASIC demonstrator with attack vectors (crossed swords) and monitors (lenses).

```
In each jumping window of length 100, with tolerance 0%:
reaction (RISC-V.GPIO.port[0]-port[7].0b00000001, RISC-V
    .GPIO.port[0]-port[7].0b00000010)
occurs within 39.08µs with symmetric jitter-distribution
     of {100% in [+0ns; -0ns]}
 /* Example for jitter distribution */
symmetric jitter-distribution of {15% in [-1.5ns to -1.0
    ns];
20% in [-1.0ns to -0.5ns]; 30% in [-0.5ns to +0.5ns];
20% in [+0.5ns to +1.0ns];15% in [+1.0ns to +1.5ns];}
```

Listing 1: STMo specification for untampered firmware.

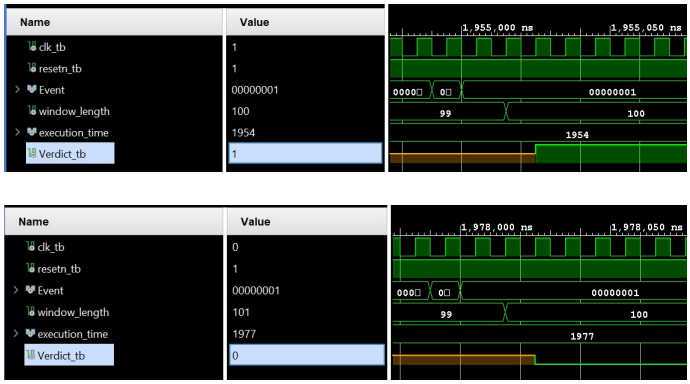


Figure 5. Untampered (top) vs. tampered (bottom) firmware.

Figure 6. Traces for a CAN IP transmission process where the data transmitted by cs_can2 trojanises cs_can in order to start an un-acknowledged sending process (tx_i).

To evaluate the STMo system, we tested firmware designed to send two-byte sensor data via CAN. A tampered version appended two extra bytes, simulating data exfiltration. The STMo system uses a baseline timing profile, derived from the untampered firmware, as a reference. Timing specifications are defined using Statistic Time Monitoring Specification Language (STMoSL), a domain-specific language for histogram-based constraints. While this paper does not provide a full overview of its syntax and semantics, yet, its features are exemplified by the specification of the untampered firmware. For instance (Listing 1), a jumping window of 100 executions with 0% tolerance captured consistent execution times of 39.08 µs (1954 clock cycles). This means, that after every 100 measurements the STMO system compares the measured run-time histogram with the STMoSL specification. STMoSL can also be used to specify timing distributions, e.g. due to jitter, using percentages of the number of measurements, and lower and upper interval-bounds relative to the previously specified mean. Characteristic events are specified by the port- & value-tuples in the reaction-phrase, referring to the RISC-V GPIO ports [0:7], and generated by the firmware using software annotation. Simulation with

both firmware versions showed STMo correctly validating the untampered case with verdict TRUE (top of Figure5) after 100 executions and flagging deviations in the tampered case with verdict FALSE (bottom of Figure5), proving its effectiveness in detecting timing discrepancies.

Hardware monitors and observers are implemented on the ASIC SoC FPGA Core (TEE), using GPIOs to capture CAN-Controller events. The objective is to avoid altering the original source code of the CAN-Controller in the ASIC design. We tested detection of a Hardware Trojan (HW-Trojan) integrated as a separate entity into the original CAN-Controller IP and controllable via a trojan_en input. Thus, Trojan logic can be bypassed when trojan_en is set to '0' without modifying the original CAN-Controller source code. The HW-Trojan listens on the CAN interface and, when triggered by a magic CAN packet, transmits previously recorded secret data (e.g. a cryptographic key) without authorisation or intervention by the software. *Aging* Monitors check timing and causality between request (event f) and send (event e). If a send occurs without a preceding request, the monitor flags a violation because of an MTSL aging pattern that specifies the backward delay: **Whenever** EventExpr **occurs then** EventExpr **has occurred within** Interval. Note that the interval detected by the pattern initiates upon the observation of the final element of an EventExpr. A waveform (Figure 6) shows two transmissions from cs_can2 to cs_can (tx2_i), where the second includes a Trojan activation code (8'h56) in register 8'd12. In the waveform, send_data_o and observer_go_tx are inputs to the cs_can monitor, while its outputs are o_pass and o_fail representing the monitor's verdict. To transmit the two sets of data the cs_can receives the send requests send_data_o2. With the trojanised data packet, cs_can transmits data (tx_i) without a proper send request. The monitor detects unauthorised transmission – send action (observer_go_tx = '1') occurs without a corresponding send request (send_data_o = '0') – and signals a violation (o_fail = '1'). This validates the monitor's ability to catch hardware-level threats.

Our runtime monitoring approach distinguishes itself from the related work by addressing multiple system vulnerabilities, including both hardware Trojans and malicious firmware, going beyond traditional software-focused solutions in terms of **target attacks** such as those by Tabakov & Vardi (2010); Havelund (2008). Unlike prior methods that primarily detect software faults, our solution provides comprehensive protection across hardware and software layers. Regarding the **time point of monitoring**, traditional approaches often focus solely on runtime monitoring, which can introduce latency and overlook hardware-level threats. In contrast, FPGA-based monitors allow us to use a uniform formal A/G contract specification across the entire life cycle. This ensures threats are detected early and reliably in both hardware and software **domains**, without incurring significant performance penalties. A major strength of our solution lies in the security and flexibility of the monitoring infrastructure: Isolation via TEE ensures the monitor's integrity, even if the main system is compromised. Dynamic **updatability** allows our monitors to evolve with emerging threats -— an advantage over fixed-function monitors such as those by Tracy et al. (2020). **Automated monitor generation** from constraint specifications accelerates deployment while maintaining correctness and security. Our monitors can observe both timing and data behaviour securely within the TEE (based on the constraint specification). Since individual participants of the supply chain are unaware of what's being observed, the risk of tampering is minimised enhancing **security**. This **logical separation** protects the monitor even in compromised environments. Our hybrid approach offers comprehensive runtime verification that ensures secure, flexible, and efficient monitoring, strengthening system resilience against both hardware and software vulnerabilities.

4. Conclusion

Unintended runtime behaviour in embedded systems, such as data leakage or malfunctions, can result from hardware Trojans or malicious firmware which can be introduced during design or manufacturing but which are often detected much later. We address this by a monitoring approach covering both hardware and software, and using TEEs for secure, out-of-band monitoring. The key features include comprehensive monitoring across hardware and software, updateability through automated constraint-based monitor generation, and logic isolation to ensure security even in compromised systems. Lightweight hardware monitors, isolated within FPGA-based TEEs, focus on timing behaviour, while STMoS handle software anomalies. The approach was successfully validated in an SoC with a RISC-V-based CAN-Controller,

detecting hardware and software tampering, including data exfiltration and altered data streams. This hybrid solution offers robust, real-time system integrity protection.

Future work could expand on this approach regarding the detection capabilities to cover more sophisticated threats, such as AI-based hardware Trojans or multi-vector attacks that combine both hardware and software vulnerabilities. Moreover, scalable deployment of these hybrid monitors could be addressed considering highly distributed systems, such as in automotive networks or industrial control systems, where centralised monitoring may not be feasible.

References

Böde, E., Damm, W., Ehmen, G., Fränzle, M., Grüttner, K., Ittershagen, P., Josko, B., Koopmann, B., Poppen, F., Siegel, M. & Stierand, I. (2019), MULTIC-Tooling, *in* 'FAT-Schriftenreihe 316', Forschungsvereinigung Automobiltechnik e.V. (FAT).

Decker, N., Dreyer, B., Gottschling, P., Hochberger, C., Lange, A., Leucker, M., Scheffel, T., Wegener, S. & Weiss, A. (2018), Online Analysis of Debug Trace Data for Embedded Systems, *in* 'Design, Automation and Test in Europe Conference and Exhibition (DATE)'.

Do Tran, D., Walter, J., Grüttner, K. & Oppenheimer, F. (2020), Towards Time-Sensitive Behavioral Contract Monitors for IEC 61499 Function Blocks, *in* 'IEEE Conference on Industrial Cyber Physical Systems (ICPS)'.

Havelund, K. (2008), Runtime Verification of C Programs, *in* '20th International Conference on Testing of Software and Communicating Systems', Springer-Verlag, Berlin, Heidelberg.

Laurent, J., Deleuze, C., Pebay-Peyroula, F. & Beroulle, V. (2021), 'Bridging the Gap between RTL and Software Fault Injection', *J. Emerg. Technol. Comput. Syst.* **17**(3).

Lu, H. & Forin, A. (2008), 'Automatic Processor Customization for Zero-Overhead Online Software Verification', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **16**.

Nanjundaswamy, N. R., Nitsche, G., Poppen, F. & Grüttner, K. (2023), RISC-V timing-instructions for open time-triggered architectures, *in* '53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)', IEEE, Porto, Portugal.

Pan, Z. & Mishra, P. (2022), Design of AI Trojans for Evading Machine Learning-based Detection of Hardware Trojans, *in* 'Design, Automation & Test in Europe Conference & Exhibition (DATE)'.

Rahman, M. M. M., Tarek, S., Azar, K. Z., Tehranipoor, M. & Farahmandi, F. (2024), 'The road not taken: efpga accelerators utilized for soc security auditing', *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* .

Solet, D., Pillement, S., Béchennec, J.-L., Briday, M. & Faucou, S. (2018), HW-based Architecture for Runtime Verification of Embedded Software on SoPC systems, *in* 'NASA/ESA Conference on Adaptive Hardware and Systems (AHS)'.

Tabakov, D. & Vardi, M. Y. (2010), Optimized Temporal Monitors for SystemC, *in* H. Barringer, Y. Falcone, B. Finkbeiner, K. Havelund, I. Lee, G. Pace, G. Roşu, O. Sokolsky & N. Tillmann, eds, 'Runtime Verification', Springer-Verlag, Berlin, Heidelberg, pp. 436–451.

Tang, X., Giacomin, E., Alacchi, A., Chauviere, B. & Gaillardon, P.-E. (2019), OpenFPGA: An Opensource Framework Enabling Rapid Prototyping of Customizable FPGAs, *in* '29th International Conference on Field Programmable Logic and Applications (FPL)'.

Tracy, T., Tabajara, L. M., Vardi, M. & Skadron, K. (2020), Runtime Verification on FPGAs with LTLf Specifications, *in* 'Formal Methods in Computer Aided Design (FMCAD)'.

Robust Fault Detection with Incomplete Data: A Weakly Supervised Approach

F. Waldhauser*, H. Boukabache**, M. Dazer***, and D. Perrin****

** European Organization for Nuclear Research (CERN), Geneva, Switzerland and Institute of Machine Components, University of Stuttgart, Stuttgart, Germany, email: felix.johannes.waldhauser@cern.ch*
*** European Organization for Nuclear Research (CERN), Geneva, Switzerland, email: hamza.boukabache@cern.ch*
**** Institute of Machine Components (IMA), University of Stuttgart, Stuttgart, Germany, email: martin.dazer@ima.uni-stuttgart.de*
***** European Organization for Nuclear Research (CERN), Geneva, Switzerland, email: daniel.perrin@cern.ch*

Abstract. Efficient and robust fault detection in PHM typically depends on high-quality run-to-failure data, yet such labeled historical data is often scarce. While unsupervised algorithms trained on data assumed to represent normal operation (such as early-stage operation data) can yield good performance, they may be limited by contamination or incomplete coverage of normal states, related to unseen operating scenarios and conditions. This study investigates the effects of incomplete coverage of normal scenarios on fault detection performance. To enhance robustness against such incompleteness without making assumptions on the dataset's composition, employing weakly supervised anomaly detection by incorporating few labeled fault samples is proposed. Evaluations on degradation testing data show that providing an autoencoder with a few fault labels significantly improves detection performance compared to a fully unsupervised autoencoder. This improvement is shown to be robust under various conditions, including changes in the number and quality of labeled fault samples and variations in the training data's composition. The proposed approach realistically simulates incomplete coverage through selective dataset composition and builds on weak supervision to enhance fault detection, making it particularly suitable when only field data—with no information on its composition—and few labeled fault samples are available.

1. Introduction

Robust fault diagnostics is central for successful Prognostics and Health Management (PHM) in industrial equipment. Yet training data-driven models typically demands extensive labeled run-to-failure data (Arias Chao *et al.*, 2021), which is often scarce due to preventive maintenance practices and high labeling costs. In contrast, large, unlabeled datasets representing normal behavior—such as data from early operation—are usually available and can be exploited by anomaly detection algorithms. In fault detection, anomaly detection algorithms aim to identify failure precursors indicative of system degradation and aging. Yet, if the training data incompletely covers normal scenarios, new but normal operating conditions may be misclassified as faults.

The central hypothesis in this study is that, if, in addition to unlabeled field data, few labeled fault samples are available, these labeled samples can be used in weakly supervised anomaly detection algorithms to increase the robustness to incomplete normal coverage of training data. This approach comes at the advantage of not requiring domain labels or making assumptions on the composition of normal data taken from field operation. To reflect realistic data-availability constraints, experimental data from analog electronic circuits is used to selectively construct training sets, which deliberately omit certain normal states. The results demonstrate that a weakly supervised autoencoder, when supplied with few fault samples, maintains similar performance across both seen and unseen normal scenarios, thus increasing fault detection robustness, while an unsupervised algorithm frequently misclassifies unseen normal data as anomalous. This study reflects a common industrial scenario—unlabeled data from early operation with unknown normal coverage, complemented by a few labeled fault samples from hardware experiments or expert labeling—and offers valuable insights into enhancing fault detection robustness and optimizing the use of labeled fault samples for fault diagnosis.

2. Related Work

Appropriate dataset composition is crucial for anomaly detection to ensure realistic training and evaluation. In classical anomaly detection, normal-state data is often split randomly between training, validation, and testing. However, fault detection studies require a more selective dataset construction that considers data availability

and underlying assumptions. For example, the widely used C-MAPSS dataset is pre-split by units to better reflect real-world conditions and evaluate transferability (Ramasso and Saxena, 2020).

In addition to the transfer between units of a system, handling different operating scenarios and conditions is also challenging for fault detection algorithms. Changes in operating conditions can shift data distributions and potentially mask fault characteristics thus affecting fault detection robustness and performance (Zio, 2022; Chao *et al.*, 2023). As industrial systems evolve, training data cannot be assumed to represent all normal states (Rahat *et al.*, 2022), which is referred to here as incomplete normal coverage. To address this, Hu et al. (2017) propose an adaptive framework that detects concept drifts and updates models and features accordingly, but it requires labeled data from known conditions, which is not available here. Alternatively, a hybrid framework using physics-based model calibration compensates for condition shifts (Chao *et al.*, 2023), yet it relies on a physical model not available for the system-level fault detection in analog electronics addressed in this study.

In contrast to the studies mentioned above, the proposed approach relies on weakly supervised learning making use of unlabeled field data and few labeled fault samples. Several frameworks for weakly supervised anomaly detection have been proposed. An anomaly scoring network in an end-to-end learning framework is proposed as DevNet in Pang *et al.* (2019), enforcing distribution differences between normal and anomalous data. However, DevNet requires normal distribution of anomaly scores and only considers samples with an anomaly score up to a certain threshold. To overcome this limitation, Kwon *et al.* (2023) propose a two-stage algorithm with an adapted loss function based on exponential deviation loss (EDL) and use an autoencoder instead of an anomaly scoring network for anomaly score generation. While most related studies using weakly supervised learning for fault detection focus on achieving stronger discrimination between normal and fault states (Pang *et al.*, 2019; Castellani *et al.*, 2021; Kwon *et al.*, 2023), this study explicitly aims at reducing misclassification of unseen normal data and thus increasing fault detection robustness to changes in operating conditions.

Typically, autoencoder-based classification of samples into normal or anomalous relies on comparing the reconstruction error to a threshold, such as the 90th percentile of reconstruction errors on the training dataset, while samples exceeding this threshold are classified as anomalies (Beggel *et al.*, 2020).

Previous studies on the presented use case of system-level fault detection in analog electronic circuits—using synthetic simulation datasets—revealed that using labeled samples from degradation experiments is preferable to manual expert labeling for informing fault detection algorithms (Waldhauser *et al.*, 2023).

3. Experiments and Dataset Composition

The source dataset used in this study originates from testbench experiments performed to study effects of component aging on analog electronics. The selected electronic board is the power supply board (see Figure 1) of CERN's next-generation radiation monitoring system CROME (Boukabache et al., 2017). It is a safety-critical instrumentation with high failure detectability and fail-safe functionality. Detection of internal faults, e.g., caused by hardware degradation, leads to interruption of accelerator operation until functionality of safety instrumentation is fully restored. The resulting high downtime costs necessitate condition-based maintenance through robust fault detection at system level, using output voltage signals to monitor component degradation throughout the circuit.

In the absence of historical failure data, simulations and hardware experiments are performed to study effects of capacitor degradation on the available monitoring signals. The dataset used in this study contains the results from experiments of one electronic board evaluating the degradation effects of a tantalum electrolytic capacitor C48 on a 5 V voltage signal (see Figure 1). To overcome the challenge of selective aging acceleration in a single component, capacitor aging is imitated through hardware modifications and the substitution of the C48 capacitor with capacitors of lower capacitance or higher series resistance. The resulting three hardware configurations are referred to as default (C48 with default properties), lowC (C48 with reduced capacitance), and highESR (C48 with increased series resistance) (see Figure 1). Each configuration is replicated several times to mitigate effects arising from component replacement, tolerances and the soldering process. These replicates are identified by a unique RepID, e.g., lowC #1, lowC #2, etc. (see Figure 2). Three RepIDs are available for the default hardware configuration and two RepIDs are available for each lowC and highESR. Data of the 5 V signal is collected with an oscilloscope (Figure 1), filtered, and processed to extract a complete

set of statistical features used for signal classification (Tsui et al., 2015): mean, standard deviation, geometric curve length, kurtosis, skewness, root mean square (RMS) and number of zero crossings. Both failure modes (lowC and highESR) primarily influence the standard deviation and curve length features. Hence, further evaluations focus on these metrics, though all features are used for fault detection. The effects are illustrated in Figure 2 for normalized data, where marker styles differentiate hardware configurations.

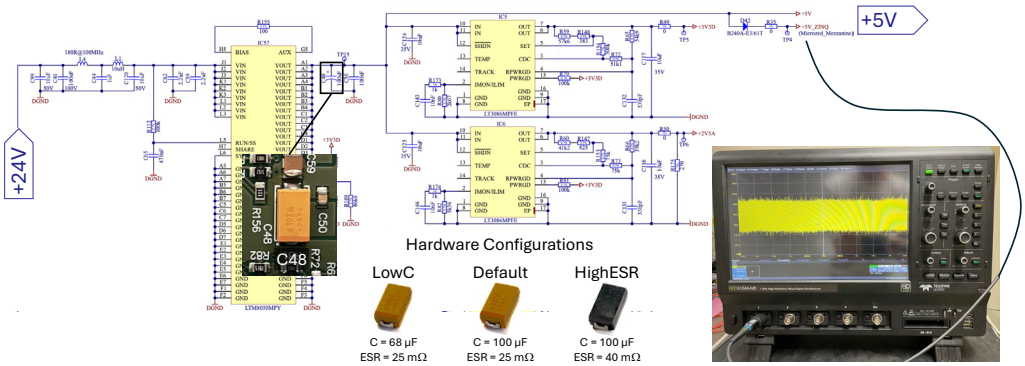


Figure 1: Schematic of the power supply board highlighting the location of capacitor C48, the three hardware configurations (default, lowC, and highESR) used to emulate capacitor degradation, and the oscilloscope measurements of the 5V output signal.

Instead of random sampling, the unlabeled training dataset X_{inc} is selectively composed based on the measurement replications indicated by their RepID. To simulate incomplete coverage, only a subset of the three default RepIDs is used to construct X_{inc}. Two settings are defined: scenario normalS1 (data from default #1 and #2) or scenario normalS2 (data from default #3), with samples randomly drawn to achieve a predefined dataset size of 15,000 samples. Labeled fault samples are similarly selected based on specified RepIDs and quantities. Figure 3 illustrates the default configuration: 250 fault samples are randomly drawn from both lowC #1 and highESR #2, and normalS1 is selected for X_{inc}. The weakly supervised learning dataset, X_{wsl}, is then formed by combining X_{inc} with these labeled fault samples, representing a scenario with unlabeled field data of incomplete normal coverage and a limited number of labeled fault samples. All data samples not used in X_{wsl} are assigned to the test dataset X_{test} (see Figure 3), which is thus also affected by the random sample selection for X_{inc} and X_{wsl}. The binary label vectors y_{field}, y_{wsl}, and y_{test}, use 0 for the default hardware configuration and 1 for lowC or highESR.

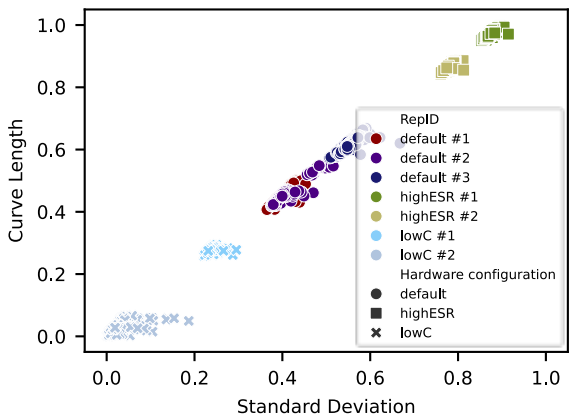


Figure 3: Scatterplot of full dataset with marker styles indicating hardware configurations and marker colors indicating different RepIDs.

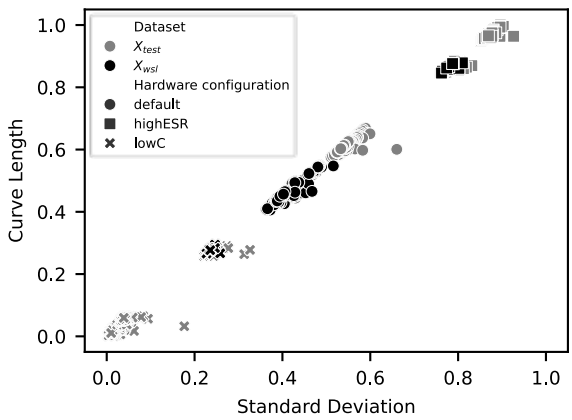


Figure 2: Scatterplot illustrating the default setting of the X_{wsl} dataset with incomplete normal coverage and few fault samples.

4. Fault Detection Algorithms and Benchmarking

To evaluate whether incorporating a few labeled fault samples enhances fault detection robustness under incomplete normal coverage, an unsupervised autoencoder (USL-AE) trained solely on the unlabeled X_{inc} dataset is compared with a weakly supervised autoencoder (WSL-AE) trained on X_{wsl}. The USL-AE is trained in one stage with mean squared error loss (MSE). The WSL-AE is trained in two-stages, as proposed in Kwon

et al. (2023): first unsupervised training on X_{inc} with MSE loss, and then training on X_{wsl} and y_{wsl} using the EDL loss function. This approach allows weakly supervised training without making additional assumptions about dataset composition or operating conditions, except that X_{inc} is fault-free. The autoencoder performance is evaluated on X_{test} using the accuracy score. Autoencoder predictions are extracted based on the MSE reconstruction error with the classification threshold set at the 95th percentile of reconstruction errors for X_{inc}. Hyperparameter tuning is performed following a Monte Carlo cross-validation process, where model performance is averaged over multiple training runs. The Optuna framework is employed using a Bayesian optimization method to identify the hyperparameters that maximize average performance on the test dataset X_{test} under default configuration of X_{wsl} (as shown in Figure 3). The resulting optimal hyperparameter set is then fixed and used consistently across all benchmarking experiments.

The performances of USL-AE and WSL-AE are evaluated on different compositions of X_{wsl} to test the robustness of the hypothesis that the WSL-AE algorithm can benefit from few labeled fault samples. First, all possible combinations of RepIDs used as source for labeled faults are evaluated under both normalS1 and normalS2 settings. Then, the effect of varying numbers of labeled faults is evaluated for the default X_{wsl} setting. For all studies, performance is quantified as the mean accuracy score from 200 training runs. To better assess the impact of incomplete normal coverage, USL-AE and WSL-AE are evaluated for reference on a scenario of full normal coverage, with X_{inc} constructed by randomly sampling from all available normal data samples.

5. Results

If trained on the reference scenario of full normal coverage, USL-AE results in a mean accuracy of 98.26% while WSL-AE results in a mean accuracy of 95.71%. Hence, in this case, the unsupervised algorithm outperforms the weakly supervised algorithm while both show high classification performance.

In contrast, WSL-AE achieves higher accuracy for all tested configurations of incomplete coverage. Figure 4 illustrates the results for various configurations of labeled fault data sources and different Xinc compositions for both USL-AE and WSL-AE. Although USL-AE training is not affected by the choice of fault samples, its performance is assessed for each configuration individually as the test set composition changes as well (shown in parentheses in Figure 4). While classification performance is generally lower for normalS2 than for normalS1 for both algorithms, WSL-AE outperforms USL-AE across all fault sample configurations. In the default labeled fault setting (using lowC #1 and highESR #2 as source for labeled fault samples), WSL-AE achieves approximately 6% higher accuracy than USL-AE for both normalS1 and normalS2 settings. Although this performance gain is generally consistent across configurations, the results vary for normalS1 depending on the selected lowC fault sample source, with lowC #1 yielding higher performance.

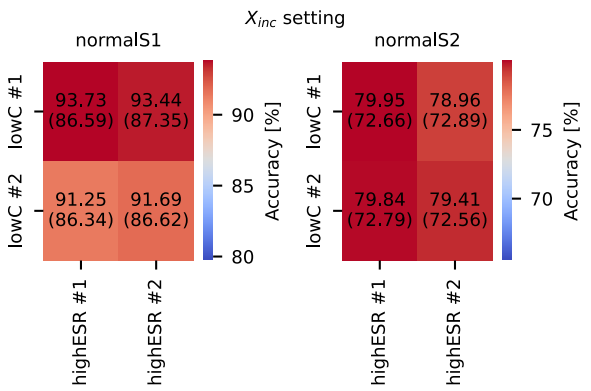


Figure 4: WSL-AE accuracy for different dataset compositions, with USL-AE accuracy in parentheses. The color bar is centered on mean USL-AE performance for the respective X_{field} settings; red indicates higher WSL-AE accuracy, and blue indicates lower.

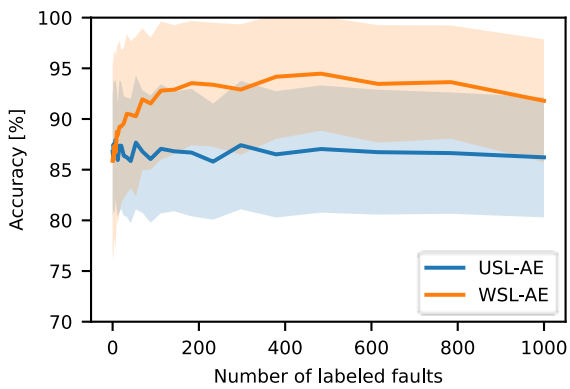


Figure 5: Autoencoder performance vs. number of labeled faults used to construct X_{wsl}. Shaded regions show ±1 standard deviation of the accuracy across 200 trainings.

Figure 5 presents the performance of WSL-AE and USL-AE across varying numbers of labeled fault samples, with all other X_{wsl} settings kept at their default values. As expected, the USL-AE performance remains

largely independent of the number of labeled faults, which affects the composition of X_{test} used for performance evaluation. In contrast, WSL-AE outperforms USL-AE starting from 10 labeled faults and reaches a stable performance level between 200 and 800 labeled samples.

6. Discussion

To interpret the results, it should be noted that anomaly detection with autoencoders ultimately relies on a distance metric between original and reconstructed data which is also used for classification into normal and non-normal based on a specified threshold. Hence, analyzing how X_{wsl} covers the feature space (e.g., data cluster locations and distances) for the different configurations can help explain performance differences.

The results demonstrate that WSL-AE is more robust to incomplete normal coverage. While USL-AE performs better under full normal coverage, its accuracy drops sharply for incomplete scenarios, making it unsuitable for most practical scenarios where complete coverage of normal states cannot be guaranteed. In contrast, WSL-AE maintains higher accuracy in these cases by learning a more generalized representation of normal states. Analysis of the reconstructed test dataset X_{test} and the distribution of reconstruction errors reveals that WSL-AE produces higher errors for normal data yet performs similarly on seen and unseen normal states, allowing correct classification of new normal states. In contrast, USL-AE's reconstruction errors differ notably for seen and unseen normal states, causing misclassification of unseen normal data. Exemplary distributions of reconstruction errors are shown in Figure 6.

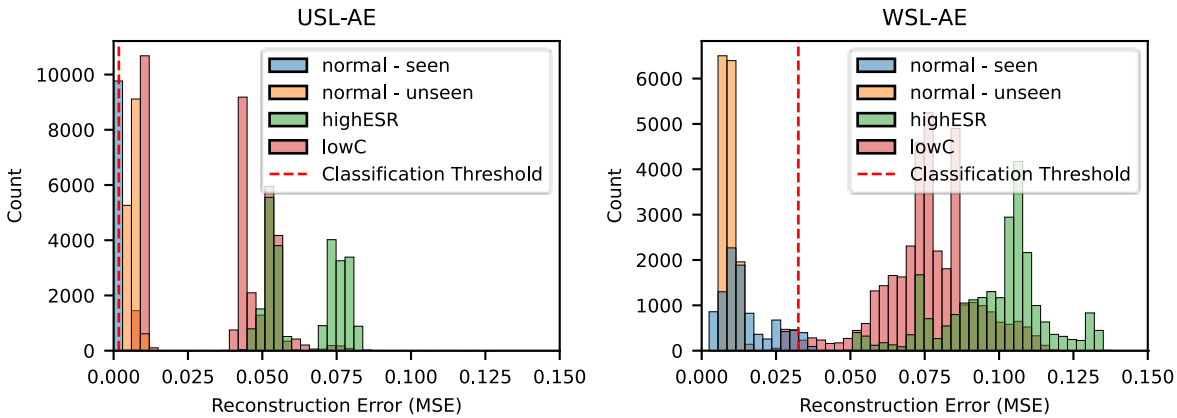


Figure 6: Exemplary distribution of reconstruction errors for different data types in X_{test}, comparing USL-AE and WSL-AE. "Seen" / "unseen" refers to whether data samples were used or not during training.

The difference in performance between normalS1 and normalS2 can be attributed to the lower coverage of normal states in the feature space for normalS2 (covering only default #3), making it more challenging to generalize unseen normal states. Furthermore, performance differences between WSL-AE and USL-AE are generally consistent across all fault sample configurations, varying only with the selected lowC fault sample source when X_{inc} is constructed under the normalS1 setting. The results suggest that samples closer to normal states (lowC #1) yield better results than those more distant (lowC #2). One possible explanation is that clear decision boundaries are harder to construct when fault samples are far from the normal distribution. Additionally, the hyperparameter optimization was conducted using lowC #1 and highESR #2—the data clusters closest to normal data—which may explain the lower performance when deviating from this configuration, which is also true for differences between normalS1 and normalS2. The effect of the number of labeled fault samples should be evaluated with respect to the similarity of those data samples. It appears that beyond a certain threshold, additional fault samples do not further improve performance, indicating a saturation point where no new information is gained. Hence, model performance can possibly be further enhanced by incorporating more fault samples from different data clusters—thus providing more diverse information. However, when the number of labeled faults exceeds 800, WSL-AE performance declines—likely because the increased supervision level (compared to the 250 labeled samples used during hyperparameter tuning) causes the model to overfit on labeled fault samples, thereby reducing its generalization capability on normal data.

Since labeled fault samples and test samples come in this study from the same electronic board, it remains to be investigated whether WSL-AE would still outperform USL-AE when labeled samples are sourced from different units. Validating this transferability would broaden practical applicability by enabling labeled samples from one unit to improve fault detection on another.

7. Summary and Outlook

Comparison of weakly supervised and unsupervised anomaly detection algorithms on a realistic dataset obtained from hardware experiments demonstrated that weakly supervised fault detection can benefit from labeled fault samples to increase robustness to incomplete normal coverage in training data. The weakly supervised autoencoder outperforms the unsupervised version in all tested cases. This performance increase is consistent across various scenarios of incomplete coverage and remains robust despite differences in the quality and number of labeled fault samples provided. The presented approach ensures robust fault detection when only unlabeled field data with unknown composition and few labeled fault samples are available.

Further studies should investigate the transferability between system units and assess the varying informative value of different labeled fault samples. This could broaden the applicability and provide insights into transitioning from real to synthetic fault samples in order to reduce experimental efforts.

Acknowledgement

This work has been sponsored by the Wolfgang Gentner Programme of the German Federal Ministry of Education and Research (grant no. 13E18CHA).

References

Arias Chao, M. *et al.* (2021) Aircraft Engine Run-to-Failure Dataset under Real Flight Conditions for Prognostics and Diagnostics, *Data* **6**(1), 5.

Beggel, L., Pfeiffer, M. and Bischl, B. (2020) Robust Anomaly Detection in Images Using Adversarial Autoencoders, in *Machine Learning and Knowledge Discovery in Databases* (U. Brefeld et al., Eds.). Cham: Springer International Publishing (Lecture Notes in Computer Science), pp. 206–222.

Boukabache, H. *et al.* (2017) Towards a novel modular architecture for CERN radiation monitoring, *Radiation Protection Dosimetry* **173**(1–3), 240–244.

Castellani, A., Schmitt, S. and Squartini, S. (2021) Real-World Anomaly Detection by Using Digital Twin Systems and Weakly Supervised Learning, *IEEE Transactions on Industrial Informatics* **17**(7), 4733–4742.

Chao, M.A. *et al.* (2023) Hybrid deep fault detection and isolation: Combining deep neural networks and system performance models, *International Journal of Prognostics and Health Management* **10**(4).

Hu, Y. *et al.* (2017) A Systematic Semi-Supervised Self-adaptable Fault Diagnostics approach in an evolving environment, *Mechanical Systems and Signal Processing* **88**, 413–427.

Kwon, M.-S. *et al.* (2023) Autoencoders with exponential deviation loss for weakly supervised anomaly detection, *Pattern Recognition Letters* **171**, 131–137.

Pang, G., Shen, C. and Van Den Hengel, A. (2019) Deep Anomaly Detection with Deviation Networks, in *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. Anchorage AK USA: ACM, 353–362.

Rahat, M. *et al.* (2022) Domain Adaptation in Predicting Turbocharger Failures Using Vehicle's Sensor Measurements, *PHM Society European Conference* **7**(1), 432–439.

Ramasso, E. and Saxena, A. (2020) Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets, *International Journal of Prognostics and Health Management* **5**(2).

Waldhauser, F. *et al.* (2023) Integrating System Knowledge in Unsupervised Anomaly Detection Algorithms for Simulation-Based Failure Prediction of Electronic Circuits, in *Proceedings of the 19th International Conference on Accelerator and Large Experimental Physics Control Systems*. JACoW Publishing, 249–256.

Zio, E. (2022) Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice, *Reliability Engineering & System Safety* **218**, 108119.

Prognostic-Driven Energy Management and Maintenance Strategies for Multi-Stack Fuel Cells with Stochastic Degradation under Dynamic Loads

M. Houjayrie*, C. Cadet**, C. Bérenguer***

*Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: mouhamad.houjayrie@.grenoble-inp.fr

**Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: catherine.cadet@grenoble-inp.fr

***Université Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, France, email: christophe.berenguer@grenoble-inp.fr

Abstract This paper presents a Prognostics and Health Management (PHM) framework for multi-stack fuel cell systems operating under deterministic, dynamic load conditions, addressing the durability challenges of Proton Exchange Membrane Fuel Cells (PEMFCs). A stochastic degradation model, based on a non-homogeneous Gamma process, employs internal resistance as a health indicator to estimate the Remaining Useful Life (RUL) distribution through Monte Carlo simulations. An Energy Management Strategy (EMS) is proposed to optimize load allocation in a two-stack PEMFC system, achieving a 6.4% extension in stack lifetime compared to an average load-split strategy. The framework incorporates predictive maintenance, using the RUL distribution to inform replacement decisions. The results highlight the potential of integrating prognostics, energy management, and maintenance strategies to enhance durability, reliability, and cost-efficiency. This work lays a solid foundation for a prescriptive maintenance approach in multi-stack fuel cell systems.

1. Introduction

Multi-stack fuel cell systems are well-suited for transportation applications due to their compact design, high power density, low operating temperature, and zero carbon emissions. Despite these advantages, high costs and limited durability hinder their widespread adoption. In transportation contexts, variable load conditions accelerate Proton Exchange Membrane Fuel Cell (PEMFC) degradation, reducing their Remaining Useful Life (RUL). Prognostics and Health Management (PHM) provides a robust framework to address these challenges by predicting system risks, optimizing control and maintenance schedules, enhancing safety and reliability, extending lifespan, and reducing operational costs. Central to PHM, prognostics focuses on extracting health indicators, modeling degradation trends, and estimating RUL to support effective system management.

Much of the existing literature examines PEMFC degradation under constant or quasi-constant load conditions, often using voltage as a direct health indicator (Liu et al., 2017). However, under variable load conditions, time-varying operations render voltage unreliable for assessing degradation. Instead, physical quantities such as internal resistance, Electrochemical Surface Area (ECSA) Ao et al. (2020), and membrane thickness Liu et al. (2019) are employed, or model parameters are used as indirect health indicators Bressel et al. (2016). While these parameters can reflect degradation trends, they are not always physical quantities and may not suit load allocation in multi-stack systems. Moreover, energy management studies often focus on hybrid systems with batteries Garcia et al. (2010) or prioritize hydrogen consumption over durability. Few address durability and lifetime constraints in multi-stack systems Bankati et al. (2024) and Zuo et al. (2024), with the latter assuming homogeneous degradation—a hypothesis unlikely to hold in practice. To address these gaps, this paper proposes a dynamic load-dependent stochastic degradation model using internal resistance as a health indicator, modeled via a non-homogeneous gamma process to account for aging and load effects. It also introduces an energy management strategy for deterministic dynamic loads to extend lifetime and a basic prescriptive maintenance approach integrating load allocation with maintenance decisions to minimize costs.

2. Prognostic Tool

The aging of Proton Exchange Membrane Fuel Cells (PEMFCs) is heavily influenced by load conditions, with both magnitude and variability driving degradation (Pei et al., 2008). Without start-stop effects, the total resistance increase is modeled as Zuo et al. (2024):

$$\Delta R = \Delta R_L + K\Delta L \tag{1}$$

where ΔR is the overall resistance rise, ΔR_L represents the load-intensity contribution, K is a proportionality constant, and ΔL indicates load variation magnitude. The load-driven increment ΔR_L is modeled using a non-homogeneous gamma process with a power-law shape function $A(t) = \alpha_L t^\beta$ and scale parameter b. For a fuel cell of age t_0, the resistance increase over δt under load L is:

$$\Delta R_L(t_0, \delta t) = R_L(t_0 + \delta t) - R_L(t_0) \sim \text{Gamma}\left(\alpha(L)(t_0 + \delta t)^\beta - \alpha(L) t_0^\beta, b\right) \tag{2}$$

This captures load and age effects, with α_L defined as Zuo et al. (2024):

$$\alpha(L) = C(L - L_{\text{nom}})^2 + \alpha_{\text{nom}} \tag{3}$$

Here α_{nom}, b, and C are fitted parameters, while β accounts for time-varying degradation rates, and α_L's parabolic function, minimizes degradation at L_{nom} and increases it for unfavorable load levels.

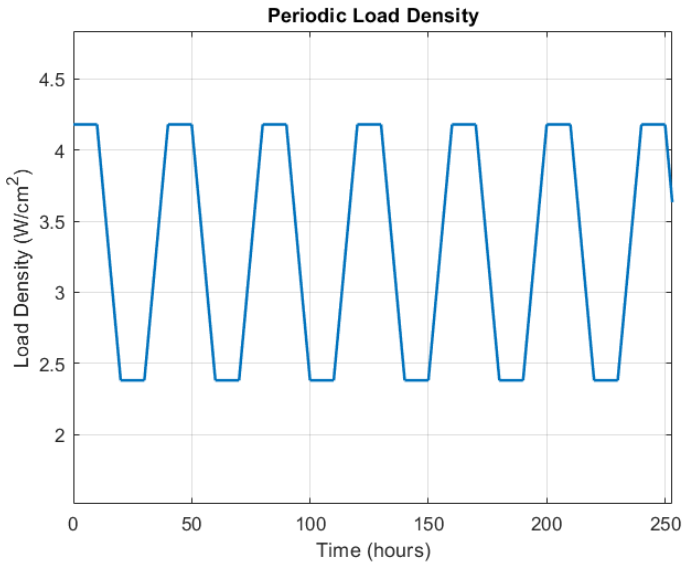


Figure 1. Dynamic periodic load profile used for simulation ($W \cdot \text{cm}^{-2}$).

Given the future load profile like in Figure 1, current age, and state of health (SOH) via resistance R, the tool estimates the RUL distribution, with failure at threshold F_T. Under variable loads, analytical RUL is complex, but Monte Carlo simulations can solve the issue, as shown in (Figure 2). The simulation provides multiple degradation paths under Figure 1's profile, yielding the RUL distribution in Figure 3.

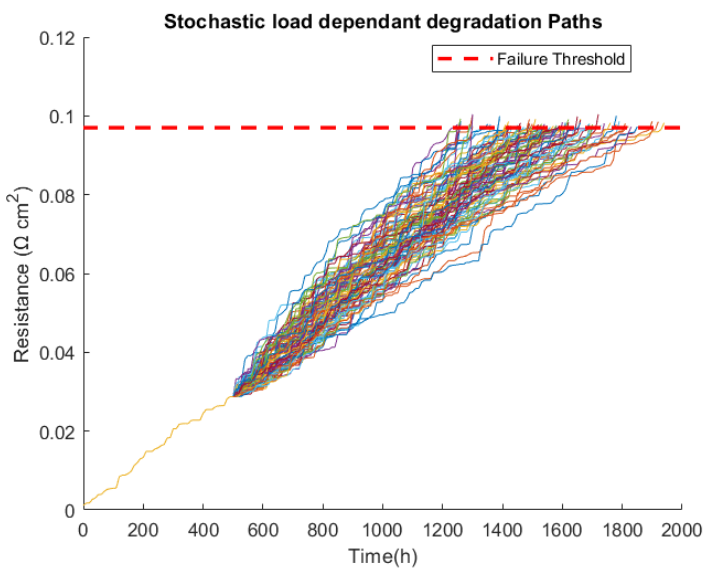


Figure 2. One hundred Monte Carlo simulated degradation paths under the load profile in Figure 1.

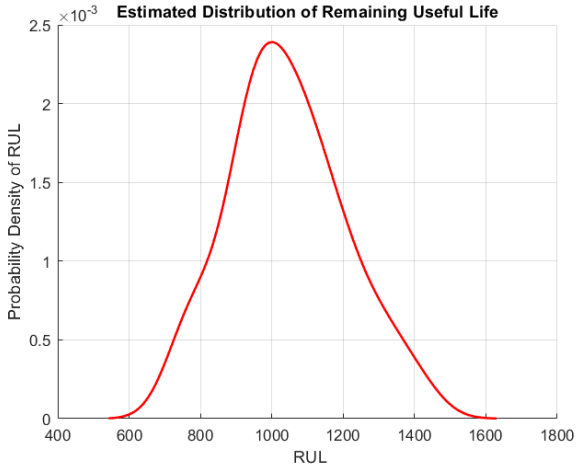


Figure 3. Estimated RUL probability density distribution under the load conditions in Figure 1.

3. Energy Management for Deterministic Dynamic Load

3.1. *Optimal Load Allocation Formulation for Durability Enhancement*

Building on the prognostic tool's degradation model, which uses total resistance as the aging indicator, this section formulates load allocation for a multi-stack fuel cell (MFC) system comprising n stacks, under a deterministic load demand known to vary over time. The goal is to enhance durability by optimally distributing this dynamic load across stacks at each instant within a decision horizon h, where the future load profile is fully specified and may fluctuate within h. At each decision time, or following a stack replacement, the varying load distribution is predefined for the entire horizon, this assigns a unique, time-varying load profile to each stack over h. To optimize this distribution, the horizon h is divided into m small time intervals of length Δt, where load changes are negligible. The expected resistance increment for a stack of age T over one interval, due to load amplitude L, is given by:

$$\begin{aligned}\Delta R_{\Delta t} &= \Delta R_L + \Delta R_{\Delta L} \\ &= \frac{\alpha(L)\left[(T+\Delta t)^{\beta} - T^{\beta}\right]}{b} + K(L - L_{\text{previous}})\end{aligned} \tag{4}$$

where L_{previous} is the load from the prior interval. For the full horizon h, the total resistance increment becomes:

$$\begin{aligned}\Delta R_h &= \sum_{i=1}^{m} \Delta R_{\Delta t_i} \\ &= \sum_{i=1}^{m} \left[\frac{\alpha(L_i)}{b}\left((t_0 + i\Delta t)^{\beta} - (t_0 + (i-1)\Delta t)^{\beta}\right) + K(L_i - L_{i-1})\right]\end{aligned} \tag{5}$$

where t_0 is the stack's age at decision time, and L_i is the load in the i-th interval. For n stacks, the optimization objective is to minimize the total expected resistance increment across all stacks:

$$\begin{aligned}Minimize: J &= \sum_{j=1}^{n} \Delta R_{h,j} \\ &= \sum_{j=1}^{n}\sum_{i=1}^{m} \left[\frac{\alpha(L_{i,j})}{b}\left((t_{0,j} + i\Delta t)^{\beta} - (t_{0,j} + (i-1)\Delta t)^{\beta}\right) + K(L_{i,j} - L_{i-1,j})\right]\end{aligned} \tag{6}$$

where $t_{0,j}$ is the age of stack j, and $L_{i,j}$ is its load in the i-th interval. This is subject to constraints ensuring the total load is met at each step, $\sum_{j=1}^{n} L_{i,j} = L_{\text{total}}$, and individual loads stay within bounds, $L_{\min} \leq L_{i,j} \leq L_{\max}$ for all i and j.

3.2. *Proposed Energy Management vs Average Load Split Method*

This section assesses the proposed energy management strategy (EMS) against a conventional average load-split approach in a basic scenario, where performance is calculated only based on the ability of the EMS to

extend the stack lifetime. A two-stack fuel cell system operates under a deterministic, dynamic load, always sufficient to run both stacks at or above their nominal power L_{nom} but below the maximum L_{max}. The load follows a square periodic profile with a 100-hour period, oscillating between $2L_{\text{nom}}$ and $L_{\text{nom}} + L_{\text{max}}$. Failed stacks are replaced instantly, assuming no downtime. Figure 4 shows the degradation paths of the two stacks under EMS, with variable load allocations that adjust at each time step yet sum to the total square-wave demand. Notably, the algorithm favors newer stacks early in their life, assigning them lighter loads, then equalizes the distribution as both stacks age, reflecting the parameter $\beta < 1$, which indicates slower degradation over time.

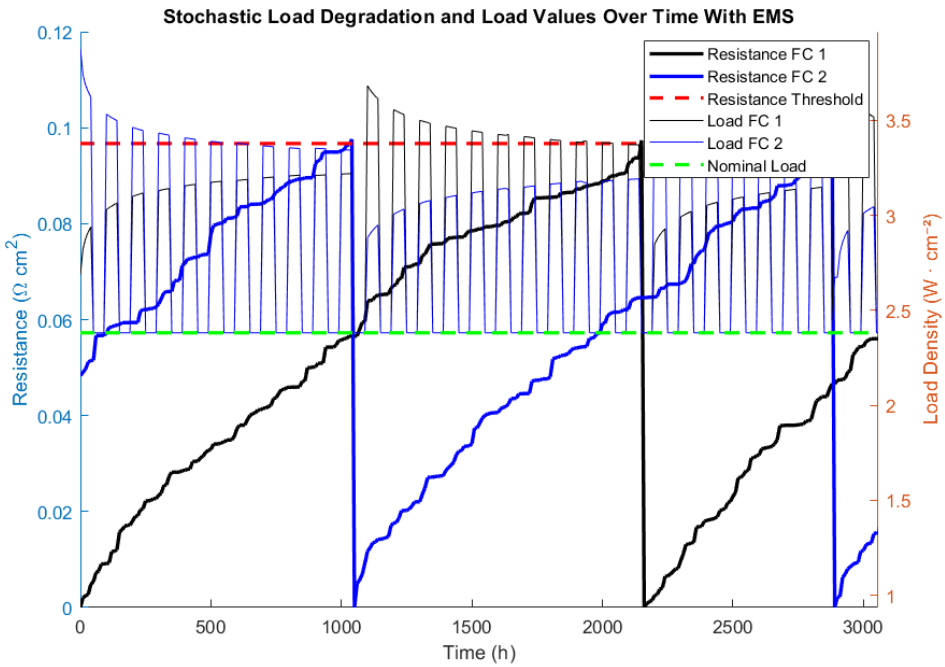


Figure 4. Degradation paths of the two stacks with variable load under the proposed energy management strategy.

Simulated over 10^6 hours with a failure threshold FT of 0.09 ohm/cm^2, the EMS outperforms the average split, as detailed in Table 1. The number of replacements drops from 1053 to 990, while the mean lifetime increases by 6.4% demonstrating enhanced durability.

Table 1. Comparison of replacements and cost under average load-split and EMS strategies.

Parameter	Value
Run Time	10^6 hours
Failure Threshold (FT)	0.09 ohm/cm^2
Mean Lifetime (EMS)	1998.91 hours
Mean Lifetime (AVG Split)	1878.96 hours

4. Maintenance and Operation Cost Optimization

Leveraging the energy management strategy (EMS) developed for optimal load allocation and the prognostic tool's RUL estimation, this section presents a predictive maintenance framework for a two-stack fuel cell system, where system failure occurs if any stack's resistance exceeds the failure threshold FT. An inspection policy measures each stack's resistance at scheduled intervals of period $T = 300$ hours, matching the optimization horizon of the EMS. At each inspection, if a stack's resistance meets or exceeds FT, corrective maintenance replaces the failed unit. For stacks below FT, the probability that the remaining useful life (RUL) is less than the inspection period, $P(\text{RUL} < T)$, is calculated; if this probability exceeds a preventive threshold P_p, preventive maintenance replaces the stack to avoid potential failure if it is unlikely to survive until the next

inspection. No action is taken if $P(\text{RUL} < T) < P_p$. Maintenance actions restore the stack to an as-good-as-new state but are restricted to inspection times, with no interventions possible between inspections, leaving failed stacks inoperative until the next scheduled check. The preventive threshold P_p is determined to minimize a cost function comprising replacement costs (set to 1 per replacement) and downtime costs at a rate of 0.01 per hour, where n_r is the number of replacements and DT is the downtime in hours, as given by

$$C = n_r \cdot 1 + 0.01 \cdot DT \tag{7}$$

Figure 5 shows the replacement rate, downtime, and cost rate for various P_p values, indicating an optimal $P_p = 0.3$ that minimizes the cost, as lower values increase unnecessary replacements while higher values increase downtime.

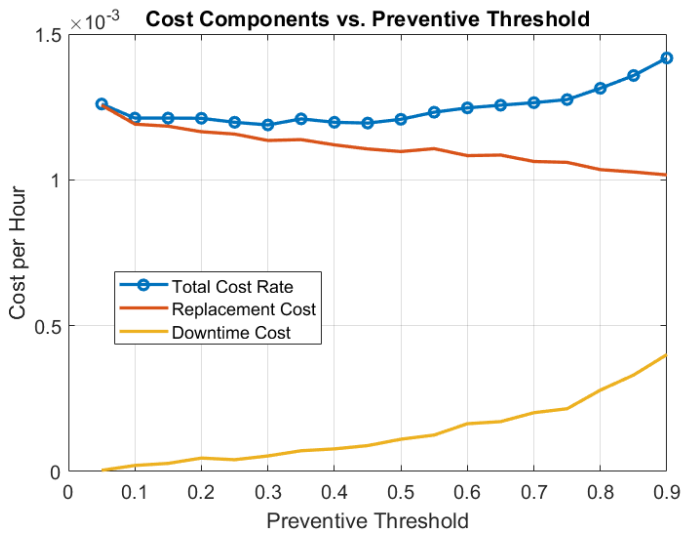


Figure 5. Replacement, downtime, and cost rate for different values of P_p.

To evaluate the cost rate of EMS versus the average load split under this maintenance policy, a simulation over 10^6 hours with $P_p = 0.3$ yields the following results in Table 2.

Table 2. Comparison of EMS and average load split under the predictive maintenance policy.

Parameter	EMS	Average Split
Cost Rate (units/hour)	0.001175	0.001264
Preventive Replacement Rate (replacements/hour)	0.001011	0.001068
Corrective Replacement Rate (replacements/hour)	0.000108	0.000124
Downtime Rate (hours/hour)	0.005693	0.007177
Mean Service Time (hours)	1777.92	1664.12

The EMS outperforms the average split, achieving a 6.84% increase in the mean service time of the stacks—defined as the total time a stack functions before failure or preventive replacement (1777.92 hours versus 1664.12 hours), and a 7.05% reduction in total cost rate (0.001175 units/hour versus 0.001264 units/hour), driven by lower replacement rates and a reduced downtime rate. These improvements highlight the EMS's ability to be combined with maintenance policy to enhance durability and minimize costs.

5. Conclusion

This paper introduced a PHM framework for multi-stack PEMFC systems operating under dynamic load conditions. By employing a non-homogeneous gamma process to model degradation—using internal resistance as a health indicator—we effectively estimated the Remaining Useful Life (RUL) of individual fuel cell stacks. Integrating this prognostic tool with an optimized Energy Management Strategy (EMS) allowed for dynamic

load allocation, which extended the operational lifetime of the stacks by 6.4% compared to conventional average load-split strategies.

Furthermore, the framework incorporates a predictive maintenance approach that leverages the RUL distribution to guide preventive maintenance decisions. The promising results demonstrate that combining prognostics, energy management, and maintenance optimization can reduce overall maintenance and operational costs.

Building on these advances, our work lays the foundation for a prescriptive maintenance strategy. This next step involves dynamically adapting operational loads and maintenance schedules based on real-time health data, ensuring that potential failures are mitigated proactively using an updated EMS, maximizing the system's resilience. Future efforts will focus on refining these methodologies and validating the framework across a broader set of operational scenarios, ultimately paving the way for more robust and sustainable fuel cell system applications.

6. Acknowledgment

The authors acknowledge the ANR – FRANCE (French National Research Agency) for its financial support of the MFC-DEPLOY project n° ANR-23-CE05-0006

References

Ao, Y., Laghrouche, S., Depernet, D. & Chen, K. (2020), 'Lifetime prediction for proton exchange membrane fuel cell under real driving cycles based on platinum particle dissolve model', *International Journal of Hydrogen Energy* **45**(56), 32388–32401.
URL: *https://www.sciencedirect.com/science/article/pii/S0360319920332158*

Bankati, W. R., Boulon, L. & Jemei, S. (2024), 'Remaining useful life prognostic-based energy management strategy for multi-fuel cell stack systems in automotive applications', *International Journal of Hydrogen Energy* **82**, 374–383.
URL: *https://www.sciencedirect.com/science/article/pii/S0360319924031367*

Bressel, M., Hilairet, M., Hissel, D. & Ould Bouamama, B. (2016), 'Extended kalman filter for prognostic of proton exchange membrane fuel cell', *Applied Energy* **164**, 220–227.
URL: *https://www.sciencedirect.com/science/article/pii/S0306261915015214*

Garcia, P., Fernandez, L. M., Garcia, C. A. & Jurado, F. (2010), 'Energy management system of fuel-cell-battery hybrid tramway', *IEEE Transactions on Industrial Electronics* **57**(12), 4013–4023.

Liu, H., Chen, J., Hissel, D., Hou, M. & Shao, Z. (2019), 'A multi-scale hybrid degradation index for proton exchange membrane fuel cells', *Journal of Power Sources* **437**, 226916.

Zuo, J., Cadet, C., Li, Z., Bérenguer, C. & Outbib, R. (2024), 'A deterioration-aware energy management strategy for the lifetime improvement of a multi-stack fuel cell system subject to a random dynamic load', *Reliability Engineering & System Safety* **241**, 109660.
URL: *https://www.sciencedirect.com/science/article/pii/S0951832023005744*

Energy pipeline individual defect growth prediction using in-line inspection data

M. Hussain*, T. Zhang* and R.A. Dwight*
**School of Mechanical, Materials, Mechatronic and Biomedical Engineering, Faculty of Engineering and Information Sciences, University of Wollongong, NSW2522, Australia + engr.mohdhussain@yahoo.com, tieling@uow.edu.au, radwight@uow.edu.au*

Abstract. Pipelines play a crucial role in the energy sector. Ensuring pipeline integrity is paramount for safe and reliable operation. Over time, local corrosion-driven defects on the exterior surface of a pipeline can 'grow' to a size that results in the catastrophic failure of the pipe. This research paper explores the possible use of well-known data analysis for practical defect-size prediction given sparse historical data available from successive in-line inspection (ILI), where Exponential, Logarithmic, Quadratic, Linear and Power Law models are tested and validated. These models are well known in industry applications; however, it is not clear what their effectiveness and performance are in the prediction of energy pipeline defect progression, given the limitation on data. In this paper, two distinct pipelines were considered as case studies to assess the generalizability and robustness of the selected degradation models. It was found that the Logarithmic and Power Law models produced acceptable results. An ILI-data analysis approach is proposed that can guide individual defect growth prediction in industrial practice. The results offer valuable insights for industry professionals, enhancing predictive capabilities related to pipeline integrity.

1. Introduction

Energy pipelines serve as crucial conduits for transporting various forms of energy, such as oil, natural gas, and refined products, over long distances between supply and distribution systems. These pipelines are vital in ensuring a reliable and efficient energy delivery from production sources to end-users (Hussain et al., 2023). Their applications range from delivering crude oil from extraction sites to refineries, transporting natural gas for electricity generation, heating, and industrial processes, to distributing refined products like gasoline and diesel for consumer use.

Typically pipeline construction cost is driven by pipe-wall design: material and thickness, resulting in the need to ensure the wall condition is closely managed over the in-service life of the pipeline. Failure to identify and eliminate local corrosion damage can lead to catastrophic failure. Hence, pipeline integrity is driven by the effective management of localized wall defects. Moreover, the behaviour of corrosion-defects in terms of generation and growth depends on specific local conditions: including soil type; geology; construction quality and cathodic protection performance.

Pipeline operators adopt reliability-based corrosion monitoring and management programs to support efforts to prevent corrosion. These programs usually include three main parts: regular, if infrequent, and detailed inspections to identify and measure corrosion defects in a specific pipeline; assessing the likelihood of pipeline failure based on the inspection findings; and, repairing any defects if needed to maintain the failure probability, integrity, of the pipeline below an acceptable level. Given access to a pipe's wall is normally limited at best, particularly for buried pipelines, non-destructive evaluation methods utilizing in-line inspection (ILI) tools are widely employed for locating and identifying anomalies within pipeline walls. On average, ILI tools have an accuracy within ±10% of the pipe wall thickness (Vanaei et al., 2017). Measurement accuracy is subject to improvement over time as the technology and its application improves or evolves.

Establishing an appropriate defect growth model is vital for predicting defect evolution and scheduling both future inspections and maintenance interventions and remaining-service-life assessments. There are two basic categories of approaches to defect-growth prediction: model-based and data-driven approaches. Model-based approaches rely on physical models like finite element analysis and data-driven approaches use condition-measurement data to investigate defect propagation. Model-based approaches are limited by the susceptibility of defect initiation and growth to specific events and local conditions. ILI is complex and costly to deploy, limiting the feasible frequency of use and resulting data volume available for a specific defect: typically, ILI tools are deployed at 5-year intervals and only for the recent life of pipelines: perhaps 20 years.

Methods of analysis and interpretation of ILI data is sought by industry. The approach involves testing and validating mathematical models to predict corrosion defect growth.

2. Summary of available models for defect-growth prediction

2.1 *Analysis Background*

To deal with this complex issue, most of the oil and gas companies are working towards sophisticated architectures with intelligent capabilities to estimate the growth of corrosion defects effectively. The progression of external corrosion in natural gas pipeline systems is characterized as a Markov process by Liu et al. (2018). They proposed a multi-level optimization strategy for integrity management to manage the total cost of natural gas pipeline systems. Similarly, Amaya-Gómez et al. (2016) utilize data acquired from ILI to enhance risk-based decision making. They suggested utilizing corrosion growth models to estimate the remaining safe life of the pipeline, relying on mechanistic models.

2.2 *Model Options*

Table 1 provides a summary evaluation of available models for pipeline condition assessment.

Table 1. Summary of basic modelling approaches available

Model type	Condition to apply	Characteristics	Limitations
Bayesian model	At least two ILI measurement datasets	Population-based defect estimation or prediction	Not accurate for individual defect growth
Common statistical models	At least one ILI measurement dataset	Population-based defect estimation or prediction	Not informative for individual defects subject to a specific growth regime
Gamma process model	At least one ILI measurement dataset	Population-based defect estimation or prediction	Computationally and data intensive
Deterministic linear growth model	At least two ILI measurement datasets	Simple and easy to interpret	May not capture complex defect growth patterns. Uncertainties cannot be represented.
Markov process	Multiple ILI datasets are required	Population-based defect estimation or prediction	Inaccurate predictions due to complexity
Dynamic linear model	Multiple ILI datasets are required	Population-based defect estimation or prediction	Not accurate for individual defect growth
Neural network model	Multiple ILI measurement data, requiring large data sample size	Population-based defect estimation or prediction	Requiring multiple datasets and larger sample size to predict defect growth
Support vector machines	Multiple ILI measurement data, requiring large data sample size	Population-based defect estimation or prediction	Requiring multiple datasets and larger sample size to predict anomaly and defect growth
Random Forest	Multiple ILI measurement data, requiring large data sample size	Population-based defect estimation or prediction	Requiring multiple datasets and larger sample size to predict anomaly and defect growth
Gradient Boosting	Multiple ILI measurement data, requiring large data sample size	Population-based defect estimation or prediction	Requiring multiple datasets and larger sample size to predict anomaly and defect growth
Simple mathematical model	At least two ILI measurement datasets	Individual defect estimation or prediction	May not capture the measurement errors, i.e., uncertainties are not represented

3. Method

3.1 *ILI Data analysis process*

Model-fitting analysis enables the development of predictive models that can capture and quantify the effects of factors such as corrosion, stress, and material degradation on the pipeline (Xiao et al., 2024). The model fitting depends on the available ILI data regarding each defect. In the model fitting analysis in this paper, we

selected the simplest models due to the ILI measurement frequency being very low in a pipeline service history. Moreover, as an initial study, the effect of uncertainty in the measurements has been ignored. Prior to such analysis, several key steps are required. The first step is to match the recorded ILI data regarding a defect. This is to make sure the measured data from different ILI runs are truly for the same defect. For pipe wall defects, the matching process includes both longitudinal and circumferential matching. Information about a method for ILI data matching can be found in Liu et al. (2019). This process was completed prior to obtaining the data used in this study. The second step is to identify the outliers in ILI data. Approaches to outlier identification have been developed in one associated work (Hussain, 2025). The third step, then, is to apply model fitting to the matched ILI data for each defect. The fourth step is to conduct the model prediction analysis, which is the main focus of this paper. To achieve this prediction analysis, each model's prediction error is verified using the ILI data measured in the next ILI run.

3.2 *Case-study data source*

The significant characteristics of pipe-wall anomalies are typically represented by overall dimension parameters: length, l, width, w, and depth, d. In this context, the anomaly is encapsulated within a rectangle, with the starting and ending points diagonally located within this boundary. The depth parameter signifies the deepest point identified within the confines of this rectangular representation.

We seek to study instantaneous and time-dependent corrosion defect-depth growth by applying degradation models to ILI defect measurement data.

ILI datasets utilized in this study were derived from two different crude oil pipelines. Pipeline 1, commissioned in 1979, has ILI data collected in the year 2000, 2005, 2010, 2015 and 2020. Pipeline 2, commissioned in 1986, has ILI data for inspections conducted in 2003, 2008, 2012, 2016 and 2022. These arguably represent typical energy pipelines. As an indication of the overall volume of data to be analysed per pipeline, following the removal of outliers from the ILI data, Pipeline 1 retained a total of 9,890 ILI dataset samples, and Pipeline 2 retained 10,289 ILI dataset samples over the pipeline lengths. Notwithstanding the overall data volume to be managed, data for individual defects is limited to a maximum of 5 measurements. Analysis of the data for Pipeline 1 is presented in this paper.

3.3 *Model selection*

Given the limited number of data points per defect, five simple models were selected for comparative analysis: Exponential; Logarithmic; Quadratic (Polynomial model with order 2); Linear; and Power Law models.

Figure 1 depicts pipeline wall loss (%) over the pipeline age (years), and sample model-fitting results, for a randomly selected defect from Pipeline 1.

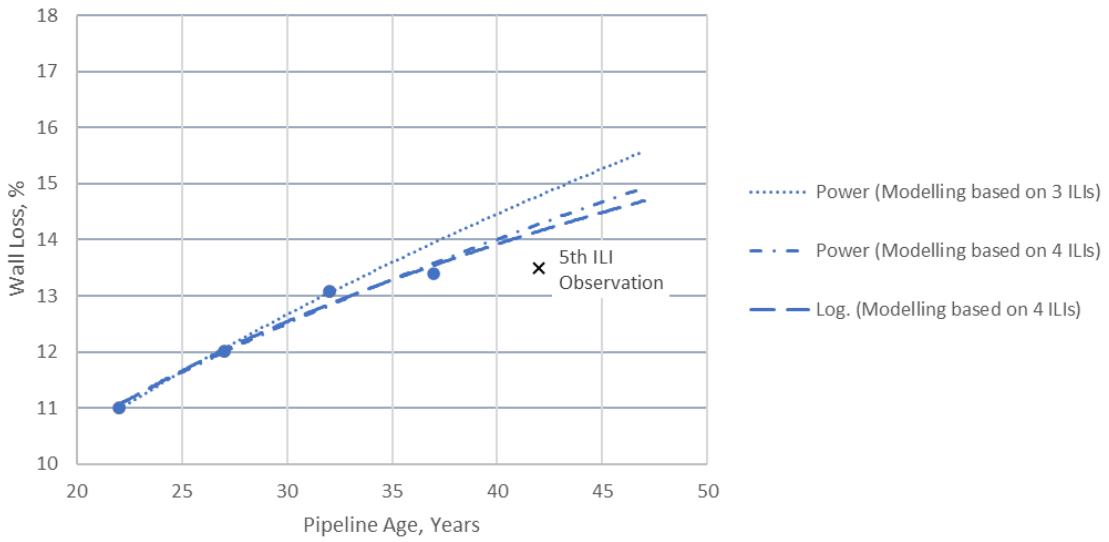


Figure 1. Model fitting using ILI data for one randomly selected defect. Prediction was simulated by excluding later points from the model development and using those to test for prediction accuracy.

In this case, Power Law and Logarithmic models were fitted to the observed data. The figure provides an illustration of the rudimentary model fitting undertaken over the first 3 and then the first 4 ILI inspections. The discrepancy between the 5th inspection point and the models obtained indicates the model-prediction-accuracy, the error.

Table 2 shows the model-prediction-accuracy analysis for the data shown in Figure 1 for various possible models. The relative error is simply calculated: (model prediction value –value indicated by the next ILI run)/ value indicated by the next ILI run. The '3 ILI runs' models are developed using the data measured in the first 3 ILI runs to predict the defect size for the 4th run. Similarly, the '4 ILI runs' models are developed using the data measured in the first 4 ILI runs to predict the defect size for the 5th run.

Table 2. Model prediction accuracy analysis based on the datasets from 3 and 4 ILI runs for one randomly selected defect illustrated in Figure 1.

Model	Dataset	Model Formula	Relative error (%)	R^2 value
Exponential	3 ILI runs	$y = 7.518e^{0.0173x}$	6.491	0.9999
	4 ILI runs	$y = 8.2809e^{0.0135x}$	8.301	0.9442
Logarithmic	3 ILI runs	$y = 5.5302\ln(x) - 6.1325$	3.335	0.9951
	4 ILI runs	$y = 4.8026\ln(x) - 3.7949$	5.012	0.9773
Quadratic	3 ILI runs	$y = 0.0012x^2 + 0.1432x + 7.2688$	6.124	1.0
	4 ILI runs	$y = -0.007x^2 + 0.5778x + 1.6354$	0.556	0.9905
Linear	3 ILI runs	$y = 0.208x + 6.414$	5.377	0.9997
	4 ILI runs	$y = 0.1648x + 7.5084$	9.705	0.9560
Power Law	3 ILI runs	$y = 2.6395x^{0.4611}$	4.193	0.9979
	4 ILI runs	$y = 3.2717x^{0.3942}$	5.918	0.9704

4. Analysis

Tables 3 to 6 present summaries of the model prediction accuracy results for randomly selected defects from Pipelines 1 and 2. Tables 3 and 4 show the results of the models built using 3 successive observations of defect-depth obtained from three successive ILI runs. Tables 5 and 6 depict the results by including a 4th observation. Through observing the tabulated results, it is found that the Logarithmic and Power Law models consistently result in relatively low prediction errors, and their prediction results are robust. Those present better predictions of individual defect depth growth than the other three models in the test. Comparing results between models built from 3 observations with the models that consider a subsequent 4th observation provides some comfort that prediction of future defect size generally improves with the number of measurements taken. Note that those measurements are generally taken at 5-year intervals as previously explained. This improvement in prediction is particularly observed in the case of the Power Law model.

An interesting finding is that, in many cases, a model with a higher R^2 value did not result in the best prediction accuracy. This observation indicates that using the R^2 value to choose a model for achieving a higher prediction accuracy may not be valid.

It is noted that this analysis assumes uncertainty in the measured values is insignificant. It is recognized that such an assumption does not generate a conservative prediction. In considering this potential deficiency in the analysis, there is some evidence that the low and consistent prediction error achieved suggests uncertainty may not be significant at least in the cases studied. The practical use of the resulting prediction method needs to consider the possible effects of uncertainty as well as the indicated prediction error.

Table 3. Model prediction errors for the randomly selected defects using 3 ILI runs from Pipeline 1, followed by the mean and standard deviation (Std Dev) of the prediction errors.

Defect ID	Model prediction relative errors for the models built				
	Logarithmic	Quadratic	Exponential	Linear	Power Law
001	0.366%	3.723%	2.435%	1.787%	0.705%
002	0.237%	3.909%	0.757%	0.833%	0.287%
003	3.100%	**5.698%**	**6.078%**	**5.003%**	3.823%
004	0.301%	4.958%	0.980%	1.059%	0.366%
005	1.100%	**11.947%**	3.639%	2.909%	1.509%
006	3.410%	**6.203%**	**6.577%**	**5.456%**	4.271%
007	-2.104%	1.837%	-1.449%	-1.410%	-2.032%
008	-3.460%	**6.924%**	-1.829%	-2.129%	-3.293%
009	-0.324%	-4.327%	0.238%	0.276%	-0.197%
010	0.270%	4.391%	1.046%	0.936%	0.304%
Mean	0.290%	4.526%	1.847%	1.472%	0.574%
Std Dev	2.067%	4.108%	2.855%	2.451%	2.301%

Table 4. Model prediction errors for the randomly selected defects using 3 ILI runs from Pipeline 2, followed by the mean and standard deviation (Std Dev) of the prediction errors

Defect ID	Model prediction relative errors for the models built				
	Logarithmic	Quadratic	Exponential	Linear	Power Law
001	1.264%	**7.074%**	2.236%	2.233%	1.346%
002	-0.144%	**16.502%**	0.257%	0.263%	-0.158%
003	0.00941%	1.0648%	0.556%	0.457%	2.572%
004	0.510%	1.309%	1.148%	1.067%	0.554%
005	-0.369%	1.389%	**53.778%**	-0.161%	-0.366%
006	0.6651%	**-18.37%**	**-5.423%**	**-5.059%**	**-5.803%**
007	-0.023%	**18.395%**	0.387%	0.236%	-0.196%
008	-0.948%	1.730%	-0.996%	-0.666%	-0.986%
009	0.201%	1.773%	0.597%	0.599%	0.244%
010	0.101%	0.903%	0.356%	0.296%	0.113%
Mean	0.127%	3.176%	5.292%	-0.073%	-0.270%
Std Dev	0.600%	10.050%	17.159%	1.911%	2.183%

Table 5. Model prediction errors for the randomly selected defects using 4 ILI runs from Pipeline 1, followed by the mean and standard deviation (Std Dev) of the prediction errors

Defect ID	Model prediction relative errors for the models built				
	Logarithmic	Quadratic	Exponential	Linear	Power Law
001	1.880%	2.833%	4.224%	3.552%	2.252%
002	0.177%	0.565%	0.926%	0.840%	0.221%
003	2.260%	**73.199%**	**5.274%**	4.116%	3.052%
004	0.223%	0.714%	1.394%	1.058%	0.313%
005	3.600%	4.184%	**6.498%**	**5.551%**	4.197%
006	3.410%	-1.078%	**6.839%**	**5.409%**	4.309%
007	-0.822%	2.799%	0.329%	0.279%	-0.711%
008	0.463%	7.281%	3.043%	2.453%	0.850%
009	0.887%	0.251%	1.926%	1.836%	0.897%
010	0.198%	0.633%	1.119%	0.942%	-3.748%
Mean	1.228%	9.138%	3.157%	2.604%	1.163%
Std Dev	1.489%	22.637%	2.241%	1.944%	2.433%

Table 6. Model prediction errors for the randomly selected defects using 4 ILI runs from Pipeline 2, followed by the mean and standard deviation (Std Dev) of the prediction errors

Defect ID	Model prediction relative errors for the models built				
	Logarithmic	Quadratic	Exponential	Linear	Power Law
001	0.486%	0.983%	2.185%	1.805%	0.658%
002	-0.255%	-0.172%	0.558%	0.505%	-0.190%
003	0.178%	0.425%	0.884%	0.950%	0.247%
004	1.120%	-0.280%	1.957%	1.967%	1.198%
005	-0.210%	0.755%	0.154%	0.150%	-0.238%
006	0.020%	0.051%	0.533%	0.363%	0.043%
007	1.267%	4.661%	2.594%	2.503%	1.368%
008	-0.099%	1.193%	0.661%	0.162%	-0.045%
009	0.022%	0.301%	0.697%	0.624%	0.044%
010	0.011%	0.900%	0.410%	0.320%	0.035%
Mean	0.254%	0.882%	1.063%	0.935%	0.312%
Std Dev	0.538%	1.418%	0.851%	0.849%	0.570%

4. Conclusion

This research paper proposes an approach to energy-pipeline individual-defect growth-prediction based on simple mathematical models: particularly Logarithmic and Power Law models have been shown to provide reasonable predictions. The prediction results achieved have been shown to be robust for the data tested. This finding underscores the practical value of simple models when dealing with limited data associated with operational systems. The proposed method can be applied to industrial operations.

Pipeline operators can benefit from employing these models to gain awareness into the current condition of their pipelines, specifically through the forecasting of corrosion growth based on sparse ILI data. When applying such approaches, the most accurate model for a particular pipeline should be verified.

It is recommended that further and ongoing broad testing of the proposed models be conducted to verify and to generalize the results observed. Further consideration should also be given to establishing upper bounds on predicted defect-size, accounting for uncertainty in the observed measurements; and, expected prediction errors.

References

Amaya-Gómez R, Sánchez-Silva M, Muñoz F (2016) Pattern recognition techniques implementation on data from In-Line Inspection (ILI). *Journal of Loss Prevention in the Process Industries* **44**, 735-747.

Hussain M, Zhang T, Seema M (2023) Adoption of big data analytics for energy pipeline condition assessment. *International Journal of Pressure Vessels and Piping* **206**, 105061.

Hussain M, (2025) Application of Big Data Analytics to Energy Pipeline Integrity Management. PhD Thesis, University of Wollongong.

Liu H, Liu Z, Taylor B, Dong HB (2019) Matching pipeline in-line inspection data for corrosion characterization. *NDT & E International* **101**, 44-52.

Liu XQ, Zheng J, Fu JJ, Ji JQ, Chen GM (2018) Multi-level optimization of maintenance plan for natural gas pipeline systems subject to external corrosion. *Journal of Natural Gas Science and Engineering* **50**, 64-73.

Vanaei H, Eslami A, Egbewande A (2017) A review on pipeline corrosion, in-line inspection (ILI), and corrosion growth rate models. *International Journal of Pressure Vessels and Piping* **149**, 43-54.

Xiao R, Zayed T, Meguid MA, Sushama L (2024) Improving failure modeling for gas transmission pipelines: A survival analysis and machine learning integrated approach. *Reliability Engineering & System Safety* **241**, 109672.

Optimal Design of Hybrid Offshore Renewable Systems with Realistic O&M Considerations

Luciana Yamada*, Masoud Golalikhani*, Flávia Barbosa*, Luís Guimarães*.

** INESC TEC, Faculty of Engineering, University of Porto, Porto, Portugal*

Email addresses: up202204404@up.pt (Luciana Yamada), masoud.golalikhani@inesctec.pt (Masoud Golalikhani), flavia.barbosa@inesctec.pt (Flávia Barbosa), lguimaraes@fe.up.pt (Luís Guimarães).

Abstract. Hybrid offshore renewable energy systems, which integrate technologies such as wind turbines, wave energy converters, and floating solar photovoltaics, present promising solutions to enhance the reliability of energy production while reducing investment and operational and maintenance (O&M) costs through shared infrastructure. Design decisions are influenced by critical O&M challenges, including operational accessibility, system reliability, availability, and maintenance strategies. These aspects must be accounted for to correctly assess the project's economic feasibility and long-term performance. Nevertheless, existing studies often overlook O&M challenges by relying on oversimplified assumptions such as fixed availability rates, maintenance costs as a percentage of overall capital expenditures, or a predetermined number of maintenances over the project's lifespan, neglecting their impact on system design decisions. To bridge this gap, this work proposes a novel framework to optimize the system design by incorporating a more holistic and accurate estimation of O&M costs. The framework incorporates design decisions and relevant maintenance strategies, including corrective and preventive approaches, while accounting for component degradation, vessel fleet selection, accessibility, availability, and energy production.

1. Introduction

Renewable energy systems (RESs) are increasingly recognized as a sustainable alternative to fossil fuel-based resources, driven primarily by global concerns regarding climate change, rising energy demand, and the depletion of natural resources. Technologies like offshore wind turbines, wave energy converters (WEC), and floating photovoltaic (FPV) panels can help this transition, offering clean, inexhaustible energy sources that significantly mitigate environmental impacts and greenhouse gas emissions compared to conventional power generation. In this context, offshore RESs offer distinct advantages over their onshore counterparts, including stronger wind conditions, improved FPV efficiency, and eliminating costly land acquisition, making ocean-based sites attractive for RES projects (Li et al., 2020). Despite these benefits, single-source RESs inherently suffer from intermittency and variability due to meteorological conditions, making it challenging to provide a stable and reliable power supply when implemented independently.

To address these challenges of single-source RESs, combining different technologies in hybrid offshore renewable energy systems (HRES) offer a promising solution. By leveraging complementary generation patterns of the resources, HRES can enhance power stability and reliability while mitigating seasonal variability and intermittency (Rasool et al., 2022). Additionally, hybridization enables the sharing of infrastructure, optimization of marine space utilization, and consolidation of Operations and Maintenance (O&M) resources, ultimately enhancing both cost and environmental benefits. However, offshore technologies generally entail higher costs than onshore systems, mainly due to the elevated expenses associated with O&M and the use of less mature technologies such as WEC and FPV. Therefore, another key motivation for exploring HRES is to assess how hybridization and shared O&M strategies can help reduce such systems' levelized cost of energy (LCOE). Given these advantages and the potential for implementing HRESs in existing offshore farms, this study focuses on their hybridization to unlock synergies between different renewable generations, optimize resource use, and further advance the transition to a sustainable energy future.

Despite significant progress in the deployment and development of commercial-scale HRES projects, several challenges affect their wider adoption. Offshore environments pose unique difficulties, including harsh weather conditions, corrosion, marine life interactions, and structural stresses induced by wind, waves, currents, and remote distances from shore, all of which threaten system reliability and survivability. Additionally, operational challenges such as varying vessel operating speeds, the availability of maintenance crews and spare parts, and accessibility constraints significantly influence the effectiveness and cost-efficiency of maintenance activities. These operational factors and the resulting system degradation directly impact

different key performance indicators (KPIs) such as availability, accessibility, and overall energy production, ultimately affecting the system´s revenue streams and O&M expenditures. Therefore, accurately estimating these KPIs by incorporating realistic assessments of system degradation and O&M factors is crucial for selecting the most economically viable combination of HRESs. A realistic estimation can also assist in making informed decisions, leading to a lower LCOE, reduced net present, and considerable cumulative savings over the system's operational lifetime. However, the existing design studies frequently overlook these critical operational considerations, relying instead on oversimplified assumptions that can lead to design decisions that compromise the economic viability and sustainability of HRESs. Recognizing this critical gap, this paper proposes a comprehensive decision-support framework to optimize the quantity and technology mix of additional renewable energy sources in an existing park. It systematically addresses the key challenges involved in developing an HRES while reflecting their real-world dependencies. In this context, the main contributions of this paper are twofold.

1) Firstly, it proposes a novel, integrated decision-support framework to support informed optimal system design decisions by incorporating strategic, tactical, and operational perspectives essential in HRES. Specifically, the framework identifies and explicitly systematizes the interactions among key O&M factors, including maintenance strategies, equipment degradation, availability, accessibility, and their combined influence on energy production, O&M costs, and optimal sizing decisions. This framework empowers researchers and industry practitioners to plan and implement economically viable HRES projects by providing a comprehensive approach.
2) Secondly, building on the proposed framework, it identifies critical research gaps and outlines a promising research agenda for addressing existing limitations in offshore HRES design optimization.

The remainder of this paper is structured as follows. Section 2 reviews existing literature on design decisions in offshore HRES, highlighting critical research gaps. Section 3 details the proposed integrated decision-support framework and describes how operational considerations are embedded into the design optimization problem. Section 4 explicitly identifies research gaps based on insights from the proposed framework and suggests possible directions to address them in the future. Section 5 summarizes the paper's contributions.

2. Background and related literature

This section reviews key studies from the literature that address the offshore design problem, with a particular focus on hybridization research. Additionally, it highlights some of the critical factors influencing sizing decisions that are commonly oversimplified in current research. Existing studies on HRESs primarily focus on analyzing the complementarity and synergy between wind and solar resources to determine optimal equipment sizing (Huang et al., 2024), yet they often overlook critical O&M considerations. The O&M costs account for up to 25–30% of total investments in an offshore project, which is twice as much as for an onshore project (Li et al., 2021). These costs typically include expenditures for personnel, spare parts, and vessels necessary for maintenance activities. However, a limitation in current sizing optimization studies is that O&M expenses are often accounted for in a simplified manner, either as a fixed percentage of the initial capital expenditures (Golroodbari et al., 2021) or using simplified representations that assume a predefined number of minor and major repairs over the project's lifespan (Castro-Santos et al., 2016). For instance, Golroodbari et al. (2021) conducted a techno-economic analysis to evaluate the feasibility of integrating an offshore solar farm into an existing offshore wind farm, assuming O&M costs to be 2% of the initial investment. Gonzalez et al. (2017) optimized hybrid wind-solar system sizing using O&M costs as the fixed parameters without considering accessibility limitations or downtime effects, which can lead to potentially misleading economic analyses.

Maintenance strategies significantly affect O&M costs and system availability, representing the time that equipment can generate electricity relative to its maximum possible uptime (Tavner, 2021). One of the key advantages of HRES is reducing O&M costs by sharing maintenance strategies, vessels, and teams. Different maintenance strategies were explored in the literature, with Preventive maintenance (PM) and corrective maintenance (CM) being the two most common approaches applied in RESs (Hameed et al., 2010). PM is performed at scheduled intervals or based on predefined criteria to reduce the likelihood of failures, while CM is conducted after a failure occurs to restore system functionality. Different optimization techniques have been developed in the literature to improve maintenance strategies, such as analytical methods, metaheuristics, and

stochastic approaches (de Jonge and Scarf, 2020). The main objective is to minimize O&M costs while maximizing system reliability and availability. It is essential to estimate the frequency of CM and PM, along with their impact on system health, O&M costs, and availability. Nevertheless, despite their significance, maintenance strategies are rarely integrated into sizing optimization, leading to inaccurate accessibility and availability assessments that affect energy production, costs, and the robustness of sizing decisions. For instance, Kazemi-Robati et al. (2024) developed a stochastic sizing model for hybridizing existing offshore wind farms with WEC and FPV. However, their study did not account for the impact of maintenance actions and the associated costs on overall system economics.

Beyond the limitations in accounting for O&M strategies, many studies also oversimplify or entirely neglect the impact of component degradation on system performance. While maintenance strategies influence the long-term health and reliability of system components, degradation affects energy production even when no outright failures occur. As components age, their efficiency declines, leading to a gradual reduction in power generation and, consequently, revenue losses. However, this aspect is often overlooked in sizing studies (Kazemi-Robati et al. 2024; Soukissian et al., 2021).

In summary, while significant advances have been made in HRES research, most studies lack an integrated, comprehensive approach that accounts for key operational factors, including realistic maintenance strategies, system degradation, availability, and accessibility. As a result, their applicability remains limited, potentially leading to suboptimal sizing decisions that reduce long-term economic viability and system performance. This highlights the need for a holistic decision-support framework that explicitly captures the interactions among these critical factors to support more accurate and informed HRES design.

3. Proposed framework for HRES design optimization

The primary contribution of this paper is the development of a comprehensive, integrated decision-support framework that systematically integrates critical O&M factors such as accessibility, availability, and maintenance strategies into sizing optimization decisions. In Figure 1, the proposed framework addresses a significant gap in existing approaches, which frequently overlook real-world interactions among these factors and their impacts on overall economic and operational viability. It consists of five interconnected major parts, each detailed in the following subsections.

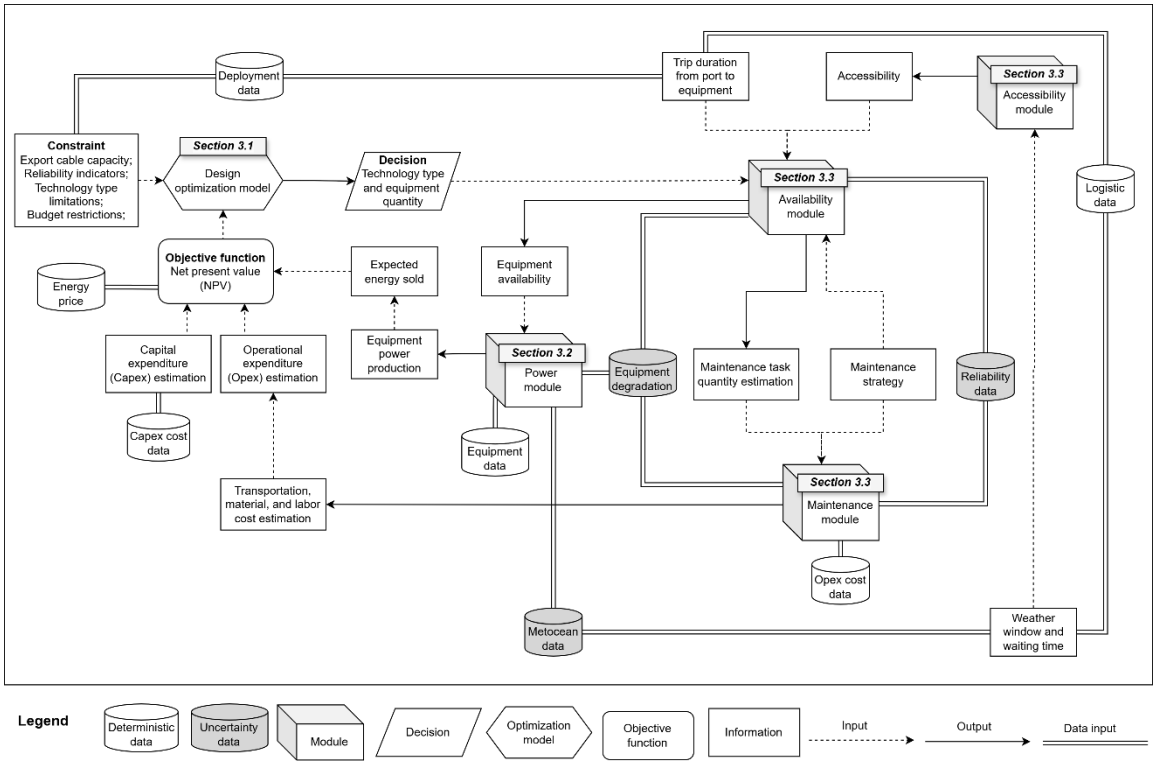


Figure 1. The proposed framework for HRES sizing with operational considerations.

3.1. Design optimization model

In an HRES design problem, key design decisions typically include selecting the optimal combination of energy technologies and determining the size and quantity of equipment. However, the design problem can be extended beyond these decisions to include considerations such as system layout, equipment placement, and cable routing, depending on the specific goals of the project. Various constraints, including export cable capacity, reliability indicators, technology type limitations, and budget restrictions, often modify these decisions. Different objective functions can be applied to the optimization problem, with economic performance indicators being the most common. Figure 1 illustrates an optimization using a net present value (NPV), which represents future energy revenue minus capital and operational costs (OPEX) over the system's lifetime. Moreover, HRES design can benefit from redundancy allocation problem (RAP) concepts, which involve selecting components to maximize reliability within cost constraints or minimize cost while meeting reliability requirements (Guan et al., 2025). By carefully balancing the trade-off between reliability and cost, more efficient and robust system design can be achieved.

3.2. Power module

The power module utilizes well-established energy production equations applied in energy studies, integrating detailed equipment characteristics and site-specific metocean data to calculate realistic hourly energy productions. These estimations are affected by system availability derived from the availability module, which, in turn, depends on the accessibility and maintenance modules. The proposed framework accounts for the health condition of the devices by incorporating the impact of device degradation on energy production, where the system remains operational but performs at a reduced capacity, ensuring a more accurate assessment of energy output.

3.3. Availability, Maintainability, and Accessibility module

The availability of any equipment directly affects energy production by accounting for expected equipment downtime due to adverse weather conditions, failure-induced outages, and chosen maintenance strategies such as CM and PM. In an offshore environment, due to adverse weather conditions, availability can range from 80% in the early stages to 93% for established offshore wind farms, compared to an average availability of 97% for onshore wind farms (Feng, Tavner, and Long, 2010). Ensuring high availability requires balancing the trade-off between O&M costs and revenue losses due to equipment downtime. The availability module should consider operational conditions, such as equipment degradation, reliability data (failure and repair rate), trip duration from port to equipment, accessibility information, and maintenance strategies. The resulting availability metric is a crucial input for accurately adjusting expected energy output, ensuring realistic and reliable sizing decisions. Another important metric is accessibility, which refers to safely reaching equipment for maintenance, considering personnel and vessel access, and the availability of spare parts. As shown in Figure 1, accessibility is heavily influenced by weather conditions. By integrating weather conditions that influence the maintenance timing with logistical factors such as vessel operating speeds, this module calculates practical weather windows and expected waiting times for various maintenance activities. Accessibility mainly impacts CM, as unscheduled events may cause delays in maintenance depending on weather conditions (Garrad, 2013). In contrast, PM can be scheduled during ideal conditions, such as periods of low energy production, minimizing the financial impact of equipment downtime. Accessibility estimations directly influence availability and maintenance activities, thus significantly impacting overall economic outcomes and sizing decisions. Finally, the maintenance module estimates the O&M costs by considering the quantity of CM and PM tasks provided by the availability module. This information is used to calculate the transportation, material, and labor costs associated with each type of maintenance strategy. By incorporating O&M information and improving OPEX cost estimation within the sizing optimization problem, the model prevents underestimating the financial burden of operation conditions, thus significantly impacting sizing decisions and overall system viability.

4. Identification of research gaps and future directions

Building on the integrated decision-support framework introduced in Section 3, this section outlines key limitations in existing HRES literature and proposes a research agenda. While considerable progress has been

made in HRES modeling, most current studies still rely on simplified assumptions and fail to capture critical operational complexities that significantly influence system design and performance.

4.1. Enhancing realism in HRES design

Integrating availability modeling: While stochastic like Markov processes (MPs) have been widely applied to model availability in renewable systems (Miao et al., 2017), these models are rarely embedded within HRES sizing optimization. MPs effectively represent system states and transitions (Čepin, 2011), offering a more accurate assessment of availability-driven impacts on expected energy output and O&M costs. Future research should incorporate MP-based availability modeling directly into sizing frameworks to better capture operational uncertainties and improve decision robustness. HRES design can benefit from such integration, similar to RAP approaches, where system redundancy is optimized alongside reliability (Guan et al., 2025).

Integrating degradation modeling into energy production module: Another critical gap lies in equipment degradation consideration. While failure-induced outages are commonly modeled, gradual degradation where equipment continues to operate but with reduced performance is often ignored. This oversight can lead to substantial underestimation of long-term losses in energy production and system profitability. To address this, advanced predictive maintenance strategies, such as condition-based maintenance and diagnostic methods based on genetic algorithms and machine learning, utilize sensor data and reliability models to improve maintenance planning and detect abnormal system behaviours (Rinaldi et al., 2021; Zio, 2022). Given the significant impact of degradation on yield and maintenance frequency, future HRES sizing methodologies should incorporate health-based degradation models within the energy and economic assessments, providing a more nuanced understanding of lifecycle performance.

Expanding uncertainty modeling and optimization techniques: While the framework considers several sources of uncertainties such as metocean conditions and reliability data, other inputs remain deterministic. Incorporating additional uncertainties, including energy prices, technological advancements, and regulatory risks, through probabilistic and possibilistic approaches like robust optimization, stochastic optimization, chance-constrained programming, Monte Carlo simulation, and scenario-based analysis, can better reflect real-world planning environments (He et al., 2023). Moreover, methods such as metaheuristic algorithms and machine learning techniques have proven effective in handling uncertainty in complex energy systems (He et al., 2022; Yang et al., 2021).

4.2. Expanding the scope of the design optimization models

Incorporating vessel selection and logistic decisions into the HRES design model: While the proposed framework effectively addresses HRES sizing decisions, it tackles some tactical and operational decisions, such as vessel selection, as input data. Future research should integrate these decisions directly into the optimization process. The optimal vessel choices could be determined based on operational needs, accessibility constraints, and cost trade-offs. Incorporating vessel characteristics (e.g., speed, capacity, cost) can improve the realism of OPEX estimations and reflect accessibility constraints more accurately.

Modeling joint O&M strategies: As shown in the framework, the maintenance module that estimates the O&M costs is a key input for optimizing the model. Nevertheless, despite one of the key advantages of HRESs being their potential to reduce O&M costs through joint maintenance activities, this benefit remains largely unexplored in current sizing studies. The cost-saving potential of shared vessels, crews, and infrastructure between different RESs has not yet been explicitly modeled. Future research should develop formulations that reflect joint O&M strategies and quantify their impact on cost efficiency and system performance.

4.3. Supporting Multi-Criteria Decision Making

Adopting multi-objective and robust optimization approaches: The proposed framework currently focuses on NPV maximization. However, the multidimensional nature of offshore HRES planning necessitates balancing multiple, and sometimes conflicting, objectives. Incorporating multi-objective optimization techniques that simultaneously consider NPV, LCOE, reliability, and environmental impact would increase the practical relevance of optimization results and support more balanced system designs.

Addressing these research gaps requires interdisciplinary collaboration between academia and industry stakeholders to access high-fidelity operational data, validate models, and ensure that research outcomes are

aligned with real-world needs. Advancing the modeling, integration, and optimization of these operational factors will ultimately contribute to developing more resilient, cost-effective, and sustainable offshore HRESs.

5. Conclusion

This paper proposes a comprehensive decision-support framework explicitly designed for optimizing HRES sizing decisions by systematically integrating key operational factors, including availability, accessibility, maintenance strategies, and equipment degradation. The proposed framework distinctly bridges existing gaps in sizing optimization research, replacing simplified assumptions with realistic estimations of O&M factors. The practical application of this framework empowers stakeholders, project developers, investors, and operators, to effectively evaluate and select the most economically viable HRES. It is important to highlight that although the primary focus of the framework is on sizing optimization, its modular design and the identified interactions among them enable adaptation to a wide range of strategic, tactical, and operational decision-making processes, thereby strengthening the overall decision-support capabilities for offshore HREs.

Acknowledgments

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101036457 and from the Portuguese Foundation for Science and Technology - FCT through PhD research grant under reference 2024.06757.BD.

References

Castro-Santos L, Martins E, Guedes Soares C (2016) Methodology to calculate the costs of a floating offshore renewable energy farm. Energies 9(5), 324.

Čepin M (2011) *Assessment of power system reliability: methods and applications*. Springer Science & Business Media.

de Jonge B and Scarf PA (2020) A review on maintenance optimization, European Journal of Operational Research, 285(3), pp. 805-824.

Feng Y, Tavner PJ, Long H (2010) Early experiences with UK round 1 offshore wind farms. *Proceedings of the Institution of Civil Engineers – Energy* 163(4), 167-181.

Garrad HGL (2013) A guide to UK offshore wind operations and maintenance. Glasgow and London, UK: Scottish Enterprise and The Crown, 42.

Golroodbari SZM, Vaartjes DF, Meit JBL, Van Hoeken AP, Eberveld M, Jonker H, Van Sark WGJHM (2021) Pooling the cable: A techno-economic feasibility study of integrating offshore floating photovoltaic solar technology within an offshore wind park. Solar Energy 219, 65–74.

Gonzalez A, Riba JR, Rius A, Puig R (2017) Optimal sizing of a hybrid grid-connected photovoltaic and wind power system. Applied Energy 201, 58–72.

Guan B, Li Z, Coit DW, Li YF (2025) Review of the Redundancy Allocation Problem to optimize system reliability. Engineering Optimization 57(1):44–68.

Hameed Z, Ahn SH, Cho YM (2010) Practical aspects of a condition monitoring system for a wind turbine with emphasis on its design, system architecture, testing and installation. Renewable Energy 35(5), 879-894.

He Y, Guo S, Zhou J, Ye J, Huang J, Zheng K, Du X (2022) Multi-objective planning-operation co-optimization of renewable energy system with hybrid energy storages. Renewable Energy 184:776–790.

He Y, Guo S, Dong P, Zhang Y, Huang J, Zhou J (2023) A state-of-the-art review and bibliometric analysis on the sizing optimization of off-grid hybrid renewable energy systems. Renewable and Sustainable Energy Reviews 183:113476.

Huang J, Iglesias G (2024) Hybrid offshore wind–solar energy farms: A novel approach through retrofitting. Energy Conversion and Management 319, 118903.

Huang LL, Fu Y, Mi Y, Cao JL, Wang P (2017) A Markov-Chain-Based Availability Model of Offshore Wind Turbine Considering Accessibility Problems. IEEE Transactions on Sustainable Energy 8, 1592-1600.

Kazemi-Robati E, Silva B, Bessa RJ (2024) Stochastic optimization framework for hybridization of existing offshore wind farms with wave energy and floating photovoltaic systems. Journal of Cleaner Production 454, 142215.

Li M, Jiang X, Negenborn RR (2021) Opportunistic maintenance for offshore wind farms with multiple-component age-based preventive dispatch. Ocean Engineering 231, 109062.

Li Y, Huang X, Tee KF, Li Q, Wu XP (2020). Comparative study of onshore and offshore wind characteristics and wind energy potentials: A case study for southeast coastal region of China. Sustainable Energy Technologies and Assessments, 39, 100711.

Miao S, Xie K, Yang H, Tai HM, Hu B (2017) A Markovian wind farm generation model and its application to adequacy assessment. Renewable Energy 113, 1447-1461.

Rasool S, Muttaqi KM, Sutanto D, Hemer M (2022). Quantifying the reduction in power variability of co-located offshore wind-wave farms. Renewable Energy, 185, 1018-1033.

Rinaldi G, Thies PR, Johanning L (2021) Current status and future trends in the operation and maintenance of offshore wind turbines: A review. *Energies* 14(9):2484.

Soukissian TH, Karathanasi FE, Zaragkas DK (2021) Exploiting offshore wind and solar resources in the Mediterranean using ERA5 reanalysis data. Energy Conversion and Management 237, 114092.

Tavner P (2021) *Offshore wind power: reliability, availability and maintenance*. 194th edn. London: IET.

Yang Z, Ghadamyari M, Khorramdel H, Seyed Alizadeh SM, Pirouzi S, Milani M, Banihashemi F, Ghadimi N (2021) Robust multi-objective optimal design of islanded hybrid system with renewable and diesel sources/stationary and mobile energy storage systems. Renewable and Sustainable Energy Reviews 148:111295.

Zio E (2022) Prognostics and Health Management (PHM): Where are we and where do we (need to) go in theory and practice. Reliability Engineering & System Safety 218:108119.

A Data-Driven Approach for Optimising Rolling Stock Fleet Using Time Series Analysis and SARIMAX Model

Md Mahabubur Rahman*, BM Alkali**, A. K. Jain***, J. M. Parrilla Gutierrez****, J. Nelson*****, C. Mcneil******

*/**/***/****Glasgow Caledonian University Glasgow, Scotland UK, mdmahabubur.rahman@gcu.ac.uk

*****/******Siemens Mobility Limited Glasgow, jim.nelson@siemens.com

Abstract: The study presented in this paper assess Electric Multiple Units (EMU) trains failure incidents and predict maintenance requirements and make predictions using the time series forecasting techniques. A Seasonal Autoregressive Integrated Moving Average with Exogenous variables (SARIMAX) model approach was utilised, and an attempt is made to forecast the patterns of failure incidents using weekly and monthly Closed Circuit Television (CCTV) failure data. The Augmented Dickey-Fuller (ADF) test is used for stationarity assessments of the failure data, while autocorrelation and partial autocorrelation function is used to determine the seasonality. Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are applied to ensure model selection balances statistical accuracy and complexity, preventing overfitting SARIMAX risk-based model approach is used to forecast the weekly and monthly CCTV failure incidents and then categorise the risk of failure of CCTV systems into low, medium and high-risk levels, represented with a traffic light system. Graphical display of the model simulation results from three geographical locations in Scotland is presented in this paper. The outcome of this paper is to support informed decisions on maintenance scheduling and the allocation of maintenance resources based on the risk of failure.

1. Introduction

The reliability and performance of railway rolling stock are critical in ensuring the efficiency and sustainability of modern rail networks. Effective maintenance strategies are required to prevent downtime and service disruptions. Siemens UK has introduced its first Desiro trains in Scotland. This paper looks at EMU units built between 2009 and 2010 at the Siemens facility in Krefeld, Germany. In-service failure also increases maintenance costs, which affects the train fleet's total availability and reliability. CCTV failures are the most frequent and disruptive technical problems which cause delays and negatively affect fleet reliability. This paper presents the location-specific failure risks of railway CCTV failure in Scotland. The highest level of technical incidents that led to a decrease in the performance of the fleet shows that 109 incidents of CCTV system failure were reported from 2020 to November 2024. The technical data for the CCTV System is classified into three types namely Primary, Secondary, and Tertiary data sets. Table 1 shows the distribution of the top 10 systems technical incidents and the CCTV incidents have the highest number of incidents recorded.

Incident Name	Incident Number
Closed Circuit Television (CCTV) systems	109
Passenger Bodyside Doors	45
Passenger Information System (PIS)	45
Autocoupler	32
Traction Control Unit (TCU)	30
Automatic Warning System (AWS)	23
Generic Components (Underframe/Bogies)	19
Driver's Safety Device (DSD)	19
On-Train Monitoring Recorder (OTMR) Unit	17
Auxiliary Converter Unit (ACU)	16

Table 1: Top 10 System Technical Incidents (January 2020- November 2024).

2. Literature Review

Chuwang, D.D. and Chen, W. (2022) developed forecasting techniques to determine daily and weekly passenger demand in Urban Rail Transit (URT) systems, emphasising the improvement of operational efficiency and customer satisfaction. Their research employs time series models, including SARIMA, and

Autoregressive Moving Average (ARMA), and Facebook Prophet, to analyse the impact of seasonality, trends, and external factors such as holidays on passenger behaviour. Dabral and M.Z. (2017) use (ARIMA) models to forecast rainfall. SARIMA proficiently models the cyclical characteristics of precipitation, demonstrating precise forecasts at monthly, weekly, and daily intervals. On the other hand, Wang, J. and Hu, J., (2015) developed a system that combines Empirical Wavelet Transform (EWT), Autoregressive Integrated Moving Average (ARIMA), and Support Vector Machines (SVM). The results indicate that integrating statistical and machine learning models can successfully capture the intricate patterns in wind speed data, resulting in more dependable forecasts. Tatarintsev, V.A. and Tolstosheev, A.K., (2021) use Big Data Risk Analysis (BDRA) to enhance understanding of risk factors, detect new threats, and assess risk levels by handling large volumes of diverse data sources.The study by Gołębiowski, P., Góra, I. and Bolzhelarskyi, Y., (2023) Conducted CCTV risk analysis for railway rolling stock, which include the use of Monte Carlo simulation method for risk assessment in rolling stock planning, which helps identify and predict specific events in order to introduce maintenance schedules in advance.

3. Data Collection and Analysis

The analysis presented in this paper explore a 4-year CCTV historical failure data. A total of 1454 data points was used for this analysis. The graph in Figure 2 shows the original time series data divided into weekly and monthly time series data. Figure 3 on the other hand shows the random nature of weekly and monthly time series after accounting for trends and seasonality.

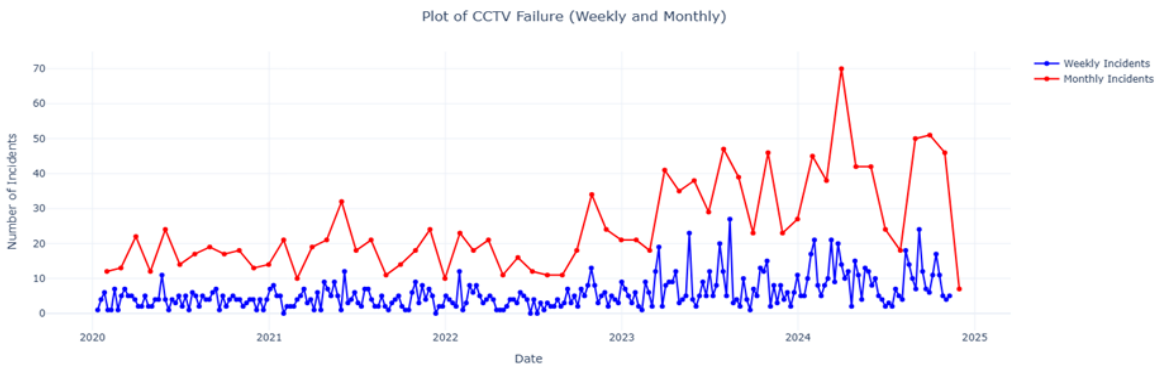


Figure 2: Plot of weekly and monthly time series data of CCTV failure incidents.

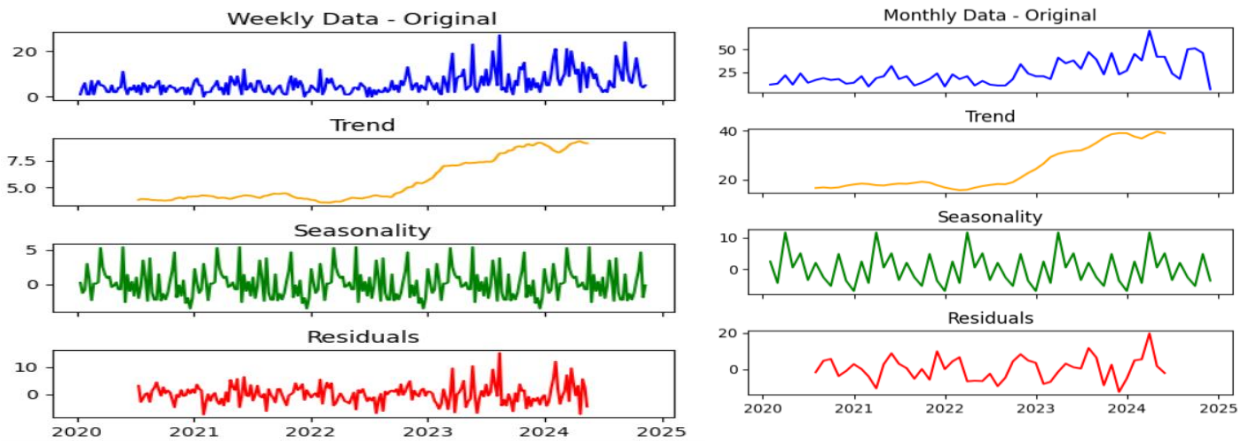


Figure 3: Weekly and monthly decomposed plot for CCTV failure incidents.

4. Methodology

The Box and Jenkins (1976) time series modelling technique is used for forecasting CCTV failure data and the Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) model is used to build weekly and monthly time series data for its suitability and flexibility and goodness of fit to our dataset. We tested the dataset for stationarity using the Augmented Dickey-Fuller test Statistics (ADF) for the daily and weekly CCTV failure dataset. The AIC, proposed by Akaike, H. (1974), is used to assess the data with the

least possible information loss, and determine whether it is appropriate to use in our predictive modelling for better accuracy of our results. In contrast, BIC, we introduced by Schwarz (1978), which includes a stricter penalty for complexity, preferring simpler models when the sample size is large, thus ensuring that the model is sparse and avoids overfitting. The process of developing the model discussed in this paper is illustrated in flow diagram in Figure 4. The risk analysis of the weekly and monthly observed data using the SARIMAX model characteristics allow us to determine the risk levels and visualise our results.

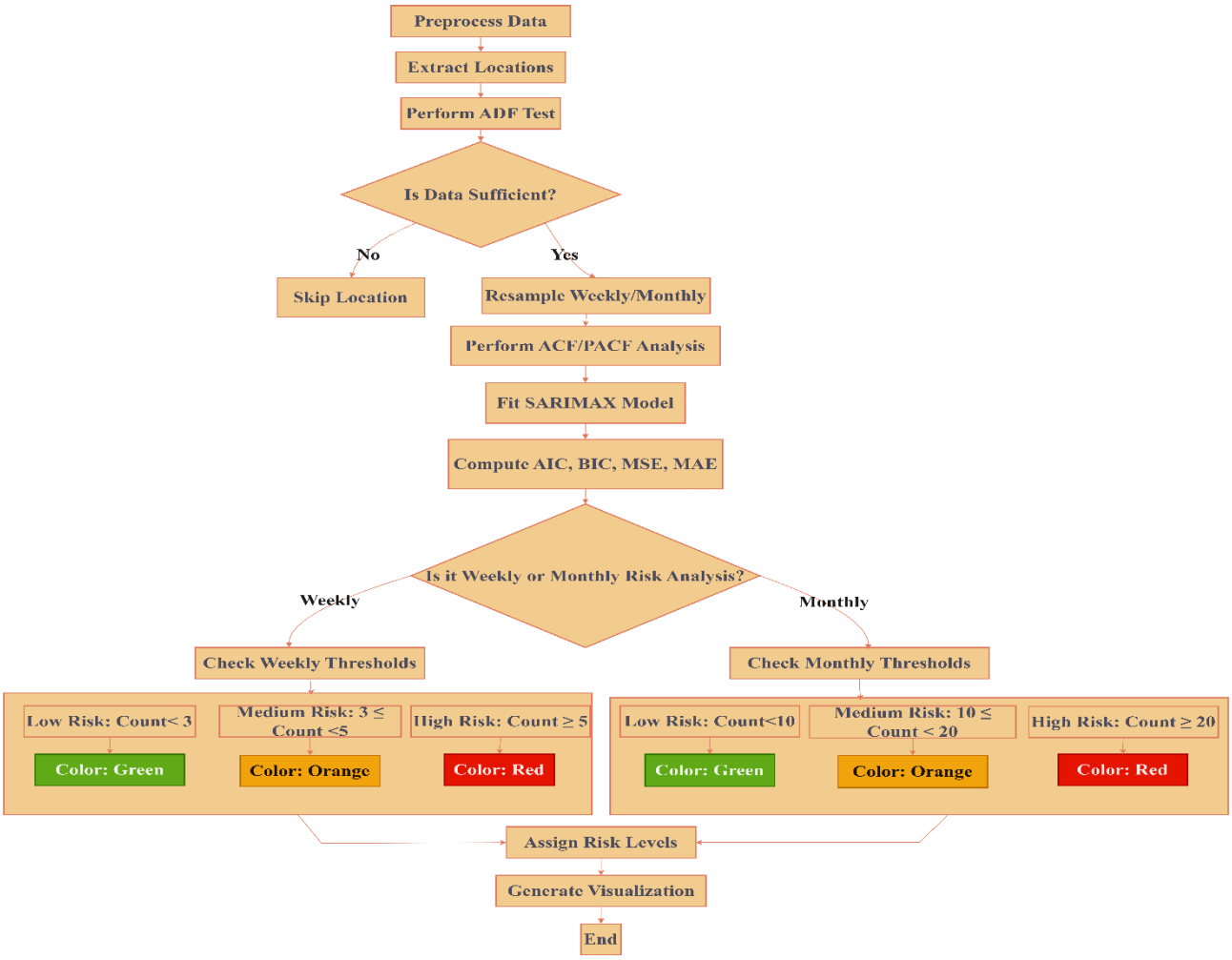


Figure 4: Flow chart of CCTV failure forecasting model

The AutoRegressive $(AR)_p$ in equation (1) considers how the current value is related to its past values

$$y_t = \emptyset_1 y_{t-1} + \emptyset_2 y_{t-2} + \cdots + \emptyset_p y_{t-p} + \epsilon_t \quad (1)$$

Here $\emptyset$ are the coefficients and ϵ_t is the white noise.

Differencing (I): There is stationarity by comparing

$$\acute{y}_t = y_t - y_{t-1} (1^{st} \text{ order differencing}) \quad (2)$$

Moving Average $(MA)_q$: Models the dependency on past forecast errors

$$y_t = \epsilon_t + \theta_{1\epsilon_{t-1}} + \theta_{2\epsilon_{t-2}} + \cdots + \theta_{q\epsilon_{t-q}} \quad (3)$$

Seasonal Component in equation (4) adds periodic AR and MA terms to capture seasonality

$$y_t = \Phi_1 y_{t-s} + \Theta_1 \epsilon_{t-s} \quad (4)$$

Where s is the seasonal period

Exogenous Variables (X): Accounts for external factors influencing the target variable

$$y_t = \beta_0 + \beta_1 X_{t,1} + \cdots + \beta_k X_{t,k} + \epsilon_t \quad (5)$$

The expression of the SARIMAX equation (6) is given as:

$$y_t = \mu + \phi(B) y_t + \theta(B)\epsilon_t + \Phi(B)^s y_t + \Theta(B)^s \epsilon_t + X_t \beta \quad (6)$$

Where:

- y_t Dependent time series variable (Incident Count)
- μ Mean

- $\Phi(B)$ Non-seasonal AR terms.
- $\theta(B)$ Non-seasonal MA terms.
- $\Phi(B)^s$ Seasonal autoregressive SAR terms with seasonality s.
- $\Theta(B)^s$ Seasonal moving average SMA terms with seasonality s.
- $X_t\beta$ exogenous regressors.
- ϵ_t error term

Akaike Information Criterion (AIC): The AIC evaluates models based on the trade-off between the likelihood function and the number of parameters.

$$\text{AIC} = -2\log L + 2k \tag{7}$$

Where:
L is the maximum likelihood.
k is the number of estimated parameters in the model
Bayesian Information Criterion (BIC): The BIC applies a stricter penalty for model complexity.

$$\text{BIC} = -2\log L + k\log n \tag{8}$$

n is the number of observations
Evaluation Metrics : The MSE Equation

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \tag{9}$$

Where:
- n the number of data points (samples) used in the calculation.
- $\hat{y}_i$ the predicted the value of the i-th data point.
- y_i the actual values for the i-th data point.
- $(y_i - \hat{y}_i)^2$the squared the difference between the predicted and actual values for the i-th data point.

The Mean Absolute Error (MAE) is the average absolute difference between the predicted and actual maintenance times across all the data samples. The MAE is calculated as shown

$$\text{MAE} = \frac{1}{\text{n}}\sum_{\text{i}=1}^{\text{n}} |y_i - \hat{y}_i| \tag{10}$$

The R-squared (R^2) statistic measures the proportion of dependent variable variability that the model explains.

$$R^2 = 1 - \frac{\sum_{\text{i}=1}^{\text{n}}(y_i - \hat{y}_i)^2}{\sum_{\text{i}=1}^{\text{n}}(y_i - \bar{y})^2)} \tag{11}$$

Where: $\bar{y}$ is the mean of the actual values y_i
Conditions for Risk Categorization:
Risk levels are assigned based on forecasted values:
1. Compare forecasted values (y_t) against predefined thresholds:
   - Weekly: Low (<3), Medium (3-4), High (≥5).
   - Monthly: Low (<10), Medium (10-19), High (≥20).
2. Visualizations are colour-coded:
   - Green (Low), Orange (Medium), Red (High)

5. Analysis and results

Analysing Stationarity with the Augmented Dickey-Fuller (ADF): The weekly and monthly data test for check to stationarity or non-stationarity. From Table 2 both datasets needed to be transformed before modelling since they were non-stationary and failed to reject the null hypothesis (p-values > 0.05). First-order differencing was used to remove trends while maintaining seasonality to make stationarity (d=1, D=1). This allowed the SARIMAX model to function well on the modified dataset.

Time Series	ADF Statistic	p-value	Critical Value (1%)	Critical Value (5%)	Critical Value (10%)	Stationary Test
Weekly Data	-2.8429	0.0524	-3.4574	-2.8735	-2.5731	No
Monthly Data	-1.1566	0.6920	-3.5577	-2.9168	-2.5962	No

Table 2: ADF test results for weekly and monthly CCTV failures

The Autocorrelation function (ACF) and partial autocorrelation function confirmed strong seasonal patterns. Weekly data Figure 5 shows spikes 7,14, and 21 lags, confirming weekly seasonal patterns for ACF and PACF plots for the 52-week seasonal cycle. For monthly data, Figure 5 shows 12 and 24 lags confirming monthly seasonality for the 12-month pattern.

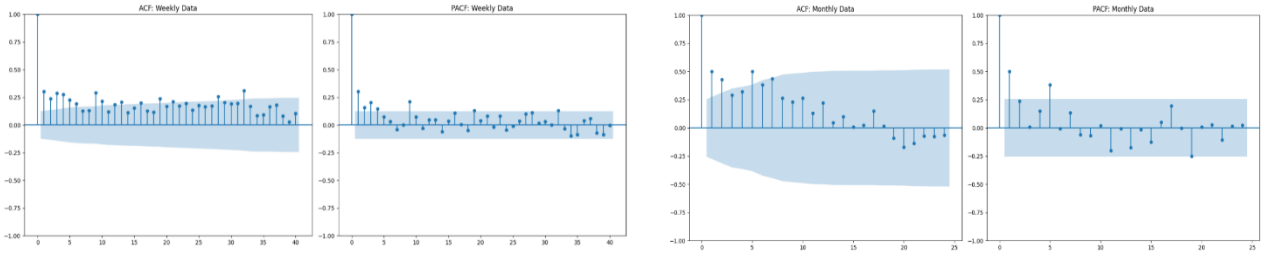


Figure 5: ACF and PACF for weekly and monthly CCTV failure

Forecasting with SARIMAX: The SARIMAX model was selected since the data showed seasonality and needed stationarity corrections. Based on differencing analysis, ACF, and PACF, model parameters were optimised in table 3:

Parameter	Weekly Value	Monthly Value
p, d, q	1, 1, 1	1, 1, 1
P, D, Q, s	1, 1, 1, 52	1, 1, 1, 12
Differencing Applied	d=1, D=1	d=1, D=1

Table 3: SARIMAX parameters for weekly and monthly CCTV failure

Differencing (d=1, D=1) retained seasonal components while effectively eliminating trends.

Model Performance and Evaluation:

Metric	Weekly data	Monthly data
MSE	23.9356	176.1724
MAE	4.0980	11.7853
R^2	0.2311	0.0854
AIC	560.3542	148.9746
BIC	576.8458	158.1178

Table 4: SARIMAX evaluation matrix for weekly & Monthly CCTV failure

The SARIMAX model results in Table 4 showed excellent predicted accuracy by effectively identifying failure trends in both weekly and monthly data. Monthly predictions offered important insights into long-term patterns and general trends in CCTV failures, whereas weekly forecasts successfully recognised short-term seasonal trends. The Bayesian Information Criterion (BIC) and Akaike Information Criterion (AIC), two model selection criteria, confirmed the model's resilience for both weekly and monthly forecasting. The model is a useful tool for railway maintenance planning since it can identify long-term seasonality and recurrent failure cycles.

Risk Analysis and Forecasting Visualization: Forecasted CCTV failures at various locations Figure 6 and 7 show during the next 52 weeks (weekly) and 12 months (monthly) using SARIMAX, classifying risk levels

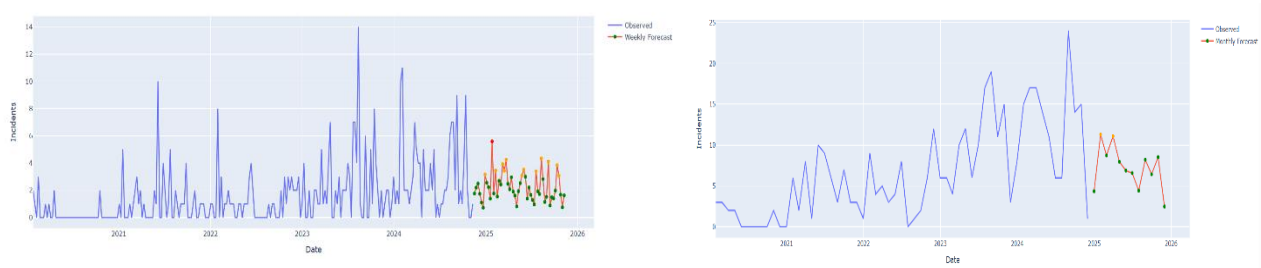


Figure 6: Forecasting weekly and monthly CCTV failures for SHIELDS

As green colour is low, orange is medium, and red represents a high risk of CCTV failure.

Figure 7: Forecasting weekly and monthly CCTV failures for LARGS

6. Conclusions and Contributions

In this paper, a forecasting framework for CCTV failure prediction utilising the SARIMAX model is presented**.** The CCTV system is acknowledged as a crucial element influencing fleet performance, and the main goal is to find event trends that lead to delays and service cancellations in various locations. The study demonstrates how predictive modelling may be used to analyse and forecast CCTV system failures, allowing for proactive maintenance planning to improve operational efficiency. The Proposed model approaches short-term (weekly) and long-term (monthly) predictions to assess maintenance risks, improve the CCTV surveillance system's dependability and facilitate well-informed maintenance schedule decision-making. By classifying failures according to risk levels, resources are allocated effectively, resulting in fewer unscheduled service interruptions and lower maintenance expenses.

Acknowledgment: This study is part of a wider study sponsored by Siemens Mobility Limited UK. The authors would like to thank Siemens for sponsoring this research.

References

Chuwang, D.D. and Chen, W., 2022. Forecasting daily and weekly passenger demand for urban rail transit stations based on a time series model approach. *Forecasting*, *4*(4), pp.904-924.

Dabral, P.P. and Murry, M.Z., 2017. Modelling and forecasting of rainfall time series using SARIMA. *Environmental Processes*, *4*(2), pp.399-419.

Wang, J. and Hu, J., 2015. A robust combination approach for short-term wind speed forecasting and analysis–Combination of the ARIMA (Autoregressive Integrated Moving Average), ELM (Extreme Learning Machine), SVM (Support Vector Machine) and LSSVM (Least Square SVM) forecasts using a GPR (Gaussian Process Regression) model. *Energy*, *93*, pp.41-56.

Tatarintsev, V.A. and Tolstosheev, A.K., 2021, April. Risk assessment and reliability of railway rolling stock elements. In *AIP Conference Proceedings* (Vol. 2340, No. 1). AIP Publishing.

Gołębiowski, P., Góra, I. and Bolzhelarskyi, Y., 2023. Risk assessment in railway rolling stock planning. *Archives of transport*, *65*(1).

Van Der Voort, M., Dougherty, M. and Watson, S., 1996. Combining Kohonen maps with ARIMA time series models to forecast traffic flow. *Transportation Research Part C: Emerging Technologies*, *4*(5), pp.307-318.

Akaike, H., 1974. A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), pp.716-723.

Schwarz, G., 1978. Estimating the dimension of a model. *The annals of statistics*, pp.461-464.

Ahmadpour, A., Mirhashemi, S.H. and Panahi, M., 2023. Comparative evaluation of classical and SARIMA-BL time series hybrid models in predicting monthly qualitative parameters of Maroon river. *Applied Water Science*, *13*(3), p.71.

Dubey, A.K., Kumar, A., García-Díaz, V., Sharma, A.K. and Kanhaiya, K., 2021. Study and analysis of SARIMA and LSTM in forecasting time series data. Sustainable Energy Technologies and Assessments, 47, p.101474.

Sirisha, U.M., Belavagi, M.C. and Attigeri, G., 2022. Profit prediction using ARIMA, SARIMA and LSTM models in time series forecasting: A comparison. *IEEE Access*, *10*, pp.124715-124727.

A study of maintenance efficacy of Gangway in a liquified petroleum gas plant

B. Alkali* and E. Pereira **

*/** *Glasgow Caledonian University, Department of Mechanical Engineering, Cowcaddense Road, Glasgow, G1 0BA*
email address: Babakalli.alkali@gcu.ac.uk

Abstract. This study presented in this paper investigates failures contributing to increased downtime in a Liquified Petroleum Gas (LPG) plant gangway system. The main goal is to identify critical failure modes, assess the effectiveness of current maintenance strategies, and propose improvement initiatives to minimise downtimes. The analysis utilises historical failure and maintenance data collected over a 4-year period. An exploratory analysis of the data is conducted using advanced statistical methods, and a failure mode and effect analysis is used when performing Fault Tree Analysis (FTA) to identify the critical failure modes and corresponding areas where the failure occur are classified as mechanical and instrumentation. The probability of failures and their impact on system reliability is calculated. Reliability analysis is conducted using the Fault Tree Analysis (FTA) approach**.** Statistical probability theory is employed to evaluate the severity of downtime events systematically. The result provides an insight into more efficient operational and maintenance planning for the LNG plants critical systems.

1. Introduction

The transition within the energy sector towards the use of Liquefied Natural Gas (LNG) has increased due to its environmental advantages and capacity to diversify energy sustainability portfolios (Botão, De Medeiros Costa and Dos Santos, 2023). As a result, the operational efficiency of LNG plants is crucial for meeting the needs of customers. Ensuring personnel safety is challenging while maintaining operational efficiency. Within this context, the Liquified Petroleum Gas (LPG) gangways play a critical role in facilitating safe personnel access, ensuring operational efficiency while enabling maintenance activities, complying with safety regulations, and facilitating emergency response protocols (Pereira, Alkali and Niculita, 2019). Downtime in industrial settings can have significant implications for productivity and efficiency. Ahmad, Saleh and Rahmaniar (2023) pointed out that downtime can result from various factors including equipment failure, disturbances in operations, or scheduled maintenance activities. It implies that higher downtime in the LPG gangway can affect the operational effectiveness of the plant. A significant number of failure incidents require the urgent need for a deeper understanding of the root causes behind failures. While historical failure analysis provides valuable insights into the nature of failure incidents, there is also the need to further investigate the maintenance requirements to reduce the occurrence of the failure incidents. This paper aims to conduct a comprehensive analysis of LPG gangway system downtime. Through the utilisation of descriptive statistical techniques, the study explores the factors influencing higher downtimes of the gangway system. The expected outcome is to provide recommendations on maintenance and operational actions towards minimising downtimes.

1.1 Gangway Structural framework

The structural framework of the gangway and its constituent part is presented in the diagram Figure 1. The gangway and its subsystem structure allow personnel to move from the ship's deck to the offshore platform, enabling safe and convenient access between the two points. The gangway type used in the LNG plant is referred to as the LPG (Liquified Petroleum Gas) Gangway and it is powered by a hydraulic power pack. The LPG gangway is about 20-30 meters high and consist of a variety of mechanical, electrical and instrumentation and control system (Pereira, Alkali and Niculita, 2019).

A review of historical failure and maintenance associated with this LPG gangway system is presented and statistical analysis of failure data is discussed later.

Figure 1: LPG Gangway (Pereira et al., 2019)

2. Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) is essential in statistical research as it lays the groundwork for understanding the dataset before engaging in more complex analytical techniques. This initial step includes cleaning, organising, and describing the collected data (Song, 2023). The EDA is instrumental in uncovering underlying patterns within the data and providing valuable insights (Hodeghatta and Nayak, 2023). Therefore, performing EDA helps to reveal trends and thereby determine the best statistical techniques to achieve the objectives specified in the study. The historical data of downtime incidents in days per year recorded from 2016 to 2019 is presented in Table 1. The highest number of downtime incidents occurred in the year 2018, which is 47/141 of the total number incidents. We consulted with maintenance personnel regarding the downtime incidents; they noted inefficiencies and missing observation in their data recording process, especially in 2019 the incidents do not reflect on the maintenance practices.

Table 1. Downtime incidents occurrences.

Incident Year	Number of incidents	Days per year	Percentage (%)
2016	12	34	16
2017	14	26	19
2018	31	47	42
2019	16	34	22
Total	**73**	**141**	**100**

However, we examine the effectiveness of the preventive maintenance and corrective maintenance actions. The total numbers of failure incidents following the maintenance actions over 4 years in presented in Table 2. Notice that the PM (48%) and CM (50%) actions following the incidents are almost equal. However, 2018 account for 42% of the downtime days, which indicate the maintenance inefficiencies.

Table 2. Maintenance Actions

Maintenance Strategy	Number of incidents	Days	Percentage (%)
CM	37	73	52
PM	36	68	48
Total	**73**	**141**	**100**

3. Downtime analysis

The gangway maintenance downtime requirements vary with time. The variation relates to the type of components. The gangway consists of mechanical, electrical and instrumentation and control systems. The total number maintenance on the systems in hours is presented in Figure 2 below. Notice that the mechanical systems have the highest number of maintenance hours followed by instrumentation and electrical systems.

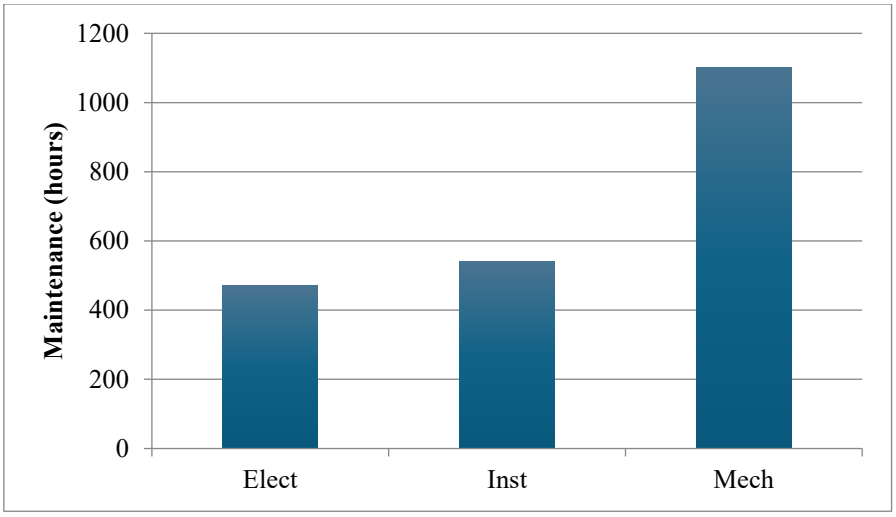


Figure 2 Critical systems maintenance duration (hours)

The graph in Figure 3 gives the downtime hours of all the electrical, instrumentation and mechanical expected normal duration to conduct maintenance and replacement and the manhours required to do the work. This reflects on the current inefficiencies in the maintenance practices and hence the complexities of the underling failure cause diagnostics in the LNG plant.

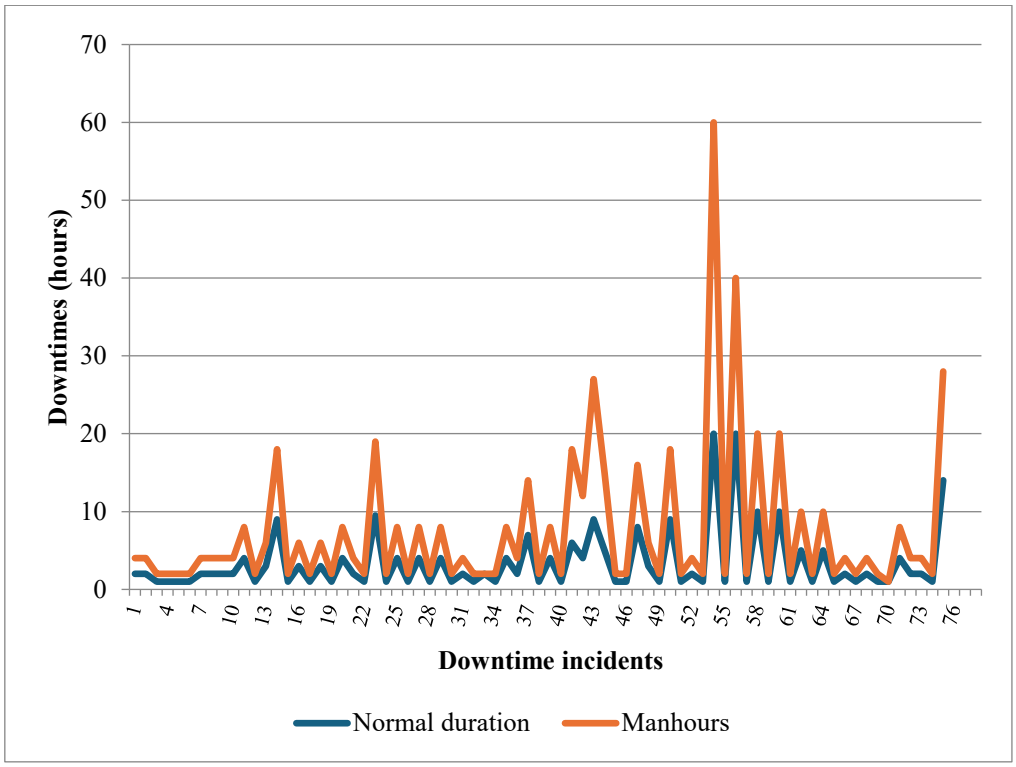


Figure 3 Maintenance downtime and manhours

The graph in Figure 3 above clearly shows the high manhours and does not have an effect nor reduce the consequences of failures. This also indicate the lack of adequate diagnostics to identify failure causes.

3.1 Gangway Failure Mode and Effect Analysis

Historical failure information from 2016 – 2019 is reviewed, (SAP Failure History Database). A broad-brush failure mode and effect analysis is conducted to complement our fault tree analysis later in section 4. Following our failure analysis, key failure modes were identified and presented in Table 3.

Table 3 Gangway Failure Modes

Failure Mode	**Count**	**Percentage (%)**
The bulwark ladder not responding	2	6.67
Gangway Rope Cut/Damage	3	10.00
Hydraulic Oil Leak on LPG Gangway	12	40.00
Intermittent loud/unusual sound	1	3.33
LPG Gangway - Provision of Spring Lines	1	3.33
LPG Gangway cable out of the drum	1	3.33
LPG Gangway Can Not Hoisted Up	3	10.00
LPG gangway console loss of power	1	3.33
LPG Gangway Malfunction	1	3.33
LPG Gangway Pin lock alarm	1	3.33
LPG Gangway Safety Pin	1	3.33
LPG Jetty Ladder Failure	1	3.33
The LPG/Cond Gangway piston plate is bent	1	3.33
The Gangway Did Not Stop Floor 2	1	3.33
Total	**30**	**100**

The statistics show that 40% of gangway failure incidents occurred due to “Hydraulic Oil Leak on LPG Gangway”, which is the highest percentage of gangway failures that occurred within the period. Some of these failure modes are attributed to both mechanical and instrumentation systems.

The statistics show that 50% of the Mechanical Failures occurred due to the “Hydraulic Oil Leak on LPG Gangway”, which is significantly higher than the occurrences of other failure modes. Based on the results in Table 4, the highest percentages of Gangway failures are Mechanical Failures, and the most occurred Mechanical Failure is “Hydraulic Oil Leak on LPG Gangway”. Accordingly, it can be concluded that “Hydraulic Oil Leak on LPG Gangway” is the most critically impacted failure mode on the LPG Gangway system.

Table 4 Gangway Failure Modes in the Mechanical Area.

Failure Mode	**Frequency**	**Percentage (%)**
The bulwark ladder not responding	1	4.17
GANGWAY ROPE CUT/DAMAGE	3	12.50
Hydraulic Oil Leak on LPG Gangway	12	50.00
intermittent loud/unusual sound	1	4.17
LPG GANGWAY - PROVISION OF SPRING LINES	1	4.17
LPG Gangway cable out of the drum	1	4.17
LPG GANGWAY CAN NOT HOISTED UP	3	12.50
LPG JETTY LADDER FAILURE	1	4.17
The LPG/Cond Gangway piston plate is bent	1	4.17
Total	**24**	

The failure modes associated with the Instrumentation systems are very negligible according to Table 5. Therefore, the occurrences of individual failure modes under the instrumentation area does not have a significant impact on the Gangway System.

Table 5. Gangway Failure Modes *in* the Instrumentation Area.

Failure Mode	**Frequency**
The bulwark ladder not responding.	1
LPG gangway console loss of power	1
LPG Gangway Malfunction	1
LPG Gangway Pin lock alarm	1
LPG Gangway Safety Pin	1
THE GANGWAY DID NOT STOP FLOOR 2	1
Total	**6**

4. Reliability Analysis

A reliability analysis is performed using the Fault Tree Analysis (FTA) approach. The failure rate of the gangway system attributed mechanical, and instrumentation is calculated, and the results are presented in Table 6 and Table 7.

Based on the failure rate calculation, the unreliability and the MTTF (Mean Time to Failure) of the Gangway System is estimated, the Isograph Reliability Workbench software is used develop an FTA logic diagram in Figure 4.

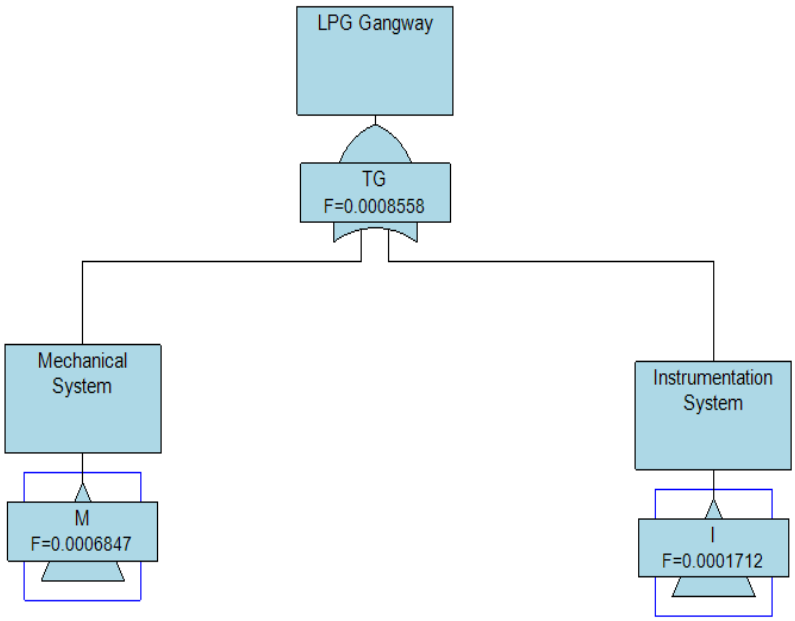


Figure 4: Fault Tree logic diagram of Gangway System

Figure 4 shows the software-generated reliability calculations. Additionally, the software is used to calculate the reliability and estimate the MTTF (hours and days) for each system as shown in Table 6 and Table 7 respectively.

Table 6: Reliability Summary

System	Unreliability	Reliability (%)
LPG Gangway System	0.0008558	99.91%
Mechanical System	0.0006847	99.93%
Instrumentation System	0.0001712	100.0%

Table 7: MTTF Summary

System	MTTF in hours	MTTF in days
LPG Gangway	1167	48.625
Mechanical	1459	60.79
Instrumentation	5838	243.25

The MTTF results imply that on average, the LPG gangway system fails about once every 48.625 days, while the Mechanical and Instrumentation systems fail about once every 60.79 days and 243.25 days respectively. The results imply that Mechanical failures critically impact on the Gangway System and the most prevalent failure is because of the "Hydraulic Oil Leak".

5. Conclusions

The analysis presented in this paper shows that downtime hours associated with CM and PM maintenance actions conducted in the gangway system are not much different. This results from the analysis presented confirmed and highlighted that the PM strategies taken towards the downtime incidents are inefficient. However, in the FTA analysis, it was found that the reliability of the Gangway is a significantly high. A possible reason for high maintenance duration despite high systems reliability is because of lack of appropriate diagnostic tool, training and systems, implying further investment and procurement of effective maintenance tools. In practice maintenance, if the preventive maintenance strategies are not effective, the frequency of failures should be higher and thereby reliability should be a lower value. This discrepancy might be due to several factors:

Severe downtime incidents: Downtime measures the duration of operational disruption, whereas reliability in the FTA measures the probability of system success without failure. CM actions associated with failure incidents. Accordingly, the incidents reported within the period is complex and might have taken a longer time to fix. However, the analysis identified 'Hydraulic Oil Leaks' as the most critical failure. Increased downtime can occur due to the severity of this failure. The recommendation is to focus on preventive maintenance using oil leak detector sensor to reduce the hydraulic oil leaks failure mode incidents.

Reporting system delay: Failure should be identified and resolved on time, since the downtime due to failures is high, the effective use of remote condition monitoring integrated with computerized maintenance management system to support effective detection of anomalies, reporting and responding to delays following a failure incident.

References

Ahmad, I., Saleh, R. and Rahmaniar, R. (2023) 'Analysis of The Effect Of Down Time on Productivity at Pt. Eastern Pearl Flour Mills Makassar', *Journal of Agriculture*, 2(01), pp. 45–53. Available at: https://doi.org/10.47709/joa.v2i01.2334.

Botão, R.P., de Medeiros Costa, H.K. and dos Santos, E.M. (2023) 'Global Gas and LNG Markets: Demand, Supply Dynamics, and Implications for the Future', *Energies*, 16(13). Available at: https://doi.org/10.3390/en16135223.

Pereira, E.D.S., Alkali, B. and Niculita, I.O., (2019) 'Liquefied petroleum gas plant maintenance: a case study of gangway failure analysis'. *Proceedings of the 16th International Conference on Condition Monitoring and Asset Management, BINDT 2019.*

Song, X. (2023) 'A Brief Introduction to Exploratory Data Analysis', *Advances in Engineering Technology Research*, 4(1), p. 420. Available at: https://doi.org/10.56028/aetr.4.1.420.2023.

Hodeghatta, U.R. and Nayak, U. (2023) *Practical Business Analytics Using R and Python*, *Practical Business Analytics Using R and Python*. 2nd ed. A press.

Shannon entropy and its variants for the maintenance of large meshed system architectures

Christian Tanguy*

*Orange Research, France, email address: christian.tanguy@orange.com

Abstract Shannon introduced entropy in his seminal theory of communications. This notion has been used in a few papers at the turn of the century, for instance for the performance assessment of intelligent machines. By contrast, in the last five years a fast growing literature has applied the Shannon entropy and its variants to system maintenance in various domains: radars, metro signaling, motor fault detection, etc. Yet the described systems are mostly either small or have a simple series or parallel architecture. The purpose of this paper is to consider the different versions of entropy used in maintenance studies and compute their values for large and possibly meshed networks. Asymptotic expansions as a function of the system size are given and explained for various architectures of practical interest. The influence of the failure distribution density — not necessarily exponential — is also addressed.

1. Introduction

Resilience has become a crucial aspect of modern day systems. It can be better ensured by an improved detection of anomalies, and adequate maintenance policies. The inherent uncertainty in the occurrence of failures makes the determination of these policies a complex task. The data uncertainty has led Shannon to introduce the notion of entropy in his theory of communication (Shannon 1948a, Shannon 1948b). Other definitions of entropies have been proposed (Rényi 1961, Havrda 1967, Tsallis 1988). The entropies were first used in reliability assessment proposals (Musto 1997) and to help determining the distributions of failures or improve Monte Carlo simulations (Hui 2005, Kroese 2007). Their expressions for various failure distributions have also been obtained. Since these developments, entropy calculations have been performed not on failure distributions of a single component, but on entire systems, albeit of small or moderate sizes. Variants of the Shannon entropy have also been proposed (Rao 2004, Di Crescenzo 2009). The range of applications has grown steadfastly in the last five to ten years (Alshehri 2024, Chan 2023, Gaona 2019, Hirata 2020, Healy 2025, Hu 2019, Iranpour 2020, Kayid 2024, Raza 2019, Sengupta 2016, Singh 2015, Wen 2022).

The paper is organized as follows. We present in Section 2 the main general formulas for the different entropies considered in this work, and their analytical expressions for series and parallel systems of n elements, the lifetime distribution of which is exponential. We expose in Section 3 the case of a large non-series-parallel system, for which the reliability has a simple, analytical expression even for an arbitrarily large size n. Different asymptotic expansions are also given, which illustrate the key parameters of the system architecture. Section 4 is devoted to linear consecutive k-out-of-n:F systems, already considered in the literature (Alshehri 2024, Kayid 2024) in the case $2k \geq n$ and $n \leq 8$. The general asymptotic expressions of three entropies are also given. Section 5 demonstrates that the lifetime distribution of the elements plays an important role in the size dependence of the entropies. The obtained results show how the shape parameter of the Weibull distribution modifies the asymptotic expressions. We then conclude and suggest a few directions for the extension of the present work.

2. Definitions and simple results

The historical definitions of entropy associated with distributions of probabilities (Shannon 1948a, Shannon 1948b, Rényi 1961, Havrda 1967, Tsallis 1988) were expressed for discrete distributions, using the logarithm of base 2. In the following, their expressions will be directly given for continuous distribution functions $f(x)$, where x will range from 0 to ∞, and the natural logarithm will be kept throughout.

The Shannon (H), Rényi (H_α), Havrda-Charvát ($S(a)$), and Tsallis (S_q) entropies are thus defined as

$$H = -\int_0^\infty f(x)\ln f(x)\,dx, \tag{1}$$

$$H_\alpha = \frac{1}{1-\alpha}\ln\int_0^\infty f(x)^\alpha\,dx \qquad (\alpha>0,\alpha\neq 1), \tag{2}$$

$$S(a) = \frac{1}{1-e^{1-a}}\left(1-\int_0^\infty f(x)^a\,dx\right) \qquad (a>0,a\neq 1), \tag{3}$$

$$S_q = \frac{k}{q-1}\left(1-\int_0^\infty f(x)^q\,dx\right) \qquad (q>0,q\neq 1). \tag{4}$$

The last three definitions reduce to H in the limits $\alpha\to 1$, $a\to 1$ and ($k=1$ and $q\to 1$), respectively. Recently, the cumulative residual entropy $\mathscr{CRE}$ (Rao 2004) and the cumulative entropy $\mathscr{CE}$ (Di Crescenzo 2009) have been proposed, where $F(x)$ is the cumulative distribution function associated to $f(x)$, and $R(x)=1-F(x)$:

$$\mathscr{CRE} = -\int_0^\infty R(x)\ln R(x)\,dx, \tag{5}$$

$$\mathscr{CE} = -\int_0^\infty F(x)\ln F(x)\,dx. \tag{6}$$

2.1. *Expressions of entropy for a system with exponential distribution*

The various entropies mentioned above have been calculated for various well-known distributions (exponential, Weibull, Inverse Gaussian, etc.). In the present paper, we focus on the result for *whole systems* consisting of statistically independent identical elements, the probability of operation of which is given by $p(t)$. In such cases, $f(t)=-p'(t)$, and the reliability of the system is given as a polynomial in $p(t)$, namely $R(p)$. The integration over time from 0 to ∞ is replaced by an integration over p (or over $q=1-p$) from 0 to 1.

When the individual lifetime distribution is exponential, $p(t)=\exp(-\lambda t)$ and $t=-\frac{1}{\lambda}\ln p$, so that

$$H = -\ln\lambda - \int_0^1 R'(p)\ln(pR'(p))\,dp = -\ln\lambda + \int_0^1 R'(q)\ln(-(1-q)R'(q))\,dq, \tag{7}$$

$$\mathscr{CRE} = -\frac{1}{\lambda}\int_0^1 R(p)\ln(R(p))\,\frac{dp}{p} = -\frac{1}{\lambda}\int_0^1 R(q)\ln(R(q))\,\frac{dq}{1-q}, \tag{8}$$

$$\mathscr{CE} = -\frac{1}{\lambda}\int_0^1 (1-R(p))\ln(1-R(p))\,\frac{dp}{p} = -\frac{1}{\lambda}\int_0^1 (1-R(q))\ln(1-R(q))\,\frac{dq}{1-q}. \tag{9}$$

2.2. *Expressions of entropy for series and parallel systems*

For a system consisting of n elements in series, $R(p)=p^n$, all these entropies decrease with n:

$$H_{\text{series}} = 1-\ln(n\lambda), \tag{10}$$

$$\mathscr{CRE}_{\text{series}} = \frac{1}{n\lambda}, \tag{11}$$

$$\mathscr{CE}_{\text{series}} = \frac{1}{n\lambda}\left(\frac{\pi^2}{6}-1\right). \tag{12}$$

For a system consisting of n elements in parallel, $R(p)=1-(1-p)^n=1-q^n$, and

$$H_{\text{parallel}} = 1-\ln(n\lambda)+\psi(n)-\psi(1) \qquad \xrightarrow[n\to\infty]{} 1-\ln\lambda+\gamma-\frac{1}{2n}+\cdots, \tag{13}$$

$$\mathscr{CRE}_{\text{parallel}} = \frac{n}{\lambda}\left(\frac{\pi^2}{6}-\sum_{i=1}^{n}\frac{1}{i^2}\right) \qquad \xrightarrow[n\to\infty]{} \frac{1}{\lambda}\left(1-\frac{1}{2n}+\frac{1}{6n^2}+\cdots\right), \tag{14}$$

$$\mathscr{CE}_{\text{parallel}} = \frac{1}{\lambda}\sum_{m=1}^{n}\frac{1}{m}\left(\psi\left(1+\frac{m}{n}\right)-\psi(1)\right) \qquad \xrightarrow[n\to\infty]{} \frac{1}{\lambda}\int_0^1 \frac{v\ln v}{(1-v)\ln(1-v)}\,dv \approx \frac{1.2577469}{\lambda}. \tag{15}$$

where ψ is the digamma function and $\gamma\approx 0.5772$ the Euler constant. All these entropies increase with n.

3. Entropies for a simple ladder network

Let us now turn to the computation of entropies for systems that cannot be described by series/parallel graphs anymore, the size of which being possibly very large. The simple ladder network considered here, where all nodes are perfect and all links have the same reliability $p(t)$, is shown in Figure 1.

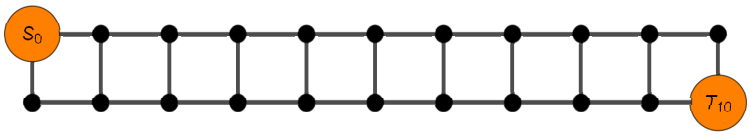


Figure 1. Simple ladder network: the source is S_0, the destination is T_{10}.

The two-terminal reliability $R_n \equiv \mathrm{Rel}_2(S_0 \to T_n)$ can be found exactly (Tanguy 2006, Tanguy 2007):

$$R_n = \alpha_0\, \zeta_0^n + \alpha_+\, \zeta_+^n + \alpha_-\, \zeta_-^n, \tag{16}$$

$$\zeta_0 = q\,(1-q), \qquad \zeta_\pm = \frac{1-q}{2}\left(1+2\,q\,(1-q) \pm \sqrt{1+4\,q^2\,(1-q)^2}\right), \tag{17}$$

$$\alpha_0 = -\frac{q}{2}, \qquad \alpha_\pm = \frac{2-q}{4} \pm \frac{2+q-4\,q^2+2\,q^3}{4\sqrt{1+4\,q^2\,(1-q)^2}}. \tag{18}$$

The variations of three entropies are displayed in Figure 2 as a function of the size n of the system in the case of an exponential distribution (with $\lambda = 1$). The calculations have been performed for $n \leq 400$.

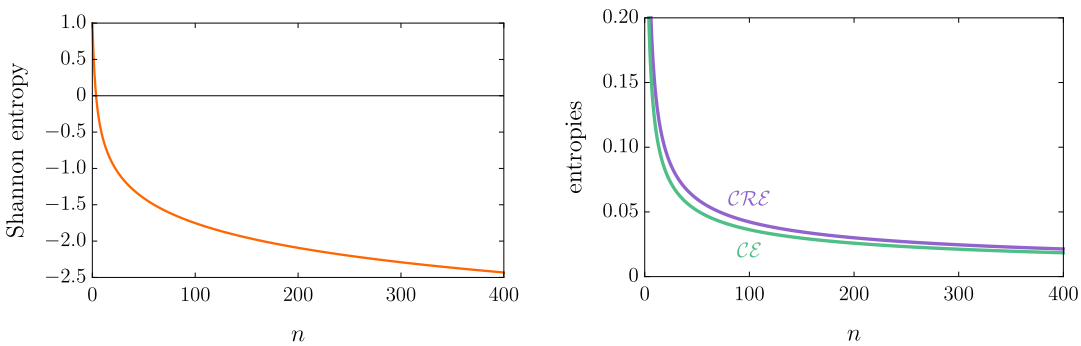


Figure 2. Entropies for the two-terminal availability $\mathrm{Rel}_2(S_0 \to T_n)$: Shannon (left); $\mathscr{CE}$ and $\mathscr{CRE}$ (right).

When n is large and q small, $R_n \approx \alpha_+ \exp(n \ln \zeta_+) \to \exp(-(n+2)\,q^2)$. This leads to

$$H_n = 1 - \ln(2\,\lambda) + \frac{\gamma}{2} - \frac{1}{2}\ln(n+2) - \frac{\sqrt{\pi}}{2\sqrt{n+2}} + \frac{29}{4\,(n+2)} + \cdots, \tag{19}$$

$$\mathscr{CRE}_n = \frac{1}{\lambda}\left(\frac{\sqrt{\pi}}{4\sqrt{n+2}} - \frac{1}{2\,(n+2)} + \cdots\right), \tag{20}$$

$$\mathscr{CE}_n = \frac{1}{\lambda}\left(\frac{C_1}{\sqrt{n+2}} - \frac{\frac{\pi^2}{6}-1}{2\,(n+2)} + \cdots\right), \tag{21}$$

where $C_1 = -\displaystyle\int_0^\infty (1-e^{-u})\ln(1-e^{-u})\,\frac{du}{2\sqrt{u}} \approx 0.3796294$.

4. Entropies for linear consecutive systems

Linear consecutive systems are a well-known configuration. Their cumulative entropies have been studied in (Alshehri 2024, Kayid 2024), assuming $2\,k \geq n$ and $n \leq 8$. In this section, we consider larger systems, with no restriction on k.

4.1. *Linear consecutive 2-out-of-n:F systems*

When $k=2$, the two-terminal reliability R_n can be explicitly expressed analytically as $R_n = \alpha_+ \zeta_+^n + \alpha_- \zeta_-^n$, where

$$\zeta_\pm = \frac{1}{2}\left(1-q\pm\sqrt{(1-q)(1+3q)}\right) \tag{22}$$

$$\alpha_\pm = \frac{1}{2}\left(1\pm\frac{1+q}{\sqrt{(1-q)(1+3q)}}\right) \tag{23}$$

The asymptotic expansions now read, since for large n the term $\alpha_+ \zeta_+^n$ prevails and $R_n \approx \exp(-(n-1)q^2)$

$$H_n \longrightarrow 1-\ln(2\lambda)+\frac{\gamma}{2}-\frac{1}{2}\ln(n-1)+\frac{\sqrt{\pi}}{\sqrt{n-1}}+\cdots, \tag{24}$$

$$\mathscr{CRE}_n \longrightarrow \frac{1}{\lambda}\left(\frac{\sqrt{\pi}}{4\sqrt{n-1}}+\frac{1}{n-1}+\cdots\right), \tag{25}$$

$$\mathscr{CE}_n \longrightarrow \frac{1}{\lambda}\left(\frac{C_1}{\sqrt{n-1}}+\frac{\frac{\pi^2}{6}-1}{n-1}+\cdots\right). \tag{26}$$

The leading terms are quite similar to those for the simple ladder network of Section 3.

4.2. *Generalization*

If the asymptotic expression of the reliability has the form $R_n \approx \exp(-\tilde{n}q^k)$ with $\tilde{n}=an+b$ (a a positive integer, and b an integer), the Shannon and cumulative entropies become

$$H_n \longrightarrow 1-\ln(k\lambda)+\left(1-\frac{1}{k}\right)\gamma-\frac{1}{k}\ln\tilde{n}+\cdots, \tag{27}$$

$$\mathscr{CRE}_n \longrightarrow \frac{1}{\lambda}\frac{1}{k^2}\Gamma\left(\frac{1}{k}\right)\frac{1}{\tilde{n}^{1/k}}+\cdots, \tag{28}$$

$$\mathscr{CE}_n \longrightarrow -\frac{1}{\lambda}\int_0^\infty (1-e^{-u})\ln(1-e^{-u})\frac{du}{ku^{1/k}}\frac{1}{\tilde{n}^{1/k}}+\cdots. \tag{29}$$

For the linear consecutive k-out-of-n:F, $\tilde{n}=n-k+1$. The decrease of the cumulative entropies as a function of n is very slow: the leading term is in $1/(n-k+1)^{1/k}$, the following term in $1/(n-k+1)^{2/k}$, etc. Furthermore, for $k\geq 4$ and n large enough, $\mathscr{CE}_n > \mathscr{CRE}_n$.

5. Beyond the exponential assumption

In the previous Sections, the lifetime distribution of all the elements of the system was assumed to be exponential. In this section, the changes brought by the use of a Weibull distribution are explained.

5.1. *Difference between Weibull and exponential distributions for the Shannon entropy*

The Shannon entropy of systems also depends of the underlying reliability of each element. In the case of the Weibull distribution, $p(t)=\exp\left(-(\lambda t)^\alpha\right)$, where the scaling factor λ has been kept and a shape factor α is chosen. This leads to

$$H_{\text{Weibull}} = -\int_0^1 R'(p)\ln\left[\lambda\,\alpha\,p\,(-\ln p)^{1-\frac{1}{\alpha}}R'(p)\right]dp, \tag{30}$$

$$H_{\text{Weibull}} - H_{\text{exponential}} = -\ln\alpha-\left(1-\frac{1}{\alpha}\right)\int_0^1 R'(p)\ln(-\ln p)\,dp. \tag{31}$$

This difference could be important in the case of large systems. If the reliability polynomial is $R(p)=\sum_i a_i p^i$, its expression is very simple

$$H_{\text{Weibull}} - H_{\text{exponential}} = -\ln\alpha+\left(1-\frac{1}{\alpha}\right)\left(\gamma+\sum_i a_i\ln i\right). \tag{32}$$

Note that for the two cumulative entropies $\mathscr{CRE}$ and $\mathscr{CE}$, using a Weibull distribution in equations (8–9) merely requires an extra factor $\frac{1}{\alpha}(-\ln(1-q))^{\frac{1}{\alpha}-1}$ in the integrals.

5.2. *Asymptotic expressions of the entropies with the Weibull distribution*

Considering again that $R_n \longrightarrow \exp(-\tilde{n}q^k)$ with $\tilde{n} = an+b$ leads to

$$H_n \longrightarrow 1-\ln(\alpha k\lambda)+\left(1-\frac{1}{\alpha k}\right)\gamma-\frac{1}{\alpha k}\ln\tilde{n}, \tag{33}$$

$$\mathscr{CRE}_n \longrightarrow \frac{1}{\lambda}\frac{1}{(\alpha k)^2}\Gamma\left(\frac{1}{\alpha k}\right)\frac{1}{\tilde{n}^{1/(\alpha k)}}, \tag{34}$$

$$\mathscr{CE}_n \longrightarrow -\frac{1}{\lambda}\int_0^\infty \frac{(1-e^{-u})\ln(1-e^{-u})\,du}{\alpha k u^{1/(\alpha k)}}\frac{1}{\tilde{n}^{1/(\alpha k)}}. \tag{35}$$

The change from exponential to Weibull distribution manifests itself by the mere multiplication of k by the shape parameter α. It shows that the size variation of the entropies depends not only on the architecture of the system, but also on the specific lifetime distribution under consideration.

6. Conclusion and outlook

An increasing number of recent publications have been using entropy in maintenance-related studies. In this work, we have shown that the Shannon entropy (Shannon 1948a, Shannon 1948b) and its cumulative variants (Rao 2004, Di Crescenzo 2009) can be calculated for very large systems of identical and statistically independent elements. The obtained asymptotic expressions can be very different, and may exhibit series-like ou parallel-like behaviors. They depend on the architecture of the system as well as on the lifetime distribution of the components, as evidenced by the difference observed between exponential and Weibull distributions. An extension of the present work will address the consequence of the dependence between elements, for instance with the β-factor model of common-cause failures. The entropy of systems made of at least two groups of equipments will also be investigated.

Acknowledgements

The author wishes to thank Guillaume Fraysse and Stéphane Gosselin for their support, and Catherine Gourdon for very useful suggestions.

References

Alshehri MA, Kayid M (2024) Cumulative entropy properties of consecutive systems. *AIMS Mathematics* **9**(11), 31770–31789.

Asadi M, Ebrahimi N (2000) Residual entropy and its characterizations in terms of hazard function and mean residual life function. *Statistics & Probability Letters* **49**(3), 263–269.

Chan J, Papaioannou I, Straub D (2023) Bayesian improved cross entropy method for network reliability assessment. *Structural Safety* **103**, 102344.

Di Crescenzo A, Longobardi M (2009) On cumulative entropies. *Journal of Statistical Planning and Inference* **139**(12), 4072–4087.

Gaona OP, González-González DS, Fuentes-Huerta MA, Praga-Alejo RJ (2019) Maximum entropy model applied to Reliability Centered Maintenance scheme for replaceable systems. In *International Conference on Mechatronics, Electronics and Automotive Engineering (ICMEAE 2019)*, pp. 140–145.

Havrda J, Charvát F (1967) Quantification Method of Classification Processes — Concept of Structural *a*-Entropy. *Kybernetica* **3**(1), 30–35.

Healy M, Cassandro R, Li ZS, Baum A, Jørgensen, TM (2019) Cost-Effective Active Learning for Cable Network Impairment Data. In *IEEE International Conference on Prognostics and Health Management (ICPHM 2025)*, pp. 1–7.

Hirata R, Arizono I, Takemoto Y (2020) On reliability analysis in priority standby redundant systems based on maximum entropy principle. *Quality Technology & Quantitative Management* **18**(1), 117–133.

Hu X, Niu R, Tao T (2019) Research on Entropy based Corrective Maintenance Difficulty Estimation of Metro Signaling. In *Intelligent Transportation Systems Conference (ITSC 2019)*, pp. 79–85.

Hui KP, Bean N, Kraetzl M, Kroese DP (2005) The Cross-Entropy Method for Network Reliability Estimation. *Annals of Operations Research* **134**(1), 101–118.

Iranpour M, Hejazi MA, Shahidehpour M (2020) A Unified Approach for Reliability Assessment of Critical Infrastructures Using Graph Theory and Entropy. *IEEE Transactions on Smart Grid* **11**(6), 5184–5192.

Kayid M, Alshehri MA (2024) Shannon differential entropy properties of consecutive k-out-of-n:G systems. *Operations Research Letters* **57**, 107190.

Kroese DP, Hui K (2007) Applications of the Cross-Entropy Method in Reliability. In *Computational Intelligence in Reliability Engineering: New Metaheuristics, Neural and Fuzzy Techniques in Reliability* (Levitin, Ed.), pp. 37–82, Springer Berlin Heidelberg.

Musto JC, Saridis GN (1997) Entropy-based reliability analysis for intelligent machines. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **27**(2), 239–244.

Rao M, Chen Y, Vemuri BC, Wang F (2004) Cumulative Residual Entropy: A New Measure of Information. *IEEE Transactions on Information Theory* **50**(6), 1220–1228.

Rausand M, Høyland A (2004) *System Reliability Theory (second ed.).* Hoboken: Wiley.

Raza A, Ulansky V (2019) Optimization of Condition Monitoring Decision Making by the Criterion of Minimum Entropy. *Entropy* **21**(12), 1193.

Rényi A (1961) On measures of information and entropy. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume I, pp. 547–561.

Sengupta DC, Sengupta JD (2016) Application of Graph Entropy in CRISPR and Repeats Detection in DNA Sequences. *Computational Molecular Bioscience* **6**(3), 41–51.

Shannon CE (1948a) A mathematical theory of communication. *The Bell System Technical Journal* **27**(3), 379–423.

Shannon CE (1948b) A mathematical theory of communication. *The Bell System Technical Journal* **27**(4), 623–656.

Singh VP, Oh J (2015) A Tsallis entropy-based redundancy measure for water distribution networks. *Physica A: Statistical Mechanics and its Applications* **421**, 360–376.

Tanguy C (2006) Exact solutions for the two- and all-terminal reliabilities of a simple ladder network. *arXiv:cs/0612143 [cs.PF].*

Tanguy C (2007) What is the probability of connecting two points? *Journal of Physics A: Mathematical and Theoretical* **40**, 14099–14116.

Tsallis C (1988) Possible Generalization of Boltzmann-Gibbs Statistics. *Journal of Statistical Physics* **52**(1–2), 479–487.

Wen T, Gao Q, Chen Y, Cheong KH (2022) Exploring the vulnerability of transportation networks by entropy: A case study of Asia-Europe maritime transportation network. *Reliability Engineering & System Safety* **226**, 108578.

Enhancing Rail Network Resilience: A Hybrid Simulation-Optimisation Model for Disruption Management Using Genetic Algorithms

Kang Rui Tan*, Rasa Remenyte-Prescott**, Darren Prescott***, Adam Cooper Watson****, Andrew Dutton*****

* *Resilience Engineering Research Group, University of Nottingham, UK + email address: evykt4@nottingham.ac.uk*
** *Resilience Engineering Research Group, University of Nottingham, UK, + email address: evzrr@exmail.nottingham.ac.uk*
*** *Resilience Engineering Research Group, University of Nottingham, UK + email address: evzdrp@exmail.nottingham.ac.uk*
**** *Network Rail, Western House, Swindon SN1 1BD, UK + email address: Adam.CooperWatson@networkrail.co.uk*
***** *Network Rail, Enterprise House, London W26JX, UK + email address: Andrew.Dutton@networkrail.co.uk*

Abstract. Disruptions on a rail network can halt operations at affected sections, leading to the rapid propagation of delays throughout the highly congested network, as observed in the UK rail network. Therefore, timely responses to disruptions are essential for minimising the delay propagation. This paper proposes a hybrid simulation-optimisation model that employs **genetic algorithms** to automate near real-time decision-making for disruption management. This research aims to determine optimal service alterations for targeted train services, facilitating faster recovery while minimising **total delay minutes**, **additional costs incurred due to disruption management** and the **number of train cancellations** across the network. The performance of the model is evaluated by comparing its outcomes with example empirical data from a historical disruption. Results demonstrate that the model effectively balances the three key objectives while adhering to operational constraints, therefore ensuring its practical applicability. The findings highlight the types of train services most beneficial to alter and quantify the improvement achieved, including reductions in service cancellations and compensation-eligible delays. In summary, this research contributes to modelling methods for strengthening the resilience of the rail network by enhancing its ability to recover from disruption efficiently.

1. Introduction

In the 2023-2024 financial year, 1.6 billion train services were scheduled, highlighting the strong reliance of the UK on rail transportation for mobility (Office of Rail and Road, 2024). Rail transport also plays a key role in sustainability efforts, offering a greener alternative to other modes of transportation. Since building new infrastructure is time-consuming and costly, the focus has shifted to optimising reliability and network capacity through efficient management to accommodate growing demand. However, the high demand has stretched the network capacity to its limit, reducing the flexibility of the timetable and making it harder to adapt to disruptions. The disruptions not only impact local operations but also swiftly extend to neighbouring stations by escalating delays to subsequent train services, ultimately affecting the entire regional network. Therefore, prompt responses to disruptions are critical to mitigate their effects before propagate extensively.

Enhancing operational efficiency through automated decision-making processes helps network operators manage disruption more effectively. To tackle this, Tan et al. (2025) introduced a real-time mathematical simulation model to analyse interdependencies among the train services in a regional rail network and predict the outcomes of mitigation plans after implementation. This model can simulate the propagation of delays across the network and estimate delays experienced by each affected train service with high detail. However, while it provides real-time estimates, operators must manually determine the best mitigation plans based on these estimates. The decision-making process can be time-consuming, as operators must consider multiple factors, such as minimising additional costs and service delays, while deciding on necessary service alterations.

Given the complexity and high interdependence of the rail network, these decisions are not always optimal. This underscores the need for a decision-making method that can support operators in disruption management and generate results in a reasonably short time. Several studies in the literature employ optimisation techniques for scheduling and disruption management, including the rolling horizon approach (Dekker et al., 2022), multi-commodity flow models (Cacchiani et al., 2014), mixed integer programming (Dollevoet et al., 2017) and genetic algorithms (Katoch, Chauhan and Kumar, 2021). However, most of these techniques have been

developed within academic research studies and have limited real-world applicability. This limitation arises from the inability of these models to simulate disruption scenarios that accurately reflect real-world conditions, as well as the challenges in modelling vast regional networks. Also, such models often overlook the knock-on effects of disruptions and the network-wide impact of service alterations. As a result, the decisions generated by these models tend to lack practicality, as the input data does not fully represent real-world situations.

To address these research gaps, this paper proposes a hybrid simulation-optimisation model that employs genetic algorithms to determine near-optimal solutions within a reasonable timeframe. This optimisation model builds upon the results of the validated simulation model proposed by Tan et al. (2025), ensuring the accuracy and practicality of the model. This research aims to identify the optimal service alterations for targeted train services, expediting service recovery while minimising the network-wide impact of disruptions.

The rest of the paper is organised as follows. Section 2 describes the methodology for developing simulation and optimisation models, including their integration. Section 3 presents a discussion of the results. Finally, Section 4 provides conclusions and possible future work.

2. Methodology

Figure 1 illustrates the sequential framework of this model, with each step detailed in the following sections.

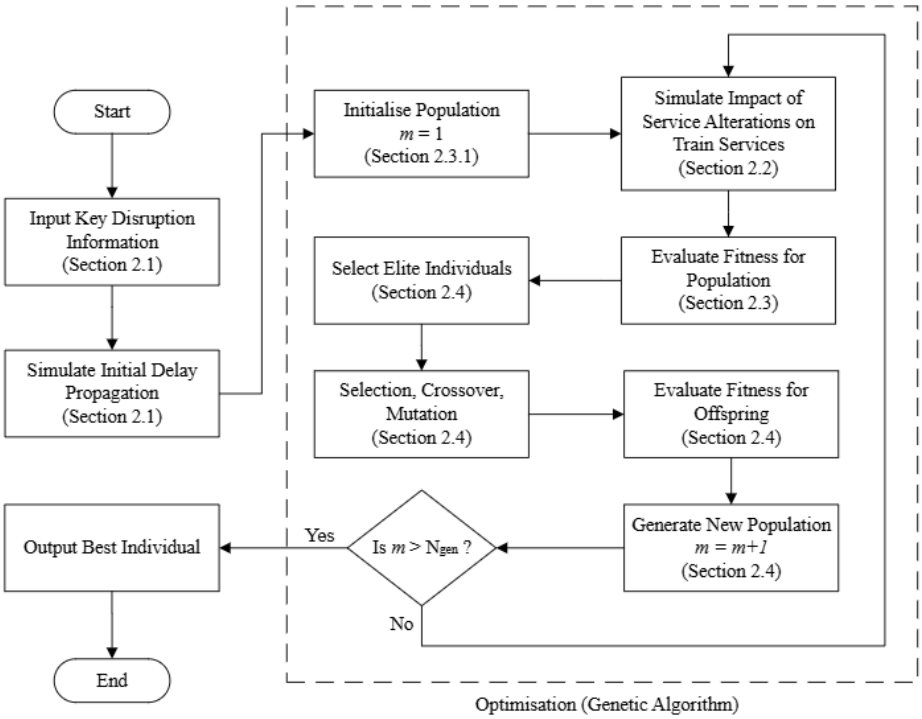


Figure 1. Sequential framework of the proposed hybrid simulation-optimisation model.

2.1 *Simulation of delay propagation*

The simulation model developed by Tan et al. (2025) mimics delay propagation in rail networks and evaluates network performance during disruptions. It incorporates key disruption parameters, including location, duration, start time, and blockage type (partial or full). The model replicates operations in the disrupted sections and calculates the delay minutes incurred by train services across the regional network. In this research, a performance evaluator will be developed from the simulation model, with its output feeding into the optimisation process. An initial simulation of delay propagation is conducted to establish baseline delay minutes under the disrupted scenario without any service alterations.

2.2 *Simulation of the effect of service alterations on subsequent train services*

Service alterations during disruptions often involve train cancellations, speed restrictions, and so on. Train cancellation is typically the least favourable option for network operators, as it directly increases the number

of affected train services and creates reactionary impacts on connecting services. However, cancelling a train can free up capacity and absorb some delay minutes that would otherwise affect other train services. Therefore, strategically delaying and cancelling selected train services can enhance the overall network performance.

The simulation model developed by Tan et al. (2025) is capable of computing the effects of train cancellations by removing the relevant train services from the timetable and re-running the simulation. However, this method is inefficient for real-time applications, particularly when navigating a large solution space, as each possible combination of train cancellations requires a full rerun to obtain results. To address this limitation, an algorithm is developed in this paper to evaluate the impact of train cancellations on subsequent services more efficiently. In this context, a trip is defined as the movement of a train between two stations, while a journey comprises multiple trips, representing the movement of a train from its departure station to its termination station. When a trip is cancelled, the algorithm simulates the downstream effects to understand how this action benefits other train services in the recovery of network operations.

Once the amount of delay absorbed by the cancellation is determined, the corresponding delay minutes are subtracted from the total delay experienced by the affected downstream trips. This approach enhances computational efficiency by limiting the simulation to only the cancelled train services, avoiding the need to rerun the entire simulation for each potential cancellation scenario.

2.3 *Optimisation in the decision-making process for disruption management*

An optimisation model is developed and integrated with the simulation model to automate decision-making in disruption management. Given the high granularity and complexity of the simulation outputs, a genetic algorithm is employed to explore the large solution space efficiently. The proposed model aims to identify the optimal combination of train cancellations that minimises total delay minutes, additional incurred costs and the number of cancellations. A binary list representing the status of each trip, where each element corresponds to a trip scheduled within the status of each trip at each station, is defined as an individual in the genetic algorithm. An initial population comprising a predefined number of individuals (n) is generated to serve as the baseline input for the optimisation model. This population is then passed to the algorithm detailed in Section 2.2 to simulate the effects of various service alterations on the recovery of the disrupted rail network.

2.3.1 *Decision Variables*

To generalise potential service alterations, each train trip is simplified into binary decisions: to continue or to cancel its operation. Initially, delay propagation is simulated without any mitigation measures. Subsequently, service alterations are applied to assess their impact on network performance. Each train is therefore assigned a decision variable that indicates its operational status at each station, either Active (1) or Cancelled (0).

2.3.2 *Objective Function*

The objective function is formulated based on performance metrics commonly adopted in the rail industry to assess network performance. This study aims to minimise a combination of three key parameters: **delay minutes (x)**, **additional costs incurred due to disruption management** (y), and the **number of train cancellations *(z)***, as presented in Equation (1).

$$\min f(x, y, z) = \omega_1 \alpha x + \omega_2 \beta y + \omega_3 \gamma z \tag{1}$$

Delay minutes (x) is a key parameter for quantifying the performance of the rail network. It reflects the deviations from the scheduled train services and serves as an indicator of network punctuality and reliability. Also, this parameter plays a crucial role in determining the compensation owed to affected passengers, as compensation is computed based on the extent to which passengers are delayed in reaching their destination.

Although various costs arise during the disruption management, such as arranging alternatives for passengers and rescheduling rolling stock and crew, the compensation to passengers forms a significant portion. These payments are calculated under the Delay Repay Compensation Scheme (Delay Repay claim, 2025). Minimising **additional costs (y)** is crucial for controlling overall expenditure and mitigating fare increases.

Historical annual ridership data is used to construct an origin-destination (OD) matrix for each route in the regional network, providing a numerical representation of passenger flow across the network. This allows the model to estimate the number of passengers travelling between each departure and destination station along a route. Using the simulation model, the delay experienced by each train service at various stations can be tracked. By identifying how much delay a train accumulates when it approaches a passenger's destination station, the delay experienced by that passenger can be estimated. With the OD matrix and simulated delay information, the total number of affected passengers and the extent of their delays can be determined. This enables the estimation of **total additional costs (y)**, which are linked to the duration of passenger delays.

Train cancellations are often employed as an effective strategy to reduce the wider operational impact of disruptions. However, a higher rate of cancellations negatively affects passenger satisfaction and reduces confidence in the reliability of rail transport. Conversely, cancelling an inadequate number of train services may lead to severe delay propagation across the network due to capacity constraints, even though the number of train cancellations appears minimal. To address this trade-off, the **number of train cancellations (z)** is included in the objective function, aiming to identify the optimal combination of service alterations that maximises overall network performance. Train cancellations are quantified by calculating the proportion of cancelled trips within a journey, ranging from 0 to 0.9 for partial cancellations, and 1.0 for full cancellations.

However, the input parameters have different units and operate with varying ranges of values. Regression-based sensitivity analysis is performed to quantify the influence of each input variable on the output of the objective function (Razavi et al., 2021). Once the sensitivity indices of the input variables are determined, scaling coefficients (α, β, γ) are introduced to equalise the contribution of each parameter. This approach determines appropriate scaling coefficients to ensure a balanced contribution from all input parameters in the objective function, thereby making them comparable. The weight coefficients (ω_i) can be adjusted to prioritise specific aspects of the objective function based on the operator's preferences.

2.4 *Genetic Algorithm*

The optimisation process in this model utilises a genetic algorithm (GA) to find a quasi-optimal solution within a reasonable duration. The GA parameters are outlined in Table 1. The GA process involves three main operations: selection, crossover and mutation. Firstly, tournament selection is performed to identify individuals based on their fitness values by selecting the best candidate from a randomly chosen subset of the population. Next, the two-point crossover is applied to generate offspring by combining features from chosen parents, allowing segments of genetic material to be exchanged. Then, the rule-based mutation is carried out to introduce genetic diversity by randomly altering certain genes while enforcing problem-specific constraints. The fitness of the offspring is evaluated, and the best individual is retained and passed to the next generation through elitism, ensuring that high-quality solutions are preserved.

Table 1. Parameters of the genetic algorithm.

Parameter	Value
Number of Generation, N_{gen}	100
Population Size, N_{pop}	50
Elitism Rate	1
Crossover Probability	0.7
Mutation Probability	0.2
Scaling Coefficients (α, β, γ)	1, 7, 2
Weight Coefficients $(\omega_1, \omega_2, \omega_3)$	1, 1, 1

The iterative process continues until the stopping criteria are met, with the best individual in each generation evolving over time. The stopping criterion is satisfied when the generation number (m) reaches a predefined number of generations (N_{gen}). Once this condition is met, a quasi-optimal solution is generated. This solution comprises a list of actions indicating the necessary decisions regarding the operational status of each train service at every station within the disruption and recovery timeframes.

2.5 *Constraints and Assumptions*

Due to the complexity of network operations and the practical considerations in disruption management, the following constraints and assumptions are incorporated into this model:

i. If a trip within a journey is cancelled, all subsequent trips in that journey are also cancelled. This is based on the operational norm that once a train is withdrawn from the track, it is moved to a depot or siding and does not continue its scheduled journey. This constraint is enforced through rule-based mutation, ensuring that once a trip is cancelled, all remaining trips in that journey are also cancelled.

ii. It is assumed that infrastructure, such as depots, sidings, and switches, is available to facilitate service alterations, enabling trains to be practically removed from the track whenever necessary.

iii. Rearrangement of crew and rolling stock is assumed to be feasible, enabling the necessary adjustments to support connecting services when a train is cancelled.

3. *Results and Discussion*

A case study is conducted on a historical disruption, *Incident 428433*, which occurred at 12:45 p.m. on 25th April 2024. This incident occurred at Watford Junction on the West Coast South Route and was caused by trespass. This disruption led to a partial blockage on the line that lasted for 35 minutes. According to empirical data, 37 passenger train services were affected by this disruption, excluding freight services. In response, contingency plans and train cancellations were implemented to recover operations. This incident is simulated in this paper, and the decisions on service alterations are optimised using the proposed model. The results are then compared against the actual empirical data to quantify the improvements achieved by applying this model.

As shown in Figure 2, the proposed model significantly reduces overall service cancellations, achieving a reduction of 43.06% compared to the empirical data. In contrast, the reduction in total delay minutes incurred by train services is relatively modest, with only a 1.55% improvement. The limited improvement in delay minutes can be attributed to the substantially reduced number of cancellations. With fewer train removals, congestion in the network persists, thereby limiting the potential for further reductions in delay minutes.

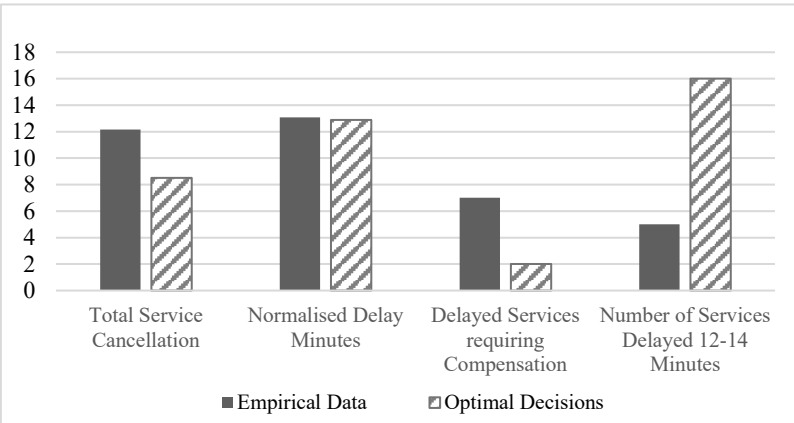


Figure 2. Comparison of empirical data and model-generated optimal results.

From the cancellation decisions, it can be observed that the train services chosen for cancellation exhibit the following characteristics:

i. Services that are directly affected by primary delays.

These trips are directly disrupted by blockage due to disruptions. Cancelling these services is highly effective in preventing the spread of reactionary delays. However, such removals are not always feasible, especially if the train is already en route and approaching the disruption zone. Through simulation, the model can predict which trains will likely experience primary delays and proactively cancel them before they enter the affected area. Ideally, such cancellations should be implemented at the early stage of the disruption to contain the propagation of delay.

ii. Services that are scheduled soon after the disruption ends.

Cancelling services during the initial recovery phase helps absorb the impact of the disruption and reduce the cascade of reactionary delays. The model effectively identifies and cancels services whose removal brings the most systemic benefit to the network.

The number of delayed services requiring compensation is reduced by 250% compared to the empirical case, indicating that the optimised cancellation strategy of the model effectively mitigates the financial impact on the network operator. This is achieved by strategically managing how delays are distributed across train services. The number of services delayed between 12 and 14 minutes in the optimal result is 128% higher than in the empirical case. Rather than allowing numerous services to incur compensable delays, the model redistributes minor delays evenly throughout the network. This approach reduces the number of services delayed beyond the compensation threshold, 15 minutes, thereby minimising overall compensation liabilities.

The model was executed on an i9-13900K processor, completing both the simulation and optimisation processes in approximately 5 minutes. It demonstrated the capability to accurately simulate the disruption and optimise the mitigation strategies within a practical computation time.

4. *Conclusion and Future Works*

This research introduces a hybrid simulation-optimisation framework designed to move towards an automated decision-making process in railway disruption management. The model integrates a mathematical simulation with a genetic algorithm-based optimisation model to evaluate and identify optimal service alteration strategies. Comparative analysis with empirical data demonstrates the effectiveness of the model in reducing delay minutes, minimising service cancellations, and lowering the number of compensation-eligible delays in a reasonably short time. These findings underscore the practical applicability and computational efficiency of the model in supporting disruption management within a highly interconnected rail network.

Future research should focus on incorporating uncertainty analysis to evaluate the robustness of the model under varying disruption scenarios and operational conditions. Additionally, the integration of machine learning techniques, such as neural networks, can be explored to enhance the optimisation speed.

Acknowledgements

The project is supported by Network Rail and the University of Nottingham. The authors gratefully acknowledge the support of these organisations. Appreciation is extended to Mr Adam Cooper Watson and Mr Andrew Dutton for their active involvement and invaluable industrial advice.

References

Cacchiani, V. et al. (2014). An overview of recovery models and algorithms for real-time railway rescheduling. *Transportation Research Part B: Methodological*, 63, 15–37. Available from https://doi.org/10.1016/j.trb.2014.01.009.

Dekker, M.M. et al. (2022). A next step in disruption management: combining operations research and complexity science. *Public Transport*, 14 (1), 5–26. Available from https://doi.org/10.1007/s12469-021-00261-5.

Delay Repay claim. (2025). Available from https://www.eastmidlandsrailway.co.uk/help-manage/manage/make-a-delay-repay-claim [Accessed 20 April 2025].

Dollevoet, T. et al. (2017). Application of an iterative framework for real-time railway rescheduling. *Computers and Operations Research*, 78, 203–217. Available from https://doi.org/10.1016/j.cor.2016.08.011.

Katoch, S., Chauhan, S.S. and Kumar, V. (2021). A review on genetic algorithm: past, present, and future. *Multimedia Tools and Applications*, 80 (5), 8091–8126. Available from https://doi.org/10.1007/s11042-020-10139-6.

Office of Rail and Road. (2024). *Passenger rail usage - October to December 2023*. Available from https://dataportal.orr.gov.uk/statistics/performance/passenger-rail-performance/ [Accessed 29 May 2024].

Razavi, S. et al. (2021). The Future of Sensitivity Analysis: An essential discipline for systems modeling and policy support. *Environmental Modelling & Software*, 137, 104954. Available from https://doi.org/10.1016/J.ENVSOFT.2020.104954.

Tan, K.R. et al. (2025). Practical predictive model for evaluating performance and delay propagation in multi-track railways during disruptions. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*. Available from https://doi.org/10.1177/09544097251344652.

From knowledge graphs to probabilistic models for system-level diagnostics

T. Nägele*, L. Barbini, G. van den Braak, M. Lipplaa and A. Piedrafita

**High Tech Campus 25, 5656 AE Eindhoven, The Netherlands*
{thomas.nagele, leonardo.barbini, gert-jan.vandenbraak, micha.lipplaa, alvaro.piedrafitapostigo}@tno.nl

Abstract. The increasing complexity of high-tech systems poses a significant challenge on service organizations tasked with the timely identification of the root causes of unexpected downtimes. While data-driven methods are effective for diagnosing frequently occurring issues, or those affecting a large number of systems, rare issues suffer from data scarcity, necessitating alternative approaches. This paper presents a two-step model-based methodology that leverages system architecture information to support diagnosing of systems for which little data is available. Firstly, system design and observability information, including diagnostic tests, is captured in a knowledge graph. Secondly, the knowledge graph is queried and transformed into a probabilistic graphical model. This is fully automated using transformation rules based on the ontology underlying the knowledge graph. The probabilistic graphical model then infers the most likely causes of failure using measurement data, and guides service engineers by suggesting cost-effective diagnostic actions. This paper outlines the proposed methodology, demonstrates its application on a small example system, and reports early-stage validation findings from high-tech system cases.

1. Introduction

Efficiently diagnosing high-tech systems is increasingly difficult. This is not only caused by the increased complexity of such systems, but also by their long lifetime and high variability due to customization (Acur & Hendriks, 2024), resulting in a variety of failure manifestations. Some failures will happen relatively frequently across the installed base, allowing the use of data-driven approaches to recognize and diagnose them quickly, provided that sufficient data is being logged. Other failures will occur more rarely, so these are hard to diagnose with data-driven methods, as such approaches require a vast amount of data for training. The diagnosis of rare failures requires one or more highly experienced service technicians and is often time consuming and costly. Growing numbers of systems in the field, combined with their increasing complexity and the scarcity of skilled personnel demand a technical solution: a digital diagnostic assistant that enables fewer technicians to handle more diagnoses effectively. This paper presents a methodology to assist in the diagnosis of rare hardware failures without depending on vast amounts of data.

Model-based diagnostics (MBD) brings automated diagnostic reasoning to the service technician, using a formal model to perform the reasoning (de Kleer & Williams, 1987). MBD formalisms to perform the reasoning include logic (Poole, 1988) and probabilistic models (Pearl, 1988). The structure and semantics of such models depend on the information they are based on. These models can be created either from expert experience (experience-based) or from the system's design information (design-based). Due to the intrinsic lack of expert experience and data for rare failures, our methodology leverages design information as source of information for these diagnostic models.

The architecture of our methodology is shown in Figure 1. On the left-hand side, system dependencies are manually extracted from design information. These system dependencies are transformed into a *diagnostic network*, which is a knowledge graph according to a fixed ontology. The diagnostic network is transformed into a reasoning model. In our methodology we use probabilistic graphical models as diagnostic reasoners for their ability to deal with arbitrary model structure and perform both deterministic and probabilistic reasoning. The reasoning model can be used in two ways. First, it can be used by an analyzer to assess the observability of failures by means of design-for-diagnostics algorithms (Barbini et al., 2021). Insights generated by the analyzer can be used to improve the system design, represented as the light grey edge back to the design information. Second, the reasoning model can be used by the diagnoser to guide the service engineer in the root cause analysis process, by iteratively suggesting the next best tests (van Gerwen et al., 2022), (van Gerwen et al., 2024). This diagnosis loop is represented on the right-hand side of Figure 1.

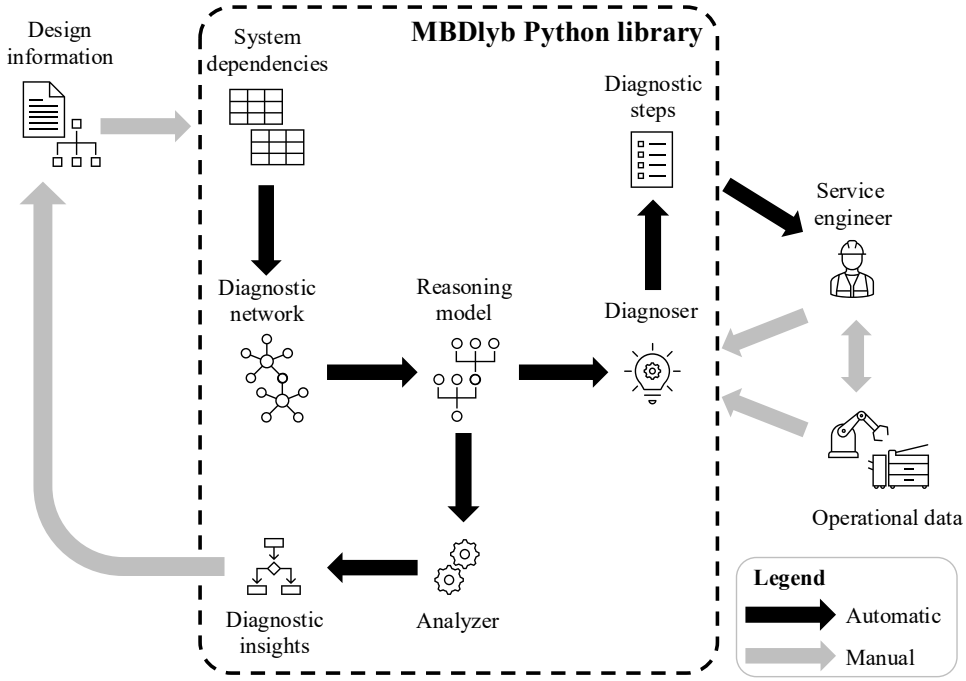


Figure 1. Architecture of our function-based diagnostic methodology, indicating transformations between different knowledge representations. Dark edges correspond to automated transformations, and light edges are (semi-)manual.

The remainder of this paper focuses on the center part of this architecture. For further details on the methodology, see (Nägele et al., 2025), and (Nägele, 2025) for the MBDlyb Python library. Section 2 introduces the ontology of the diagnostic network and describes the types of information sources used to instantiate it. Section 3 specifies the formalism of the reasoning model and explains how the diagnostic network can be transformed into this formalism. Section 4 shows how the model can be used to support the diagnostic process, and the paper concludes in Section 5.

2. Diagnostic network: a knowledge graph

The proposed methodology uses system architectural information for diagnostic purposes. This information is stored in the diagnostic network. The methodology follows a model-based systems engineering (MBSE) approach starting from the system functional decomposition, in which the high-level system functions are iteratively decomposed into subsystem functions, module functions, and so on until an elementary level of functionality is reached. Once the functional decomposition is specified, a structural decomposition is made and system functions are allocated to structural parts, i.e., hardware components. Inter-functional dependencies can be specified to indicate what system functions are required for fulfilling another system function. Figure 2 (right) illustrates how modeling functional dependencies using design information enables automatic derivation of hardware failure observability, eliminating the need for manual expert linking of observations to hardware failures (left).

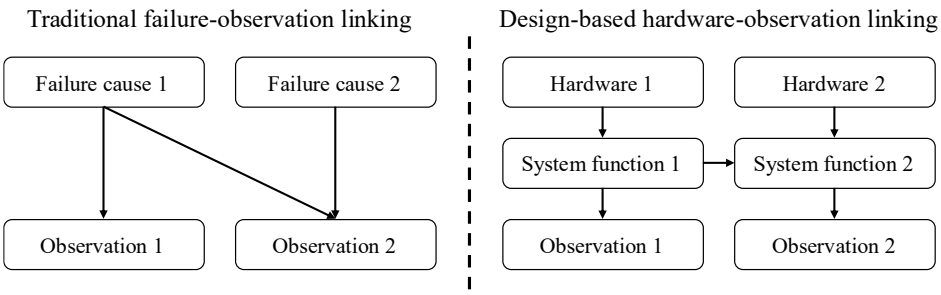


Figure 2. (Left) Traditional constructs in which observations are linked to possible failure causes based on experience. (Right) The proposed idea of a design-based approach: hardware failures are observable due to system functions failing.

As shown in Figure 2, including observations in the diagnostic network is crucial for diagnosing hardware failures. These observations represent system parameters, such as key performance indicators (KPIs) or events logged by software. For example, software may log an error if it is unable to fulfill the function to move a robotic arm to a specified position. Another type of observation is one resulting from an action triggered by the operator, such as taking a physical measurement from the system. We refer to the first type of observation as a *DirectObservable*, and to the latter as a *DiagnosticTest*.

The diagnostic network follows a hierarchical representation of the system according to the structural decomposition. This tree-like structure is organized into elements called *Clusters*. A *Cluster* may contain other clusters, hardware parts, system functions or observables. While the diagnostic network is structured hierarchically, it can be represented as a flat knowledge graph following a strict ontology to support systematic generation of the reasoning model. Table 1 shows the node types in our ontology and a brief description of what these types represent in the domain of high-tech systems. Table 2 shows the relation types in the ontology, as well as the types of nodes the relation connects and a brief description of what the relation represents.

Node	**Represents**
Cluster	Hierarchical layer in the structural decomposition, acts as container for other nodes
Hardware	Hardware component that may fail
Function	System function
DirectObservable	Automated observation, such as KPIs, parameters or event logs
DiagnosticTest	Action (test) that can be executed to acquire additional data from the system
DiagnosticTestResult	One of the findings resulting from conducting a diagnostic test

Table 1. Types of nodes in the diagnostic ontology and a short description of what these represent.

Relation	**Connects**	**Represents**
part_of	Any node to a Cluster	The containment of a specific node in the hierarchy
realizes	Hardware to Function	The functional allocation of the function on the hardware. Indicates that the hardware part is necessary for realizing the system function
subfunction_of	Function (X) to Function (Y)	A functional decomposition of function Y into function X
required_for	Function (X) to Function (Y)	A functional dependency of function Y onto X
observed_by	Function or Hardware to DirectObservable	A dependency relation for an automated observable onto the function or hardware component
results_in	DiagnosticTest to DiagnosticTestResult	A containment relation indicating that a reading for the connected result will be acquired by conducting the test
indicated_by	Function or Hardware to DiagnosticTestResult	A dependency relation for a test observation onto the function or hardware component

Table 2. Types of relations in the diagnostic ontology, including the relation scope and representation.

The ontology connects to the Arcadia method (Voirin, 2017), which has been implemented in Eclipse Capella (Roques, 2017). The logical and physical architectures in this MBSE method contain relevant information, such as the functional decomposition (*subfunction_of*), structural decomposition (*part_of*), functional deployment (*realizes*) and functional exchanges (*required_for*). Since an increasing number of high-tech companies are adopting such model-based approaches to formalize their system's design, Capella models have been used as basis for the diagnostic models in our validation.

3. Transformation to reasoning model

Unlike simulation tasks, the diagnostic task focuses on inferring the health state of unobserved variables – top layer in Figure 2 – from a set of uncertain observations – bottom layer in Figure 2. This formulation of diagnostic reasoning as an inference problem makes probabilistic graphical models (PGMs) a natural choice for its formalization. While many diagnostic approaches use Bayesian networks as their PGM of choice, our methodology uses Markov random field (MRF) (Koller & Friedman, 2009), as it can cope with cyclic relations.

In a MRF, every node represents a probabilistic variable with one or more discrete states. Edges between the nodes represent probabilistic dependencies between the variables. Figure 3 shows an example Markov random field, in which 'Function 1' depends on two hardware components, and 'Function 2' depends on one hardware component and on 'Function 1'. 'Function 1' is testable, and 'Function 2' can be observed automatically. In our approach not all types of nodes in the diagnostic network are transformed into nodes of a MRF. Only the *Hardware*, *Function*, *DirectObservable* and *DiagnosticTestResult* nodes are transformed. The remaining node types are used solely to structure the system knowledge. In the MRF all nodes are random variables with two states. *Function*, *DirectObservable* and *DiagnosticTestResult* nodes can be *Ok* or *NOk*. *Hardware* nodes can be *Healthy* or *Broken*.

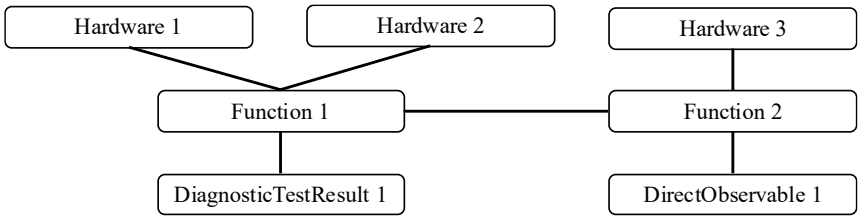


Figure 3. Visualization of an example Markov random field.

Conditional dependencies between nodes in the MRF are expressed as factors. The structure and parameters of each factor are determined by the type of the corresponding node in the diagnostic network and the types of its neighbors in the knowledge graph. Hardware nodes, which do not depend on other variables, are associated with unary factors representing prior probabilities: 0.99 for the *Healthy* state and 0.01 for the *Broken* state. All other factors in the MRF are modeled using noisy AND-gates (Pearl, 1988). To finetune the dependency — e.g., how strongly a function depends on another function—relations in the diagnostic network have weights. These weights are incorporated into the parameters of the corresponding factors in the MRF. Table 3 shows the factor for 'Function 2' as shown in Figure 3 for the default case of all weights set to 1.

Hardware 3	**Function 1**	**Function 2 = Ok**	**Function 2 = NOk**
Healthy	*Ok*	1.0	0.0
Healthy	*NOk*	0.0	1.0
Broken	*Ok*	0.0	1.0
Broken	*NOk*	0.0	1.0

Table 3. Factor for 'Function 2' as shown in Figure 3.

This factor encodes that each Hardware and Function on which 'Function 2' depends must be *Healthy* or *Ok* for 'Function 2' itself to be *Ok*; otherwise, it will be *NOk*. Table 4 shows the factor when different weights are assigned to these dependencies. The specific weights are indicated in the header of the table.

Hardware 3 (0.8)	**Function 1 (0.6)**	**Function 2 = Ok**	**Function 2 = NOk**
Healthy	*Ok*	1.0	0.0
Healthy	*NOk*	0.4	0.6
Broken	*Ok*	0.2	0.8
Broken	*NOk*	0.08	0.92

Table 4. Factor for 'Function 2' if its dependency on 'Hardware 3' would have a weight of 0.8 and its dependency on 'Function 1' a weight of 0.6.

For *DirectObservable* and *DiagnosticTestResult*, false positive and false negative rates can be specified on the corresponding nodes in the diagnostic network. This allows the reasoning model to account for uncertainty in observations.

The diagnoser uses the reasoning model to compute posterior probabilities for all hardware nodes in the model after inserting evidence gathered from the system, see (Nägele et al., 2025) for further details. Evidence is inserted into the MRF by adding unary factors to the *DirectObservable* and *DiagnosticTestResult* nodes,

reflecting their observed values. After every step, these posteriors represent the likelihood of each of the hardware components to have failed. The conditional entropy between each *DiagnosticTest* (as collection of *DiagnosticTestResults*) and suspected hardware nodes is computed and used to rank the utility of the remaining diagnostic tests. This ranking is then presented to the service technician. After a new test is executed, the resulting evidence is added to the model, and updated diagnoses are computed (van Gerwen et al., 2022). This is akin to dynamically generating a fault tree based on the available evidence. As such, it supersedes static fault tree analysis.

4. Application

This section applies the methodology to a simple system: a CD-radio player. For illustrative purposes, we assume that this system consists of four main modules: a power system, a CD reader, a radio receiver, and an audio system (i.e., the speaker). The transformation from the system architecture model to the diagnostic network is depicted in Figure 4. The figure shows only the highest level of the model. Each module has been further detailed out (not shown here), resulting in a final model with depth 5. After this automatic transformation, the diagnostic network does not yet contain any observability information. For this simple system, it was assumed that the lowest-level functions and hardware components could be tested, and *DiagnosticTests* were therefore added at this lowest level. At the higher level, two *DirectObservables* were added: one for the 'ReadCD' function and one for the 'ReceiveRadio' function. The diagnostic network can be automatically transformed into a reasoning model. This resulting reasoning model (not shown here) is topologically very similar to the diagnostic network of the system. Below we show the usage of a reasoning model of the CD-radio player system under an hypothetical scenario. Hypothetical findings in accordance with the scenario will be inserted in the model by following the diagnostic tests suggested by the diagnoser.

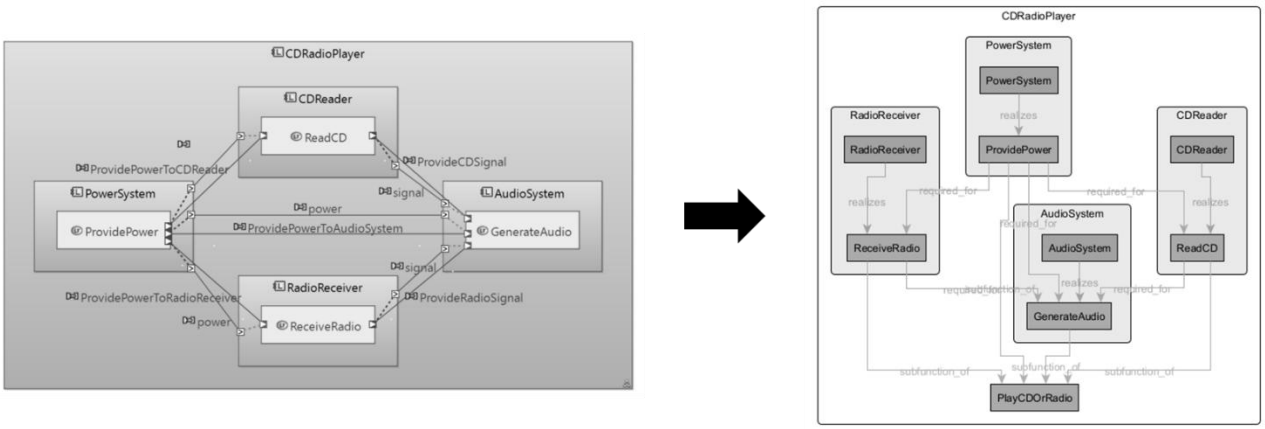


Figure 4. Transformation from the system architecture model in Capella (on the left) to the diagnostic network (on the right).

Let us assume the hypothetical failure to diagnose is a broken motor (in the CD reader). This broken motor causes the listener to be unable to hear music, while attempting to play the CD. In this scenario, it is assumed that the power system, which also supplies all other modules, is functioning correctly, and that there are no issues in the audio system itself. Evidence inserted in the model are according to these assumptions. Table 5 shows the diagnostic steps and their findings for this hypothetical situation. The model correctly identifies the CD reader's motor as being faulty.

Step	New finding	Diagnoses / likely suspects (likelihood)	Suggested diagnostic test
1	CD error showing	All components in the power system and the CD reader modules (~10%)	Check whether the CD reader's motor is spinning
2	CD reader's motor is not spinning	CD motor and all components in the power system (~15%)	Check temperature of the power supply
3	Power supply temperature is Ok	CD motor and all electrical components in the power system (~20%)	Check whether power LED is on
4	Power LED is on	CD motor (~100%)	*Reached a clear diagnosis.*

Table 5. Diagnostic steps to diagnose the hypothetical failure of the CD motor in the CD-radio player.

We applied the proposed methodology also in an industrial setting. Together with ASML we created a diagnostic network based on high-level system architecture information, covering 10 subsystems in various levels of detail. Three solved cases from customer support were selected to be diagnosed retrospectively with the model. This single model was able to reach a satisfactory diagnosis for all three cases, also according to domain experts from ASML, confirming the diagnostic capabilities of this MBD approach.

5. Conclusions and future work

In this paper we have presented a methodology for aiding service technicians in identifying the root cause of system failures. We have shown that diagnostic reasoning can be effectively executed based on available design information in the form of architectural descriptions of the system. The diagnostic models presented here are conceptually very similar to those built by system engineers following MBSE principles, and, as shown, can be partially derived from them. We have identified that a key challenge lies in integrating observability information (KPIs, errors, tests) with MBSE models, a functionality currently not sufficiently provided by MBSE approaches. Using knowledge graphs as a diagnostic knowledge base has proven to be effective. It introduces a clear separation of concerns, in which the knowledge graph is meant to formalize knowledge, while the reasoning model can be of any structure that is relevant for supporting diagnosis. The use of a knowledge graph is not limited to diagnostics: it may also be one central knowledge base, storing much more information than for diagnostic purposes. As long as it also contains the diagnostic-relevant information, the formal transformation to the reasoning model may be done. Together with industry, more large-scale validation is needed to increase confidence in this MBD approach and to identify edge-cases for which extensions or modifications are required. Future research may focus on the inclusion of software behavior into the model, as the current approach is tailored to identify hardware components.

Acknowledgement

The research is carried out as part of the SD2Act program under the responsibility of TNO-ESI in cooperation with ASML. The research activities are co-funded by TKI HTSM via the PPP Innovation Scheme (PPP-I) for public-private partnerships.

References

Acur, S., & Hendriks, T. (2024). *PMC SA / SE Vision and Outlook for Systems Architecting and Systems Engineering in the High-Tech Equipment Industry*.

Barbini, L., Bratosin, C., & Nägele, T. (2021). Embedding Diagnosability of Complex Industrial Systems Into the Design Process Using a Model-Based Methodology. *PHM Society European Conference*, *6*(1), 9.

de Kleer, J., & Williams, B. C. (1987). Diagnosing multiple faults. *Artificial Intelligence*, *32*(1). https://doi.org/10.1016/0004-3702(87)90063-4

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Nägele, T. (2025). *MBDlyb: Python library for model-based diagnostics*. https://esi.nl/research/output/tools/mbdlyb

Nägele, T., Barbini, L., van den Braak, G.-J., Piedrafita Postigo, A., & Lipplaa, M. (2025). *Guided diagnosis of functional failures in cyber- physical systems*. https://resolver.tno.nl/uuid:52f484ea-2638-4403-88a4-65305b245697

Pearl, J. (1988). Probabilistic Reasoning in Intelligent Systems. In *Probabilistic Reasoning in Intelligent Systems*. https://doi.org/10.1016/c2009-0-27609-4

Poole, D. (1988). Representing Knowledge for Logic-Based Diagnosis. In *Proceedings of the International Conference on Fifth Generation Computer Systems 1988*.

Roques, P. (2017). Systems Architecture Modeling with the Arcadia Method: A Practical Guide to Capella. In *Systems Architecture Modeling with the Arcadia Method: A Practical Guide to Capella*. https://doi.org/10.1016/C2016-0-00854-9

van Gerwen, E., Barbini, L., Borth, M., & Passmann, R. (2024). Efficient Differential Diagnosis using Cost-aware Active Testing. *International Journal of Prognostics and Health Management*, *15*(3), 1–11. https://doi.org/10.36001/ijphm.2024.v15i3.3849

van Gerwen, E., Barbini, L., & Nägele, T. (2022). Integrating System Failure Diagnostics Into Model-based System Engineering. *Insight*, *25*(4), 51–57. https://doi.org/10.1002/inst.12412

Voirin, J. L. (2017). Model-based system and architecture engineering with the Arcadia method. In *Model-based System and Architecture Engineering with the Arcadia Method*. https://doi.org/10.1016/c2016-0-00862-8

Optimal manufacturing-remanufacturing- transport planning in low carbon supply chain: Incorporating a carbon tax strategy

Basma Abassi*, Sadok Turki, Sofiene Dellagi****

** Laboratoire de Génie Informatique, de Production et de Maintenance*
Université de Loraine France (basma.abassi@univ-lorraine.fr)
*** Laboratoire de Génie Informatique, de Production et de Maintenance*
Université de Loraine France
{sadok.turki, sofiene.dellagi }@univ-lorraine.fr)

Abstract: This study develops a Mixed-Integer Linear Programming (MILP) model to optimize a manufacturing-remanufacturing-transport supply chain (MRTSC). The model integrates raw material procurement, transportation, inventory limits, and vehicle constraints which are factors often overlooked in previous research. It distinguishes between new and remanufactured products, leading to tailored manufacturing, transportation, and warehousing decisions. The model optimizes raw material procurement, production, remanufacturing, transportation, and returned product management under a carbon tax strategy, using carbon footprint as a sustainability metric. Implemented in CPLEX, the profit-maximizing model is validated through numerical results, demonstrating its effectiveness for MRSC (Manufacturing-Remanufacturing Supply Chain) optimization.

Keywords: MILP, manufacturing, remanufacturing, supply chain, transport, carbon tax, profit.

1. Introduction

Sustainability is becoming more and more a priority for companies these days. This is mainly because of growing environmental challenges and the constant need to stay competitive in the market. One big step in this direction is shifting towards low-carbon supply chains. In this context, reverse logistics is getting serious attention from both researchers and practitioners. This is mainly due to environmental concerns, global competition, rising customer awareness, increased product returns, and changing regulations (Okorie et al., 2021). All this highlights how crucial reverse logistics is for greener supply chains.

A lot of studies already focused on remanufacturing, like the ones by (Liu et al., 2022) and (Feng and Li, 2024). But when it comes to transportation, it's surprising that it's still not fully explored in MRSC studies, even though transportation has a major impact on sustainability. Reducing transport-related emissions means making decisions on modes, vehicle types, unit sizes, and routing (Yin et al., 2024). (Turki and Rezg, 2019) actually showed how transportation issues could affect profits in closed-loop supply chains. Their study highlighted that ignoring transportation risks could reduce efficiency and profitability. (Ko et al., 2017) also worked on this by proposing an inventory model that includes transportation risk and cost, aiming to minimize total costs while maintaining efficiency.

Apart from transportation, procurement, inventory management, and warehousing are also big contributors to carbon emissions. These involve energy-consuming tasks like product movement, drop-offs, collections, and storage conditioning. Surprisingly, this area hasn't been properly addressed in the literature yet, making it a real research gap. (Karim and Nakade, 2022) reviewed sustainable EPQ models, focusing on carbon emissions and recycling. Their study pointed out the need to integrate these factors into production planning and identified important research gaps.

For companies aiming to reduce their environmental impact, managing carbon footprints across their supply chains has become essential. (Shabir et al., 2023) defined carbon footprint as the total greenhouse gas emissions, including CO_2, produced over the product's life cycle, from manufacturing to distribution. Managing MRSCs effectively now means finding ways to reduce these emissions without hurting profitability.

In this direction, (Reddy et al., 2019) developed an MILP model for reverse logistics, considering both vehicle type selection and carbon footprint impacts, showing how it's possible to balance sustainability and cost.

At the same time, policymakers are pushing stricter emission regulations, forcing industries to invest more in carbon reduction efforts. Luckily, these investments often pay off quickly and bring long-term economic and environmental benefits, even for sectors like transportation and remanufacturing where the immediate carbon savings seem small (Luo et al., 2022). Among these measures, carbon taxes remain one of the most applied and studied, encouraging cleaner production (Konstantaras et al., 2021).

Recently, research on MRSCs under emission constraints has really gained momentum. A lot of studies are now focusing on joint optimization problems, where several conflicting objectives need to be balanced carefully. More models are including carbon emissions as a key factor in industrial decision-making. (Ndhaief et al., 2019) worked on integrating environmental aspects into production and maintenance planning for unreliable systems, showing how to balance efficiency and environmental impact. Similarly, (Lahmar et al., 2024) developed joint optimization models that also integrate carbon emissions into manufacturing-remanufacturing planning.

This study fills important gaps in MRSC research by proposing a MILP model that optimizes procurement, production, remanufacturing, transportation, and returns to maximize profit. Unlike earlier models, it distinguishes between new and remanufactured products by considering price and quality differences. It also accounts for limited inventory with upstream and downstream buffers. Two transportation strategies, including vehicle return management, are integrated to improve efficiency. Finally, a carbon tax policy is included to support sustainability and make the model more realistic for real-world MRSC scenarios.
To end this section, here's how the document is structured: Section 2 describes the system under study and the problem setup. Section 3 details the mathematical model and assumptions. Section 4 presents a numerical example, and Section 5 discusses the results and suggests future research directions.

2. Problem description

This section describes the proposed system shown in Figure 1. It's built around two parallel machines: M1 for manufacturing new products and M2 for remanufacturing end-of-life items. M1 uses raw materials to supply the primary market, while M2 processes returns for the secondary market. The system combines both forward and reverse production flows.
Sales inventories (V_n and V_r) are set up to meet constant demands for new and remanufactured products ($d_n(t)$ and $d_r(t)$) at any time t. Inventory capacities are limited: stock S holds raw materials for M1, and stock G stores returned products for M2. Outgoing products ($T_n(t)$ and $T_r(t)$) are shipped from manufacturing stock W and remanufacturing stock Z to the city. Sales are recorded as sn(t) for new products and sr(t) for remanufactured ones. Returned products Rq(t) are collected and stored in G for future remanufacturing.
Production planning happens in equal time intervals across the planning horizon. The goal is to maximize total profit, considering revenues from both product types and deducting all related costs: production, storage, shortages, transportation, procurement, return handling, and carbon emissions.

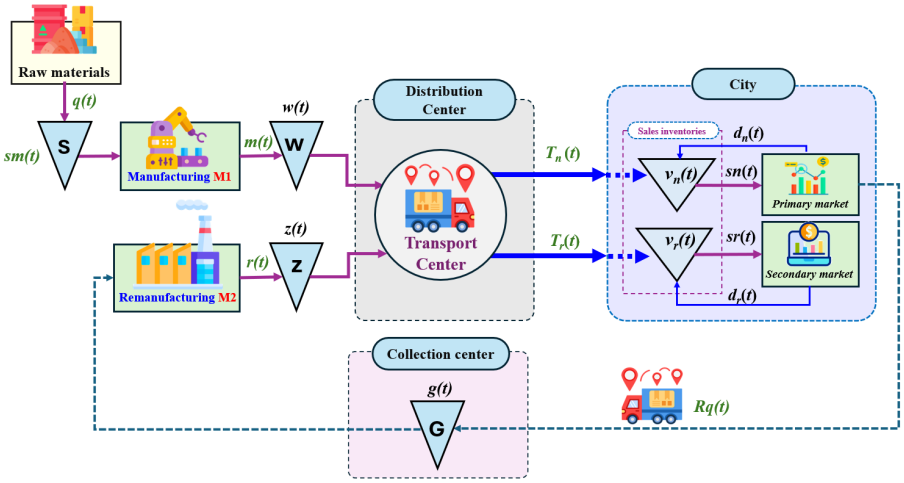


Figure 1. Transportation-Manufacturing-Remanufacturing system.

3. Analytical study

3.1 Notations

T. Δt	the duration of the finite planning horizon
$PR_{CXi}(T)$	total profit under carbon tax strategy and strategy i
q(t)	quantity of raw materials ordered in period t
m(t)	quantity produced by machine M1 in period t
r(t)	quantity produced by machine M2 in period t
$T_n(t)$	quantity of new products transported in period t
$T_r(t)$	quantity of remanufactured products transported in period t
Rq(t)	quantity of used products returned from the primary market and transported in period t
Sm(t)	inventory of raw materials remaining at the end of period t.
W(t)	inventory of new products in manufacturing storage at the end of period t
Z(t)	inventory of remanufactured products in remanufacturing storage at the end of period t.
$V_n(t)$	inventory of new products available for sale at the end of period t
$V_r(t)$	inventory of remanufactured products available for sale at the end of period t
G(t)	inventory of returned (collected) used products at the end of period t
$d_n(t)$	customer demand for new products in period t
$d_r(t)$	customer demand for remanufactured products in period t
lcn (t)	quantity of unmet demand for new products.
lcr (t)	quantity of unmet demand for remanufactured products.
rm	cost to acquire one unit of raw material.
cpn	cost to produce one unit of a new product (cpn > cpr)
cpr	cost to produce one unit of a remanufactured product.
crm	cost of holding one unit of raw material in inventory.
chn	cost of holding one unit of a new product in manufacturing inventory (chn = chr)
chr	remanufacturing inventory holding cost of one remanufactured product unit
ctrv	transport cost of one new or remanufactured product
cvn	inventory holding cost of one new product in the selling inventory Vn
cvr	cost of holding one remanufactured product unit in the sales inventory Vr
ccq	cost of holding one unit of a collected product in inventory
ctn	penalty cost for unmet demand of one unit of new product (ctn > ctr)
ctr	penalty cost for unmet demand of one remanufactured product unit
crp	acquisition or processing cost of one returned (used) product
RTC	fixed round trip cost
TCX (T)	total carbon tax cost over the horizon
sn(t)	sales quantity of new products during period t
sr(t)	sales quantity of remanufactured products during period t
sn	selling price per unit of new product
sr	selling price per unit of remanufactured product
ct	unit carbon tax price
CF	carbon footprint over the horizon
S_m^{max}	maximum capacity of the stock S
W^{max}	maximum capacity of the stock W
Z^{max}	maximum capacity of the stock Z
V_n^{max}	maximum capacity of the sales inventory Vn
V_r^{max}	maximum capacity of the sales inventory Vr
Ctv	maximum load capacity of the transport vehicle per trip
$\alpha_i(t)$	round trip variable in the strategy i
Rp	Percentage of returns of new products to be reconditioned
m_{max}	maximum production capacity of machine M1in period t
r_{max}	maximum production capacity of machine M2 in period t
CEm	unit carbon emissions from acquisition of raw materials
CEn	unit carbon emissions from manufacturing new product
CEr	unit carbon emissions from remanufacturing product
CEs	unit carbon emissions from holding raw materials (stock S)
CEw	unit carbon emissions from holding new product (stock W)
CEz	unit carbon emissions from holding remanufactured product (stock Z)
CEg	unit carbon emissions from collected product holding (stock G)
CEvn	unit carbon emissions from holding new product in the sales inventory Vn

CEvr	unit carbon emissions from holding remanufactured product in the sales inventory Vr	CE_t	unit carbon emissions from transporting products
		Cfix	fixed quantity of carbon emitted if the vehicle is empty round trip.

This study focuses on maximizing total profit $PR_{CXi}(T)$ by optimizing key decision variables: Q for raw material orders, M for manufacturing on M1, R for remanufacturing on M2, T_n and T_r for transporting new and remanufactured products, and Rq for collecting returns from the primary market. By finding the optimal set (Q*, M*, R*, T_n*, T_r*, Rq*), the study determines the maximum achievable profit PRCXi(T)*.

3.2 Formulation of the total expected profit with carbon tax strategy

This study proposes a complete framework for managing production, inventories, transportation, costs, and environmental impacts using the following set of equations. A round-trip decision variable, $\alpha_i(t)$, is introduced to optimize transport choices. It's a binary variable: 1 if a vehicle makes a round trip, 0 otherwise. Two trip management strategies (i = 1 and i = 2) are considered: in $\alpha_1(t)$, all transports ($T_n(t)$, $T_r(t)$, and Rq(t)) equal zero, and in $\alpha_2(t)$, only Rq(t) is zero. Together with constraints on inventory, production capacity, and transport limits, this help reduce unnecessary trips and control costs. Inventory balances for manufacturing (W) and remanufacturing (Z) are given in Equations 1 and 2, returns (G) in Equation 3, raw material ordering with lead time in Equation 4, and sales inventories (V_n and V_r) in Equations 5 and 6.

$W(t+1) = W(t) + m(t+1) - (T_n(t+1) \times \alpha_i(t))$ With: $W(1) = W_0 + m(1) - (T_n(1) \times \alpha_i(t))$ $W(t) \leq W^{max} \quad \forall\, t = 0,\ldots, T-1$	(1)
$Z(t+1) = Z(t) + r(t+1) - (T_r(t+1) \times \alpha_i(t))$ With: $Z(1) = Z_0 + r(1) - (T_r(1) \times \alpha_i(t))$ $Z(t) \leq Z^{max} \quad \forall\, t = 0,\ldots, T-1$	(2)
$G(t+1) = G(t) + R_q(t+1) - r(t+1)$ $\forall\, t = 0,\ldots, T-1$ With: $G(t) = G_0 + R_q(1) - r(1)$	(3)
$S_m(t+1) = S_m(t) + q(t) - m(t+1)$ $\forall\, t = 0,\ldots, T-1$ With: $S_m(1) = S_{m0} + q(0) - m(1)$ $S_m(t) \leq S_m^{max}$	(4)
$V_n(t+1) = V_n(t) + (T_n(t+1) \times \alpha_i(t)) - sn(t+1)$ $\forall\, t = 0,\ldots, T-1$ $V_n(t) \leq V_n^{max}$	(5)
$V_r(t+1) = V_r(t) + (T_r(t+1) \times \alpha_i(t)) - sr(t+1)$ $V_r(t) \leq V_r^{max} \quad \forall\, t = 0,\ldots, T-1$	(6)

Production limits for new and remanufactured products, along with vehicle capacity constraints (Ctv), further refine the model, as shown below.

$$0 \leq m(t) \leq m_{max} \;\; \forall\, t = 1,\ldots T$$
$$0 \leq r(t) \leq r_{max} \;\; \forall\, t = 1,\ldots T$$
$$Tr(t) + Tn(t) <= Ctv$$
$$Tn(t) <= W(t)$$
$$Tr(t) <= Z(t)$$
$$Rq(t) <= Ctv$$

Sales quantities for new and remanufactured products are defined in Equations 7 and 8 below.

$sn(t) = \min(d_n(t), V_n(t-1))$	(7)
$sr(t) = \min(d_r(t), V_r(t-1))$	(8)

The expressions of costs for production, shortages (with $lcn(t) = d_n(t) - sn(t)$, and $lcr(t) = d_r(t) - sr(t)$), inventory holding, raw material acquisition, returned products, transport, raw material acquistion and returned

products are shown successively as following: $c_{pn} \times \sum_{t=1}^{T} m(t) + c_{pr} \times \sum_{t=1}^{T} r(t)$; $c_{tn} \times \sum_{t=1}^{T} lcn(t)$; $ctr \times \sum_{t=1}^{T} lcr(t)$; $c_{hn} \times \sum_{t=1}^{T} W(t)$
; $c_{hr} \times \sum_{t=1}^{T} Z(t)$; $c_{rm} \times \sum_{t=1}^{T} S_m(t)$; $c_{vn} \times \sum_{t=1}^{T} V_n(t)$
; $c_{vr} \times \sum_{t=1}^{T} V_r(t)$; $c_{cq} \times \sum_{t=1}^{T} G(t)$; $c_{trv} \times \sum_{t=1}^{T} ((T_n(t) + T_r(t) + R_q(t)) \times \alpha_i(t))$; $r_m \times \sum_{t=1}^{T} q(t)$; $c_{rp} \times \sum_{t=1}^{T} Rq(t)$

The returned product quantity is modeled in Equation 9, and revenue from both product types is given in Equation 10 below.

$Rq(t) = sn(t-1) \times R_p$	(9)
$s_n \times \sum_{t=1}^{T} sn(t) + s_r \times \sum_{t=1}^{T} sr(t)$	(10)

Carbon emissions from production, storage, and transport which including fixed emissions from round trips, are calculated in Equation 11, while total carbon costs under the tax strategy are in Equation 12.

$CF_i(T) = \sum_{t=0}^{T} \Big((CE_m \times q(t)) + (CE_n \times m(t)) + (CE_r \times r(t)) + (CE_m \times S_m(t)) + (CE_w \times W(t)) + (CE_z \times Z(t)) + (CE_G \times G(t)) + + (CE_{vn} \times V_n(t)) + (CE_{vr} \times V_r(t)) + \big(CE_t \times (T_n(t) + T_r(t) + Rq(t)) \times \alpha_i(t)\big)\Big) + (C_{fix} \times \alpha_i(t))$	(11)
$TCX_i(T) = ct \times CF_i(T)$	(12)

The total profit, incorporating the carbon tax strategy, is defined as follows:

$$PR_{CXi}(T) = \Big(s_n \times \sum_{t=1}^{T} (sn(t))\Big) + \Big(s_r \times \sum_{t=1}^{T} (sr(t))\Big) - \Big(c_{pn} \times \sum_{=1}^{T} (m(t))\Big) - \Big(c_{pr} \times \sum_{t=1}^{T} (r(t))\Big) - \Big(c_{hn} \times \sum_{t=1}^{T} (W(t))\Big) - \Big(c_{hr} \times \sum_{t=1}^{T} (Z(t))\Big) - \Big(c_{rm} \times \sum_{t=1}^{T} (S_m(t))\Big) - \Big(c_{vn} \times \sum_{t=1}^{T} (V_n(t))\Big) - \Big(c_{vr} \times \sum_{t=1}^{T} (V_r(t))\Big) - \Big((c_{cq} \times \sum_{t=1}^{T} (G(t))\Big) - \Big(c_{trv} \times \sum_{t=1}^{T} ((T_n(t) + T_r(t) + Rq(t) \times \alpha_i(t))\Big) - \Big(c_{tn} \times \sum_{t=1}^{T} (lcn(t))\Big) - \Big(ctr \times \sum_{t=1}^{T} (lcr(t))\Big) - \Big(r_m \times \sum_{t=1}^{T} (q(t))\Big) - \Big(c_{rp} \times \sum_{t=1}^{T} (Rq(t))\Big) - RTC \times \alpha_i(t) - TCX_i(T)$$

4. Numerical example

This section focuses on optimizing procurement, production, remanufacturing, transportation, and returns. The MILP model was solved using CPLEX 22.1.1 on a laptop with an Intel Core i3-6006U (2.00 GHz) and 3.87 GB RAM. CPLEX was chosen for its robustness and ability to handle complex problems efficiently. The next section presents the dataset used for implementation and validation.

4.1 The data

The first part of this section presents the model through a case study on an MRTSC for new and remanufactured microwaves. Data were collected from a home appliance sales company, covering prices, costs, and demand. Table 1 shows historical demand for both product types and return rates. The data were considered realistic enough for the case study. In the model, the round-trip variable was set to 1 for all periods except t=7 and t=18. Table 2 lists cost and capacity data needed for the model, and initial parameters are given in Table 3.

4.2 The results

This section shows the results of applying the carbon tax strategy and analyzing trip cancellation methods. Canceling trips when tn=tr=rq=0 gives a profit of 1.55E+05 and emissions of 150092 units, matching the optimal plans in Table 4. For periods like 9 and 20, no returns are transported, but new or remanufactured products are still delivered. Using the α2(t) strategy, which cancels trips when rq=0, increases profit to 1.56E+05 and cuts emissions to 145574 units, as shown in Table 5. In this case, when no returns exist but deliveries are needed, the trip is canceled and products are rescheduled to nearby periods, as seen again in periods 9 and 20. This rescheduling improves both profit and emission levels.

5. Conclusions

This study focused on managing manufacturing, remanufacturing, and transportation in the supply chain while applying a carbon tax strategy from raw material purchase to product sales. An integrated MILP model was developed to optimize procurement, production, sales, transportation, and returns. It considers raw material sourcing, transport costs, inventory limits, and vehicle constraints, using carbon footprint as the sustainability

metric. CPLEX was used to solve the model and generate optimal plans for sustainable decision-making. Results show that investing in carbon reduction can improve profits. For policymakers, it's more effective to cancel trips only when no returned products are transported, rather than when no products at all are moved. For future work, the model could be extended to cover multiple cities to tackle inter-city logistics. Also, using a fleet instead of a single vehicle would make the model more realistic for large-scale operations. Finally, studying how delivery times and returned product quality affect planning and emissions could help improve both system performance and sustainability.

Table 1. Demands, percentages of returns and round-trip variable and (H = 25)

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
$d_n(t)$	40	44	50	63	11	10	78	0	48	42	57	88	45	73	52	98	0	12	0	15	92	12	1	36	76
$d_r(t)$	28	37	67	40	39	0	58	0	74	38	71	95	49	12	23	67	21	9	0	14	0	54	71	32	45
R_p	0.83	0.72	0.94	0.73	0.87	0.79	0.94	0.76	0.91	0.94	0.96	0.74	0.66	0.83	0.79	0.94	0.73	0.87	0.79	0.94	0.76	0.91	0.94	0.87	0.79
$\alpha_i(t)$	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1

Table 2. Initial values

vn[1]	vr[1]	z[1]	w[1]	sm [1]	g[1]	tn[1]	tr[1]	q[1]	m[1]	r[1]	sn[1]	sr[1]	rq[1]	lcn[1]	lcr[1]
40	30	80	80	80	80	80	80	50	100	60	40	28	30	0	0

Table 3. Necessary data for coding

S_m^{max}	W^{max}	Z^{max}	G^{max}	V_n^{max}	V_r^{max}	*mmax*	*rmax*	*Ctv*	
160	220	200	240	200	190	150	140	100	
rm	*cpn*	*cpr*	*crm*	*chn*	*chr*	*ctrv*	*cvn*	*cvr*	
5	40	12	0.1	0.5	0.4	2	1.5	1	*RTC* 300
CEm	*CEn*	*CEr*	*CEs*	*CEw*	*CEz*	*CEg*	*CEvn*	*CEvr*	
1	60	10	1	2	2	1	3	3	
ccq	*ctn*	*ctr*	*crp*	*ct*	s_n	s_r	*Cfix*	*CEt*	
0.2	280	210	4	0.05	140	105	2500	1	

Table 4. Optimal plans with α_1(t) strategy (H = 25)

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
m	100	29	101	0	78	0	48	42	57	88	45	73	52	98	0	12	0	15	92	12	1	36	76	0	0
r	60	27	39	0	58	0	74	38	71	66	46	12	23	67	21	9	0	14	0	54	71	32	45	0	0
q	50	0	78	0	48	42	57	88	45	73	52	98	0	12	0	15	92	12	1	36	76	0	0	0	0
tn	80	45	63	102	0	78	0	48	42	57	88	45	73	52	98	0	12	0	15	92	12	1	36	76	0
tr	80	67	40	39	0	58	0	74	38	71	66	46	12	23	67	21	9	0	14	0	54	71	32	45	0
rq	30	29	38	33	55	81	0	59	0	45	40	42	58	37	53	49	72	0	9	0	11	84	11	1	28

Table 5. Optimal plans with α_2(t) strategy (H=25)

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
m	100	28	102	0	78	0	90	0	57	88	45	73	52	98	0	12	0	107	0	12	1	36	76	0	0
r	60	27	39	0	58	0	80	0	71	95	49	12	23	67	21	9	0	14	0	54	71	32	45	0	0
q	50	0	78	0	90	0	57	88	45	73	52	98	0	12	0	107	0	12	1	36	76	0	0	0	0
tn	80	45	63	102	0	78	0	90	0	57	88	45	73	52	98	0	12	0	107	0	12	1	36	76	0
tr	80	67	40	39	0	58	0	80	0	71	95	49	12	23	67	21	9	0	14	0	54	71	32	45	0
rq	30	29	38	33	55	81	0	59	0	45	40	42	58	37	53	49	72	0	9	0	11	84	11	1	28

References

FENG, Y. & LI, X. 2024. Remanufacturing closed-loop supply chain decision comparison under multiple carbon policies. *Applied Mathematics and Nonlinear Sciences,* 9.

KARIM, R. & NAKADE, K. 2022. A Literature Review on the Sustainable EPQ Model, Focusing on Carbon Emissions and Product Recycling. 6**,** 55.

KO, M. D., TU, M. & HO, T. C. Supply chain inventory model considering transportation risk and cost. 2017 4th International Conference on Industrial Engineering and Applications (ICIEA), 21-23 April 2017, 2017. 126-130.

KONSTANTARAS, I., SKOURI, K. & BENKHEROUF, L. 2021. Optimizing inventory decisions for a closed–loop supply chain model under a carbon tax regulatory mechanism. *International Journal of Production Economics,* 239.

LAHMAR, H., DAHANE, M., NADIA, M. & MOHAMMED, H. 2024. Multi-objective sustainable production planning for a hybrid multi-stage manufacturing-remanufacturing system with grade-based classification of recovered and remanufactured products. *Journal of Intelligent Manufacturing*, 1-23.

LIU, K., LI, Q. & ZHANG, H. 2022. Analysis of the Impact of Remanufacturing Process Innovation on Closed-Loop Supply Chain from the Perspective of Government Subsidy. *Sustainability (Switzerland),* 14.

LUO, R., ZHOU, L., SONG, Y. & FAN, T. 2022. Evaluating the impact of carbon tax policy on manufacturing and remanufacturing decisions in a closed-loop supply chain. *International Journal of Production Economics,* 245.

NDHAIEF, N., REZG, N., HAJJI, A. & BISTORIN, O. 2019. Environmental issue in an integrated production and maintenance control of unreliable manufacturing/remanufacturing systems. *International Journal of Production Research,* 58, 1-19.

OKORIE, O., OBI, M., RUSSELL, J., CHARNLEY, F. & SALONITIS, K. 2021. A triple bottom line examination of product cannibalisation and remanufacturing: A review and research agenda. *Sustainable Production and Consumption,* 27, 958-974.

REDDY, K. N., KUMAR, A. & BALLANTYNE, E. E. F. 2019. A three-phase heuristic approach for reverse logistics network design incorporating carbon footprint. *International Journal of Production Research,* 57, 6090-6114.

SHABIR, I., DASH, K. K., DAR, A. H., PANDEY, V. K., FAYAZ, U., SRIVASTAVA, S. & R, N. 2023. Carbon footprints evaluation for sustainable food processing system development: A comprehensive review. *Future Foods,* 7, 100215.

TURKI, S. & REZG, N. Impact of the Transport Activities within a Closed-loop Supply Chain: Study of theLost Profit Risk. 2019 IEEE 6th International Conference on Industrial Engineering and Applications, ICIEA 2019, 2019. Institute of Electrical and Electronics Engineers Inc., 800-804.

YIN, C., ZHANG, Z.-A., FU, X. & GE, Y.-E. 2024. A low-carbon transportation network: Collaborative effects of a rail freight subsidy and carbon trading mechanism. *Transportation Research Part A: Policy and Practice,* 184, 104066.

Comparative Performance of Machine Learning Architectures for Fault Detection and Diagnosis in Chemical Processes

R. AMMAR KHODJA [1, 2], A. VOISIN [2], V. COSTA [1], B. CELSE [1], F. CASTERAN [1], B. IUNG [2]

[1] IFPEN, Rond-point de l'échangeur de Solaize, France, *rayane.ammar-khodja@ifpen.fr*, *victor.costa@ifpen.fr*, *fanny.casteran@ifpen.fr*, *benoit.celse@ifpen.fr*

[2] *Université de Lorraine, UMR CNRS 7039, CRAN, F-54000, Nancy, France*, *alexandre.voisin@univ-lorraine.fr*, *benoit.iung@univ-lorraine.fr*

Abstract. Fault Detection and Diagnosis (FDD) is crucial for both maintenance and control in chemical industries, where early fault detection can prevent costly failures and optimize operations. This is particularly critical in pilot units like those at IFP Energies Nouvelles (IFPEN), which operate under short-term experimental conditions with frequently varying operational parameters.

This study conducts an extensive benchmarking analysis using the Tennessee Eastman Process (TEP), a widely used simulated chemical process dataset, to evaluate multiple approaches for fault detection and diagnosis, featuring numerous continuous operations with different sensors and faults. Several methods were implemented and compared including Multi Scale PCA (MS-PCA), AutoEncoder, Ensemble Learning, and LSTM models for fault detection, alongside Random Forest, XGBoost, and BLSTM (Bidirectional LSTM) for fault diagnosis.

Using the TEP dataset, our results demonstrate that Ensemble Learning achieves detection rates ranging from 80% to 100% across various fault scenarios for the fault detection task. For the fault diagnosis task, BLSTM achieved a diagnosis accuracy of 98.76%. The study reveals that ensemble-based approaches consistently outperform individual models in handling the complex, multivariate nature of chemical process data, due to its robustness in combining multiple perspectives, comprehensive data capture, and localized detection capabilities. Furthermore, the superior performance of the BLSTM is due to its ability to capture both past and future temporal dependencies in the sequential data, particularly important in chemical processes where fault patterns may manifest with complex temporal relationships.

This research contributes to the field of Fault Detection and Diagnosis by providing empirical evidence for the effectiveness of ensemble methods and Bidirectional LSTMs in order to address industrial FDD applications on chemical processes.

1. Introduction

Chemical processes present unique challenges for fault detection and diagnosis due to their inherently complex, nonlinear dynamics and strong interdependencies among process variables. In these environments, undetected faults can lead to product quality deterioration, equipment damage, environmental hazards, or catastrophic safety incidents with potential loss of life. Traditional FDD approaches in chemical processing have relied on first-principles models, statistical process control, and expert systems, which often require extensive process knowledge and struggle with the high dimensionality of modern chemical plants that may incorporate hundreds of sensors.

The evolution of artificial intelligence has introduced advanced computational methods that offer significant improvements over classical approaches. Machine learning and deep learning algorithms demonstrate remarkable capabilities in processing large volumes of data to identify latent patterns and adapt to novel anomaly types that conventional methods might overlook. This advancement is particularly valuable for chemical reactors, where early anomaly detection and accurate diagnosis are critical for forecasting undesired events and enhancing process efficiency.

While various machine learning approaches have been proposed for Fault Detection and Diagnosis (FDD) applications [Schmidl & Wenig], there remains a significant gap in comprehensive testing of ensemble models and deep learning architectures for this specific application [Zamanzadeh Darban et al, 2025]. These advanced methods have shown promise in other domains but require thorough evaluation in the context of chemical process monitoring, where data is often multivariate, non-linear, and exhibits complex temporal dependencies.

2. Material and Methods

2.1 *Tennessee Eastman Process Dataset*

The TEP dataset is derived from a real industrial process that was computationally modelled in 1993 by Downs and Vogel [Downs & Vogel, 1993]. It has become a widely used benchmark for studying industrial process control, particularly in FDD. The process involves 8 chemical components (4 reactants, 2 products, 1 by-product, and 1 inert component) and 5 main process units (reactor, condenser, separator, compressor, and stripper).

The dataset provides 53 different variables tracking system properties, including flow rates, pressures, temperatures, levels, mole fractions, and compressor power outputs. Process variables are sampled every 3 minutes for 25 hours in the training set and 48 hours in the testing set. The dataset contains both fault-free data and 20 different simulated process faults, making it ideal for benchmarking anomaly detection algorithms.

2.2 *Fault Detection Methods*

Anomaly detection in industrial processes involves identifying deviations from normal operating conditions by establishing baseline patterns of healthy system behaviour and detecting significant departures from these patterns. Most advanced methods employ a reconstruction-based approach [Zamanzadeh Darban et al, 2025], where models are trained on fault-free data to learn normal process behaviour, and anomalies are detected when the reconstruction error exceeds a predetermined threshold. In our study, we evaluated five distinct approaches:

- Statistical Thresholding (ST) establishes control limits at ±3 standard deviations from the mean for each sensor based on fault-free data, flagging values outside this range as potential faults.
- Multi-Scale PCA (MS-PCA) transforms high-dimensional process data into a lower-dimensional space to capture the principal patterns, detecting anomalies when the reconstruction error between original and PCA-reconstructed data exceeds a statistical threshold.
- AutoEncoder (AE) employs a neural network architecture to compress input data into a latent representation and then reconstruct it, identifying anomalies when the reconstruction error is abnormally high.
- Ensemble Learning (EL) constructs multiple linear regression models where each model predicts one specific feature based on all others, detecting faults when the collective prediction error exceeds expected values.
- Long Short-Term Memory (LSTM) networks analyse temporal sequences using a sliding-window approach to capture time-dependent patterns in the process data, recognizing anomalies when the predicted sequence deviates significantly from actual measurements.

2.3 *Fault Diagnosis Methods*

Fault diagnosis extends beyond detection to accurately classify and identify the specific type of fault occurring in the system, enabling appropriate corrective actions and maintenance procedures. Once an anomaly is detected, diagnosis techniques analyse the pattern of deviations across multiple sensors to determine the root cause from among possible fault categories. For this critical classification task, we evaluated three promising machine learning approaches:

- Random Forest Classifier constructs multiple decision trees during training and determines the fault class through majority voting, providing robust classification by aggregating predictions across a diverse ensemble of trees.
- XGBoost Classifier implements gradient boosting techniques that sequentially build decision trees to correct errors made by previous trees, creating a powerful ensemble that excels at capturing complex nonlinear relationships in process data.
- Bidirectional LSTM (BLSTM) processes time-series data in both forward and backward directions to capture temporal dependencies from both past and future contexts, making it particularly effective for fault patterns with complex sequential characteristics.

2.4 *Methodology*

All models were trained exclusively on fault-free data to learn normal process behavior patterns. For the reconstruction-based detection methods, data normalization was performed using z-score standardization to ensure equal contribution from each variable.

Hyperparameter optimization employed grid search with 5-fold cross-validation to identify optimal configurations. Neural networks benefited from an exponential decay learning rate schedule starting at 10^{-3} and decreasing by a factor of 0.9 every 1000 steps.

Detection thresholds were established at $\mu + 3\sigma$ of the reconstruction loss distribution from fault-free validation data, providing statistical significance at the 99.73% confidence level.

For classification models, training involved label encoding of the 20 fault types as categorical targets, an 80:20 train-validation split for hyperparameter tuning, and techniques to address the imbalanced nature of fault occurrences to prevent bias toward majority classes.

2.4 Metrics used to compare methods

Our comprehensive evaluation framework employed multiple complementary metrics to thoroughly assess both detection and diagnosis capabilities. For fault detection performance, we utilized Fault Detection Rate (FDR) to measure the proportion of correctly detected faults relative to all actual faults, providing insight into the models' sensitivity. Accuracy measured the overall correctness of predictions, counting both true positives and true negatives. To address potential imbalances in the dataset, we incorporated Matthews Correlation Coefficient (MCC), which provides a balanced assessment by considering all confusion matrix components even when classes are disproportionately represented.

Receiver Operating Characteristic (ROC) curves enabled visualization of the trade-off between true positive and false positive rates across varying detection thresholds, with Area Under the Curve (AUC) serving as a single comprehensive metric of detection quality.

For fault diagnosis evaluation, we employed confusion matrices to visualize classification performance across all fault types, highlighting both successful classifications and misclassifications between similar fault patterns. Classification accuracy provided an overall performance metric, while t-SNE dimensionality reduction enabled visualization of the models' internal representations, offering valuable insights into how effectively each approach separated different fault classes in the feature space.

3. Results and Analysis

3.1 *Fault Detection Results*

The performance of fault detection methods was evaluated using multiple metrics across 200-day simulations. Ensemble Learning demonstrated superior performance with accuracy rates ranging from 80% to 100% across most fault scenarios (Table 1). LSTM also showed strong performance, particularly after optimizing window size and stride parameters.

Analysis of the ROC curves (not showed here) revealed that choosing a threshold of $\mu + 3\sigma$ consistently resulted in the smallest False Positive Rate (FPR) and the highest True Positive Rate (TPR) for all five fault detection methods. This threshold effectively minimizes false alarms while maximizing the detection of actual faults.

The strengths of Ensemble Learning included robustness through combining multiple models, effective identification of deviations in specific sensors, and computational efficiency suitable for real-time monitoring. Additionally, it demonstrated superior performance in MCC scores, indicating balanced classification across various fault scenarios.

3.2 *Fault Diagnosis Results*

For fault classification over 400-day simulations, the evaluated models demonstrated increasing levels of accuracy, with Random Forest achieving 91%, XGBoost reaching 93%, and BLSTM delivering 99% accuracy.

The BLSTM model significantly outperformed both RF and XGBoost classifiers, demonstrating its superior capability to capture temporal dependencies and sequential patterns critical for accurate fault diagnosis. The t-SNE visualization of the BLSTM's intermediate layer outputs (Figure 1) showed clear

clustering of different fault classes with minimal overlap, confirming the model's ability to learn highly discriminative features for fault classification.

No.	Fault Description	ST	MS-PCA	AE	EL	LSTM
1	A/C feed ratio, B comp. const. (Stream 4)	86.77	85.41	85.83	**99.37**	99.05
2	B comp., A/C ratio const. (Stream 4)	86.25	84.79	85.10	**98.54**	99.26
4	Reactor cooling water inlet temp.	87.08	85.41	85.94	**99.69**	98.52
5	Condenser cooling water inlet temp.	41.04	16.35	65.73	**99.69**	28.63
6	A feed loss (Stream 1)	87.08	85.62	86.04	**99.69**	98.63
7	C header pressure loss (Stream 4)	87.08	85.62	86.04	**99.69**	98.63
8	A, B, and C comp. (Stream 4)	85.41	80.10	84.16	96.25	**97.89**
10	C feed temp. (Stream 4)	49.27	19.06	19.48	**66.14**	20.74
11	Reactor cooling water inlet temp.	72.29	48.13	51.88	69.79	**72.32**
12	Condenser cooling water inlet temp.	86.45	82.91	84.38	**98.13**	98.11
13	Reaction kinetics	82.50	79.06	80.00	93.33	**94.95**
14	Reactor cooling water valve	87.08	79.06	85.94	**99.69**	98.63
16	Unknown	42.70	29.47	25.83	**65.41**	19.68
17	Unknown	77.71	75.72	76.66	**89.16**	**89.16**
18	Unknown	81.35	80.21	80.83	93.22	**94.42**
19	Unknown	42.29	8.85	10.41	**69.37**	22.84
20	Unknown	51.14	46.87	45.41	**67.71**	67.16

Table 1. Comparison of Fault Detection Methods using Accuracy (Percentage %).

Real-time fault classification using BLSTM further validated this performance, showing high-confidence predictions with distinct probability distributions for each fault type, reinforcing the model's reliability for industrial fault diagnosis applications.

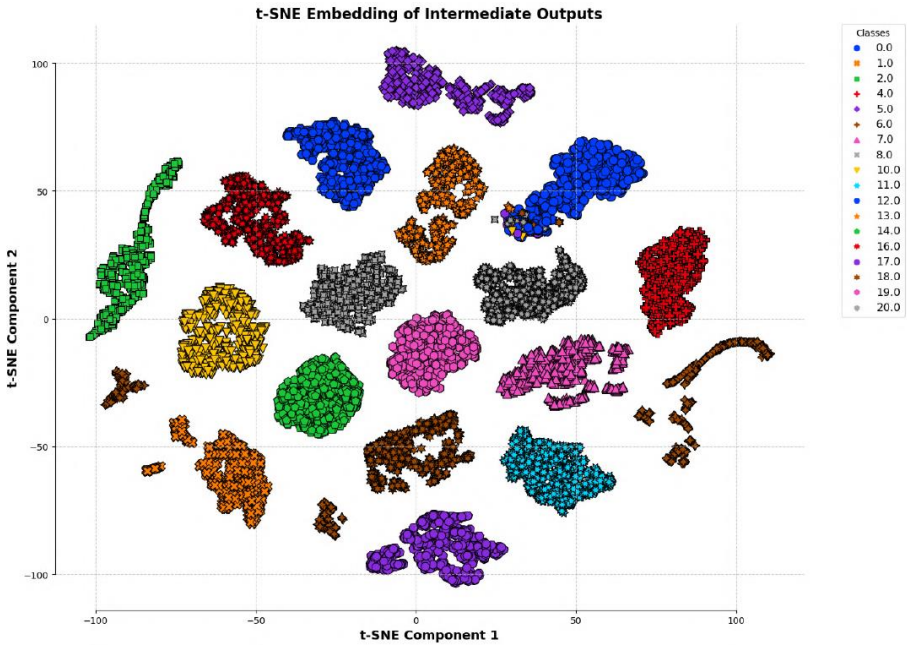


Figure 1. t-SNE plot for the classes of the trained data in the BLSTM for 400 days simulations.

To enhance fault interpretability and provide deeper diagnostic insights, we implemented a variable contribution analysis method for detected anomalies, as illustrated in Figure 2. This visualization approach not only identifies when a fault occurs but also determines which specific process variables contribute most significantly to the anomaly detection. The upper plot displays the reconstruction loss trajectory, clearly marking when it exceeds the detection threshold, while the lower plot quantifies the individual contribution of each sensor variable at a selected timestep after fault detection. This addition proves particularly valuable for maintenance personnel, as it pinpoints the most affected process components, effectively bridging the gap between detection and diagnosis. In the case shown, the reactor cooling water flow variable (xmvs_10) exhibits

a substantially higher contribution compared to other variables, directly corresponding to the nature of Fault 4, which involves changes in reactor cooling water inlet temperature.

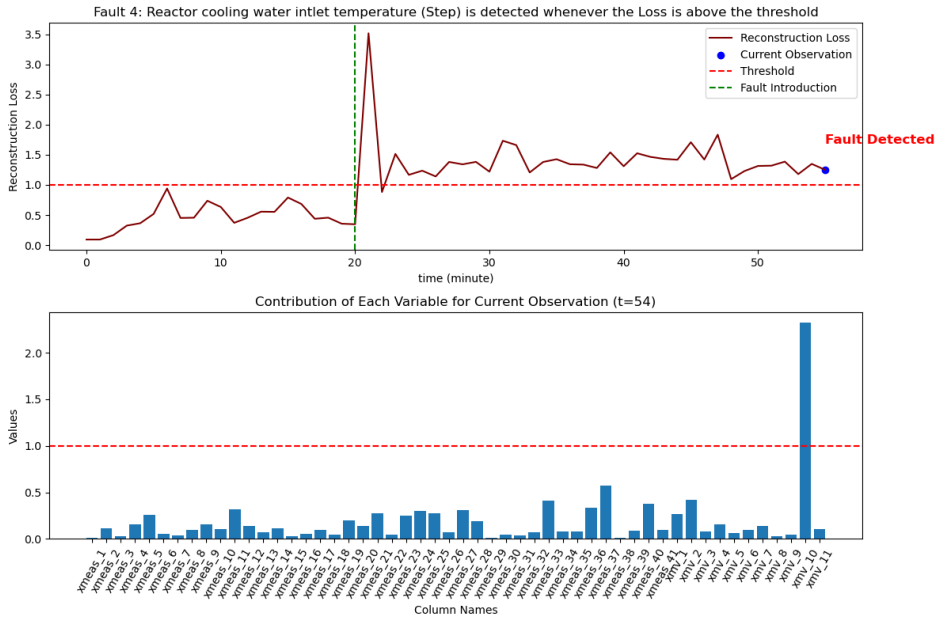


Figure 2. Anomaly detection in the Tennessee Eastman process: Analysis of reconstruction loss and identification of contributing variables: evolution of reconstruction loss over time (to identify when the fault was introduced) and analysis of individual variable contributions at time t=54 (162 minutes from the start of simulation) for anomaly diagnosis.

4. Conclusion

This study contributes to the field of fault detection and diagnosis by demonstrating the effectiveness of ensemble-based and deep learning approaches for industrial chemical processes. Our comprehensive evaluation framework established that ensemble methods for detection and bidirectional sequence models for diagnosis offer significant advantages in handling the complex, multivariate nature of process data. The complementary strengths of these approaches provide a foundation for robust fault management systems that can significantly enhance industrial safety and efficiency.

Future work will focus on developing predictive capabilities to forecast faults several days in advance and addressing the significant challenge of applying these methods to real pilot plants characterized by frequently changing operating conditions. Unlike production facilities, research pilot plants deliberately modify process parameters to explore new chemical pathways and optimizations, requiring FDD systems that can distinguish between intentional parameter changes and actual faults, while facing the additional challenge that real-life chemical pilot plants typically lack labeled data distinguishing between faulty and non-faulty operations or identifying specific fault types. This dynamic environment necessitates adaptive models capable of functioning effectively when baseline "normal" operation continuously evolves.

References

Schmidl, S., Wenig, P., & Papenbrock, T. (2022). Anomaly detection in time series. *Proceedings of the VLDB Endowment*, *15*(9), 1779–1797.

Zahra Zamanzadeh Darban, Geoffrey I. Webb, Shirui Pan, Charu Aggarwal, and Mahsa Salehi. 2024. Deep Learning for Time Series Anomaly Detection: A Survey. ACM Comput. Surv. 57, 1, Article 15 (January 2025), 42 pages.

Downs, J.J., Vogel, E.F., 1993. A plant-wide industrial process control problem. Comput. Chem. Eng. 17, 245–255.

Carla Martin-Villalba, Alfonso Urquia, Guodong Shao, Implementations of the Tennessee Eastman Process in Modelica, IFAC-PapersOnLine, Volume 51, Issue 2, 2018, Pages 619-624, ISSN 2405-8963.

Joint maintenance, mission abort and repairpersons assignment optimization problem in systems under random operating environment

R. O'Neil[*,**], C. Diallo[*], A. Khatab[*,**], E.-H. Aghezzaf[***]

[*] *Dalhousie University, Halifax, Canada: ry855314@Dal.ca*

[*] *Dalhousie University, Halifax, Canada: Claver.Diallo@Dal.ca*

[**] *Lorraine University, LGIPM-Metz, France: abdelhakim.khatab@univ-lorraine.fr*

[***] *Ghent University, Belgium: ElHoussaine.Aghezzaf@UGent.be*

Abstract Despite the increasing number of studies dealing with mission abort policies (MAP), very few references considered the impact of maintenance on mission-abort decisions. This paper presents a novel model to jointly optimize selective maintenance scheduling and mission-abort decisions for mission-critical systems. The system is assumed to operate in an uncertain environment, which impacts its performance. Given that maintenance resources are limited, this paper develops a new integrated optimization approach where the classical SMP is extended to include mission-abort policies. In this approach, SM and mission-abort decisions are integrated into a single optimization model. A mixed-integer non-linear programming model is formulated to minimize expected total maintenance, mission failure, and penalty costs. A solution method is developed, and numerical experiments are then conducted to validate the proposed model and highlight its added value.

Acronyms

CM	Corrective maintenance
PM	Preventive maintenance
MSP	Mission success probability
SSP	System survival probability
RSP	Rescue success probability
RBD	Reliability block diagram
SM	Selective maintenance
SMP	Selective maintenance problem
MAP	Mission abort policy
MCS	Mission-critical system

1. Introduction

Mission-critical systems (MCS) are ubiquitous in several application areas such as transportation and military. To assess their performance, the main metric used is the probability of successfully operating a mission (MSP). To increase the MSP of a MCS to operate subsequent missions, its components usually undergo maintenance actions during scheduled breaks between two consecutive missions. However, maintenance resources (i.e., spare parts, time, budget) are usually limited. Therefore, only a subset of the MCS's components can be maintained. In the literature, this resource-constrained maintenance problem is known as the selective maintenance problem (SMP). Selective maintenance (SM) has been demonstrated to be suitable for mission-oriented systems; systems that operate a sequence of alternating missions and breaks (Rice 1999). Its main objective is to identify the optimal subset of components and maintenance actions that either (i) minimize maintenance costs while ensuring a required level of system reliability or MSP, or (ii) maximize the system reliability subject to a limited amount of maintenance resources (Al Jabouri et al. 2024).

According to the literature review conducted, it is shown that existing SM approaches deal only with MSP assessment (Al Jabouri et al. 2024). Nevertheless, in many real-world situations, the survival of MCS, such as submarines, unmanned aerial vehicles, and chemical power plants, is prioritized over the

completion of the mission; indeed, a failure of a MCS can have serious economic, environmental, and human consequences . To avoid such undesirable situations, one can resort to mission-abort policies (MAP) which contribute to improving the system's survivability (Levitin et al. 2018). A mission is aborted when specific conditions are satisfied and an immediate rescue/recovery procedure is initiated. An optimal MAP shall ensure an appropriate trade-off between mission success versus system survival.

An increasing number of studies have dealt with MAP optimization problem. However, only three papers (Levitin et al. 2021, Liu et al. 2022, Zhao et al. 2023) studied the impact of maintenance on mission abortion decisions. Levitin et al. (2021) investigated the joint optimization of maintenance planning and mission abort decisions for a dual-unit standby system where, during the mission, one unit must operate while the other can undergo preventive maintenance (PM). An optimization model is developed to maximize the MSP under a required minimum level of the SSP. Meanwhile, this approach is inapplicable to many mission-critical systems as it assumes that maintenance is allowed during missions. The work in (Liu et al. 2022) deals with the problem of jointly optimizing task termination and PM decisions for phased-task systems. At the end of each phase, PM can be performed to reduce system degradation. The resulting optimization problem is formulated as a Markov decision process (MDP) to identify the optimal PM level and the number of phases to carry out. This approach relies on single-component systems, and the duration of the rescue procedure is negligible, which makes the rescue procedure always successful. Limited maintenance resources such as time and repairpersons, which often need to be considered when planning maintenance tasks, are also not accounted for. Zhao et al. (2023) considered multicomponent MCS operating under random shocks impacting their performance. To increase the MSP and system survivability, the mission is performed by performing multiple attempts. At a given number of failed components, the mission is aborted, the rescue procedure is initiated, and the system is either imperfectly maintained or completely renewed. At the component level, only replacement is available as a maintenance option, and imperfect maintenance refers to the case where not all components are replaced. Furthermore, time to replace a component is negligible, maintenance actions are performed during the mission, and no maintenance resource constraints are accounted for.

This paper presents a novel selective maintenance model to jointly optimize maintenance planning and mission-abort decisions for mission-critical systems. The proposed model applies to MCS whose RBD is a three-block switching structure where the main block is in series with a two-block cold standby unit. The system is assumed to operate in an uncertain environment which impacts its performance. Given that maintenance resources are limited, a new integrated optimization approach where the classical SMP is extended to include mission-abort policies is developed. A mixed-integer non-linear programming model is formulated to minimize expected total maintenance, mission failure, and penalty costs. A solution method is developed, and numerical experiments are conducted to validate the proposed model and highlight its added value.

2. System description and problem formulation

The system under consideration is tasked with running a series of alternating missions and scheduled breaks. The system is assumed to have just completed a given mission and is shut down to start the subsequent break of duration $\mathscr{T}_0$ during which a maintenance plan can be performed. After the break, the system will be operated during the next mission of duration W. The system is composed of three blocks, each of which consists of series-parallel subsystems(see Figure 1). The ith $(i = 1,2,3)$ block is composed of N_i series subsystems S_{i_j} $(j = 1,\ldots,N_i)$ which in turn are composed of N_{i_j} elements E_{ijk} $(k = 1,\ldots,N_{i_j})$ arranged in parallel (see Figure 2). Time to failure T_{ijk} of an element E_{ijk} is generally distributed with a baseline reliability denoted as $R_{0,ijk}(t)$. At the beginning and at the end of a break, each element E_{ijk} is characterized by two binary status variables U_{ijk} and V_{ijk} defined as in (1) and

(2). Furthermore, we let A_{ijk} and B_{ijk} be the respective ages of E_{ijk} at the beginning and at the end of the scheduled break.

$$U_{ijk} = \begin{cases} 1, & \text{if } E_{ijk} \text{ is functioning at the start of the break,} \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

$$V_{ijk} = \begin{cases} 1, & \text{if } E_{ijk} \text{ is functioning at the end of the break} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

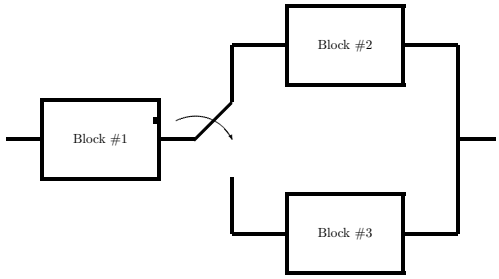


Figure 1. RBD of the MCS under investigation

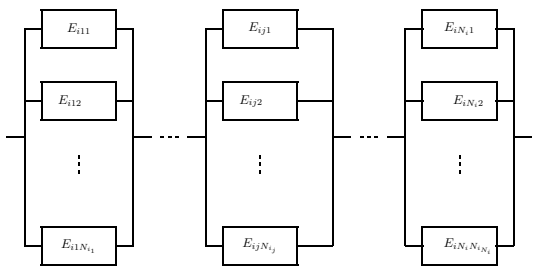


Figure 2. RBD of block #i ($i = 1,2,3$) in the MCS under investigation

To successfully operate the next mission, all subsystems in the two blocks 1 and 2 must be functioning during the mission. If at least one of these subsystems fails before achieving the mission, the mission fails. However, if all subsystems from both blocks 1 and 3 are still functioning upon the mission failure, a rescue operation can be launched to save the system from being lost. The duration $\Psi(t)$ of the rescue operation depends on the time of its activation. Obviously, the rescue operation fails if at least one of the subsystems from blocks 1 and 3 fails, resulting then in the system lost.

The system is assumed to operate under random operation environments (OE) the stresses of which may considerably increase their risk of failure. In the present work, shock theory is used to model the effect of the random OE on the failure process of the system's components. Shocks are assumed to occur according to a homogeneous Poisson process (HPP). Two operating environments are considered. The first, denoted as OE_1, is related to the primary mission period where shocks occur according to a HPP $\mathscr{P}_1 = \{N_1(t), t \geq 0\}$, with rate $\lambda_1(t)$, and $N_1(t)$ is the number of shocks that occurred during $[0,t]$. The second operating environment, denoted as OE_2, characterizes the rescue period where shocks occur according to a HPP $\mathscr{P}_2 = \{N_2(t), t \geq 0\}$, with rate λ_2. Accordingly, each element E_{ijk} is characterized by internal and external failures. The former is due to the inherent aging process, while the latter is due to shocks. An external failure results from shock occurrence with a given probability that increases by the increase in the number of experienced shocks. In the present work, the probability $\mathscr{Q}_{ijk}(m)$ that the element E_{ijk} will survive the mth shock is defined as:

$$\mathscr{Q}_{ijk}(m) = \left(\mathscr{Q}_{ijk}(1)\right)^m \tag{3}$$

where $\mathscr{Q}_{ijk}(1)$ is the probability that element E_{ijk} survives the first shock.

3. Mission success probability and system survivability computation

3.1. *Mission success probability computation*

The system successfully operates the next mission if all series subsystems in blocs 1 and 2 do not fail during the mission of length M. The MSP $\mathscr{R}_{pm}(W,s,OE_1)$ is computed as:

$$\mathscr{R}_{pm}(W,\mathbb{B},s,OE_1) = \mathscr{R}_1(W,\mathbb{B}_1,s,OE_1) \times \mathscr{R}_2(W,\mathbb{B}_2,s,OE_1), \tag{4}$$

where $\mathscr{R}_i(W,\mathbb{B}_i,s,OE_1)$ is the MSP of the ith block ($i \in \{1,2\}$) under operation environment OE_1:

$$\mathscr{R}_i(W,\mathbf{B},s,OE_1) = \prod_{j=1}^{N_i}\left[1-\prod_{k=1}^{N_{i_j}}\left(1-\mathscr{R}_{ijk}\left(W\Big|_{B_{ijk}},s,OE_1\right)\cdot V_{ijk}\right)\right], \tag{5}$$

where $\mathscr{R}_{ijk}\left(W\Big|_{B_{ijk}},s,OE_1\right)$ is the probability of element E_{ijk} to survive the next mission of length W under the operating environment OE_1 given its initial age is B_{ijk}. This probability is defined as the conjunction of the three events: (1) no internal failure occurs, (2) less than s shocks have occurred, (3) and element E_{ijk} survives all shocks. Accordingly, if T_{ijk} denotes the failure time of E_{ijk}, $\mathscr{R}_{ijk}(W,s,OE_1)$ is computed as:

$$\begin{aligned}\mathscr{R}_{ijk}\left(W\Big|_{B_{ijk}},s,OE_1\right) &= \Pr\left(T_{ijk} > W\Big|_{T_{ijk}>B_{ijk}},\mathscr{N}_1(W) < s\right) \\ &= \frac{\mathscr{R}_{0,ijk}(W+B_{ijk})}{\mathscr{R}_{0,ijk}(B_{ijk})}\sum_{n=0}^{s-1}\Pr(\mathscr{N}_1(M)=n)\prod_{m=0}^{n}\mathscr{Q}_{ijk}(m) \\ &= e^{-\lambda_1 W}\frac{\mathscr{R}_{0,ijk}(W+B_{ijk})}{\mathscr{R}_{0,ijk}(B_{ijk})}\sum_{n=0}^{s-1}\frac{(\lambda_1 W)^n}{n!}\left(\mathscr{Q}_{ijk}(1)\right)^{\frac{n(n+1)}{2}}\end{aligned} \tag{6}$$

$$\tag{7}$$

3.2. *Rescue procedure successful probability computation*

The rescue procedure is launched whenever the system fails to achieve the next mission. The system fails at time t during the mission, and the rescue procedure is successfully initiated at that time if one subsystem from the second block fails while all subsystems in blocks 1 and 3 are still functioning. The probability that all subsystems from the first block survive up to time t and one subsystem among those of the 2nd block fails within the time interval $[t,t+\mathrm{d}t)$ is given as:

$$-\mathscr{R}_1\left(t\big|_{\mathbb{B}_1},s,OE_1\right)\mathrm{d}\mathscr{R}_2\left(t\big|_{\mathbb{B}_2},s,OE_1\right) \tag{8}$$

where $\mathscr{R}_i(t\big|_{\mathbb{B}_i},s,OE_1)$ is the probability of the ith ($i=1,2$) block to survive the mission up to time t given an initial age vector $\mathbb{B}_i$. These probabilities are straightforwardly obtained in a similar way as in Equations (5) and (6). These quantities are computed as :

$$\mathscr{R}_i(t\big|_{\mathbb{B}_i},s,OE_1) = \prod_{j=1}^{N_i}\left[1-\prod_{k=1}^{N_{i_j}}\left(1-\mathscr{R}_{ijk}\left(t\Big|_{B_{ijk}},s,OE_1\right)\cdot V_{ijk}\right)\right] \tag{9}$$

and, $\mathscr{R}_{ijk}\left(t\Big|_{B_{ijk}},s,OE_1\right)$ can be computed from (6). Given that the rescue procedure of length $\Psi(t)$ is initiated within the time interval $[t,t+\mathrm{d}t)$, the probability to successfully be performed is computed as:

$$-\mathscr{R}_1(t\big|_{\mathbb{B}_1},s,OE_1)\mathscr{R}_1(\Psi(t)\big|_{\mathbb{B}_1+t},OE_2)\mathscr{R}_3(\Psi(t)\big|_{\mathbb{B}_3},OE_2)\mathrm{d}\mathscr{R}_2(t\big|_{\mathbb{B}_2+t},s,OE_1) \tag{10}$$

where $\mathscr{R}_1(t\big|_{\mathbb{B}_1},s,OE_1)$ and $\mathscr{R}_2(t\big|_{\mathbb{B}_2+t},s,OE_1)$ can be derived from (5) and (6). $\mathscr{R}_i(.,OE_2)$ is the probability of the ith block ($i \in \{1,3\}$) to survive the rescue phase while subjected to the operating envi-

ronment OE_2. These probabilities are evaluated as:

$$\mathscr{R}_1(\Psi(t)\,|_{\mathbb{B}_1+t}\,, OE_2) = \prod_{j=1}^{N_1}\left[1-\prod_{k=1}^{N_{1_j}}\left(1-\mathscr{R}_{1jk}(\Psi(t)\,|_{\mathbb{B}_1+t}\,, OE_2)\cdot V_{1jk}\right)\right] \tag{11}$$

$$\mathscr{R}_3(\Psi(t)\,|_{\mathbb{B}_3}\,, OE_2) = \prod_{j=1}^{N_3}\left[1-\prod_{k=1}^{N_{3_j}}\left(1-\mathscr{R}_{3jk}(\Psi(t)\,|_{\mathbb{B}_3}\,, OE_2)\cdot V_{3jk}\right)\right] \tag{12}$$

where $\mathscr{R}_{1jk}(.,OE_2)$ and $\mathscr{R}_{3jk}(.,OE_2)$ are the respective probabilities of element E_{1jk} and E_{3jk} to survive the rescue phase under the operating environment OE_2. These probabilities are defined as a conjunction of the two events: (1) no internal failure occurs, (2) and element E_{ijk} survives all shocks:

$$\mathscr{R}_{1jk}\left(\Psi(t)\,|_{\mathbb{B}_1+t}\,, OE_2\right) = e^{-\lambda_2\Psi(t)}\mathscr{R}_{0,1jk}\left(\Psi(t)\,|_{\mathbb{B}_1+t}\right)\sum_{n=0}^{\infty}\frac{(\lambda_2\Psi(t))^n}{n!}\left(\mathscr{Q}_{1jk}(1)\right)^{\frac{n(n+1)}{2}} \tag{13}$$

$$\mathscr{R}_{3jk}(\Psi(t), OE_2) = e^{-\lambda_2\Psi(t)}\mathscr{R}_{0,3jk}\left(\Psi(t)\,|_{\mathbb{B}_3}\right)\sum_{n=0}^{\infty}\frac{(\lambda_2\Psi(t))^n}{n!}\left(\mathscr{Q}_{3jk}(1)\right)^{\frac{n(n+1)}{2}} \tag{14}$$

From the above equations, it follows that the probability $\mathscr{R}_{rp}(W,\mathbb{B},\Psi,s)$ of the system to successfully operate the rescue procedure is obtained by integrating along the mission duration as:

$$\mathscr{R}_{rp}(W,\mathbb{B},\Psi,s) = -\int_0^W \mathscr{R}_1(t\,|_{\mathbb{B}_1}\,, s, OE_1)\mathscr{R}_1(\Psi(t)\,|_{\mathbb{B}_1+t}\,, OE_2)\mathscr{R}_3(\Psi(t)\,|_{\mathbb{B}_3}\,, OE_2)\mathrm{d}\mathscr{R}_2(t\,|_{\mathbb{B}_2+t}\,, s, OE_1). \tag{15}$$

Survivability of the system is ensured upon successfully completing either the next mission or the rescue procedure. From Equations (4) and (16), the survivability $\mathscr{R}_{sv}(L,\Psi,s)$ of the system for the upcoming mission is computed as:

$$\mathscr{R}_{sv}(W,\mathbb{B},\Psi,s) = \mathscr{R}_{pm}(W,\mathbb{B},s,OE_1) + \mathscr{R}_{rp}(W,\mathbb{B},\Psi,s). \tag{16}$$

4. Total maintenance cost and duration, and penalty cost

During the break, each element E_{ijk} can be subjected to either corrective replacement (CR) or preventive replacements (PR) after which the element becomes new. A CR concerns failed elements while a PR is executed on elements that are still functioning at the beginning of the break. When a CR is performed on E_{ijk} it induces a cost c^c_{ijk} and consumes t^c_{ijk} units of time from the break duration. Similarly, when a PR is performed on E_{ijk} it induces a cost c^p_{ijk} and consumes t^p_{ijk} units of time from the break duration. Obviously, when an element is not selected for maintenance, neither maintenance cost nor time is incurred. To compute the total maintenance cost and time during the break, the following binary decision variable z_{ijk} is introduced:

$$z_{ijk} = \begin{cases} 1, & \text{if } E_{ijk} \text{ is replaced} \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Accordingly, the total maintenance cost incurred by corrective and preventive replacements is respectively denoted $\mathscr{C}_{CR}$ and $\mathscr{C}_{PR}$, and computed as:

$$\mathscr{C}_{CR} = \sum_{k=1}^{3}\sum_{i=1}^{N_k}\sum_{j=1}^{N_{k_i}} c^c_{ijk}\cdot\left(1-U_{ijk}\right)\cdot z_{ijk} \quad \mathscr{C}_{PR} = \sum_{k=1}^{3}\sum_{i=1}^{N_k}\sum_{j=1}^{N_{k_i}} c^p_{ijk}\cdot U_{ijk}\cdot z_{ijk} \tag{18}$$

Finally, the total maintenance cost is the sum $\mathscr{C}_M = \mathscr{C}_{CR} + \mathscr{C}_{PR}$.

$$\mathscr{C}_M = \sum_{k=1}^{3}\sum_{i=1}^{N_k}\sum_{j=1}^{N_{k_i}} c^c_{ijk}\cdot\left(1-U_{ijk}\right) + c^p_{ijk}\cdot U_{ijk}\cdot z_{ijk} \tag{19}$$

Similarly, the total maintenance time $\mathscr{T}_M$ spent to carry out maintenance activity is given by:

$$\mathscr{T}_M = \sum_{i=1}^{3}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{i_j}} t_{ijk}^c \cdot \left(1 - U_{ijk}\right) + t_{ijk}^p \cdot U_{ijk} \cdot z_{ijk} \tag{20}$$

5. Mathematical programming model

The goal is to jointly determine the set of system's elements to be maintained and the number s of shocks to initiate the rescue procedure. Depending on the objective of the decision maker, several objective functions can be considered. In this paper, the goal is to maximize the survivability of the system while subjected to the limited maintenance budget $\mathscr{C}_0$ as well as to the limited duration $\mathscr{T}_0$ allotted to the intermission break.

$$\text{Max } \mathscr{R}_{sv}(W, \mathbb{B}, \Psi, s) \tag{21}$$

$$\text{S.t.:} \sum_{i=1}^{3}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{i_j}} c_{ijk}^c \cdot \left(1 - U_{ijk}\right) + c_{ijk}^p \cdot U_{ijk} \cdot z_{ijk} \leq \mathscr{C}_0 \tag{22}$$

$$\sum_{i=1}^{3}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{i_j}} t_{ijk}^c \cdot \left(1 - U_{ijk}\right) + t_{ijk}^p \cdot U_{ijk} \cdot z_{ijk} \leq \mathscr{T}_0 \tag{23}$$

$$V_{ijk} = U_{ijk} + \left(1 - U_{ijk}\right) \cdot z_{ijk} \tag{24}$$

$$\sum_{i=1}^{3}\sum_{j=1}^{N_i}\sum_{k=1}^{N_{i_j}} z_{ijk} \leq 1 \tag{25}$$

In the above optimization model, Constraints (22) and (23) are, respectively, the allotted maintenance budget and the break duration constraints. For each component, Constraints (25) ensure that at most one replacement can be performed if an element is to be maintained. Constraints (24) update the status of elements at the end of the intermission break.

References

Rice WF (1999) Optimal selective maintenance decisions for series systems. MAsC Thesis. Mississippi State University.

Al-Jabouri H, Saif A, Khatab A, Diallo C, Venkatadri U (2024) A critical review of selective maintenance for mission-oriented systems: challenges and a roadmap for novel contributions. *International Journal of Production Research*,62,4980-5015.

Levitin G, Xing L, Dai Y (2021) Mission aborting in n-unit systems with work sharing. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 52, 4875-4886.

Levitin G, Xing L, Dai Y (2021) Joint optimal mission aborting and replacement and maintenance scheduling in dual-unit standby systems. *Reliability Engineering & System Safety*, 216, 107921.

Liu B, Huang H, Deng Q (2022) On optimal condition based task termination policy for phased task systems. *Reliability Engineering & System Safety*, 221, 108338.

Zhao X, Lv Z, Qiu Q, Wu Y (2023) Designing two-level rescue depot location and dynamic rescue policies for unmanned vehicles. *Reliability Engineering & System Safety*, 233, 109119.

Federated Multi-source Domain Adaptation via Barycenter for Fault Diagnosis of Machine Group

Bin Yang*, Yaguo Lei*, Xiang Li*, Naipeng Li*

Key Laboratory of Education Ministry for Modern Design and Rotor-Bearing System, Xi'an Jiaotong University, Xi'an, 710049, China.

Abstract A dynamic barycenter bridging network (DBBN) is proposed to tackle challenges in federated multi-source domain adaptation in fault diagnosis of machine group. These challenges include the neglect of conditional distributions during multi-domain adaptation and the limitations of rough intermedia distribution bridging to address domain gaps. DBBN addresses these issues by introducing a server that collects key distribution parameters, such as statistical metrics, rather than raw data, from multiple domains to ensure data privacy. A dynamic distribution barycenter is computed using fixed-point iteration, representing a common distribution reference point. This barycenter is shared with all domain nodes, guiding the adaptation of local data distributions based on their associated labels. By aligning both marginal and conditional distributions through collaborative training between the server and domain sides, DBBN effectively reduces domain discrepancies. Experimental results on a multi-source transfer fault diagnosis case using machine-used bearings demonstrate DBBN's ability to improve adaptation accuracy and achieve superior fault diagnosis performance.

1. Introduction

Machine groups play a pivotal role in modern industries such as wind farms and flexible production factories. These systems enable collaborative production through multiple interconnected machine nodes, promoting enhanced efficiency and quality in swarm production. Ensuring the reliable operation of these nodes is vital, making intelligent fault diagnosis a cornerstone for maintaining their health and functionality (Lei et al. 2020). The rapid development of deep transfer learning has further revolutionized fault diagnosis by enabling the transfer of knowledge from well-studied machine nodes (source domains) to less-explored nodes (target domains) (Weiss et al. 2016). This eliminates the need for extensive retraining on the target domain and underscores the importance of improving model generalization in fault diagnosis for machine group.

However, machine group nodes often exhibit unique characteristics and varying degradation patterns, creating significant challenges. In many cases, multiple source domains are available, but none individually offer sufficient, highly transferable knowledge to the target domain. Multi-source transfer learning has emerged as a promising solution, focusing on extracting domain-invariant features from multiple sources and applying them to the target domain (Yang et al. 2022b). Techniques such as maximum mean discrepancy (MMD) and generative adversarial networks (GANs) have been employed to address domain discrepancies, enhancing the generalization of diagnosis models across diverse domains.

Despite their effectiveness, existing multi-source transfer methods often assume centralized domain data for distribution adaptation. This assumption is impractical in real-world engineering scenarios due to two primary challenges. First, data privacy concerns hinder data centralization, as monitoring data often contains sensitive information, including production quality metrics and manufacturing processes. This is particularly relevant for machine nodes belonging to different organizations, where sharing sensitive data is restricted by competitive and privacy considerations. Second, high costs associated with data centralization, including the need for high-bandwidth industrial internet and powerful data centers to process large volumes of data, conflict with the cost-efficiency goals of most enterprises.

Federated learning offers a promising alternative, enabling global model training across multiple clients without requiring centralized data (Li et al. 2020). Instead, lightweight, encrypted data—such as model parameters—is aggregated by a central server and shared with clients for further training. This architecture preserves data privacy while reducing transmission and computational costs. Federated learning approaches have recently been explored in multi-source transfer learning for fault diagnosis (Chen et al. 2022; Lu et al. 2022; Yu et al. 2023). For instance, adversarial networks have been employed to align domain distributions (Peng et al. 2019),

while intermediate distributions (Dai et al. 2021) or public datasets (Li and Wang 2019) have been used to bridge domain gaps.

However, existing federated methods face two significant limitations: First, most methods focus on marginal distribution alignment while neglecting conditional distributions in decentralized settings. This oversight is particularly problematic in scenarios involving distinct but related machines, where conditional distribution discrepancies remain unresolved. Second, the intermediate distributions used in prior works are often unrelated to the given domains or fixed, making them unsuitable for dynamic domain adaptation during federated training.

To overcome these challenges, we propose a dynamic barycenter bridging network (DBBN) for federated multi-source transfer fault diagnosis in machine groups. DBBN introduces a federated learning framework that utilizes a central server to collect lightweight distribution parameters (e.g., means and covariances) from multiple source domains instead of raw data. Using these parameters, a dynamic distribution barycenter is calculated through fixed-point iteration and serves as an intermediate distribution to bridge diverse machine nodes. This barycenter is broadcast to all domain nodes, enabling targeted alignment of both marginal and conditional distributions during collaborative training. Unlike the static intermediates used in previous methods, DBBN dynamically adjusts the barycenter throughout the training process, ensuring effective adaptation to domain shifts. The contributions of this work are twofold.

First, we propose a federated multi-source transfer learning framework, called DBBN, to facilitate lightweight and secure diagnosis knowledge transfer in machine groups. DBBN achieves directional alignment of multiple domain distributions by transmitting extremely lightweight data, such as distribution means and covariances, rather than high-dimensional features or raw data. Second, we introduce the Wasserstein barycenter as an intermediate distribution to bridge gaps between domains. By dynamically adjusting the barycenter during training, DBBN ensures its proximity to all domains, improving model generalization and transferability.

2. The proposed Method

2.1. *Multi-source domain feature distribution collection*

The architecture of DBBN is presented in Fig. 1. The function of each part is detailed in the following subsections.

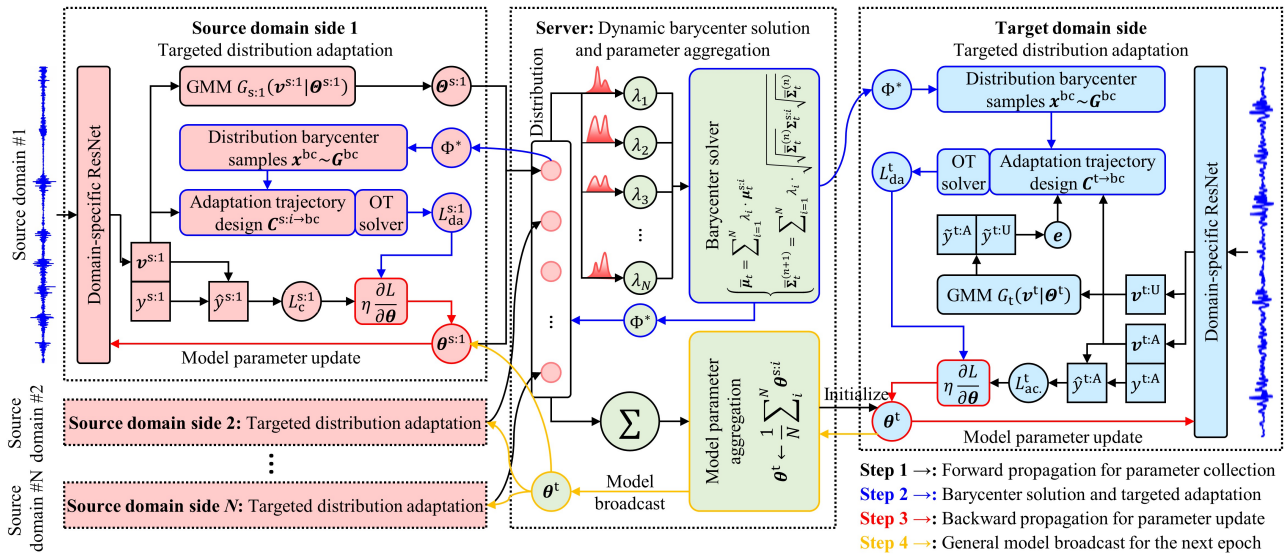


Figure 1. Architecture of the proposed DBBN.

In each domain, a deep residual network (ResNet) is independently designed to extract high-level features from the local data. For the ith source domain, the feature extractor, denoted as $g(\cdot|\theta_f^{s:i})$, is composed of stacked residual blocks. The extracted high-level features, $\boldsymbol{v}^{s:i} = g(\boldsymbol{x}^{s:i})$ are subsequently passed through a fully connected layer, $g(\cdot|\theta_c^{s:i})$, for classification. Since all the source domain samples are labeled, Since all samples in the source domains are labeled, the ResNet for each source domain is trained by minimizing the cross-entropy loss $L_c^{s:i}$.

To capture the distribution characteristics of high-level features in each source domain, a Gaussian Mixture Model (GMM) (Sarkar et al. 2020) is employed. During the GMM fitting process, the number of Gaussian components, $k_{\mathrm{s}:i}$, corresponds to the number of health states $\left|Y^{\mathrm{s}:i}\right|$ in the ith source domain. The GMM parameters, $\Theta^{\mathrm{s}:i} = \{\, \alpha_t^{\mathrm{s}:i}, \mu_t^{\mathrm{s}:i}, \Sigma_t^{\mathrm{s}:i} \big| t = 1,\, 2, \cdots, k_{\mathrm{s}:i}\}$ include the weights, means, and covariances of the Gaussian components. The proposed DBBN framework incorporates a central server that aggregates the model parameters $\Theta^{\mathrm{s}:i} = \Theta_f^{\mathrm{s}:i}, \Theta_c^{\mathrm{s}:i}$ and the feature distribution parameters $\Theta^{\mathrm{s}:i}$ from each source domain.

2.2. *Dynamic distribution barycenter solution*

The distribution barycenter serves as an intermediary in Wasserstein space (Agueh and Carlier 2011), aiming to approximate the feature distributions of all source domains simultaneously. This involves solving a barycenter task for multiple GMMs, as each source domain transmits its distribution through GMM parameters (Delon and Desolneux 2020). After the server collects the multi-source distribution parameters, the barycenter computation is divided into subtasks based on label information. Each subtask aggregates Gaussian models corresponding to the same health state to generate sub-category barycenter, incorporating label information into the overall barycenter. Given the multi-source GMM parameters $\{\Theta^{\mathrm{s}:i} | i = 1, 2, \cdots, N\}$, the sub-category Gaussian model $\psi_t^{\mathrm{s}:i} \sim N(\mu_t^{\mathrm{s}:i}, \Sigma_t^{\mathrm{s}:i})$ represents the distribution of the t-th health state. The resulting distribution barycenter Ξ^* is defined as a collection of these sub-category barycenter.

$$\Xi := \left\{ \inf_{\varphi_t \in \Pi(\Psi_t^{\mathrm{s}:1}, \cdots, \mathrm{s}:1_t)} \sum_{i=1}^{N} \lambda_i W_2^2(\Psi_t^{\mathrm{s}:i}, \varphi_t) \middle| t = 1, 2, \cdots \right\} \tag{1}$$

The sub-category barycenter, φ_t, is computed using weights λ_i for each source domain, where $\lambda_i \in [0,1]$ and $\sum_i \lambda_i = 1$. A higher weight brings the barycenter closer to the corresponding local distribution of source domain. $\Pi(\psi_t^{\mathrm{s}:1}, \cdots, \psi_t^{\mathrm{s}:N})$ represents the set of probability distributions with marginals $\psi_t^{\mathrm{s}:1}, \cdots, \psi_t^{\mathrm{s}:N}$, and $W_2(\cdot)$ measures the Wasserstein distance between φ_t and $\psi_t^{\mathrm{s}:i}$. The distribution barycenter is derived from the multi-source domains, with the weights controlling the similarity to each domain. The barycenter Φ^* could be computed by following the fix-point iteration (ÁLvarez-Esteban et al. 2016).

2.3. *Targeted adaptation with distribution barycenter*

In each domain, intermediate samples are randomly generated according to the barycenter distribution parameters. The dataset X^{bc} contains n^{bc} labeled virtual samples, drawn from the Gaussian mixture distribution $G^{\mathrm{bc}} = \sum_t N\left(\bar{\mu}_t, \bar{\Sigma}_t\right)$, where $t \in Y^{\mathrm{bc}}$. The marginal and conditional distribution discrepancies across multiple domains are then minimized using a targeted adaptation approach, with the adaptation trajectory guided by the associated labels.

1) In each source domain sides: For the ith source domain, the high-level features $\boldsymbol{v}^{\mathrm{s}:i}$ are obtained from local samples. A relationship matrix $\boldsymbol{A} \in \{0,1\}^{n^{\mathrm{s}:i} \times n^{\mathrm{bc}}}$ is created to indicate whether a source sample and an intermediate sample share the same label, while 0 represents no match. The cost function could be designed as follows.

$$C_{j,k}^{\mathrm{s}:i \to \mathrm{bc}} = \begin{cases} \left\| \boldsymbol{v}_j^{\mathrm{s}:i} - \boldsymbol{x}_k^{\mathrm{bc}} \right\|, & if\ a_{j,k} = 1 \\ \varepsilon, & if\ a_{j,k} = 0 \end{cases} \tag{2}$$

where $a_{j,k} \in \boldsymbol{A}$ indicates the label relationship between the jth source domain sample and kth intermedia sample. ε is constant, and $\varepsilon \to +\infty$ if the sample couple has different labels. In this article, ε is set as 10. Based on Eq. 2, the optimal transport (OT) divergence is calculated as follows (Peyré et al. 2019).

$$L_{\mathrm{da}}^{\mathrm{s}:i} = OT(P^{\mathrm{s}:i} \| G^{\mathrm{bc}}) = \sum_j \sum_k \Gamma_{j,k}^{\mathrm{s}:i \to \mathrm{bc}} \cdot C_{j,k}^{\mathrm{s}:i \to \mathrm{bc}} \tag{3}$$

where $\Gamma_{j,k}^{\mathrm{s}:i \to \mathrm{bc}}$ is the transport policy from the source domain to the barycenter, and it could be optimized by the Sinkhorn-Knopp fixed point iteration algorithm (Cuturi 2013).

2) In the target domain side: The diagnosis model is initialized using FedAvg, averaging the parameters of

multi-source domain models (McMahan et al. 2017). The general model is trained on labeled target domain anchors with the loss function:

$$L_{\text{anchor}}^{\text{t}} = -\frac{1}{|X^{\text{t:A}}|}\sum_{i=1}^{|X^{\text{t:A}}|}\sum_{j}^{R} I(y_i^{\text{t:A}} = j) + \max\left\{\frac{1}{n^{\text{bc}}}\sum_j b_{j,i}\cdot\left\|\boldsymbol{x}_j^{\text{bc}} - \boldsymbol{v}_i^{\text{r:A}}\right\|_2^2\right\} \tag{4}$$

where $b_{j,i}$ indicates the label relationship between the jth intermedia samples $\boldsymbol{x}_j^{\text{bc}}$ and the high-level features of the ith labeled sample $\boldsymbol{v}_i^{\text{t:A}}$, and it is defined by

$$b_{j,i} = \begin{cases} 0, & if\ I(y_j^{\text{t:A}} \neq y_i^{\text{bc}}) \\ 1, & if\ Otherwise \end{cases} \tag{5}$$

Given limited labeled target samples, anchors are used to guide the adaptation trajectory. The relationship between anchor and intermediate sample labels is propagated to unlabeled samples through cluster analysis. Target domain samples are fitted with a GMM $G_t(\cdot|\Theta^t)$, where the number of Gaussians equals the number of health states. Pseudo-labels for unlabeled samples are predicted based on their cluster alignment with anchors:

$$e_{k,j} = \begin{cases} 0, & if\ I(\widetilde{y}_k^{\text{t:U}} \neq \widetilde{y}_j^{\text{t:A}}) \\ 1, & if\ Otherwise \end{cases} \tag{6}$$

The cost function for target domain adaptation is defined as:

$$C_{k,i}^{\text{t}\to\text{bc}} = \begin{cases} \left\|\boldsymbol{v}_k^{\text{t:}U} - \boldsymbol{x}_i^{\text{bc}}\right\|, & if\ e_{k,j}\cdot b_{j,i} = 1 \\ \varepsilon, & if\ e_{k,j}\cdot b_{j,i} = 0 \end{cases} \tag{7}$$

From this, the OT divergence is calculated as:

$$L_{\text{da}}^{\text{t}} = OT(P^{\text{t}}\|G^{\text{bc}}) = \sum_k\sum_i \Gamma_{k,i}^{\text{t}\to\text{bc}}\cdot C_{k,i}^{\text{t}\to\text{bc}} \tag{8}$$

where $\Gamma_{k,i}^{\text{t}\to\text{bc}}$ is determined by solving the OT problem (Peyré et al. 2019).

2.4. *Collaboratively training among multiple domains*

The proposed DBBN is trained through collaborative interactions between the server and multiple domain sides, which is described as follows:

Step 1: The server collects the distribution parameters of high-level features from multiple source domains.

Step 2: Using the collected parameters, the distribution barycenter is computed according to Eq. 1. The barycenter parameters are then broadcast to each domain side for designing the adaptation trajectory.

Step 3: On the source domain side, intermediate samples are generated based on the received barycenter parameters. The local diagnosis model is trained by minimizing:

$$\min_{\theta^{\text{s:}i}} L_{\text{c}}^{\text{s:}i} + \alpha\cdot L_{\text{da}}^{\text{s:}i} \tag{9}$$

where α is the tradeoff parameter balancing classification loss and targeted distribution adaptation.

Step 4: The server aggregates the diagnosis models from multiple source domains by averaging the model parameters:

$$\theta^{\text{t}} \leftarrow \frac{1}{N}\sum_i^N \theta^{\text{s:}i} \tag{10}$$

The aggregated parameters are sent to the target domain side for initializing the diagnosis model. The general model is further trained using target domain samples by minimizing:

$$\min_{\theta^{\text{t}}} L_{\text{anchor}}^{\text{t}} + \beta\cdot L_{\text{da}}^{\text{t}} \tag{11}$$

Table 1. Details of multiple machine-used bearing datasets

Datasets	Designations	Sample amount	Working conditions	Sampling frequency
A	ER-8K	2048 (512×4)	2100 r/min	12.8 kHz
B	NSK/NFC 6205	2184 (546×4)	1500 r/min, 10 N·m	25.6 kHz
C	ER-16K	2160 (540×4)	1800 r/min	12.8 kHz
D	552732QT	1092 (273×4)	490-530 r/min	12.8 kHz

where β is the tradeoff parameter. In the next training epoch, the updated general model parameters are transmitted back to the source domain sides via the server to continue the collaborative training process.

3. Case Study

3.1. *Dataset description*

The proposed DBBN is demonstrated on four bearing datasets, which are collected from different machines. The datasets are detailed in Table 1.

Dataset A originates from a rotor-bearing test rig (Yang et al. 2022a). Bearings (ER-8K) with four health conditions—normal (N), inner race fault (IF), outer race fault (OF), and roller fault (RF)—are installed on the left end of the rotor-bearing system. The motor operates at 2100 r/min, and vibration signals are sampled at 12.8 kHz, resulting in 2048 samples per condition, balanced across all health states.

Dataset B is collected from a motor-driven-end bearing test rig, illustrated in Fig. 2. Data are acquired at a motor speed of 1500 r/min and an output torque of 10 N·m (provided by a magnetic eddy brake). This dataset also includes four bearing health conditions, with 546 samples for each state.

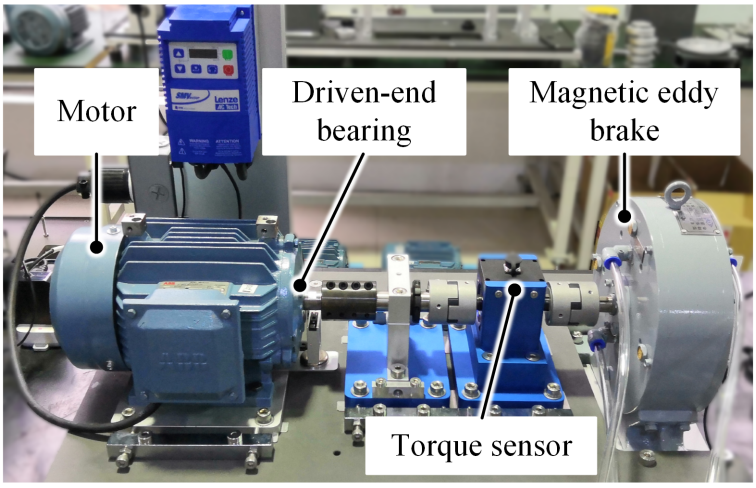


Figure 2. Picture of the rotor-bearing test rig.

Dataset C comes from a water pump test rig, as shown in Fig. 3. Bearings are positioned between the motor and the water pump. At a motor speed of 1800 r/min and a sampling frequency of 12.8 kHz, an accelerometer mounted on the bearing housing collects vibration data. The dataset contains 2160 samples, covering the same four bearing states: N, IF, OF, and RF.

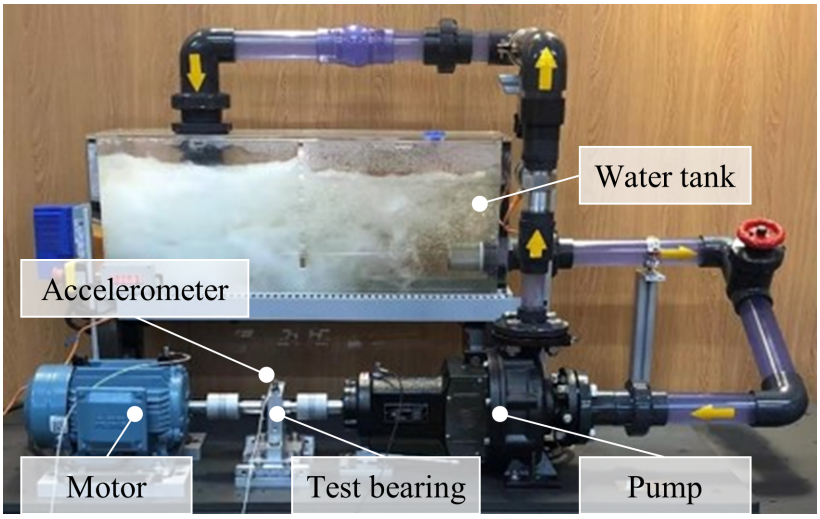


Figure 3. Picture of the water pump test rig.

Dataset D is derived from a locomotive bearing test rig driven by a hydraulic motor (Yang et al. 2022a). During testing, the inner race rotates at approximately 500 r/min, and vibration signals are sampled at 12.8 kHz.

Using these four datasets, we design various multi-source transfer diagnosis tasks to evaluate the performance of the proposed DBBN in the machine group. The datasets, collected from different machines, exhibit discrepancies not only between the source and target domains but also among the multiple source domains.

3.2. *Diagnosis results and comparisons*

Based on the datasets in Table I, four multi-source transfer diagnosis tasks were designed to evaluate the proposed DBBN. For each task, 25% of target domain samples were randomly selected for training, with the remainder used for testing. In the training set, only one labeled sample per health state was included to guide the adaptation trajectory. The DBBN was run ten times under optimal hyperparameters, and the mean and standard deviation of test accuracy were recorded in Table II. DBBN achieved an average accuracy of 97.82%, with a maximum of 99.29%, demonstrating its effectiveness in federated multi-source transfer diagnosis.

The proposed method was compared to data-centralized methods (Only-T, S+T, DeepJDOT (Damodaran et al. 2018), and MDAN(Zhao et al. 2018)) and data-decentralized methods (FedAvg (McMahan et al. 2017), Finetune, FADA (Peng et al. 2019), and FTL-PD (Zhang and Li 2022)). Only-T, using only four target domain samples, suffered from overfitting, achieving 56.85% accuracy. S+T ignored inter-domain discrepancies and yielded 42.54% accuracy. DeepJDOT reduced joint distribution discrepancies but failed to address multi-source discrepancies, performing poorly on tasks (A,B,D)→C and (A,B,C)→D. MDAN used adversarial learning to adapt source-target pairs, achieving 64.33% accuracy.

Data-decentralized methods generally outperformed centralized ones due to federated training, which mitigates multi-source discrepancies. FedAvg, unable to address cross-domain gaps, achieved the lowest accuracy (33.67%). Finetune used limited labeled target samples but showed minimal improvement. FADA reduced cross-domain discrepancies via adversarial learning but struggled with severe multi-domain differences, achieving 62.67% accuracy. FTL-PD generated intermediate distributions but faced challenges in simultaneous domain adaptation.

In contrast, DBBN dynamically calculated a distribution barycenter across source domains, simplifying multi-domain adaptation. By leveraging OT theory, it optimized adaptation trajectories, achieving superior diagnostic accuracy across tasks.

4. Conclusions

This article introduces the Dynamic Barycenter Bridging Network (DBBN) for federated multi-source transfer fault diagnosis in machine groups, where different but related machines act as nodes. DBBN aims to integrate knowledge from multiple source nodes and transfer it effectively to the target node. To achieve this, the distribution barycenter is employed as an intermediary to bridge the target domain and multiple source domains.

Table 2. Diagnosis accuracy (%) by using different methods on multi-source transfer diagnosis tasks [mean ± standard deviation]

Category	Methods	Multi-source transfer diagnosis tasks				Average
		(B,C,D)→A	(A,C,D)→B	(A,B,D)→C	(A,B,C)→D	
Data centralization	Only-T	59.59± 2.90	58.40± 5.45	55.14± 8.48	54.27± 6.29	56.85
	S+T	66.05± 4.84	58.59± 8.51	26.06± 1.40	19.47± 3.98	42.54
	DeepJDOT	80.40± 18.3	67.28± 22.7	27.28± 7.99	32.75± 3.98	51.93
	MDAN	71.82± 10.9	89.10± 13.7	51.09± 3.59	45.31± 6.56	64.33
Data decentralization	FedAvg	58.03± 3.20	37.72± 10.9	25.79± 1.58	13.14± 3.12	33.67
	Finetune	65.42± 4.61	54.30± 8.36	32 .28± 4.76	31.53± 5.59	45.88
	FADA	71.07± 3.16	76.72± 4.39	49.48± 3.20	53.41± 4.29	62.67
	FTL-PD	77.90± 5.30	71.15± 6.69	44.35± 4.72	48.63± 2.56	60.51
	DBBN (Ours)	99.28± 0.72	99.91± 0.17	99.19± 0.68	92.92± 1.81	97.82

The DBBN framework utilizes a central server to aggregate light-encrypted key parameters, rather than raw data or high-level features, for barycenter computation. The dynamically generated barycenter, derived from multiple source domains, can be manually adjusted and is broadcast back to the nodes for targeted distribution adaptation using local data. Guided by a carefully designed adaptation trajectory, feature distributions are aligned toward the common barycenter. Through collaborative training between the central server and individual domain nodes, the model ensures consistent feature distributions across source and target domains, achieving high diagnostic accuracy under decentralized data settings.

References

Agueh, M., Carlier, G., 2011. Barycenters in the wasserstein space. SIAM Journal on Mathematical Analysis 43, 904–924.

Álvarez-Esteban, P.C., Del Barrio, E., Cuesta-Albertos, J., Matrán, C., 2016. A fixed-point approach to barycenters in wasserstein space. Journal of Mathematical Analysis and Applications 441, 744–762.

Chen, J., Li, J., Huang, R., Yue, K., Chen, Z., Li, W., 2022. Federated transfer learning for bearing fault diagnosis with discrepancy-based weighted federated averaging. IEEE Transactions on Instrumentation and Measurement 71, 1–11.

Cuturi, M., 2013. Sinkhorn distances: Lightspeed computation of optimal transport. Advances in neural information processing systems 26.

Dai, Y., Liu, J., Sun, Y., Tong, Z., Zhang, C., Duan, L.Y., 2021. Idm: An intermediate domain module for domain adaptive person re-id, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 11864–11874.

Damodaran, B.B., Kellenberger, B., Flamary, R., Tuia, D., Courty, N., 2018. Deepjdot: Deep joint distribution optimal transport for unsupervised domain adaptation, in: Proceedings of the European conference on computer vision (ECCV), pp. 447–463.

Delon, J., Desolneux, A., 2020. A wasserstein-type distance in the space of gaussian mixture models. SIAM Journal on Imaging Sciences 13, 936–970.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., Nandi, A.K., 2020. Applications of machine learning to machine fault diagnosis: A review and roadmap. Mechanical systems and signal processing 138, 106587.

Li, D., Wang, J., 2019. Fedmd: Heterogenous federated learning via model distillation. arXiv preprint arXiv:1910.03581 .

Li, T., Sahu, A.K., Talwalkar, A., Smith, V., 2020. Federated learning: Challenges, methods, and future directions. IEEE signal processing magazine 37, 50–60.

Lu, S., Gao, Z., Xu, Q., Jiang, C., Zhang, A., Wang, X., 2022. Class-imbalance privacy-preserving federated learning for decentralized fault diagnosis with biometric authentication. IEEE Transactions on industrial informatics 18, 9101–9111.

McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A., 2017. Communication-efficient learning of deep networks from decentralized data, in: Artificial intelligence and statistics, PMLR. pp. 1273–1282.

Peng, X., Huang, Z., Zhu, Y., Saenko, K., 2019. Federated adversarial domain adaptation. arXiv preprint arXiv:1911.02054 .

Peyré, G., Cuturi, M., et al., 2019. Computational optimal transport: With applications to data science. Foundations and Trends® in Machine Learning 11, 355–607.

Sarkar, S., Melnykov, V., Zheng, R., 2020. Gaussian mixture modeling and model-based clustering under measurement inconsistency. Advances in Data Analysis and Classification 14, 379–413.

Weiss, K., Khoshgoftaar, T.M., Wang, D., 2016. A survey of transfer learning. Journal of Big data 3, 1–40.

Yang, B., Lei, Y., Li, X., Roberts, C., 2022a. Deep targeted transfer learning along designable adaptation trajectory for fault diagnosis across different machines. IEEE Transactions on Industrial Electronics 70, 9463–9473.

Yang, B., Xu, S., Lei, Y., Lee, C.G., Stewart, E., Roberts, C., 2022b. Multi-source transfer learning network to complement knowledge for intelligent diagnosis of machines with unseen faults. Mechanical Systems and Signal Processing 162, 108095.

Yu, Y., Guo, L., Gao, H., He, Y., You, Z., Duan, A., 2023. Fedcae: A new federated learning framework for edge-cloud collaboration based machine fault diagnosis. IEEE Transactions on Industrial Electronics 71, 4108–4119.

Zhang, W., Li, X., 2022. Data privacy preserving federated transfer learning in machinery fault diagnostics using prior distributions. Structural Health Monitoring 21, 1329–1344.

Zhao, H., Zhang, S., Wu, G., Moura, J.M., Costeira, J.P., Gordon, G.J., 2018. Adversarial multiple source domain adaptation. Advances in neural information processing systems 31.

Optimizing budget allocation for multimission selective maintenance planning

F. Falahaty*, R. O'Neil*, Z. Liu**, A. Khatab*,***, C. Diallo*

*Dalhousie University, Industrial Engineering, Halifax, Nova Scotia, Canada, {fr411343, ry855314, cd}@dal.ca

**Cape Breton University, Shannon School of Business, Sidney, Nova Scotia, Canada, Zhuojun_Liu@cbu.ca

*** LGIPM, University of Lorraine, Metz, France, abdelhakim.khatab@univ-lorraine.fr

Abstract Mission-critical systems in sectors such as aerospace, defence, transportation, petrochemistry, and power generation require high reliability to prevent failures causing major economic losses, environmental damages, and safety risks. For such systems, solving the selective maintenance problem (SMP) yields optimal maintenance planning decisions during scheduled breaks. Its extension, the multi-mission SMP (MMSMP), focuses on optimizing component maintenance, maintenance levels, and repairperson assignments over multiple consecutive missions interspersed with maintenance breaks. While recent advances integrate predictive, resource-constrained, and fleet-wide strategies, they rely on the unrealistic assumption of fixed budgets, ignoring the reality of fluctuating and tight financial constraints faced by planners. This study investigates how different maintenance budget allocations across missions affect system performance. Using a two-phase decomposition model and binary integer programming, it explores various budget distribution strategies: uniform, linearly increasing, and inverted-V. The goal is to determine how allocating resources differently across missions can enhance asset reliability within fixed budget limits. The findings aim to guide maintenance planners in making budget decisions to improve overall system reliability while balancing resource constraints.

1. Introduction

Many engineered systems operate in cycles of missions and breaks. These mission-oriented systems (MOS) are common in many industries such as airlines, transportation, nuclear energy production, mining and military operations. For example, truck-drone tandems used for package deliveries work during the day and undergo maintenance overnight. Fighter jets undergo maintenance between two flying missions. The selective maintenance problem (SMP) has been widely investigated and used to determine maintenance planning for such MOS. The SMP focuses on selecting a subset of components for repair or replacement during constrained maintenance breaks between consecutive missions (Rice et al., 1998).

Since its introduction, the SMP framework has been extensively extended to address a wide range of practical and theoretical complexities. These include general system configurations beyond series structures (Diallo et al. 2018), multistate component models (Liu et al., 2018), component dependence and failure propagation, fleet-level maintenance coordination, environments with multiple concurrent repair channels, and uncertain mission times, system downtime, and maintenance durations (Al-Jabouri et al., 2024). Additional developments have incorporated probabilistic maintenance quality and hybrid models integrating selective maintenance with combinatorial optimization problems, such as orienteering (O'Neil et al., 2023b).

SMP can feature a single mission or multiple missions, which is less covered in the academic literature although more prevalent in practice. Furthermore, traditional SMP models assume static budget allocation and synchronized maintenance actions, which may not adapt to changing resource availability (Chaabane et al., 2020). Multimission SMP (MMSMP) involves multiple intermission breaks for scheduled maintenance. The budget available at each break may not be identical, thus impacting the maintenance activities to be carried out, which in turn affect the system reliability achieved in the next mission. Thus, this study aims to investigate budget allocation strategies and their impact on the system reliability of multi-mission MOS. Three budget distribution strategies are considered: uniform, linearly increasing, and inverted-V shape.

The remainder of this paper is structured as follows. System description and the optimization model are described in Sections 2 and 3 respectively. Section 4 presents the results of the numerical experiments conducted. Conclusions and extensions are presented in Section 5.

2. System description, reliability and cost computations

2.1. *Notation and description*

$\mathbb{I}$	Set of subsystems, $\mathbb{I} = \{1,2,...,N\}$ indexed by i
$\mathbb{J}_i$	Set of components in subsystem i, $\mathbb{J}_i = \{1,2,...,N_i\}$ indexed by j
E_{ij}	The jth component of subsystem i
$\mathbb{L}_{ij}$	Set of PM levels available for component E_{ij}, $\mathbb{L}_{ij} = \{0,1,...,L_{ij}\}$ indexed by l
$\mathbb{M}$	Set of missions, $\mathbb{M} = \{1,2,...,M\}$ indexed by m
$\mathbb{P}_i$	Set of maintenance patterns generated for subsystem i, $\mathbb{P}_i = \{1,2,...,P_i\}$ indexed by p
U_m	Duration of mission m
D_m	Duration of break m
B	Total maintenance budget available over all missions
B_m	Maintenance budget available for break and mission m
X_{ijm}	Age of component E_{ij} at the start of break m
Y_{ijm}	Age of component E_{ij} at the end of break m
$\lambda_{ij}(t)$	Failure rate function of component E_{ij}
t_{ijl}^{PM}	Duration of PM level l on component E_{ij}
c_{ijl}^{PM}	Cost of PM level l on component E_{ij}
c_{ij}^{mr}	Cost of minimal repair on component E_{ij}
C_{ipm}	Expected total maintenance cost for subsystem i during break and mission m in pattern p
K_i	Minimum number of components that must be functioning in subsystem i
R_{ij}^c	Conditional reliability of component E_{ij} during mission m
R_{im}^s	Reliability of subsystem i during mission m
R_m	Overall system reliability during mission m

This paper investigates the MMSMP for a multicomponent system operating over M missions. The system has N subsystems, each being a K_i-out-of-N_i configuration (*i.e.,* subsystem i works if at least K_i out of its N_i components work). After each mission m of length U_m, a maintenance break of duration D_m occurs, allowing PM actions to be performed on each component E_{ij}. During a mission, minimal repair is performed at failure.

2.2. *System Reliability Computation*

As shown in Figure 1, the recursive relationship given by $X_{ij,m+1} = Y_{ijm} + U_m, \quad \forall m \in \mathbb{M} \setminus \{M\}$ links the components ages before and after breaks. Given that component E_{ij} has effective age Y_{ijm} at the end of maintenance break m, its reliability for the next mission of duration U_m is:

$$R_{ijm}^c = \frac{R_{ij}(Y_{ijm} + U_m)}{R_{ij}(Y_{ijm})}, \tag{1}$$

where $R_{ij}(t)$ refers to the unconditional reliability function of component E_{ij}. Thus, the reliability R_{im}^s of the ith subsystem is obtained from the exact formulation proposed in Arulmozhi (2002).

$$R_{im}^s = \sum_{j_{k_i}=1}^{N_i} \sum_{j_{k_i-1}=1}^{j_{k_i}-1} \cdots \sum_{j_1=1}^{j_2-1} \left(\prod_{v=j_1}^{j_{k_i}} R_{ivm}^c \right) \cdot \prod_{\substack{u=1 \\ u \neq j_1,\cdots,j_{k_i}}}^{j_{k_i}} (1 - R_{ium}^c). \tag{2}$$

Subsystems are in series, hence the overall system reliability for mission m is given by:

$$R_m = \prod_{i \in \mathbb{I}} R_{im}^s. \tag{3}$$

2.3. *Maintenance levels and maintenance patterns*

For each component E_{ij}, there is a list $\mathbb{L}_{ij} = \{0, \cdots, L_{ijk}\}$ of PM levels that can be selected during breaks. Do-Nothing ($l = 0$) refers to no maintenance performed on the component. The perfect repair level ($l = L_{ij}$) resets the component's age to 0, while an IM level $0 < l < L_{ij}$ returns the component's age between its age at

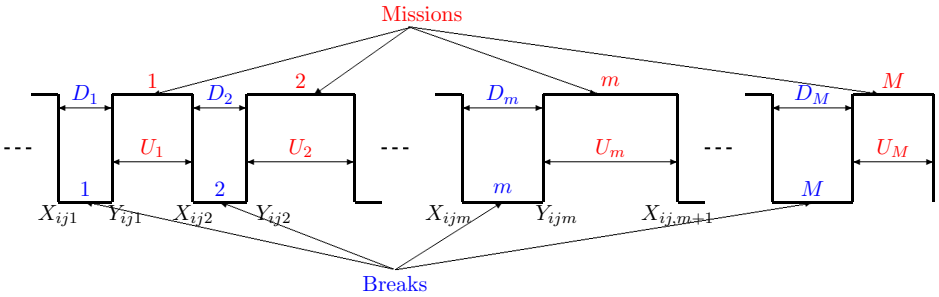


Figure 1. Alternating sequence of missions and maintenance breaks (O'Neil et al., 2023a).

the start of the break and 0. When carried out on a component E_{ij}, a PM of level l requires t^P_{ijl} time units, and costs c^P_{ijl} monetary units. Using the age reduction by Malik (1979), each IM level l available for component E_{ij} is characterized by an age reduction coefficient γ_{ijl} ($0 \leq \gamma_{ijl} \leq 1$). Thus, when level l PM is performed on E_{ij}, its effective age becomes $Y_{ijm} = \gamma_{ijl} \cdot X_{ijm}$.

The proposed formulation relies on a full enumeration of all maintenance patterns for each subsystem i as in O'Neil et al. (2023a). A maintenance pattern $p \in \mathbb{P}_i$ is represented as a matrix $\mathbf{A}_{ip}$ of M rows and N_i columns, where element a^i_{mj} is the maintenance level performed on the jth component of the ith subsystem during maintenance break m.

$$\mathbf{A}_{ip} = \begin{bmatrix} a^i_{11} & \cdots & a^i_{1j} & \cdots & a^i_{1N_i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a^i_{m1} & \cdots & a^i_{mj} & \cdots & a^i_{mN_i} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a^i_{M1} & \cdots & a^i_{Mj} & \cdots & a^i_{MN_i} \end{bmatrix} \tag{4}$$

To illustrate the generation of maintenance patterns, let us consider a 1-out-of-2:G subsystem (*i.e.,* $K_i = 1$ and $N_i = 2$) with two maintenance levels available: Do nothing ($l = 0$) and perfect repair ($l = 1$). If this subsystem performs a sequence of two missions alternating with two breaks, the 16 possible maintenance patterns are:

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \cdots \quad , \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

where the third pattern means that component E_{i1} is perfectly repaired during both breaks. The fourth pattern means that E_{i1} is replaced during the first break, and E_{i2} is replaced during the second break.
The following auxiliary variable x^p_{ijlm} is defined as follows:

$$x^p_{ijlm} = \begin{cases} 1, & \text{if } a^i_{mj} = l \text{ under pattern } p \\ 0, & \text{otherwise.} \end{cases}$$

2.4. *Maintenance cost and time computation*

The total expected maintenance cost per subsystem i per mission m in pattern p denoted C_{ipm} is the sum of the cost of PM actions performed during break m and the cost of minimal repairs carried out at failure during mission m on all components of subsystem i. Thus,

$$C_{ipm} = \sum_{j \in J_i} \left(\sum_{l \in L_{ij}} c^{PM}_{ijl} \cdot x^p_{ijlm} + c^{mr}_{ij} \int_{Y_{ijm}}^{Y_{ijm}+U_m} \lambda_{ij}(t)\, dt \right).$$

It is assumed that the duration of minimal repair is negligible compared to mission length. Thus, the total maintenance time per subsystem i during break m in pattern p denoted T_{ipm} is given by:

$$T_{ipm} = \sum_{j \in J_i} \sum_{l \in L_{ij}} t^{PM}_{ijl} \cdot x^p_{ijlm}.$$

3. Mathematical Model

A binary decision variable z_{ip} such that $z_{ip} = 1$ when pattern p is selected for subsystem i, and $z_{ip} = 0$ otherwise, is introduced to formulate the selection of maintenance patterns for all subsystems. For each subsystem i, all feasible maintenance patterns P_i are enumerated and associated cost C_{ipm}, time T_{ipm}, and system reliability R^s_{ipm} computed. The MMSMP is then given by:

$$\max \quad \sum_{m\in\mathbb{M}}\sum_{i\in\mathbb{I}} \ln\left(\sum_{p\in\mathbb{P}_i} R^s_{ipm}\cdot z_{ip}\right) \tag{5}$$

$$\text{s.t.} \quad \sum_{p\in\mathbb{P}_i} z_{ip} = 1, \qquad \forall i\in\mathbb{I} \tag{6}$$

$$\sum_{i\in\mathbb{I}}\sum_{p\in\mathbb{P}_i}\sum_{m\in\mathbb{M}} C_{ipm}\cdot z_{ip} \le B_m \tag{7}$$

$$\sum_{i\in\mathbb{I}}\sum_{p\in\mathbb{P}_i} T_{ipm}\cdot z_{ip} \le D_m, \qquad \forall m\in\mathbb{M} \tag{8}$$

$$z_{ip}\in\{0,1\}, \qquad \forall i\in\mathbb{I},\ p\in\mathbb{P}_i \tag{9}$$

The objective function in (5) maximizes the logarithm of the system reliability, which is equivalent to maximizing the system reliability (Diallo et al., 2018). Constraint (6) ensures that only one pattern is selected per subsystem. Constraints (7) and (8) guarantee that the maintenance costs and times do not exceed the maintenance budget and break duration available per mission and break m respectively.

3.1. *Uniform budget allocation (UNIBA)*

When the overall budget B is allocated uniformly over all M missions, the budget per mission m is $B_m = \frac{B}{M}$.

3.2. *Linearly increasing budget allocation (LIBA)*

The linearly increasing budget is in the form $y = ax + b$, which can be written as $B_m = a(m-1) + b$ where a is the slope and b is the intercept with the y-axis (*i.e.,* $B_1 = b$). $\sum_m B_m = B$ and $B_m \ge 0\ \ \forall m$ leads to:

$$-\frac{2B}{M(M-1)} \le a \le \frac{2B}{M(M-1)}.$$

The optimal values of a and b are determined to maximize the system reliability obtained.

3.3. *Inverted V-shaped budget allocation (IVBA)*

Here, mission budgets are set to follow an inverted V pattern: budgets for missions $1,2,\ldots,M$ are set such that the first and last missions receive lower budgets, while the intermediate missions receive higher budgets (*i.e.,* $B_1 = B_M < B_2,\ldots,B_{M-1}$). The overall budget constraint is maintained by ensuring that $\sum_{m=1}^{M} B_m = B$.

4. Numerical Experiments

The three modeled budget allocations are used on a MMSMP with $M = 5$ missions, each with duration $U_m =$ 100 t.u. and break duration $D_m = 60$ t.u. The system considered has $I = 4$ subsystems with $N_1 = N_3 = 3$, and $N_2 = N_4 = 2$ components. Three PM levels are available: 0 (Do-Nothing), 1 (Imperfect), and 2 (Replacement).

Table 1 lists the Weibull shape and scale parameters (β_{ij}, η_{ij}) of the component lifetimes, along with other parameter values. The optimization model is run on an Intel™i5 2.9GHz desktop computer with 12GB of RAM running Windows 11™. All algorithms were coded in Python 3.9. The optimization runs were carried out by Gurobi 12.1 using gurobipy. The model is run with $B \in \{200, 230, 250, 280, 300, 325, 350, 375, 400, 439.28\}$.

Table 2 gives the Reliability per budget used ratio and the overall system reliability obtained after 5 missions for the three budget allocations under consideration (UNIBA, LIBA, and IVBA) under varying budget values. LIBA achieves the highest system reliability across the highest number of budget values. The IVBA yields better reliability per budget used. Overall, UNIBA is the worst performing allocation. As the system components are increasing failure rate (IFR), it is reasonable to expect more maintenance to be done over time, thus

Table 1: Component Parameters

Components	E_{11}	E_{12}	E_{13}	E_{21}	E_{22}	E_{31}	E_{32}	E_{33}	E_{41}	E_{42}
η_{ij}	250	380	280	400	280	340	260	280	260	350
β_{ij}	1.5	2.4	1.6	2.6	1.5	2.4	2.5	2.0	1.2	1.4
X_{ij1}	110	150	170	120	180	100	130	170	150	120
t_{ij1}^{PM}	5.3	3.0	4.5	2.3	5.3	3.0	4.5	2.3	5.3	3.0
t_{ij2}^{PM}	6.0	3.8	5.3	3.0	6.0	3.8	5.3	3.0	6.0	3.8
c_{ij1}^{PM}	6	6	6	5	5	6	6	6	5	5
c_{ij2}^{PM}	8	8	8	7	7	8	8	8	7	7
c_{ij}^{mr}	5	5	5	4	4	5	5	5	4	4

Table 2: System reliability and efficiency comparison across budget allocation strategies.

	Reliability to Budget Used Ratio (%/$)			**Overall Reliability (%)**		
Budget ($)	**IVBA**	**UNIBA**	**LIBA**	**IVBA**	**UNIBA**	**LIBA**
200	27.09	27.01	**28.62**	51.60	51.92	**56.68**
230	27.72	28.36	**28.51**	60.68	60.75	**62.63**
250	**27.52**	27.31	27.47	65.85	65.71	**66.62**
300	24.79	**25.00**	24.69	71.39	70.75	**71.94**
325	**23.54**	23.08	23.12	73.51	73.68	**73.91**
350	**21.91**	21.83	21.82	75.17	**74.82**	75.21
400	**19.63**	19.43	19.59	**76.51**	76.32	**76.51**
450	**18.21**	17.60	17.60	77.09	**77.30**	**77.30**
500	**17.60**	**17.60**	**17.60**	**77.30**	**77.30**	**77.30**

resulting in increasing costs. Hence, the LIBA allocation is the best performer overall. Another result is that, as expected, increasing overall budget B yields higher overall reliability. Figure 2 depicts the budget utilization efficiency across the three allocation strategies. Although all three strategies comply with the overall budget constraint, their effectiveness in converting available resources into reliability outcomes varies notably. Under the reliability per budget used ratio metric, LIBA performs better for low budget values, whereas the IVBA performs better to less constrained budget values. The UNIBA is rarely the best strategy. This is also confirmed by the plots on Figure 3 which show the differences between the Uniform and both LIBA and IVBA.

5. Conclusions

This study evaluated three budget allocation strategies (*Linearly increasing*, *Inverted V*, and *Uniform*) within a MMSMP framework. The preliminary numerical results across various budget levels showed that the *Linearly increasing* strategy consistently outperforms the other two in terms of both total system reliability and budget utilization efficiency. Extensive testing of the budget allocations over key parameters based on design of experiments would allow to draw conclusive insights for maintenance planners and decision makers. Future extensions will include more budget allocation schemes such as Inverted-U, and other non-decreasing profiles.

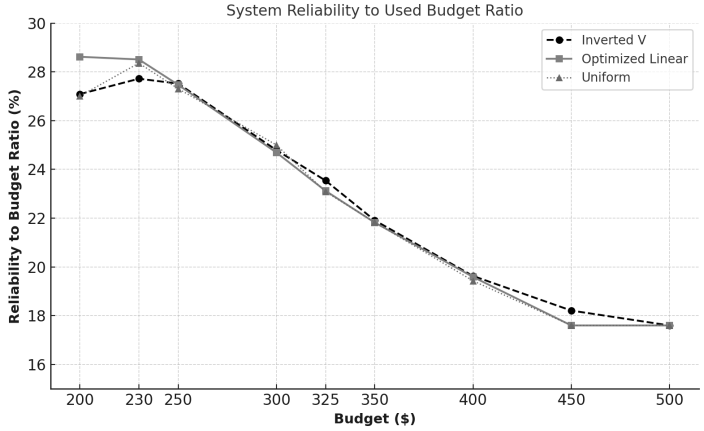


Figure 2. Budget usage efficiency comparison by budget levels

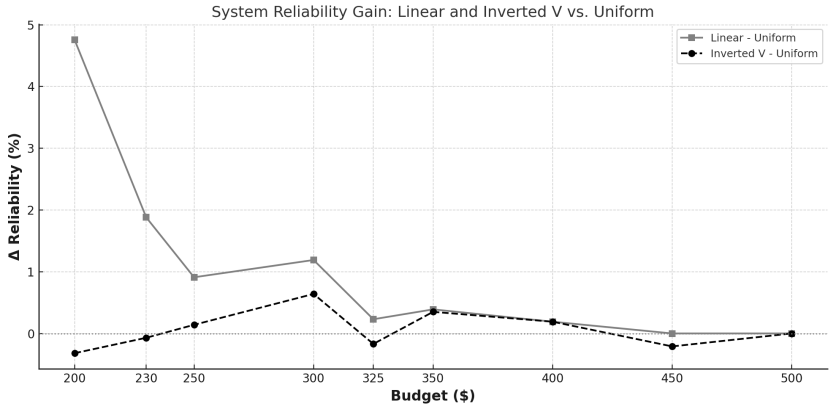


Figure 3. Reliability gain of LIBA and IVBA relative to UNIBA by budget.

References

Al-Jabouri H, Saif A, Khatab A, Diallo C, Venkatadri U (2024) A critical review of selective maintenance for mission-oriented systems: challenges and a roadmap for novel contributions. *International Journal of Production Research*, **62**(13), 4980–5015.

Arulmozhi G (2002) Exact equation and an algorithm for reliability evaluation of k-out-of-n: G system. *Reliability Engineering & System Safety* **78**(2) 87-91.

Chaabane K, Khatab A, Diallo C, Venkatadri U, Aghezzaf EH (2020) Selective maintenance optimization problem in systems under repairpersons availability *Proc. 7th Int. Conf. Control, Decis. Inf. Technol. (CoDIT)*, Prague, Czech Republic, pp. 687-692.

Diallo C, Venkatadri U, Khatab A, Liu Z (2018) Optimal selective maintenance decisions for large serial k-out-of-n: G systems under imperfect maintenance. *Reliability Engineering & System Safety* **176**, 249–258.

Liu Y, Chen Y, Jiang T (2018) On sequence planning for selective maintenance of multi-state systems under stochastic maintenance durations. *European Journal of Operational Research*, **268** (1), 113-127.

Malik M (1979) Reliable preventive maintenance scheduling. *AIIE transactions* **11**(3), 221–228.

O'Neil R, Diallo C, Khatab A, Aghezzaf EH (2023a) A hybrid column-generation and genetic algorithm approach for solving large-scale multimission selective maintenance problems in serial K-out-of-n:G systems. *International Journal of Production Research* **61**(9), 3071-3087.

O'Neil R, Khatab A, Diallo C, Venkatadri U (2023b) Optimal joint maintenance and orienteering strategy for complex mission-oriented systems: A case study in offshore wind energy *Computers & Operations Research* **149**, 106020.

Rice WF, Cassady CR, Nachlas JA (1998) Optimal maintenance plans under limited maintenance time. *Proceedings of the seventh industrial engineering research conference*, pp. 1-3.

Low-Power Time-Sensitive Behavioral Contract Monitoring for Hardware

Behnam Razi Perjikolaei*, Friederike Bruns**, Dennis Kempf**, Jörg Walter*

*OFFIS e.V. Institut für Informatik, Oldenburg, Germany, email: {behnam.razi.perjikolaei, joerg.walter}@offis.de

**Department of Computing Science, Carl von Ossietzky Universität Oldenburg, Germany, email:{friederike.bruns, dennis.kempf}@uol.de

Abstract Reliable execution of embedded systems is crucial and, thus, their validation and verification during design is essential. However, such methods cannot fully address unexpected issues, such as environmental factors, arising during operation. To enhance system reliability, runtime monitoring is widely used to detect and address misbehavior in real-time and increase trustworthiness. Most runtime monitoring solutions, however, focus solely on functional or extra-functional aspects. This work introduces an assertion-based runtime monitoring approach using Time-Sensitive Behavioral Contracts (TSBCs), implemented as hardware components within a versatile VHDL library. The library provides low-level building blocks and ready-to-use parameterizable monitor entities that simplify integration. To ensure minimal impact on the monitored system, the approach was evaluated for resource usage, performance, and power efficiency. In a motor inrush current monitoring scenario, the hardware solution reduced power consumption by 0.4% and achieved a frequency over 800 times higher than its software counterpart and comparable hardware solution.

1. Introduction

Embedded systems are crucial for handling safety-critical tasks in avionics, automotive, and manufacturing, where reliable execution is essential to prevent harm Marwedel (2021). Risks stem from design flaws, manufacturing defects, wear and tear, and even malicious manipulation or counterfeit chips Guin et al. (2014). Despite significant efforts, achieving absolute certainty of error-free operation remains practically unfeasible Kneuper (1997); Kern & Greenstreet (1999); Mitra et al. (2010).

Runtime monitoring, commonly realized in software Havelund (2008); Tabakov & Vardi (2010), observes functions inputs and outputs to detect specification violations. For industrial control systems, Tran et al. (2020) introduced Time-Sensitive Behavior Contracts (TSBCs) to monitor Function Blocks (FBs) based on the IEC 61499 standard Zoitl & Lewis (2014). However, software-based monitoring introduces performance overhead and depends on trusted hardware and software infrastructures -— an increasing challenge in light of complex CPU architectures and vulnerable supply chains. To improve performance, hybrid hardware/software solutions (e.g., Solet et al. (2018)) extract execution traces evaluated by hardware monitors, but still incur overheads of 16.2–33%. Alternatives such as NUVA Nassar et al. (2015), and fully hardware-based monitors using temporal logic (Tracy et al. (2020); Lu & Forin (2008)) offer promising directions.

This work aims to develop a more energy-efficient, high-performance hardware-based monitoring solution integrated into the Contract-Based Design (CBD) process. CBD formally specifies component behavior using assumption-guarantee pairs (A, G), enabling modularity, reusability, and early error detection Benveniste et al. (2015); Sangiovanni-Vincentelli et al. (2012). We formalize contracts as TSBCs and propose a VHDL library of configurable monitoring components that can be flexibly integrated (pre- or post-implementation) at various design stages. These components, implemented on FPGAs, observe system behavior at runtime to validate conformance with TSBC specifications, enabling efficient and adaptable hardware-based verification.

2. Pre- and Post-Implementation Parameterization of Time-Sensitive Behavioral Contract Monitors

The specification of TSBCs Tran et al. (2020) relies on assume-guarantee contracts. For example, assuming that a timing condition is fulfilled, the system is supposed to guarantee a specified output behavior. Consequently, this can be used to determine whether certain data conditions are adhered to at a given point in time. The top-level syntax of TSBCs is defined by the following grammar given in Backus–Naur form notation:

$$
\begin{aligned}
\langle contract \rangle \ ::= \ & \langle data_condition_true \rangle \\
| \ & \langle time_condition \rangle \ \ '?' \ \langle data_condition_true \rangle \\
| \ & \langle time_condition \rangle \ \ '?' \ \langle data_condition_true \rangle \ ':' \ \langle data_condition_false \rangle
\end{aligned}
$$

Here, $\langle time_condition \rangle$ specifies an open or closed interval condition over the current time. It is used to determine which of the data-dependent conditions ($\langle data_condition_true/false \rangle$) must be asserted: Whenever $\langle time_condition \rangle$ is satisfied or absent, $\langle data_condition_true \rangle$ must be satisfied. Contrary, when $\langle time_condition \rangle$ is present and not satisfied, $\langle data_condition_false \rangle$ must be satisfied. Thus, we can categorize TSBCs into three types: **Type I** solely operates on data conditions. **Type II** and **III** additionally require time conditions. Moreover, **Type II** specifies only one data condition as $\langle data_condition_true \rangle$, while **Type III** has a second data condition ($\langle data_condition_false \rangle$). Constraints are realized by comparing the current value of the condition's time or data variable. Point comparisons check whether a variable's value is *equal* ('$=$') to or *not equal* ('$\neq$') to a specific point or number, where interval comparisons check whether the variable's value is *within* ('$\in$') or *outside* ('$\notin$') of a specific interval.

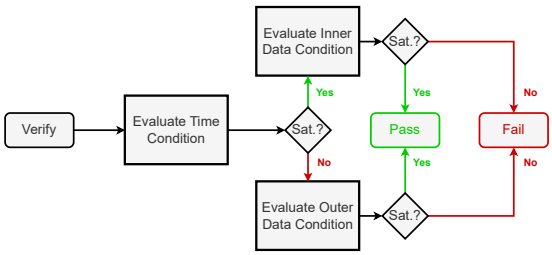


Figure 1. Generalized TSBC-based monitoring algorithm.

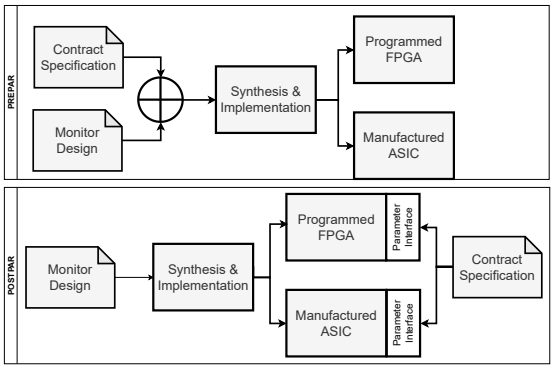


Figure 2. Parametrization Pre-implementation PREPAR vs. post-implementation POSTPAR.

TSBC-based hardware monitors require protocols to derive monitors from contract specifications, determine the methodology for hardware realization of the *Time t* and *Data d* variables, and define a mechanism for validating final verdict. The monitors constrain the system's behavior triggered by events such as system startup or environmental stimuli, termed *requests*. Monitors synchronize with actions by receiving the same request events, aligning their execution with $t = 0$ relative to a triggering event. They operate in two states: idle when no requests are pending or actions are complete, and active when executing actions upon request arrival. Each subsequent request in the active state resets the time to indicate the start of a new monitoring cycle. The data variable d represents a system property's value, which, thus, has to be measured typically facilitated by interfaces named *observers*. In the original TSBC specification, d is real-valued ($d(t) \in \mathbb{R}$). Since floating-point arithmetic consumes a large amount of hardware resources and typical hardware-level properties (e. g. ADC register values) are of a fixed-point nature anyways, our monitors use a fixed-point representation for d. Assuming the presence of both the time and second data conditions (**Type III**), typical in normalized contracts, monitoring proceeds as visualized in Figure 1: Firstly, the monitor evaluates the time condition. If satisfied, it proceeds to assess the inner data condition, otherwise, the outer data condition is evaluated. If the evaluated data condition is satisfied, the contract itself is deemed satisfied; otherwise, it is considered violated. Upon a detected violation, monitors may continue in the hope of future contract satisfaction, termed *recoverable*, or pause until a subsequent request, termed *unrecoverable*. The recoverable approach assumes monitors may pass after initial failure within the same request providing additional information for decision-making. For instance, filters could discard intermittent violations from noisy sensors, propagating stable monitor results. Contrary, the unrecoverable approach presumes that contracts, which are violated once persist as violated for the remainder of the ongoing request, convenient for critical fault conditions that are intolerant of any violations.

TSBCs constrain the permissible value ranges of system properties within certain time intervals (time and data bounds). Determining when contract specifications are passed to monitors significantly impacts the resulting hardware design and, thus, requires careful consideration. We identify two options depicted in Figure 2: (1) Parameterizing monitors either *prior to* their hardware realization (pre-implementation parameterization: PREPAR) or (2) *afterwards* (post-implementation parameterization: POSTPAR). The PREPAR approach offers

potential for design optimizations before monitor designs are implemented in hardware. This allows for improvements in resource utilization, performance, and energy efficiency. However, a significant drawback of this approach is that modifications in the contract specification necessitate the implementation of new hardware. This is an issue particularly when the precise properties of the monitored system are not fully understood, as may occur during early design and testing phases. In contrast, the POSTPAR approach involves adapting monitors to contract specifications after their implemention in hardware. An adjustment is possible after programming the bitstream containing the design onto FPGA devices or after the manufacturing process for ASICs. The more flexible POSTPAR approach enables specifications to be altered without requiring new hardware, utilizing interfaces through which the parameters can be dynamically passed. On the downside, these monitors have to evaluate any contract specification they receive at runtime, making it challenging to optimize their implementation to the same extent as monitors with fixed contract specifications. Moreover, additional hardware is necessary for storing, transferring and processing specifications.

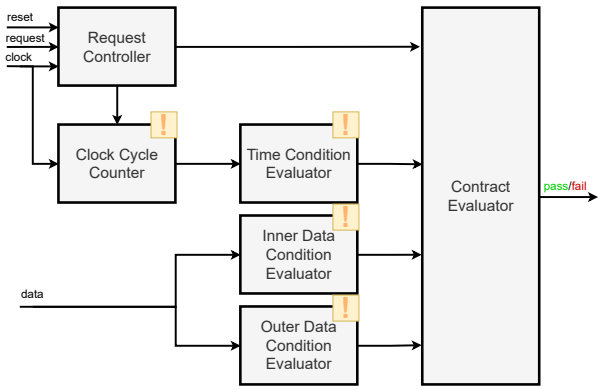
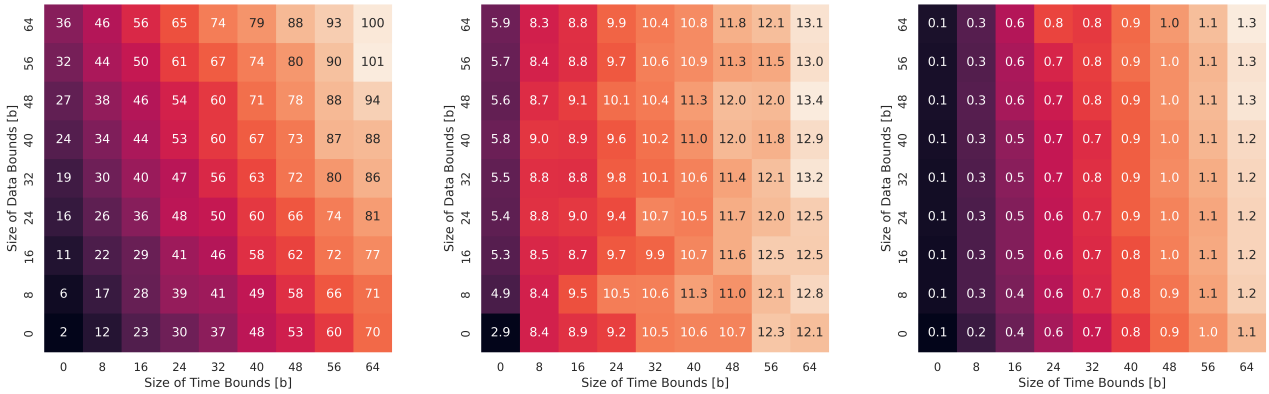


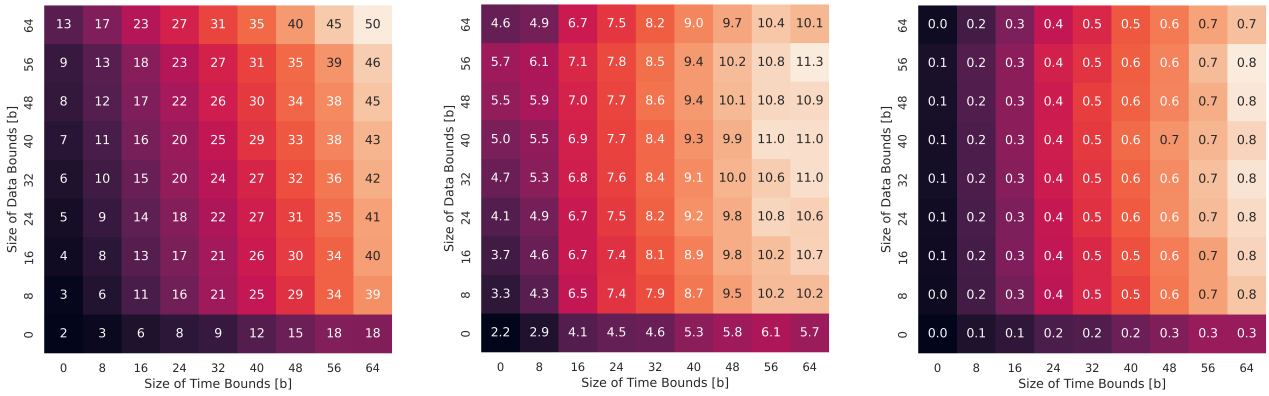
Figure 3. Monitor architecture design with relevant components for parameterization (yellow symbol).

Figure 3 depicts the VHDL monitor architecture: **Request Controller** manages contract execution, resets the **Clock Cycle Counter**, and interfaces with external observers to input data. Implemented as a state machine at Register Transfer Level (RTL), it tracks monitoring activity based on mutual request signals. During idle states or reset events, output signals remain inactive. The controller transitions to active monitoring upon a new request unless overridden by a reset signal. The **Clock Cycle Counter** tracks time by counting clock cycles to determine the monitor's time variable t. The counter's size, parametrized via the generic parameter g_size (referring to the number of bits of the counter), accommodates the required time bounds $[a,b]$: Left-bounded conditions (g_size $= \lceil log_2(a+1) \rceil$) and right-bounded or both-sided conditions (g_size $= \lceil log_2(b+2) \rceil$). Overflow handling ensures the count remains valid and the counter resets to zero as needed. The **Evaluator Entities** implement the TSBC monitoring algorithm by evaluating time and data conditions against contract semantics and processing the values received by the request controller. Utilizing VHDL standard operators, they support technology-optimized synthesis while enabling modular contract evaluation at various levels. The **Condition Evaluators** determine whether the values meet the corresponding TSBC conditions. When the time condition is satisfied, the inner data condition determines whether the result signals (in)valid behavior. Otherwise, the result signal represents whether the outer data condition is satisfied. Data condition evaluation employs a speculative approach to maintain simplicity in the monitor implementation. Based on the evaluation results and execution state, a **Contract Evaluator** component decides whether the contract as a whole is satisfied or not. Lastly, an **Adapter Module** can be used for post-processing of the final result to take into account the current state of the monitor as reported by the controller. Its RTL implementation distinguishes between recoverable and non-recoverable monitors derived from *active* and *result* inputs: Reset status upon new requests, enabling subsequent evaluations (recoverable monitors) in contrast to retaining violation status throughout the active monitoring phase (non-recoverable monitors). Logical AND operations ensure outputs are asserted only during active monitoring cycles. As an example, *fail* is returned only if a violation has occurred at least once during an active monitoring phase. Contrary, the signal for recoverable monitors is reset upon the initiation of a new request, regardless of whether the contract was previously violated.



(a) Slice utilization. (b) Longest path delay $[ns]$. (c) Power consumption $[mW]$.

Figure 4. POSTPAR



(a) Slice utilization. (b) Longest path delay $[ns]$. (c) Power consumption $[mW]$.

Figure 5. Mean PREPAR

3. Evaluation

The functionality and efficiency of TSBC monitors were evaluated on a PYNQ-Z2 development board using Xilinx Vivado for implementation and synthesis, providing detailed resource and performance reports. Key metrics included **resource utilization**, **latency**, and **dynamic power consumption** to assess practical applicability. Designed to operate at full system clock speed without introducing runtime overhead, the monitors run concurrently with the observed system; however, they may affect the maximum achievable clock frequency, which depends on the longest data path and register timing constraints. Dynamic power consumption was analyzed using Vivado's Vectorless Power Analysis, focusing on switching activity. For POSTPAR monitors, we evaluated the influence of time m and data n bounds over the set $0, 8, \ldots, 64$, synthesizing and analyzing 81 monitors. PREPAR monitors, which depend on specific TSBCs at compile time, required a broader evaluation: 256 randomly generated specifications per POSTPAR configuration were synthesized, and 8192 additional evaluations were conducted for the representative case $(32, 16)$, resulting in 28532 unique monitors. PREPAR **Type II** & **III** showed significantly different characteristics compared to time-invariant **Type I** monitors, which were excluded to avoid skewing the results and to maintain focus on time-sensitive contracts.

On average, PREPAR **slice utilization** (Figure 5a) is about 40 % of POSTPAR (Figure 4a). A linear regression model estimates POSTPAR utilization (time m and data n bound) as: $SLICE_{post}(m,n) \approx 1.016m + 0.516n + 5.291$, $R^2 \approx 0.993$. Here, m carries roughly twice the weight of n, indicating that the size of time bounds has twice the impact on POSTPAR utilization compared to data bounds. Since monitors evaluate two data conditions but only one time condition, clock cycle counters consume approximately three times the resources of condition evaluators of similar size. For PREPAR utilization, the model is: $SLICE_{pre}(m,n) \approx 0.526m + 0.242n - 1.628$, $R^2 \approx 0.939$, which shows that the weight disparity is even more pronounced, fur-

ther emphasizing the dominant impact of clock cycle counters on resource utilization compared to condition evaluators. Figure 4b shows the **data path length** for POSTPAR with delays in nanoseconds (ns) for various configurations. The shortest path, is observed in the $(0,0)$ configuration with $DELAY_{post}(0,0) = 2.9$. The longest path is recorded at $DELAY_{post}(64,48) = 13.4$ instead of the expected $(64,64)$, likely due to noise. For PREPAR monitors (Figure 5b) delays range from $DELAY_{pre}(0,0) = 2.2$ to $DELAY_{pre}(64,56) = 11.3$. On average, PREPAR delays are about 77 % of POSTPAR delays with mean values of 7.68 ns for PREPAR and 10.01 ns for POSTPAR. This rather small difference is due to the consistent delay introduced by the clock cycle counter in both implementations. Figure 4c depicts **dynamic power consumption** for various POSTPAR configurations in milliwatts (mW). The lowest consumption is $POWER_{post}(0,0) = 0.1$, while the highest is $POWER_{post}(64,64) = 1.3$. For PREPAR configurations (Figure 5c), the dynamic power ranges from $POWER_{pre}(0,0) = 0.04$ to $POWER_{pre}(64,56) = 0.8$, with deviations from the expected maximum $(64,64)$, likely due to noise from stochastic distribution. Although static power consumption introduces a large bias, total power consumption differences between PREPAR and POSTPAR are negligible. However, the results show that PREPAR implementations are more efficient, consuming only about 58 % of POSTPAR's average power(427 µW vs. 728 µW).

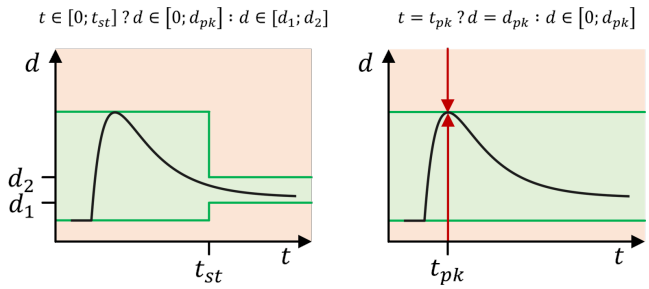


Figure 6. Different TSBC monitors for motor inrush current.

To **compare hardware- and software-based** TSBC monitor implementations, a motor inrush current monitoring system was developed. The software monitor was implemented as a natively compiled C++ FB and deployed on the PYNQ board's PS running PetaLinux OS. The hardware monitor, using PREPAR designs, was implemented in the PL, synthesized with Vivado's default settings at 100 MHz, and could achieve up to 150 MHz with optimization. Using the simulation environment Simulink, ideal sensors measured the motor's current draw, focusing on detecting the inrush current peak. Monitors extracted data from the simulation to validate system behavior. Two monitoring principles are illustrated in 6 with the permissible current draw (green enclosure), valid example traces (black line), and point conditions (red arrows). The first monitor (left, Figure 6) intends to precisely validate the location and amplitude of the inrush current peak. However, software-based TSBC FBs, running at low execution frequencies, struggled with accurate timing for point conditions like $t = 0.101\,154$ s. In contrast, hardware monitors operated reliably at 100 MHz. For a fair comparison, a shared contract specification (right, Figure 6), was used: $t \in [0.0; 0.110]\ ?\ d \in [0.0; 4.696732] : d \in [0.75; 1.25]$. Performance tests revealed that the software monitor on the PS (ARM Cortex A9 at 650 MHz) achieved a maximum frequency of around 120 kHz, while a modern AMD Ryzen 5 3600 desktop CPU (4.2 GHz boost clock) reached 3 MHz—still 50 times slower than the hardware monitor. Power consumption measurements highlighted stark differences. Using a USB power meter, 500 hardware monitors running at 100 MHz increased power consumption by 266 mW, translating to just 532 µW per monitor. In contrast, a single software monitor running at 120 kHz with full CPU load increased power consumption by 137 mW, over 250 times higher than a hardware monitor operating at more than 800 times the frequency. For comparison, the hardware-based solution NUVA Nassar et al. (2015), with an average total power consumption of 140 mW, is similar to the software monitor in energy usage.

4. Conclusion

Runtime monitoring complements design-time verification by detecting operational misbehavior caused by environmental factors, thereby enhancing the trustworthiness and reliability of embedded systems. We propose

a versatile VHDL library offering configurable, high-performance, low-power hardware monitors, utilizing two parameterization strategies – PREPAR and POSTPAR. Key findings include: PREPAR monitors consistently outperformed POSTPAR implementations, achieving approximately double the efficiency in resource use, performance, and energy consumption. However, PREPAR's fixed contract specifications necessitate reevaluation for ASICs that require hardware premanufacturing. Hardware monitors revealed significant performance gains over software implementations by a factor of 800 and energy efficiency gains by a factor of 250. Despite limitations in time and data variable accuracy due to clock frequency and observer capabilities, the proposed hardware monitors effectively balance performance, efficiency, and flexibility, advancing the state of runtime monitoring.

Future work could include a combination with ML-driven diagnostics and predictive analytics to enhance monitoring capabilities and system reliability.

References

Benveniste, A., Caillaud, B., Nickovic, D., Passerone, R., Raclet, J.-P.-B., Reinkemeier, P., Sangiovanni-Vincentelli, A., Damm, W., Henzinger, T. & Larsen, K. (2015), Contracts for Systems Design: Theory, Research Report RR-8759, Inria Rennes Bretagne Atlantique ; INRIA.

Guin, U., Huang, K., DiMase, D., Carulli, J., Tehranipoor, M. & Makris, Y. (2014), 'Counterfeit Integrated Circuits: A Rising Threat in the Global Semiconductor Supply Chain', *Proceedings of the IEEE* **102**(8), 1207–1228.

Havelund, K. (2008), Runtime Verification of C Programs, *in* 'Proceedings of the 20th IFIP TC 6/WG 6.1 International Conference on Testing of Software and Communicating Systems: 8th International Workshop', Springer-Verlag, Berlin, Heidelberg, p. 7–22.

Kern, C. & Greenstreet, M. (1999), 'Formal Verification in Hardware Design: A Survey', *ACM Transactions on Design Automation of Electronic Systems* **4**, 123–193.

Kneuper, R. (1997), 'Limits of formal methods', *Formal Aspects of Computing* **9**(4), 379–394.

Lu, H. & Forin, A. (2008), 'Automatic Processor Customization for Zero-Overhead Online Software Verification', *IEEE Transactions on Very Large Scale Integration (VLSI) Systems* **16**(10), 1346–1357.

Marwedel, P. (2021), *Embedded System Design Embedded Systems Foundations of Cyber-Physical Systems, and the Internet of Things.*

Mitra, S., Seshia, S. & Nicolici, N. (2010), Post-silicon validation opportunities, challenges and recent advances, *in* '47th Design Automation Conference on (DAC)', ACM Press, Anaheim, CA.

Nassar, A., Kurdahi, F. & Elsharkasy, W. (2015), NUVA: Architectural Support for Runtime Verification of Parametric Specifications over Multicores, *in* 'International Conference on Compilers, Architecture and Synthesis for Embedded Systems (CASES)', IEEE Press, p. 137–146.

Sangiovanni-Vincentelli, A., Damm, W. & Passerone, R. (2012), 'Taming Dr. Frankenstein: Contract-Based Design for Cyber-Physical Systems', *European Journal of Control* **18**(3), 217–238.

Solet, D., Pillement, S., Béchennec, J., Briday, M. & Faucou, S. (2018), HW-based Architecture for Runtime Verification of Embedded Software on SoPC systems, *in* 'NASA/ESA Conference on Adaptive Hardware and Systems (AHS)', pp. 249–256.

Tabakov, D. & Vardi, M. (2010), Optimized Temporal Monitors for SystemC, *in* H. Barringer, Y. Falcone, B. Finkbeiner, K. Havelund, I. Lee, G. Pace, G. Roşu, O. Sokolsky & N. Tillmann, eds, 'Runtime Verification', Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 436–451.

Tracy, T., Tabajara, L., Vardi, M. & Skadron, K. (2020), Runtime Verification on FPGAs with LTLf Specifications, *in* 'Formal Methods in Computer Aided Design (FMCAD)', pp. 36–46.

Tran, D. D., Walter, J., Grüttner, K. & Oppenheimer, F. (2020), Towards Time-Sensitive Behavioral Contract Monitors for IEC 61499 Function Blocks, *in* 'IEEE Conference on Industrial Cyber-Physical Systems (ICPS)'.

Zoitl, A. & Lewis, R. (2014), *Modelling Control Systems Using IEC 61499*.

From Single Output Centralized Learning to Federated Learning: Introducing a Cross-Teaching Multi-Output Model for Physical Assets

S. Kamei* and S. Taghipour**

* *Department of Mechanical, Industrial, and Mechatronics Engineering, Toronto Metropolitan University*
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3, jkamei@torontomu.ca
** *Department of Mechanical, Industrial, and Mechatronics Engineering, Toronto Metropolitan University*
350 Victoria Street, Toronto, Ontario, Canada, M5B 2K3, sharareh@torontomu.ca

Abstract. Predicting the remaining useful life (RUL) is central for effective prognostics and health management (PHM). The current prognostics approaches for a network of assets are centralized, where aggregated data makes a richer dataset. When assets are located at different enterprises, data sharing is not preferred. One solution like federated learning, the local client data and training is on-site rather than being shared with a central server. For **Phase I**, we created benchmark results for centralized and decentralized learning approaches for RUL prediction. Benchmark studies are important since it gives researchers a benchmark and the ability to produce the same reported results. The results indicated that decentralized setups using Transformer as the final prediction model performed similar to centralized. Furthermore, previous prognostic prediction models have a single output RUL. **Phase II**, we developed a single model cross-teaching multi-input-multi-output (MIMO) neural network framework, which accounts for more than one fault mode and produces simultaneously fault classifications and RUL predictions. The results indicated this framework performed better against a non-cross-teaching MIMO, and single output approach. These findings help with proactive maintenance strategies, minimizing operational and maintenance disruptions while extending the shelf-life of assets.

1. Introduction

Maintaining optimal operational efficiency of assets is one of the goals of industries, and prognostics and health management (PHM), employs various strategies to achieve this. A proactive maintenance approach should be taken to reduce the risks and costs associated with unforeseen operational complications (Zhang *et al.,* 2018; Zio, 2022). Prognostics has been employed in machinery maintenance, in which the current state of the machinery is monitored for degradation patterns, with the aim of predicting the remaining useful life (RUL). RUL is the remaining time before a machinery or system failure (Pektas and Pektas, 2018; Xia *et al.*, 2022). Developing an accurate prognostics model for RUL prediction presents several challenges. Assets can have more than one failure mode, exhibit unforeseen degradation, and some possessing more degradation information than others (Chadha *et al.,* 2022; Jiao *et al.,* 2020; Kundu *et al.*, 2019).

Current centralized prognostics approaches for a network of assets are reliant on the availability of assets' sensors, failures, and anomaly data. While aggregating data from assets make for a richer dataset for prognosis, assets located at different enterprises, business owners may not be willing to share their raw data with each other. Federated learning (FL) is a promising decentralized solution. In FL, data remains local to each client, model training occurs on-site, with only model updates being shared with the central server. Since FL theoretically addresses the challenges faced by the traditional centralized learning approaches, its performance needs to be investigated and compared with the centralized methods (Lo *et al.,* 2022; McMahan *et al.,* 2017; Rahman *et al.,* 2020).

Furthermore, recent studies in medical image segmentation have highlighted the benefits of applying Transformers in combination with CNN through a technique called cross-teaching. Cross-teaching utilizes the advantages and different learning behaviors of two distinct deep learning (DL) networks. In this framework, for a better performance, each DL network compensates for the characteristics that maybe lacking in other network by forming a consistency regularization between the two DL models (Luo *et al.*, 2022). This study will utilize **Phases I and II** to reduce epochs and communication rounds in final model training. By utilizing these strategies within the prognostics model, an accurate assessment of RUL facilitates the development of effective maintenance strategies, leading to a reduction in unforeseen mechanical issues, and preserve

operational safety, all with the goal of extending the lifespan of the asset (Dhada *et al.*, 2019; Ellefsen *et al.* 2019; Zhang *et al.*, 2022).

2. Motivation

Many prognostic studies create a framework consisting of multiple models for a single output. However, we also noticed limited prognostics studies with multi-output approaches, and even less studies using single model that can handle multiple outputs. For further prediction improvements, we looked into other fields outside of prognostics and noted that in medical image segmentation studies, they use a concept called cross-teaching between two DL models that has yet to be demonstrated for prognostics to the best of our knowledge. These were the motivations for this study (Jiao *et al.*, 2020; Luo *et al.*, 2022; Xue *et al.*, 2021). **Phase I,** we created benchmark results for centralized and decentralized learning approaches for RUL prediction. We believe that performance values for benchmark studies are important since it gives researchers a benchmark and the ability to produce the same reported results. Future prognostic studies involving decentralized methods, can evaluate how much their novel methodologies can improve the performance comparing to this benchmark to make a fair comparison in performances. **Phase II,** we developed a single model cross-teaching multi-input-multi-output (MIMO) neural network framework with the capacity to recognize asset's fault mode and predict RUL simultaneously.

3. Methodology

This study was conducted in two phases. **Phase 1** utilized the NASA Commercial Modular Aero-Propulsion System Simulation (C-MAPSS) turbofan engine degradation simulation dataset (FD001-FD004). Each subset dataset contains 20,000 to over 61,000 data samples. Sensors readings within each engine capture internal and external conditions, indicating the current health (Ellefsen *et al.*, 2019; Pillai, and Vadakkepat, 2021; Yuan *et al.*, 2016). Engines initially operate normally but degrade over time, exhibiting either one high-pressure compressor (HPC), or two (HPC and fan) fault modes under one or six operating conditions. These collectively affect the engines' life cycles (Krokotsch *et al.*, 2021; Zhang *et al.*, 2023). In this phase, we conducted two separate benchmark prediction models, long short-term memory (LSTM), and the Transformer framework. These models were chosen because LSTM's popularity in time series, and to explore Transformer's potential beyond natural language processing. We then compared the performances of the two standard FL algorithms, Federated Averaging (FedAvg) and Federated Proximal (FedProx), under different decentralized scenarios against the centralized (Chandha *et al.*, 2022; Kamei and Taghipour, 2023; Liu *et al.*, 2022). Centralized setup, node (N) = 1 involved all the trajectories grouped. Decentralized setups, (N = 2, 5, and 10) involved randomly created trajectory clusters (as illustrated in Figure 1 for N = 2). The trajectory division for FD001-FD004 are detailed in Table 1. The data was randomly split, where 90% of the engines in the training dataset were allocated for training, and the remaining 10% for validating the models. This approach is aimed to improve the global model prediction in the FL after sufficient communication rounds (Llasag, *et al.*, 2020; Lo *et al.*, 2022).

Phase II focused on C-MAPSS FD003 and FD004, but also expanded to include New-CMAPSS (N-CMAPSS) dataset DS01- DS08, the successor to C-MAPSS. N-CMAPSS datasets contain millions of samples and provided fault mode labels. Ideally it would be better if it was possible that personal computers had the capability to run and process all the data samples. However, due to computational limits, a reduction in data samples was required. We reduced the sampling frequency from 1 Hz to 0.05 Hz (i.e., we took every 20th data sample for the training and testing). Despite the reduction in the training and testing datasets, the 0.05Hz DS02 dataset is still significantly larger than C-MAPSS, and it still retains long-range data trends that capture enough degradation information necessary for prognostics unlike C-MAPSS which records data at the end of a cycle. We specifically used DS02 which features two fault modes: high pressure turbine (HPT), and HPT and low-pressure turbine (LPT) together (Koutroulis *et al.*, 2022; Xiong *et al.*, 2023). In this phase, we developed the MIMO framework with a single final prediction model, evaluating it against a single output model using the same test data selection as **Phase I.** To further enhance fault classification and RUL predictions, we implemented a single model cross-teaching between two DL models within the MIMO framework and compared its performance to the non-cross-teaching multi-output framework, and the single output approach.

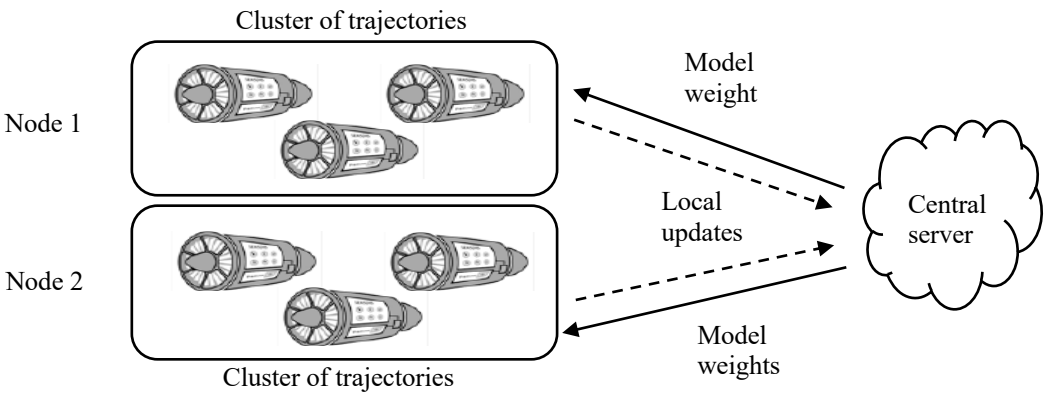


Figure 1: **Phase I** Federated learning using clustering of trajectories (N = 2), where each cluster represents one node. The diagram is adapted from Dhada *et al.*, 2019; Lo *et al*. 2022.

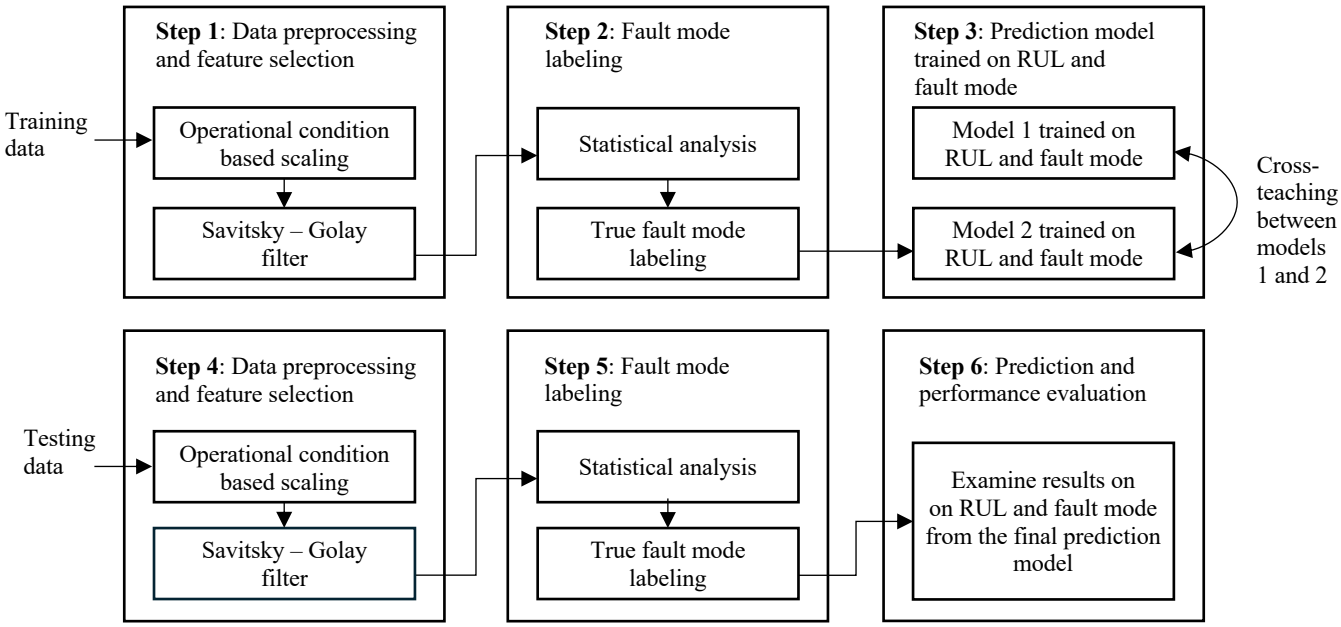


Figure 2: **Phase II** Flowchart for cross-teaching multi-output framework for predicting the fault mode and estimating the RUL simultaneously.

Table 1: Centralized and decentralized setups for FD001-FD004 (Kamei and Taghipour, 2023).

Dataset	Clustering of trajectories (N)	Number of flight trajectories per node
FD001	1	100
	2	50
	5	20
	10	10
FD002	1	260
	2	130
	5	52
	10	26
FD003	1	100
	2	50
	5	20
	10	10
FD004	1	249
	2	124, and 125
	5	50 each for 4 nodes, 49 for 1 node
	10	25 each for 9 nodes, 24 for 1 node

4. Results

Firstly, the choice of DL model impacted the performance across all tested scenarios. The Transformer model consistently outperformed the LSTM on the C-MAPSS, and N-CMAPSS subset datasets, irrespective of a whether a centralized, decentralized, non-cross-teaching or the cross-teaching MIMO framework was used. The Transformer from herein is referred as the final prediction models for all results below.

For **Phase I**, we created a single output RUL where we provided complete centralized performance results across all four subset datasets of C-MAPSS for each prediction model, LSTM and the Transformer as seen in Table 2. For example, with FD004, the Transformer achieved an RMSE = 20.38 after 192 epochs compared to RMSE = 22.71 after 100 epochs for LSTM. We also provided completed decentralized results in Table 3. Between the two standard FL algorithms, the results indicated that FedProx performed better than FedAvg overall. We illustrated the decentralized N = 2 setup using FedProx where the results were shown for the two prediction models. For example, with FD004, the Transformer had an RMSE = 18.71, whereas the LSTM RMSE = 22.38 (Tables 2 and 3). These potential findings demonstrated that decentralized models could perform just as well as centralized (Kamei and Taghipour, 2023).

For **Phase II**, Table 4 shows the performance metrics for the MIMO framework against the single RUL model from Table 2. For example, for FD004, the MIMO achieved a RMSE = 16.16 after 46 epochs, compared to RMSE = 20.38 after 192 epochs for the single RUL prediction model. The MIMO results were very encouraging; so, we decided to extend this contribution by utilizing the cross-teaching. With this method, there were slight improvements for FD004 where RMSE = 16.07 after 37 epochs. Even though the RMSE showed improvements, the fault classification accuracy scores were the same between the two frameworks.

We also wanted to observe if the results from C-MAPSS would be reflective with N-CMAPSS DS02. Although we did not use DS02 during **Phase I**, instead we used MIMO framework as a benchmark. Similarly, we observed that the cross-teaching MIMO performed better than the non-cross-teaching MIMO framework with RMSE = 3.05 after 8 epochs versus 3.14 after 7 epochs respectively. Although the fault classification accuracy scores didn't show results similar like C-MAPSS, these findings could be due to having fewer testing data for fault mode 1.

Regardless of the dataset used, these results were promising since the MIMO framework in itself decreased the number of epochs and improved the performance metrics compared to the single output approach. Furthermore, the cross-teaching multi-output framework performed better against the non-cross-teaching MIMO.

Table 2: Results for LSTM and Transformer (TF) prediction models for centralized (Kamei and Taghipour, 2023).

Dataset	**Epochs LSTM / TF**	**RMSE LSTM / TF**	**Score LSTM / TF**
FD001	100 / 164	14.51 / 13.52	318.59 / 287.07
FD002	60 / 56	21.20 / 19.32	4098.68 / 11436.74
FD003	200 / 131	15.34 / 13.44	411.99 / 263.64
FD004	100 / 192	22.71 / 20.38	5546.04 / 2784.62

Table 3: Results for LSTM and Transformer (TF) prediction models for N = 2 using FedProx (Kamei and Taghipour, 2023).

Dataset	**Epochs LSTM / TF**	**Number of Rounds LSTM / TF**	**RMSE LSTM / TF**	**Score LSTM / TF**
FD001	5 / 20	11 / 22	14.31 / 13.06	277.28 / 254.84
FD002	10 / 10	3 / 48	21.04 / 19.30	2495.34 / 1180.93
FD003	10 / 5	10 / 22	14.05 / 13.55	304.52 / 302.41
FD004	10 / 30	6 / 28	22.38 / 18.71	4143.27 / 2552.31

Table 4: Performance of the non-cross-teaching (Non-CT) and cross-teaching (CT) multi-output neural network framework using the Transformer as the final prediction model on the subset datasets of C-MAPSS and N-CMAPSS (0.05Hz sampling frequencies).

Datasets	**Epochs Non-CT / CT**	**Accuracy score (%) Non-CT / CT**	**RMSE Non-CT / CT**	**Score Non-CT / CT**
FD003	48 / 42	80 / 80	11.75 / 11.63	195.22 / 174.81
FD004	46 / 37	81.5 / 81.5	16.16 / 16.07	1264.92 / 1868.57
DS02	7 / 8	59.5 / 59.5	3.14 / 3.05	20767.82 / 21147.86

5. Conclusion

Phase I summarized and compared the performance of the decentralized FL and centralized learning approaches. The results indicated that using Transformer as the final prediction model, FedProx performed better than FedAvg across the C-MAPSS datasets compared to the centralized. However, this work has limitations. It would be interesting to observe how the estimations performance varies under different operational conditions for specific fault modes. With more and more state-of-the-art algorithms being developed, it will be interesting to assess how these algorithms can provide better predictions for RUL, with the intent of increasing the health and longevity of industrial systems and assets.

Phase II, the MIMO and the cross-teaching multi-output neural network frameworks simultaneously produced fault mode classification and RUL predictions for C-MAPSS's FD003 and FD004, and N-CMAPSS's DS02. From this, it was determined that the single model cross-teaching MIMO framework slightly outperformed the non-cross-teaching MIMO and single output approach. Even though the fault mode classification showed promising results on FD003 and FD004, it performed less on N-CMAPSS's DS02. However, future research may see whether improvements in accuracy scores of fault mode classifications are possible for these datasets.

Extending the contributions of **Phases I and II**, for **Phase III** we want to create a single model cross-teaching MIMO FL framework that allows for simultaneous fault classification and RUL prediction at the fleet-wide and local client level using local client data. Granular level (i.e., client level or engine level) predictions provide additional insights and analysis for proactive maintenance scheduling of assets. This methodology will consider the variations among each individual client and allowing inter-asset knowledge exchange using local client data. With this cross-teaching MIMO FL framework, the aim is to create a methodology where less epochs and reduced number of communication rounds are observed compared to the single RUL prediction model from **Phase I,** as well as improved performance metrics against the cross-teaching multi-output framework (used as centralized benchmark) from **Phase II**. With the implemented strategies, these innovations in prognostics are carried out with the goal of maintaining the health, safety, and durability of assets.

References

Chadha G, Bin Shah SR, Schwung A, Ding SX (2022) Shared temporal attention transformer for remaining useful life estimation. *IEEE Access* **10**, 74244-74258.

Dhada M, Palau AS, Parlikad AK (2019) Federated learning for collaborative prognosis. *International Conference on Precision, Meso, Micro, and Nano Engineering (COPEM 2019)*, 1-6.

Ellefsen AL, Bjørlykhaug E, AEsøy V, Ushakov S, Zhang H (2019) Remaining useful life predictions for turbofan engine degradation using semi-supervised deep architecture. *Reliability Engineering and System Safety* **183**, 240-251.

Jiao R, Peng K, Dong J, Zhang C (2020) Fault monitoring and remaining useful life prediction framework for multiple fault modes in prognostics. *Reliability Engineering and System Safety* **203**, 107028.

Kamei S, Taghipour S (2023) A comparison study of centralized and decentralized federated learning approaches utilizing the transformer architecture for estimating remaining useful life. *Reliability Engineering and System Safety* **233**, 109130.

Koutroulis G, Mutlu B, Kern R (2022) Constructing robust health indicators from complex engineered systems via anticausal learning. *Engineering Applications of Artificial Intelligence* **113**, 104926.

Krokotsch T, Knaak M, Gühmann C (2021) Improving semi-supervised learning for remaining useful life estimation through self-supervision. *Int. J. Progn. Health Manag* **13(1)**, 1-19.

Kundu P, Chopra S, Lad BK (2019) Multiple failure behaviors identification and remaining useful life prediction of ball bearings. *J. Intell. Manuf.* **30**, 1795-1807.

Liu L, Song X, Zhou Z (2022) Aircraft engine remaining useful life estimation via a double attention-based data-driven architecture. *Reliability Engineering and System Safety* **21**, 108330.

Llasag Rosero RH, Silva C, Ribeiro B (2020) Remaining useful life estimation in aircraft components with federated learning. *Int. J. Progn. Health Manag.* **13(1),** 1-8.

Lo SK, Lu Q, Zhu L, Paik HY, Xu X, Wang C (2022) Architectural patterns for the design of federated learning systems. *Journals of Systems and Software* **191**, 111357.

Luo X, Hu M, Song, T, Wang G, Zhang S (2022) Semi-supervised medical image segmentation via cross teaching between CNN and Transformer. In *Proc. 5th International Conference on Medical Imaging With Deep Learning*, PMLR **172**, pp. 820-833.

McMahan B, Moore E, Ramage D, Hampson S, Agüera y Arcas B (2017) Communication efficient learning of deep networks from decentralized data. In *Proc. Int. Conf. Artif. Int. Stat. (AISTATS),* **54,** pp.1273-1282.

Pektas A, Pektas E (2018) A novel scheme for accurate remaining useful life prediction for industrial IoTs by using deep neural network. *International Journal of Artificial Intelligence and Applications (IJAIA*) **9(5)**, 17-25.

Pillai S, Vadakkepat P (2021) Two-stage deep learning for prognostics using multi-loss encoder and convolutional composite features. *Expert Systems With Applications* **171,** 114569.

Rahman SA, Tout H, Ould-Slimane H, Mourad A (2020) A survey on federated learning: The journey from centralized to distributed on-site learning and beyond. *IEEE Internet of Things Journal* **8(7)**, 5476-5497.

Xia J, Feng Y, Teng D, Chen, J, Song Z (2022) Distance self-attention network method for remaining useful life estimation of aeroengine with parallel computing. *Reliability Engineering and System Safety* **225**, 108636.

Xiong J, Fink O, Zhou J, Ma Y (2023) Controlled physics-informed data generation for deep learning-based remaining useful life prediction under unseen operation conditions. *Mechanical Systems and Signal Processing* **197**, 110359.

Xue B, Chang B, Du D (2021) Multi-output monitoring of high-speed laser welding state based on deep learning. *Sensors* **21**, 1626.

Yuan M, Wu Y, Lin L (2016) Fault diagnosis and remaining useful life estimation of aero engine using LSTM neural network. In *Proc. 2016 IEEE International Conference on Aircraft Utility Systems (AUS)*, pp.135-140.

Zhang J, Wang P, Yan R, Gao, RX (2018) Long short-term memory for machine remaining life prediction. *Journal of Manufacturing Systems* **48**, 78-86.

Zhang Z, Song W, Li Q (2022) Dual aspect self-attention based on transformer for remaining useful life prediction. *IEEE Transactions on Instrumentation and Measurement* **71**, 2505711.

Zhang Z, Song W, Li Q, Guo H (2023) Multiscale global and local self-attention-based network for remaining useful life prediction. *Meas. Sci., Technol* **34**, 125154.

Zio E (2022) Prognostics and Health Management (PHM): Where are we and where we (need to) go in theory and practice. *Reliability Engineering and System Safety* **218**, 108119.

Rare failure analysis - challenges and perspectives from high-tech industry

M. Skarupski*, M. Veenvliet**, M. Wouters***, B. Dielissen****, and S. Kapodistria *****

*Department of Mathematics and Computer Science, Eindhoven University of Technology, The Netherlands & Wrocław University of Science and Technology, Poland, m.skarupski@tue.nl

**ASML, Veldhoven, The Netherlands, miklas.veenvliet@asml.com

***ASML, Veldhoven, The Netherlands, maarten.wouters@asml.com

****ASML, Veldhoven, The Netherlands, bas.dielissen@asml.com

*****Department of Mathematics & Computer Science, Eindhoven University of Technology, The Netherlands, email address: s.kapodistria@tue.nl

Abstract High-tech machines are highly complicated machines containing many interacting parts and exhibiting few (rare) failures, with high heterogeneity (partly due to the plethora of underlying failure causes). This makes predicting the expected number of future failures per time unit in an operational fleet of machines challenging, and thus hinders a cost-effective stocking and maintenance strategy. This article explores the benefits and drawbacks of popular statistical models within the high-tech industry for modeling rare failures. In this effort, the article should be viewed as a point of practical inspiration and not as a complete literature overview of the status quo.

1. Introduction

In the semiconductor industry and other high-tech industries, the downtime of the machines is often very expensive. For example, for ASML's† customers a couple of hours of production loss typically results in tens of millions of euros of loss. So not being able to repair a machine purely due to an insufficient stock of spare parts is extremely costly and should be avoided as much as possible. Naively one could try to prevent these costs by placing sufficient spare parts on stock with a very wide safety margin. However, the parts on stock will most likely become obsolete before they are used due to the rapid pace of innovation, resulting in a lot wasted capital investment. Having a cost-effective stocking and maintenance strategy is therefore vital.

The complicated machines in the high-tech industry can often contain thousands of spare parts. The majority of these parts typically have at most a hand full of observed failures over the entire life span of all fielded machines. Forecasting the expected number of failures per unit of time can be very challenging with such few data points, but it is of vital importance for a cost-effective stocking and maintenance strategy, since the majority of the components falls in the category of rare (failure) events.

Another complicating factor is the variability of the time required to replenish stock levels. This variability can be (partly) attributed to the significant amount of time required to manufacture such complicated and specialized service parts, which can often last from several months up to even a year. Therefore, it is important not only to predict the expected failure rate for each spare part accurately, it also is important to accurately model the variation (cf. Chen et al. (2018), van Jaarsveld and Dekker (2011)). With variation you can model the expected range of service parts needed, which in turn help determine the risk of over- or under-stocking. For example, if you expect a usage of 3 in the coming year with a 95 percentile of 4, that is different from expecting 3 and having a 95 percentile of 6. The stocking level really differs if you want to reduce the risk of not having stock and consequently loosing millions because of unexpected down. And even if one would correctly predict the expected usage of 3 parts in the coming year for the entire fleet of operational machines, it also matters if the 3 parts are used at the same time unit or if their use is spread uniformly over the year (impact also depending on the replenishment time of the component). In the latter case, it might suffice stocking a single part under the assumption that the lead time is short enough. This policy, however, for the former case, would result in a long downtime, since the service engineer would need to wait for the manufacturing of the two required spare parts.

In an ideal situation, the statistician has homogeneous data, and the data is rich enough that model match-

† ASML Holding BV, headquartered in Veldhoven, Netherlands, is a world leader in the design and manufacturing of photolithography systems, crucial for producing integrated circuits. See: https://www.asml.com

ing is not a problem. However, in practice this is difficult to achieve (cf. Hogg at al. (2010), Stauffer (2014)). Machines are put into service at irregular intervals, so they do not have the same start date. In addition, technological processes are constantly developing, which leads to a situation where newer machines are structurally different from older ones, although in practice, the same designs are used for them. Another element that spoils homogeneity is the different use of machines. The same type of machine is sent to many different customers, who use their machines in different environmental conditions (e.g., ambient temperature or humidity) and under different operational protocols (e.g., throughput or weekend downtime). Also, globalization contributes to the inhomogeneity, as machines reach different countries with different work cultures and sociocultural conditions (Matten and Geppert (2004)).

Inaccuracies can of course, occur in any type of data gathering as a result of human errors (cf. Bosu and MacDonell (2021). Here the understanding of the domain knowledge plays a vital role for the statistician. For example, as a result of the huge financial impact of machine downtime, the replacement of a component might not necessarily be caused by the failure of the component itself. It could also be the result of a proactive swap while the machine is down, where the increased hardware costs are justified in an effort to reduce the total impact of downtime (i.e. opportunistic maintenance). In other cases, a component could have been swapped as the result of a misdiagnosis due to the huge time pressure and the high degree of complexity of the machine, as well as due to the interactions between the individual components. Furthermore, the event data database might contain events not relevant to the objective at hand of trying to forecast the behavior. For example, components can be replaced due to a failure, or due to the pro-active roll-out of redesign. The former can be seen as a continuous process triggering events in the entire operational fleet of machines, whereas the latter is only triggered by human intervention and will cause any model to overestimate the rate of events.

Note that such data pollutions might not cause problems with the quality of customer service directly, but can cause complications when later analyzing the data post hoc (e.g., the preventive swap of a non-failed part). This requires meticulous reconstruction of the course of events only to supplement the missing information. Unfortunately, this can be very time-consuming and the effect is not always satisfactory. Automating the data cleaning might be a possibility, but achieving high enough accuracy can be challenging given the many different factors polluting the data.

This work presents a range of models and places them in the context of high-tech industries such as ASML. According to the knowledge of authors, there is no single comprehensive review that covers these models.

2. Statistical inference for reliability analysis

In the previous section, we highlighted the importance of making the right business decisions with respect to securing the optimal quantity of spare parts to guarantee customer service, and we described various challenges in analyzing the underlying data. However, it is important to note that in order to achieve optimal stocking (and more generally to reach optimal business decision making), one first needs to model the data accurately. There are a huge number of theoretical models, but to achieve useful estimations, predictions, and decisions, the underlying assumptions of those models need to be in line with the characteristics of the operational data and processes. This is where the field of applied mathematics and the industrial world come together. Choosing and optimizing the right statistical models and dealing with the inherent imperfections in the data needs to be performed hand-in-hand with a thorough understanding of how the machines are used in the field and which boundary conditions are valid.

In this section, we review several mainstream models from the literature and discuss the challenges encountered when modeling real-world data originating from a fleet of highly complex, high-tech machines. The primary challenge lies in analyzing and modeling a multitude of components, each with only a limited number of failures across the fleet. A secondary challenge is selecting or adapting models that accurately reflect the essential characteristics observed in practice. This often requires modifying existing models to better capture these critical elements. A further significant difficulty arises in estimating model parameters given the scarcity of available data. Overall, the process of analyzing and modeling such data using statistical methods highlights both the power of applied mathematics in the high-tech domain and the intricate challenges that can elevate this science to an art form.

Understanding and predicting the occurrence of discrete events over a certain period of time or over usage cycles is very important for high-tech companies, as detailed in the previous sections. We shall discuss these

discrete events (or counts) in terms of component failures, but as noted earlier they may also be replacements, maintenance actions, or warranty claims for example.

Mathematicians usually consider either the lifetime distribution (e.g. Weibull or lognormal) or the condition process (e.g. Poisson or Gamma process). However, the former relies on individual mathematical modeling per component. This is costly in terms of data requirements and in practice data is scarce especially when failures are rare. By aggregating the data and looking at the number of failures, we have more data (for analysis and model fitting), but this comes at the expense of the model being less detailed. This is a necessary trade-off, and therefore we will focus on counts of the events. For industry, such count models provide the easiest framework for analyzing failure frequencies as they are usually available as functions in mainstream data analysis and visualization tools, making them very easy and quick to implement.

2.1. *Count distribution models*

The foundational model for count data is the Poisson (regression) model. In this model, events occur independently at a constant average rate per time unit. It is suitable when sufficient failures occur. In the regression context, the failure rate can be considered as a function of covariates (e.g. operating temperature, age). One of the underlying assumptions of this model is equidispersion, i.e. that the mean equals the variance.

However, in our practice, we observe that the variance across the number of counts might exceed the mean (leading to overfitting). This can be attributed to population heterogeneity (even when considering data for the same component) caused by variations in the production processes of the component or by changes in the design or by usage and environmental factors.

Even in a highly reliable environment that adheres to all the challenges mentioned in the previous section, we are often dealing with only a small number of failures over the entire population or a fleet of similar components. E.g., a specific component only broke down in four machines out of thousands of machines where the part did not (yet) break. Due to the low intensity of parts with few failures, there will be many periods without any event (zeros), thereby leading to overdispersion, which makes the Poisson model inadequate. This overdispersion is seen in the data as the number of zeros being greater than would be expected from a Poisson distribution. Data with such an excess of zero counts can be described as zero-inflated, and as such the Zero Inflated Poisson (ZIP) model is better suited.

Zero-inflated models are generalizations of the classical models in which the zero-inflated data can arise as a mixture of two distributions. The first distribution generates zeros. The second distribution, which is associated with the Poisson distribution generates counts, some of which may be zeros. These excess zero counts are a logical result of a very robust component and/or production process, which is resistant to common cause variation. For example, when manufacturing equipment is in a normal operating state, defects may be nearly impossible. But when an extraordinary event occurs, this special cause variation causes defects according to a Poisson distribution. Such a situation has been modeled by Lambert (1992). She showed its usefulness in the statistical analysis of soldering defects on printed wiring boards. Since then, the model has been used in a wide range of research, including health insurance (eg. Mouatassim and Ezzahid (2012)), natural calamities' data (Beckett et al.(2014)), or software reliability engineering (Khoshgoftaar et al. (2001)). Denks et al. (2020) used a ZIP model for the microbubble counts in super-resolution voxels. In their work, they propose a formula for estimating the model parameters based on the Maximum Likelihood Estimation (MLE) method.

In our work submitted for MIMAR2025 (Skarupski et al.(2025)), we applied this method and estimated the model parameters following the proposed formula. However, the challenge we were confronted with was that the estimation of the zero-inflated probability was negative. This astonishing result indicated to us the limitations of the method, which to the best of our knowledge were not previously reported in the literature. The computed negative probability is the result of not meeting the assumptions of the estimation procedure. First, we showed what conditions must be met to perform the estimation procedure. We also proposed a modification of the estimators to enable estimation in extreme cases of rare and single events.

2.2. *Stochastic process models*

The Poisson models described above assume a constant behavior of the components over time. However, in industry we often deal with changes over time in the underlying process as described earlier. Typically, the fleet

of machines is manufactured over many years and can remain operational for more than some decades even. During the production of the individual components, the final characteristic properties of the components may be vary due to changes of suppliers, variations within the production processes, or small redesigns for example. Furthermore, customers continuously try to squeeze the most performance out of their machines, causing an ever increasing operational load on the components and thereby a slightly increasing failure rate over time.

Assuming there are sufficient data points within the population to model a growth factor of the underlying process, a simple adjustment is to use, among others, (weighted) moving-average or auto-regressive models. When this change is slow enough over time, and the time window is chosen correctly, such models are reasonably able to account for the change over time. However, when there are abrupt changes these models inherently fail.

Another commonly used alternative for logistical planning is Croston's method for intermittent demand forecasting (see Croston (1972)). The Croston model is a statistical method used for forecasting time series data with intermittent demand. Intermittent demand is characterized by periods of zero demand interspersed with occasional periods of non-zero demand. This pattern is common in industries dealing with slow-moving items, spare parts, or new product launches. However, it can also not account for trends or seasonality, and will also underestimate the behavior when the rate of increase is on a faster timescale than the timescale used in the model. Teunter et al. (2011) have explored its utility in predicting maintenance needs for components with low and sporadic failure rates

Failed parts are usually treated as unrepairable systems in mathematical models and there is a fairly extensive literature review on such systems, eg. Lawless (2002), O'Conor and Kleyner (2025). This approach allows for determining basic reliability characteristics for a given part. In addition, repairable system models constitute a separate group of models. The differences and similarities between the two approaches were discussed by, among others, Thompson (1981). However, both approaches have their drawbacks in the case of rare events. One solution to deal with these rare events is to aggregate all machines of an entire fleet. When the component is present in all machines of the fleet, then these machines effectively create a single system, in which we consider only the failure of the observed component as an event. In this way, even if the failures of parts in the machine are rare events, the proposed aggregation creates a repairable system to model.

This can then be done by the Crow-AMSAA process, which is often used in the field of reliability engineering to model the growth of the reliability of a system over time. Note that by applying the Crow-AMSAA model to a fleet of machines, we loose the ability to predict which machine will be impacted and we can only predict the average number of failures to be expected within the entire fleet.

The Crow-AMSAA model is based on a non-homogeneous Poisson process $N(t)$ with intensity $\lambda(t) = \lambda\beta t^{\beta-1}, \lambda, \beta > 0$. This model allows for a simplification of reality and a fairly accurate adjustment of actual data. The model allows for the prediction of future failure occurrences, enabling proactive maintenance and preventing costly downtime. It provides a quantitative basis for making decisions related to maintenance, testing, and system improvements. The β parameter allows us to conclude at what stage of operation the system whose failure process we are modeling is. If $\beta > 1$ this indicates a deterioration in system reliability, and if $0 < \beta < 1$ it indicates reliability growth. In the case of $\beta = 1$ we are dealing with constant reliability, which indicates complete randomness of failures.

Parameter estimation is a key point of data modeling and can be done in different manners for the Crow-AMSAA model. The basic model is the graphical plot (cf. Duane (1962)). The moments of failures are marked on a log-log scale, and then a trend line is determined using linear regression. The slope of this line is the value of the parameter estimator β. This method is characterized by extraordinary simplicity and is therefore eagerly used in the engineering environment. Extrapolation of the trend line also allows for the prediction of the cumulative number of failures in the future, which facilitates the management of spare parts. The linear behavior of the data results from the law of large numbers. Therefore, in the case of small amounts of data, the model assumptions are violated. Also the general sensitivity of least squares regression to outliers, potentially amplified by the logarithmic scale, remains a concern.

The second estimation method is the maximum likelihood method. There are two ways to determine the parameters. The first is based on the knowledge of the times of subsequent reports in the process $0 = t_0 < t_1 < \ldots < t_n$. The joint distribution of this vector is determined based on the conditional density of the variable t_i: $f(t_i|t_{i-1}) = \lambda\beta t^{\beta-1}\exp(-\lambda t_i^{\beta} + \lambda t_{i-1}^{\beta}), i = 1, \ldots, n$ (cf. Crow (1975)). The second approach assumes that we

do not have the times of subsequent calls, but we only have the total number calls grouped within specific time intervals. That is, we have a vector of process observation times $0 = s_0 < s_1 < ... < s_n$. Then we know that the increments of the Poisson process are independent, and the magnitudes of these increments are Poisson distributed with parameter $\lambda(s_i^\beta - s_{i-1}^\beta), i = 1, ..., n$. Therefore, depending on the data available, a different form of likelihood function is used to estimate the model parameters. It is important that the Crow-AMSAA model shows some flexibility with respect to the type of data analyzed. Maximum likelihood estimators are biased estimators. Adjustment for unbiasedness is possible and, in the case of small data, improves the accuracy of the estimation somewhat. Unfortunately, this does not solve the problem entirely.

As mentioned earlier, it is important to not only predict the failure rate for each spare part but also to model its variation accurately. When components are present in a single machine, or when multiple machines have a component failure within the same time interval, multiple events can be registered in the data. The Crow-AMSAA model for grouped data includes different jump sizes, but it does not allow for the variability analysis mentioned above. The non-homogeneous compound Crow-AMSAA process does account for this, where the cumulative number of reports is modeled using the process $Y_t = \sum_{i=1}^{N_t} X_i$, where N_t is the Crow-AMSAA process and the jump sizes $X_1, ..., X_n$ are independent random variables with the same discrete distribution and independent of the process N_t. Note that special cases of this process are: the Hermite Process, the Negative-Binomial Process or the Polya-Aeppli Process (see Wimmer and Altmann (1996)).

Despite the aggregation of machines into one fleet, it is not always possible to obtain satisfactory results. In a highly reliable environment, it may still turn out that we do not have enough data after aggregation. Then the statistical analysis is highly distorted. Simulating CrowAMSAA for different scale and shape parameters, shows that the error diverges for small data sets up to a an error of 140% when having only 5 data points. A solution that helps improve estimation quality is using a priori knowledge. This leads to the introduction of Bayesian methods. Newby (1991) considered the case when the prior for the shape parameter is deterministic and the prior distribution for the scale parameter is gamma; however, after finding the posterior distribution, he claimed that it is impossible to find an analytic form of the Bayesian estimator. For the Bayesian framework, we refer to eg. Guida et al.(1989), Bar-Lev et al. (1992). Those methods lead use of multidimensional prior distribution and therefore can produce complexity in the calculation of posterior parameters. This problem is usually solved by numerical methods, eg. Monte Carlo Markov Chains or Gibbs sampling. The use of Bayesian methods significantly improves the results. Simulation studies have shown that the parameter estimation error has decreased by at least twofold, and in extreme cases even sevenfold.

3. Conclusion

We discussed the main challenges for modeling the occurrence of rare events in a fleet of high-tech machines, focusing of predicting the number of events within a given time period. We discussed various causes of heterogeneity and data imperfections resulting from the operational context of the individual components. Furthermore, we investigated various different models ranging from simple counting models such as Poisson distribution, to more advanced models such as Crow-AMSAA (non-homogeneous Poisson process) which aim to also model changes in the temporal data. To support an optimal business impact it is of vital importance to combine expertise on both the domain knowledge (the industry) with expertise on the statistical modeling (the statistician). The perfect model still provides poor insights when fed with poor data, and the perfect data will still fail to give the correct insights when fed into a model of which the boundary conditions do not match to the actual problem. Finally, also a proper understanding of how the insights will be used is important, where different models could all predict a comparable mean value, but some fail to accurately predict the variability.

Acknowledgements

The work of Marek Skarupski has been partially funded by the NWO grant NWA.1160.18.238 (PrimaVera).
The work of Stella Kapodistria is supported by NWO through the Gravitation-grant NETWORKS-024.002.003.
We want to thank the reliability and sourcing and supply chain departments of ASML for their insights and fruitful discussions.

References

Bar-Lev, S., Lavi, I., Reiser, B. (1992) Bayesian inference for the power law process. *Annals of the Institute of Statistical Mathematics*, 44(4), 623—639

Beckett, S., Jee, J., Ncube, T., Pompilus, S., Washington, Q., Singh, A., Pal, N. (2014) Zero-inflated Poisson (ZIP) distribution: parameter estimation and applications to model data from natural calamities. *Involve: A Journal of Mathematics*, 7(6), 751 – 767.

Bosu, M.F., MacDonell, S.G. (2021) Data Quality in Empirical Software Engineering: A Targeted Review, *arXiv 2105.10895* , https://arxiv.org/abs/2105.10895

Chen, T., Lin, W., Sheu, Ch., (2018) A Dynamic Failure Rate Forecasting Model for Service Parts Inventory *Sustainability*, 10(7), 2408

Croston, J.D.: Forecasting and Stock Control for Intermittent Demands.(1972) *Operational Research Quarterly (1970-1977)*, Vol. 23, No. 3 (Sep., 1972), pp. 289-303

Crow, L.: Reliability analysis for complex, repairable systems. *Tech. rep., Army Materiel Systems Analysis Activityberdeen Proving Ground MD* (1975)

Dencks, S., Piepenbrock, M., Schmitz, G. (2020) Assessing Vessel Reconstruction in Ultrasound Localization Microscopy by Maximum Likelihood Estimation of a Zero-Inflated Poisson Model. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 67(8), 1603–1612

Duane, J.T., Learning Curve Approach To Reliability Monitoring, *IEEE Transactions on Aerospace*, Vol. 2, pp. 563-566, 1964.

Guida, M., Calabria, R., Pulcini, G. (1989) Bayes inference for a non-homogeneous Poisson process with power intensity law (reliability). *IEEE Transactions on Reliability* 38(5), 603–609

Hogg, D. W., Bovy, J., Lang, D. (2010) Data analysis recipes: Fitting a model to data. *arXiv 1008.4686*, https://arxiv.org/abs/1008.4686

Jaarsveld, W. van, and Dekker, R. (2011) Spare parts stock control for redundant systems using reliability centered maintenance data, *Reliability Engineering & System Safety*, vol. 96, Issue 11, pp. 1576–1586

Khoshgoftaar, T., Gao, K., Szabo, R. (2001) An application of zero-inflated Poisson regression for software fault prediction. *In: Proceedings 12th International Symposium on Software Reliability Engineering*, pp. 66–73

Lambert, D. (1992) Zero-inflated Poisson Regression, with an Application to Defects in Manufacturing. *Technometrics* 34(1), 1–14

Lawless, J.F. (2002) Statistical Models and Methods for Lifetime Data. *John Wiley & Sons, Inc.* 2 ed.

Matten, D. and Geppert, M. (2004) Work systems in heavy engineering: the role of national culture and national institutions in multinational corporations, *Journal of International Management*, Vol. 10, Issue 2, pp. 177–198

Mouatassim, Y., Ezzahid, E.H. (2012) Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. *European Actuarial Journal* 2, 187–204

Newby, M. (1991) Repairable systems and growth models. In: P. Sander, R. Badoux (eds.) *Bayesian Methods in Reliability*, 101–117. Springer Netherlands, Dordrecht

O'Connor, P. D. T., Kleyner, A. V. : Practical Reliability Engineering, *Wiley* 6th Edition, 2025

Skarupski, M., Veenvliet, M., Wouters, M., Kapodistria, S.: Zero inflated model for rare failure analysis, *MIMAR* 13th, 2025

Stauffer, R. (2014). Of Processes and Project Baselines: Why homogeneity matters. *iSixSigma*. Retrieved from https://www.isixsigma.com/statistical-analysis/of-processes-and-project-baselines-why-homogeneity-matters/

Teunter, R.H., Syntetos, A.A., Zied Babai, M., (2011) Intermittent demand: Linking forecasting to inventory obsolescence, *European Journal of Operational Research*, vol. 214, Issue 3, pp. 606–615

W.A. Thompson (1981), On the Foundations of Reliability, *Technometrics* 23 no. 1, 1-13

Wimmer, G., Altmann, G. (1996). The multiple Poisson distribution, its characteristics and a variety of forms. *Biometrical journal*, 38(8), 995-1011.

Evaluation of the health state of oil-immersed transformer based on dissolved gas analysis: An empirical study

Huy Vu Tran*, Kim Anh Nguyen*, Duc Hanh Dinh**, Dinh Duong Le*, Huu-Truong Le***, Phuc Do***

**Faculty of Electrical Engineering, The University of Danang - University of Science and Technology, Da Nang, Vietnam: thvu@ncs.dut.udn.vn, nkanh@dut.udn.vn (Corresponding author), ldduong@dut.udn.vn*

*** Faculty of Electrical Engineering, The University of Danang - University of Science and Technology, Da Nang, Vietnam: ddhanh@dut.udn.vn*

**** Université de Lorraine, CRAN, CNRS, UMR-7039, Nancy 54000, France: huu-truong.le@univ-lorraine.fr, phuc.do@univ-lorraine.fr*

Abstract. Power transformers are vital for grid reliability, but insulation degradation remains a major failure risk. Dissolved gas analysis (DGA) is widely used for diagnostics, yet traditional methods lack sensitivity to early faults and face field implementation challenges. This study proposes the Dissolved Gas Analysis Factor (DGAF), a novel health index integrating concentrations of seven key gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, CO_2) with optimized weights to improve fault detection. Using 871 DGA samples from 240 transformers, optimization boosted sensitivity to 89.36% (97.59% accuracy, 98.06% specificity) and identified 16 latent faults among 824 normal samples, validated by Duval methods showing T2 thermal trends. A logistic regression model links DGAF to failure probability, aiding risk-based maintenance. By enabling early fault detection, DGAF supports proactive interventions to minimize outages and costs.

1. Introduction

Power transformers are vital for electrical grids, managing voltage and power distribution. Failures from insulation degradation, arcing, or thermal faults cause costly outages, environmental risks, and grid instability. Dissolved Gas Analysis (DGA) monitors health by analyzing gases like H_2, CH_4, C_2H_4, and C_2H_2, indicating faults such as partial discharge (PD), discharges (D1, D2), and thermal faults (T2, T3). However, traditional methods like Duval's Triangle and Rogers' Ratio miss early faults and struggle with mixed classifications, limiting proactive maintenance.

Recent studies enhance diagnostics with health indices. Jahromi *et al.*, (2009) combined DGA with oil and electrical tests, but it needs extensive data, impractical for routine use. Wattakapaiboon and Pattanadech (2016) used gas trends, though subjective weighting reduced applicability. Machine learning (Li *et al.*, 2023) improves fault detection but requires large datasets and lacks field-friendly interpretability, highlighting the need for a simple, sensitive DGA-based index.

This study introduces the DGAF, a new health index from 871 DGA samples of 240 transformers (ages 5–35 years, 110 kV). DGAF enhances fault detection with optimized weights and modeling, raising sensitivity from 72.34% to 89.36%, and detects 16 latent T2 faults in 824 normal samples, confirmed by traditional methods. A logistic regression model estimates failure probability for risk-based maintenance. Goals include: *(i)* validating DGAF for accuracy, *(ii)* enhancing early fault detection, and *(iii)* prioritizing high-risk transformers. Sections 2–5 cover methods, methodology, results, and conclusions.

2. Materials and Methods

This study evaluates the health state of oil-immersed transformers using DGA through an empirical approach. The methodology encompasses data collection, gas analysis, statistical processing, and health index formulation based on real-world transformer data. This section details the materials, procedures, and analytical techniques employed, ensuring the study's reproducibility and robustness.

2.1. Dataset description and performance evaluation

This research employed a dataset comprising 871 DGA samples derived from 240 oil-immersed power transformers managed by electrical utilities. The transformers exhibited a diverse range of characteristics, with ages spanning from 5 to 35 years (averaging 22 years), voltage ratings between 110 kV and 220 kV, and capacities from 25 MVA to 63 MVA. Each transformer utilized mineral oil as the insulating medium, adhering

to IEC 60296 standards. Seven key gases were analyzed and reported in parts per million (ppm): hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2). Within the dataset, 824 samples were classified as Normal (N), while 47 samples were identified as faulted, categorized as follows: 6 samples exhibited Partial Discharge (PD), 2 showed Low-Energy Discharge (D1), 4 indicated High-Energy Discharge (D2), 4 displayed Discharge-Thermal (DT) faults, 7 experienced Medium-Temperature Overheating (T2), and 24 demonstrated High-Temperature Overheating (T3). Analysis of 871 DGA samples showed clear gas patterns in Figure 1. These align with IEEE C57.104-2019, supporting DGAF diagnostics.

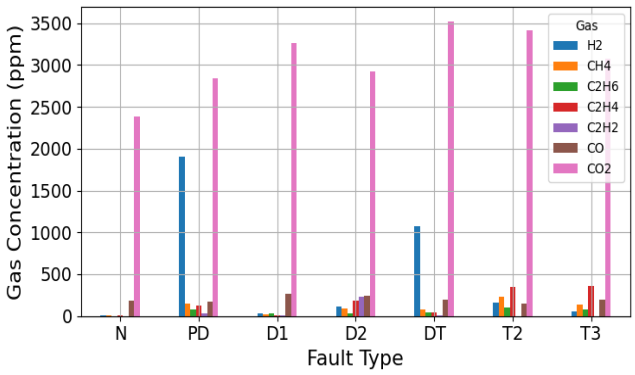


Figure 1. Mean gas concentrations by fault type.

The study evaluates the proposed method using key metrics from Zhu *et al.*, (2010). These metrics ensure a robust assessment of the method's performance in transformer health diagnostics

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}; \quad (1)$$

$$Sensitivity = \frac{TN}{TP+FN}; \quad (2)$$

$$Specificity = \frac{TN}{TN+FP}. \quad (3)$$

Where, *TP*, *TN*, *FP*, and *FN* are the counts of true positives, true negatives, false positives, and false negatives, respectively.

2.2. *DGAF formulation*

To enhance the accuracy of transformer condition assessment, a health index termed the DGAF was developed based on a weighted scoring system, assigning scoring factors (S_i) and weighting coefficients (W_i) to individual gases. The condition score for each gas is determined based on IEC 60599-2007/ IEEE C57.104-2019 standards (Prasojo *et al.*, 2020; Tamma *et al.*, 2021; *Li et al.*, 2023). These parameters were optimized to improve DGAF's sensitivity in distinguishing between minor and severe fault conditions (Haema *et al.*, 2013; Murugan *et al.*, 2019). DGAF was calculated as follows:

$$DGAF = \frac{\sum_{i=1}^{7} S_i.W_i}{\sum_{i=1}^{7} W_i}. \quad (4)$$

Where S_i is the score assigned to each gas based on its concentration (1 to 6) and W_i is the weighting factor reflecting its diagnostic importance (1 to 5).

These materials and methods provide a robust foundation for DGA-based transformer health assessment, leveraging a real-world dataset, standardized sampling, and rigorous statistical analysis.

3. Proposed method

This study introduces DGAF, an empirical health index from DGA data, using gas trends, weights, and modeling for reliable transformer assessment. It details DGAF's development, optimization, and validation with 871 samples from 240 transformers.

3.1. *DGAF framework*

The DGAF quantifies transformer health by combining the concentrations of seven key gases, namely hydrogen (H_2), methane (CH_4), ethane (C_2H_6), ethylene (C_2H_4), acetylene (C_2H_2), carbon monoxide (CO), and carbon dioxide (CO_2), into a single index (Eq. (4)). The DGAF value is then optimized and mapped to a condition rating (Table 1 and 2).

3.2. *Optimization of weighting factors*

Analysis of the 871-sample dataset revealed that the initial W_i values did not fully capture fault-specific gas trends. For instance, H_2 showed extreme elevation in partial discharge (PD) cases (e.g., 10,055.22 ppm), while

C_2H_4 dominated in high-temperature overheating (T3) samples (e.g., 2809.364 ppm). To improve sensitivity, the weighting factors were optimized as follows (Table 1).

Table 1. Optimized scoring and weighting actors.

Gas	Score (S_i)						Init. W_i	Opt. W_i	Noted
	1	2	3	4	5	6			
H_2	≤100	100-200	200-300	300-500	500-700	>700	2	3	Increased for PD, D1, D2
CH_4	≤75	75-125	125-200	200-400	400-600	>600	3	2	Reduced, less impact in severe faults
C_2H_6	≤65	65-80	80-100	100-120	120-150	>150	3	2	Reduced, similar to CH_4
C_2H_4	≤50	50-80	80-100	100-150	150-200	>200	3	4	Increased for T2, T3, D2
C_2H_2	≤3	3-7	7-35	35-50	50-80	>80	5	6	Increased for arcing
CO	≤350	350-700	700-900	900-1100	1100-1400	>1400	1	3	Increased for insulation aging
CO_2	≤2500	≤3000	≤4000	≤5000	≤7000	>7000	1	2	Increased, less diagnostic alone

The optimization was qualitative, guided by fault distributions (e.g., 824 normal vs. 47 faulted samples) and gas concentration patterns. For example, in a T3 sample (DGAF = 3.090909091), C_2H_4 reached 2809.364 ppm, justifying its higher weight. The total weight denominator increased from 18 to 22, normalizing the DGAF scale. The DGAF value is then mapped to a proposed condition rating (Table 2)

Table 2. Transformer condition ratings based on optimized DGAF.

Rating Code	Condition	Initial Wi	Optimized Wi
A	Good	DGAF < 1.2	DGAF < 1.2
B	Acceptable	1.2 ≤ DGAF < 1.5	1.2 ≤ DGAF < 1.5
C	Need	1.5 ≤ DGAF < 2	1.5 ≤ DGAF < 2
D	Poor	2 ≤ DGAF < 3	2 ≤ DGAF < 2.5
E	Very	DGAF ≥ 3	DGAF ≥ 2.5

3.3. *Failure probability estimation*

To enhance risk-based maintenance, a logistic regression model estimates the failure probability *P(Fail)* of transformers using DGAF values. The model uses the failure probability and is defined as follows:

$$P(Fail) = \frac{e^{(-\beta_0+\beta_1.\mathrm{DGAF})}}{1+e^{(-\beta_0+\beta_1.\mathrm{DGAF})}} \tag{5}$$

Where, β_0 and β_1 are regression coefficients and determined through statistical fitting using the 871-sample dataset (824 normal, 47 faulted). Using python, the logistic regression algorithm iteratively minimized the log-loss function to optimize β_0 andβ_1, yielding β_0=-13.665 and $\beta_1 = 7.8342$. The negative β_0 reflects a low baseline failure probability for normal transformers (low DGAF), while the positive β_1 indicates that higher DGAF values exponentially increase *P(Fail)*, aligning with fault severity trends (e.g., PD, T3). This probabilistic approach empowers utilities to prioritize maintenance effectively, reducing outage risks and operational costs.

3.4. *Application procedure*

The proposed method for evaluating transformer health follows a structured application procedure. Initially, gas concentration inputs are obtained by measuring seven key gases (H_2, CH_4, C_2H_6, C_2H_4, C_2H_2, CO, and CO_2) via gas chromatography, reported in parts per million (ppm). Next, scores (Si) are assigned to each gas based on predefined concentration ranges outlined. These scores are then used to compute the DGAF by applying optimized weighting factors (Wi) specified in Table 1, resulting in a single health index. The transformer's condition is subsequently classified into one of five categories (A to E) according to Table 2, reflecting its health status from "Good" to "Very Poor."

Following this, the failure probability *P(Fail)* is estimated using a logistic regression model, providing a quantitative risk assessment. Based on the DGAF value, maintenance decisions are recommended: inspection is advised for DGAF ≥ 1.5, and immediate action is required for DGAF ≥ 2.5.

4. Results

This section presents the outcomes of applying the DGAF to assess the health of oil-immersed transformers using a dataset of 871 DGA samples from 240 transformers. The results compare the performance of initial and optimized DGAF parameters, evaluate fault detection accuracy, analyze failure probability predictions, and explore practical implications. Limitations and future research directions are also discussed to contextualize the findings.

4.1. *DGAF performance*

The performance of the Dissolved Gas Analysis Fault (DGAF) method was evaluated by comparing its initial and optimized parameters across various fault types, using a dataset of 871 transformer samples.

The initial and optimized DGAF results for 871 transformer samples are compared in Table 3. Initially, 84 samples were "Acceptable," 753 "Good," 7 "Need Caution," 10 "Poor," and 17 "Very Poor." Normal (N) samples (824) included 75 "Acceptable" and 749 "Good," with 824 true negatives (TN) and 0 false positives (FP). Faulted samples showed: PD (6) with 3 "Good," 1 "Need Caution," 2 "Very Poor" (3 true positives (TP), 3 false negatives (FN)); T3 (24) with 4 "Acceptable," 4 "Good," 3 "Need Caution," 6 "Poor," 7 "Very Poor" (16 TP, 8 FN). Total initial TP was 34, FN 13.

After optimization, classifications shifted to 103 "Acceptable," 710 "Good," 31 "Need Caution," 15 "Poor," and 12 "Very Poor." Normal samples had 102 "Acceptable," 706 "Good," and 16 "Need Caution" (808 TN, 16 FP). PD improved to 4 "Need Caution," 1 "Poor," 1 "Very Poor" (6 TP, 0 FN); T3 had 4 "Good," 7 "Need Caution," 8 "Poor," 5 "Very Poor" (20 TP, 4 FN). Total optimized TP rose to 42, FN dropped to 5, but FP increased to 16.

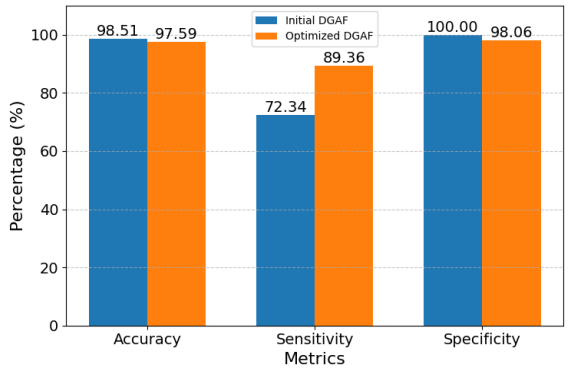


Figure 2. Performance comparison between initial dgaf and optimized DGAF methods.

Figure 2 compares the performance metrics of the initial and optimized DGAF methods. The initial DGAF achieved an accuracy of 98.51%, a sensitivity of 72.34%, and a specificity of 100.00%. In contrast, the optimized DGAF had a slightly lower accuracy of 97.59%, a higher sensitivity of 89.36%, and a reduced specificity of 98.06%. The decrease in accuracy and specificity indicates that the optimized method flagged more Normal samples as needing caution (16 FP), but the significant increase in sensitivity highlights its improved ability to detect latent faults, particularly in samples initially classified as Normal (N).

Table 3. Optimized DGAF results by fault type.

Fault Type	Optimized DGAF										Initial DGAF			
	Acceptable	Good	Need caution	Poor	Very poor	Total	TP	FN	TN	FP	TP	FN	TN	FP
D1	1		1			2	1	1			1	1		
D2			1	2	1	4	4				4			
DT			1	1	2	4	4				3	1		
N	102	706	16			824			808	16			824	0
PD			4	1	1	6	6				3	3		
T2			1	3	3	7	7				7			
T3		4	7	8	5	24	20	4			16	8		
Total	**103**	**710**	**31**	**15**	**12**	**871**	**42**	**5**	**808**	**16**	**34**	**13**	**824**	**0**

Table 4 details the 16 Normal samples flagged as "Need Caution" by the optimized DGAF, revealing elevated gas levels that suggest latent issues. For instance, one sample showed C_2H_4 at 36.131 ppm and CH_4 at 18.666 ppm, levels indicative of potential thermal faults. Validation using traditional methods, such as the Duval Triangle and Pentagon, confirmed T2 fault tendencies in all 16 samples, aligning with the DGAF's assessment. The optimized DGAF's ability to identify these 16 potential issues within 824 Normal samples underscores its enhanced reliability for maintenance decisions. By detecting latent faults early, it enables proactive interventions, reducing outage risks and associated costs.

Table 4. Details of 16 "Need caution" cases.

H2	CH4	C2H6	C2H4	C2H2	CO	CO2	Actual Fault	DGAF	Duval Triangle1	Pentagon1
5.744	4.535	1.416	8.099	0	446.836	7841.675	N	1.590909	N	T2
0	2.619	1.192	1.562	0	406.105	7623.438	N	1.590909	N	T2
7.92	7.744	5.026	28.465	0	614.962	12110.64	N	1.590909	N	T2
…	…	…	…	…	…	…	…	…	…	…
10.321	5.217	2.524	3.825	0	634.073	7042.789	N	1.590909	N	T2
12.37	10.148	4.275	14.241	0	587.658	7488.928	N	1.590909	N	T2
7.777	15.123	5.787	36.131	0	441.52	8053.905	N	1.590909	N	T2

The optimized DGAF method demonstrates a notable improvement in sensitivity (89.36% vs. 72.34%), allowing it to better identify latent faults in transformers initially diagnosed as Normal by traditional methods. However, this comes at the cost of slightly lower accuracy (97.59% vs. 98.51%) and specificity (98.06% vs. 100.00%), as it introduces some false positives. The ability to flag 16 out of 824 Normal samples for further inspection, as detailed in Table 4, highlights the optimized DGAF's potential to enhance transformer health monitoring and support proactive maintenance strategies. By detecting latent faults early, it enables proactive interventions, reducing outage risks and associated costs.

4.2. *Failure probability analysis*

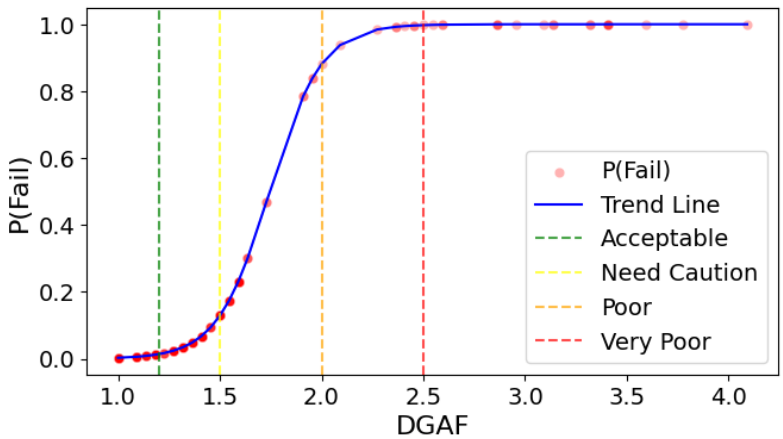


Figure 3. P(Fail) probability distribution chart according to DGAF on 871 samples.

The logistic regression model for the optimized DGAF, as depicted in Figure 3, effectively estimates the failure probability *P(Fail)* based on 871 samples, plotting the probability curve against DGAF values alongside data points for normal and faulted samples. Normal samples exhibited a low *P(Fail)* of 0.0029 at DGAF = 1.07, indicating negligible risk. In contrast, faulted samples showed significantly higher probabilities: a PD sample with DGAF = 3.11 had P(Fail) = 0.9987, reflecting severe risk. A threshold analysis revealed *that* a P(Fail) cutoff of 0.5 (corresponding to DGAF ≈ 1.8) correctly separated 95% of faulted samples (44/47) from normal ones, supporting robust classification. Table 5 details P(Fail) for representative samples.
This method enhances maintenance prioritization by quantifying risk, enabling proactive interventions to mitigate transformer failures.

Table 5. Example P(Fail) values for selected samples.

Fault Type	DGAF (Optimized)	*P(Fail)*	Maintenance Action Recommended
N	1.07	0.0029	Routine monitoring
PD	3.11	0.9987	Immediate inspection
D1	1.95	0.8385	Urgent inspection
D2	2.56	0.9946	Priority inspection

Fault Type	DGAF (Optimized)	*P(Fail)*	Maintenance Action Recommended
DT	3.59	0.9999	Immediate inspection
T2	3.41	0.9999	Immediate inspection
T3	2.71	0.9973	Priority inspection

5. Conclusion

This study introduces the DGAF, a novel index for transformer health assessment, leveraging 871 DGA samples from 240 oil-immersed transformers. By synthesizing seven gas concentrations into a single metric, DGAF simplifies diagnostics while enhancing early fault detection. The optimized DGAF achieves 89.36% sensitivity, identifying 16 latent T2 thermal faults among 824 normal samples—outperforming conventional methods like the Duval Triangle. Initial accuracy (98.51%) and specificity (100%) slightly declined post-optimization (97.59%, 98.06%) but balanced improved fault detection with minimal false positives. A logistic regression model translates DGAF values into failure probabilities, distinguishing low-risk (e.g., 0.0029 probability at DGAF 1.07) from critical cases (0.9946 at DGAF 2.56) to prioritize maintenance. Designed for field use, DGAF supports proactive, risk-based strategies to reduce outages. Future integration with real-time sensors and machine learning could automate monitoring, addressing current limitations in dataset diversity and weighting methods. DGAF advances transformer diagnostics, offering a practical, scalable tool for grid resilience.

References

Jahromi A, Piercy R, Cress S, Service J, Fan W (2009) An approach to power transformer asset management using health index. *IEEE Electrical Insulation Magazine* **25**, 20-34.

Wattakapaiboon W, Pattanadech N (2016) The new developed health index for transformer condition assessment. In *Proc. 2016 International Conference on Condition Monitoring and Diagnosis*, pp. 32–35.

Li S, Li X, Cui Y, Li H (2023) Review of transformer health index from the perspective of survivability and condition assessment. *Electronics* **12**, 2407.

Zhu W, Zeng N, Wang N (2010) Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. In *Proc. NESUG: Health Care and Life Sciences*, Baltimore, Maryland, **19**:67.

Prasojo RA, Gumilang H, Suwarno, Maulidevi NU, Soedjarno BA (2020) A fuzzy logic model for power transformer faults' severity determination based on gas level, gas rate, and dissolved gas analysis interpretation. *Energies* **13**, 1009.

Haema J, Phadungthin R (2013) Development of condition evaluation for power transformer maintenance. In *Proc. Fourth International Conference on Power Engineering, Energy and Electrical Drives (POWERENG)*, Istanbul, Turkey, 13–17 May 2013, pp. 620–623.

Murugan R, Ramasamy R (2019) Understanding the power transformer component failures for health index-based maintenance planning in electric utilities. *Engineering Failure Analysis* **96**, 274–288.

Brandtzæg G (2015) Health indexing of Norwegian power transformers. Master's thesis, Norwegian University of Science and Technology (NTNU).

Tamma WR, Prasojo RA, Suwarno (2021) High voltage power transformer condition assessment considering the health index value and its decreasing rate. *High Voltage* **6**(2), 314–327.

Guo H, Guo L (2022) Health index for power transformer condition assessment based on operation history and test data. *Energy Reports* **8**, 9038–9045.

Threshold-Based Maintenance for Systems with Unobservable Degradation: A Hybrid Preventive-Corrective Approach

H. Jiménez*, P. Do**, A. Voisin***, and C. Franciosi****

** Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France hanser-steven.jimenez-gonzalez@univ-lorraine.fr*
*** Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France phuc.do@univ-lorraine.fr*
**** Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France alexandre.voisin@univ-lorraine.fr*
***** Université de Lorraine, CNRS, CRAN, F-54000 Nancy, France chiara.franciosi@univ-lorraine.fr*

Abstract. In many industrial applications—particularly those involving failures that are difficult to predict—condition monitoring via sensors or inspections may be desirable. However, such approaches are often impractical due to harsh environments, limited accessibility, or high associated costs. This raises a key question: how can traditional strategies such as time-based preventive maintenance (PM) and run-to-failure corrective maintenance (CM) be effectively combined to balance cost efficiency and system reliability in systems where degradation data is unavailable? This work considers a scenario in which multiple production line stops require corrective interventions on a critical component—either through partial rejuvenation (repairs) or full replacement. Moreover, the effectiveness of repairs diminishes over time due to repeated assembly and disassembly. In this context, the objective is to determine the optimal timing for each type of maintenance. We propose a threshold-based hybrid maintenance strategy, in which imperfect repairs are performed up to a predefined number of consecutive stops. Once this threshold is reached, a preventive replacement is triggered. The approach relies on the history of discrete events to dynamically manage the trade-off between CM and PM, without the need for real-time monitoring. This strategy is especially suitable for systems with unobservable degradation and high uncertainty, such as offshore oil and gas components. Through numerical simulations we show that the proposed policy significantly outperforms conventional CM and PM in settings characterized by high degradation variability.

1. Introduction

In industrial systems, effective maintenance is essential to ensure operational continuity, safety, and cost efficiency. Maintenance strategies are typically designed based on a system's criticality, the predictability of its degradation, and the availability of monitoring technologies.

Condition-based maintenance (CBM) has gained considerable attention in recent years due to its ability to reduce unnecessary interventions and improve reliability (Teixeira et al. 2020; Lee & Mitici, 2023). As highlighted by De Jonge et al. (2020), the growing availability of sensors has led researchers to shift their focus toward predictive maintenance planning, moving away from traditional strategies such as PM or CM.

Yet, several barriers continue to limit the applicability of CBM. Harsh operating environments, component inaccessibility, and the high cost of sensors frequently preclude continuous condition monitoring and inspections (Lee, Bagheri, & Kao, 2015). Furthermore, as noted by De Jonge et al. (2020), even in settings where sensors can be feasibly installed, CBM implementation may still be challenging. Fitting a reliable stochastic degradation model often requires large datasets—something rarely available in practice, especially when conservative maintenance policies have been used for extended periods or when equipment is new and lacks historical data.

Although some research has explored partially observable degradation due to sensor unreliability (see Zhao & Smidts, 2022), there remains a clear gap in how to maintain systems subject to high degradation uncertainty, limited observability without sensors, and a lack of inspection feasibility. In such contexts, the frequency of discrete degradation events—such as system stops—can provide critical insights for improving maintenance decisions even within traditional frameworks.

To address this challenge, we propose a hybrid maintenance strategy based on a repair-count threshold that integrates elements of both PM and CM. In this strategy, the component undergoes imperfect corrective repairs at each system stop, representing low-cost, partial rejuvenations. After a predefined number of such stops, a full preventive replacement is triggered. This approach does not rely on real-time degradation monitoring but

instead leverages the system's failure history to dynamically guide maintenance actions. As such, it offers a practical alternative for systems where inspections or continuous monitoring are infeasible.

The aim of this study is to evaluate the effectiveness of repair-threshold-based policies in environments characterized by high degradation variability. We model degradation using a gamma process, which captures both the stochastic and nonlinear progression of wear commonly seen in industrial components. Through numerical simulations, we benchmark the proposed policy against traditional strategies, including perfect and imperfect CM as well as fixed-interval PM.

This work contributes in two key ways: (1) it introduces a flexible, event-driven maintenance policy tailored for systems without degradation monitoring capabilities; and (2) it demonstrates, through simulation-based evidence, that repair-threshold-based strategies can significantly outperform classical approaches in settings with high uncertainty and limited observability.

2. System description

2.1 System degradation

Consider a single-component system for which the deterioration level can be represented by a scalar variable $X(t)$ with initial value zero. The deterioration process of the system is stochastic and strictly increasing. The system is considered to stop when the degradation level reaches a failure threshold F, which is self-announcing.

The stochastic and non-linear nature of wear can be represented by a gamma deterioration process (Van Noortwijk, 2009) characterized by independent, non-negative increments that follow a Gamma distribution. Given this, it is assumed that the deterioration of the system between two maintenance actions evolves like a gamma stochastic process $\{X(t), t \geq 0\}$, with parameters $\alpha > 0\ (shape)$ and $\beta > 0\ (rate)$, with the following characteristics:

- Initial condition: $X(0) = X(t_k^+)$ is the condition just after the kth maintenance action, either a repair or a replacement.
- Independent degradation increments: For $0 \leq s < t$, the degradation increment $X(t) - X(s)$ is independent of the past.
- Gamma-distributed degradation increments: $X(t) - X(s) \sim Gamma(\gamma = \alpha(t-s), \beta)$, where the Gamma distribution has the probability density function outlined in Eq (1).

$$f(x; \gamma, \beta) = \frac{\beta^{\gamma} * x^{\gamma-1} * e^{\beta x}}{\Gamma(\gamma)}, \qquad x > 0. \tag{1}$$

2.2 System maintenance

It is assumed that continuous monitoring or inspections at discrete times are not possible due to harsh environments, inaccessibility, or excessive costs. Hence, the system can be maintained in three ways: 1) correctively, through a perfect replacement when the system fails; 2) through preventive replacement to avoid failures, when reaching a scheduled replacement time; and 3) correctively, through imperfect repairs (IR) that can fast recover the system to an operational state with reduced maintenance costs but do not fully recover the asset's capacity.

Corrective or preventive replacements are perfect, meaning that they can recover the system to as good-as-new state (AGAN), i.e., $X(t_k^+) = 0$ after which, the deterioration behavior evolves with time according to the speed $v = \alpha/\beta$. IRs do not restore the system to AGAN condition but instead reduce the accumulated degradation from $X(t_k^-)$ to a level $X(t_k^+)$, such that:

$$0 < X(t_k^+) < X(t_k^-), \text{ with } X(t_k^+) = \zeta_k * X(t_k^-). \tag{2}$$

Where: $X(t_k^-)$ is the degradation level just before the kth repair; $X(t_k^+)$ is the degradation level just after the kth repair; the inequality $0 < X(t_k^+) < X(t_k^-)$ captures the partial effectiveness of the IRs where the system is partially rejuvenated and; $\zeta_k \in [0,1]$ is the effectiveness factor of the kth repair (with $\zeta = 0$ meaning AGAN, and $\zeta = 1$ meaning no improvement at all).

Note that when such an IR is carried out, maintenance effectiveness, i.e., the capacity to reduce the accumulated degradation, is reduced for the next repair, meaning that components damage cumulates as the component ages. Therefore, to model the progressive loss of maintenance quality, we define:

$$\zeta_{k+1} = \zeta_k * (1 - \lambda_k). \tag{3}$$

Where $\lambda_k \in (0,1)$ represents the stochastic decay factor of maintenance effectiveness. It is assumed to be a continuous random variable that captures uncertainties related to maintenance quality due to factors such as variability in maintenance team performance, system complexity, and environmental stressors. To reflect this randomness, λ_k is sampled from a truncated normal distribution $\lambda_k \sim TN(\mu, \sigma^2, 0, \lambda max)$ with density:

$$f(\lambda) = \frac{1}{\sigma} * \frac{\phi\left(\frac{\lambda - \mu}{\sigma}\right)}{\Phi\left(\frac{\lambda_{max} - \mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)}, \qquad for\ \lambda \in [0, \lambda_{max}]. \tag{4}$$

Where: $\phi(\cdot)$ is the standard normal PDF; $\Phi(\cdot)$ is the standard normal CDF; μ is the mean decay rate; σ^2 is the variance of decay rate; and λ_{max} is the upper bound to ensure $\lambda_k < 1$ to avoid extremely ineffective repairs.

Given the advantages and disadvantages of the three feasible maintenance actions for the described system, a natural question arises: how can these actions be smartly combined to perform cost-effective maintenance? In the next section, we propose a hybrid policy that combines CM and PM to address this question.

3. Hybrid preventive-corrective threshold-based model

We propose a policy that performs a sequence of IRs followed by a preventive replacement. Specifically, IRs are carried out up to a predefined number referred to as the repair threshold L, which signals the need for a perfect replacement. This policy is based on the idea that IRs provide short-term benefits by partially restoring the system at a lower cost with a low perturbation of the production as their duration is short; however, over time, these repairs may become ineffective due to the progressive deterioration of maintenance quality.

A repair threshold-based policy adapts to changing degradation patterns in uncertain environments, where failures occur stochastically due to variable operating conditions, usage, or environmental factors. By relying on failure frequency rather than fixed schedules or purely reactive actions, it offers a flexible alternative when condition monitoring is not feasible. However, for systems where failures are primarily age-related and predictable, time-based strategies may be more effective.

An additional feature of the proposed policy is the inclusion of a spare parts replenishment period, denoted by δ, which must be respected to ensure the feasibility of preventive replacements. As a result, preventive maintenance is scheduled with a delay of δ from the moment the repair threshold is reached. During this replenishment period, IRs continue to be performed. Algorithm 1 is the optimization algorithm based on simulation that presents the decision logic. In each simulation step i of Algorithm 1, we assess the system's state and determine what type of maintenance is triggered (early repair, waiting repair, or replacement). Each of these has an associated C_i cost, which is composed of the cost of the actions and the cost of downtime.

Here, C_{replac} includes spare parts and labor for replacements, and $P_{planned_replac}$ accounts for planned downtime. , C_{repair} covers repair costs, while downtime penalties for repairs vary: $P_{unplanned}$ applies before the threshold, and $P_{planned_repair}$ applies during δ, when replacement is already scheduled. $P_{planned_repair}$ is considered to be significantly lower than $P_{unplanned}$ because once the failure threshold is reached, a replacement has already been scheduled, and the system is simply awaiting spare part delivery. During this interim period δ, maintenance teams benefit from increased visibility into the upcoming replacement and can proactively prepare for potential failures. Such preparation may include pre-positioning personnel, implementing temporary workarounds, or rerouting operations—measures that effectively reduce actual downtime. Although interventions are still needed when failures occur during δ, the availability of a plan enables quicker, less diagnostic-intensive responses, justifying the reduced cost.

The efficiency of the maintenance policy over time is represented by the cost rate R_n in Eq (5), which is accumulated dynamically.

$$R_n = \frac{\sum_{i=1}^{n} C_i}{T_n}. \tag{5}$$

Algorithm 1. Decision logic of threshold-based policy

Step	Instruction
1	Input: Fixed delay δ, list of failure thresholds to evaluate
2	**For** each $failure\ threshold$ in list:
3	Initialize: $current_time \leftarrow 0$, $age \leftarrow 0$, $repairs_count \leftarrow 0$
4	$preventive_replacement_time \leftarrow \infty$
5	$age_reduction_rate \leftarrow$ initial value

```
        total_cost ← 0
        While current_time < total simulation time:
            Advance current_time and increase age
            Sample degradation ~ Γ(ρ · age, θ)
            If current_time == preventive_replacement_time:
                Perform perfect replacement
                Reset age, repairs_count, age_reduction_rate
                preventive_replacement_time ← ∞
                Add cost: C_replac + P_planned_replac
            Else if degradation ≥ failure threshold:
                Perform imperfect repair
                repairs_count ← repairs_count + 1
                Update age and decay age_reduction_rate
                If repairs_count < repair threshold:
                    Add cost: C_repair + P_unplanned
                Else if repairs_count == repair threshold:
                    Schedule preventive_replacement_time ← current_time + δ
                    Add cost: C_repair + P_planned_repair
                Else:
                    Add cost: C_repair + P_planned_repair
            End if
            Record cost rate = total_cost / current_time
        End while
        Save final average cost rate for this threshold
    End for
    Output: Optimal threshold and cost rate
```

Under a simulation approach, the optimization of the proposed threshold-based policy implies finding a value for L under a certain value of δ, that minimizes the cost rate when taking maintenance decision n times with a total simulation time T_n. The objective function is defined formally in Eq (6):

$$L^* = argmin_{L \in \Theta} R(L), \qquad L \in \Theta = \{ L \in N \mid L_{min} \leq L \leq L_{max} \}. \tag{6}$$

4. Benchmarking

The proposed policy was benchmarked with a pure corrective policy based on perfect replacements, a corrective policy based on IRs, and a hybrid policy combining IRs and time-based preventive replacements.

The cost of the pure corrective policy when a failure occurs is defined by $C_{PC} = c_{replac} + c_d$, where c_d is the maximum cost of downtime incurred by the system when it fails unexpectedly without any preparation.

The cost of the corrective policy when the system fails based on IRs is defined by $C_{IR} = c_{repair} + P_{unplanned}$, and it is similar to that incurred by the proposed policy before reaching the repair threshold.

For the cost of the hybrid policy combining IRs and time-based preventive replacements, we have four possible scenarios depending on whether the scheduled replacement was achieved and the system fails. The scenarios and their respective costs are represented by Fig.1. Under a simulation approach similar to that in Algorithm 1, the optimization of the decision taken in Fig.1 corresponds to finding the replacement interval I that minimizes the cost rate.

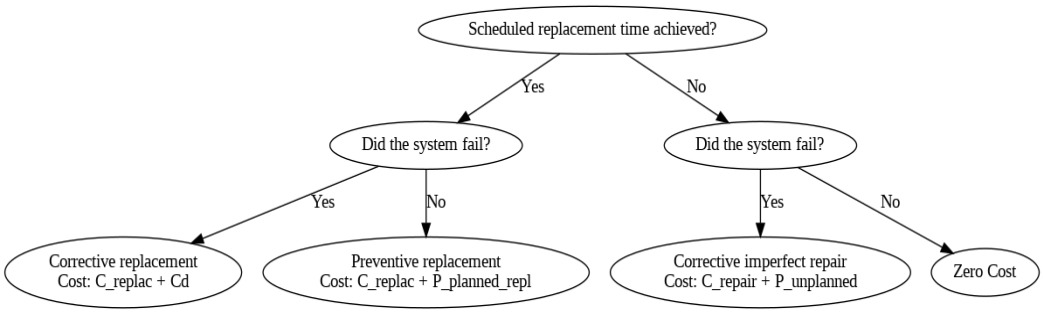


Fig.1. Decision tree for hybrid time-based policy

5. Numerical simulation and results

The hybrid threshold-based policy was benchmarked with the policies mentioned in section 4. The simulation was performed using Python 3.7 and an Intel(R) Core (TM) Ultra 7 155H, 1.40 GHz, 16,0 GB RAM. Input parameters are presented in Table 1-3, while results of the numerical application are summarized in Table 4. Figure 2 shows the convergence of the cost rate for the proposed policy.

Table 1. Degradation parameters (moderate degradation speed and degradation increment uncertainty)

Parameter	Value
Scale parameter	7
Degradation rate	0.8
Failure threshold	300
Shape parameter	Degradation rate*age

Table 2 presents the parameters for the proposed threshold-based policy. The first three parameters correspond to the maintenance quality decay in IRs which are also valid for the pure imperfect corrective policy based on IRs, and the hybrid policy combining imperfect corrective and time-based policy.

Table.2 Parameters for threshold-based maintenance

Parameter	Value
Initial age reduction rate	0.9
Minimal age reduction rate	0.2
Age reduction rate decay	$X \sim \mathcal{TN}(\mu = 0.1, \sigma = 0.03, a = 0)$
Delta	2

Table 3. Costs that compose different benchmarking policies

Cost	Value
C_d	25
C_{replac}	15
$P_{planned_replac}$	5
C_{repair}	2.5
$P_{unplanned}$	7.5
$P_{planned_repair}$	0.5

Table 4. Cost rate of threshold-based policy compared to benchmarking policies

Policy	Cost rate	Optimal variables
Corrective	0.8735	-
Imperfect corrective	7.9786	-
Hybrid time-based	0.2697	$I = 340$
Hybrid threshold-based	0.2487	$L = 81$

Simulation results demonstrate that the proposed hybrid threshold-based maintenance policy significantly outperforms all benchmark strategies in terms of long-run cost rate. The optimal configuration characterized by a repair threshold of L=16 emerges after 14,000 simulation steps. Compared to the corrective-only strategy, the hybrid policy achieves a cost reduction of 71.5%. It also surpasses the imperfect corrective policy by 96.8% and improves upon the hybrid time-based strategy by 7.8%.

These findings highlight the policy's ability to dynamically adapt to the stochastic nature of failure processes. By triggering a perfect preventive replacement after a series of IRs, the proposed strategy strikes a balance between short-term cost-effectiveness and long-term reliability. In contrast, the pure IRs policy exhibits the worst performance underperforming even the basic corrective policy by 813.4%. This drastic inefficiency supports the theoretical assumption that although IRs offer immediate savings, their effectiveness deteriorates over time due to declining maintenance quality, eventually leading to accelerated degradation and higher failure costs.

The hybrid time-based policy, while incorporating scheduled perfect replacements, remains decoupled from the real-time stochastic degradation behavior. Its static timing ignores the actual accumulation of failures, resulting in suboptimal intervention points and diminished cost-efficiency. In contrast, the proposed threshold-based approach remains responsive to the observed failure dynamics, which is particularly advantageous in the absence of continuous condition monitoring.

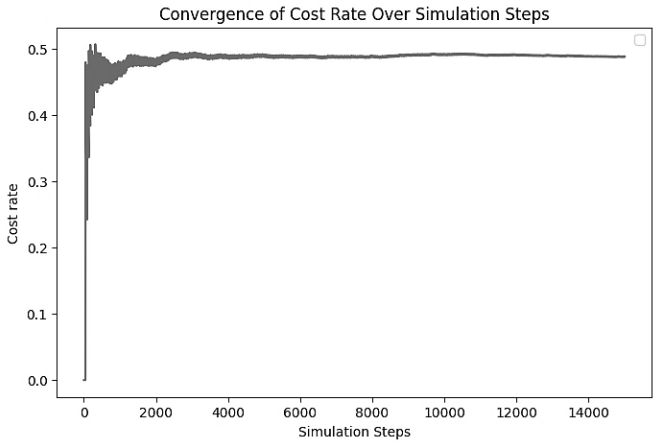


Figure 2. Convergence of Cost Rate for the Hybrid Threshold-Based Policy

6. Conclusion

This study tackled maintenance optimization for systems where condition monitoring and inspections are not feasible. We introduced a threshold-based hybrid policy that performs a set number of imperfect repairs before triggering a preventive replacement. This approach relies solely on failure events to guide decisions, making it suitable for uncertain, hard-to-monitor environments. Simulations using gamma-process degradation show that our policy consistently outperforms corrective-only, imperfect-only, and time-based strategies, achieving lower cost rates by adapting to failure behavior and maintenance quality decline. The inclusion of a replenishment delay ensures spare part logistics are respected with short service interruptions.

These results support our central claim: maintenance aligned with failure frequency is more effective than fixed schedules in uncertain conditions. The proposed strategy offers a robust, cost-efficient alternative for industries like aerospace, offshore operations, and heavy machinery. Future research may extend this framework to multi-component systems or integrate adaptive threshold learning.

Acknowledgement

This work was supported by the MODAPTO project (MODULAR MANUFACTURING AND DISTRIBUTED CONTROL VIA INTEROPERABLE DIGITAL TWINS) funded by the European Union's Horizon 2022 (grant agreement No 101091996).

References

De Jonge, B., & Scarf, P. A. (2020). A review on maintenance optimization. *European journal of operational research, 285(3), 805-824.*

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing letters*, 3, 18-23.

Lee, J., & Mitici, M. (2023). Deep reinforcement learning for predictive aircraft maintenance using probabilistic remaining-useful-life prognostics. *Reliability Engineering & System Safety*, 230, 108908.

Teixeira, H. N., Lopes, I., & Braga, A. C. (2020). Condition-based maintenance implementation: a literature review. *Procedia Manufacturing*, 51, 228-235.

Van Noortwijk, J. M. (2009). A survey of the application of gamma processes in maintenance. *Reliability Engineering & System Safety*, 94(1), 2-21.

Zhao, Y., & Smidts, C. (2022). Reinforcement learning for adaptive maintenance policy optimization under imperfect knowledge of the system degradation model and partial observability of system states. *Reliability engineering & system safety*, 224, 108541.

Evaluating reliability and maintenance effectiveness of a system modelled by a Continuous Time Hidden Markov Model

M.L. Gámiz*, N. Limnios** and M.C. Segovia***

**Universidad de Granada, Granada, Spain, mgamiz@ugr.es*
*** Université de Technologie de Compiègne, Compiègne, France, nikolaos.limnios@utc.fr*
**** Universidad de Granada, Granada, Spain, msegovia@ugr.es*

Abstract. Hidden Markov Models appear in many real-world estimation problems where a process with unobservable states produces observable outputs (indicators or signals) that are related to the process hidden state. In this work we consider a Continuous Time Hidden Markov Model to model the degradation process of a system of interest. The current state of the system is not directly observable, but it is estimated based on the available signals, for example indicators of temperature or pressure provided by a sensor.

The system considered is maintained throughout its life. We are interested in knowing how well the maintenance action performed recovers the system. We proposed some measures to estimate the effectiveness of the maintenance as well as reliability measures. The proposed measures are based on estimators of the model parameters obtained via a discretization strategy, that is, we consider that the signals are produced continuously but only observed at regular intervals of time.

1. Introduction

The advances in sensing and automated data collection from multiple sensors has given rise to the availability of large data sets providing useful information (signals or indicators) about certain aspects of the real health of industrial equipment (Duan et al. (2022), Wang et al. (2002)). The correct interpretation of these signals or indicators leads to an appropriate diagnosis of the degradation level of the system of interest, and therefore, appropriate maintenance actions can be performed to restore the system to a good level of performance when required.

In the literature many models try to estimate the new condition of the system after the maintenance action. These models mainly consider that the system is restored to a previous level of degradation and try to quantify the improvement of the system condition. Some classical models are the ARI and ARA models (Doyen and Gaudoin (2004)) and Kijima (1989)).

In this paper we consider a stochastic model based on a Continuous Time Hidden Markov process to describe the degradation level of the system of interest. The estimation of the current state of the system is based on indicators or signals as the true system state cannot be observed directly. The signals are produced continuously but only observed at regular intervals of time. This system was studied in Gámiz et al (2023b). Here we consider that the system is maintained throughout its life. Depending on the maintenance action performed, the system would be restored to a different level of performance. For this system we proposed some measures to estimate the effectiveness of the maintenance as well as reliability measures.

HMM appear in many real-world estimation problems where a process with unobservable states produces observable outputs normally referred as signals (see Mor et al. (2021) for a recent review of the different applications of these models). In this paper we consider a stochastic system according to a continuous-time Markov. The states of the system present different levels of degradation up to failure. The true state of the system is non observable at any time t. The system is monitored, and observations of a random process are available providing certain information about the true state of the system. Additionally, the considered system is maintained at specific points in time. Our aim is to propose some reliability measures for the system and for the efficiency of the performed maintenance action by considering observations of the process Y.

The remainder of the paper is organized as follows: First an Introduction to Continuous-Time Hidden Markov Model (CTHMM) is given, then in Section 3 a discretization strategy is presented to estimate the parameters of the model in continuous time. In Section 4 we proposed some reliability, and the maintenance efficiency measures based on the estimators of the CTHMM parameters, then, in Section 5 an illustration example is provided. Finally, some conclusions are given.

2. The continuous time Hidden Markov Model

Let (X, Y) be a coupled process in continuous time, where X is an unobserved process, i.e. a hidden process, and Y is the observed process. The law of the coupled process is estimated based on the observations. In general, the hidden process X is a Markov chain, and Y is a process whose distribution depends on the value of X_t. Specifically, let $\{X_t;\ t \geq 0\}$ be a continuous-time Markov chain with finite state space $E = \{1, 2, \ldots, d\}$ and generating matrix $A = (a_{ij}\ ,;\ i, j \in E)$. Let $\{Y_t;\ t \geq 0\}$ an $\mathcal{Y}$-valued sequence such that given $X_t = i$, Y_t is conditionally distributed with density $g(i, y)$ with respect to some σ-finite measure μ on $\mathcal{Y}$, and fixed $i \in E$. Therefore, the considered model is an M1-M0.

Both, the generating matrix $\boldsymbol{A}$ as well as the family of densities $\{g(i,\cdot);\ i \in E\}$, depend on a vector of parameters θ, that is $A(i, j) = a_{ij}(\theta)$ and $g(i,\cdot) = g_\theta(i,\cdot)$, for all i, j ∈ E. The set of possible values of the vector θ is denoted $\Theta \subset R_k$, and θ must be estimated from a set of observations of the process $\{Y_t\}$. The vector θ usually includes the transition rates and some parameters characterizing the densities g.

For the rest of the paper, we will consider the process (X, Y) defined as follows.

Definition 1. Let $X = \{X_t;\ t \geq 0\}$ be an irreducible homogeneous Markov process in a finite set E and with generating matrix $A = (a_{ij}\ ;\ i, j \in E)$, and $Y = \{Y_t;\ t \geq 0\}$ is a homogeneous process in a general set $\mathcal{Y} \subset \mathbb{R}^q$, with $q \geq 1$ such that, for $t > 0$, the distribution of Y_t is determined $G(i,\cdot)$, over the event $X_t = i$, for $i \in E$. Then we say that (X, Y) is a two-dimensional process with dependence structure M1-M0. That is, if $B_\mathcal{Y}$ denotes the set of Borel subsets of $\mathcal{Y}$, then for any $i, j \in E$, $y \in \gamma$ and B ∈ $B_\mathcal{Y}$, and for all s, t > 0,

$$P(X_{t+s} = j, Y_{t+s} \in B | X_s = i, Y_s = y, X_u, Y_u, 0 \leq u \leq s) = P(X_t = j, Y_t \in B | X_0 = i, Y_0 = y)\, P_{ij}(t) G(j, B). \tag{1}$$

Using the homogeneity of processes X and Y. Additionally for $t = 0, P(X_t = j, Y_t \in B | X_0 = i, Y_0 = y) = 1$ if $i = j$ and $y \in B$; and *0* otherwise.

Proposition 1. Let *(X, Y)* be a stochastic process where X is a CTMC with finite state space E and generating matrix $A = (a_{ij};\ i, j \in E)$, and let the transition probabilities of the process *(X, Y)* be given by (1). The generator of the two-dimensional process (X_t, Y_t) can be written as

$$\tilde{A} f(i, y) = \sum_{j \in E \setminus \{i\}} a_{ij} \int_\mathcal{Y} G(j, du)[f(j, u) - f(i, y)]. \tag{2}$$

for all $i \in E$ and $y \in \mathcal{Y}$, (see Gámiz et al (2023b) for a proof of Proposition 1).

In particular, for a given state space $E = \{1, 2, \ldots, d\}$ of the hidden MC and given state space $\mathcal{Y} = \{y_1, y_2, \ldots, y_s\}$ for the observable process, the set of states of the two-dimensional process (X, Y) can be written as $\tilde{E} = E \times \mathcal{Y}$, in particular $\tilde{E} = \{(1, y_1), \ldots, (1, y_s), (2, y_1), \ldots, (2, y_s), \ldots, (d, y_1), \ldots, (d, y_s)\}$.

Then, the expression given in (2) for matrix $\tilde{\boldsymbol{A}}$ can be written as follows:

$$\tilde{A} f(i, y) = \sum_{i \in E} \sum_{y_r \in \mathcal{Y}} f(j, y_r) \left(\lim_{t \to 0} \frac{P(X_t = j, Y_t = y_p | X_0 = i, Y_0 = y_r) - \delta_{ij} \delta_{y_r y_p}}{t} \right). \tag{3}$$

We have that $P(X_t = j, Y_t = y_p | X_0 = i, Y_0 = y_r) = P_{ij}(t) G(j, y_p)$, and we assume that $P(X_t = j, Y_t = y_p | X_0 = i, Y_0 = y_r) \to 0$ as $t \to 0$ when $y_r \neq y_p$. We consider the following cases:

- If $i = j, y_r = y_p, (1/t)[P_{ij}(t) G(j, y_p) - 1] \to a_{ii}$;
- If $i = j, y_r \neq y_p, (1/t)[P_{ij}(t) G(j, y_p)] \to 0$; and,
- If $i \neq j,\ (1/t)[P_{ij}(t) G(j, y_p)] \to a_{ij} G(j, y_p)$.

In summary, the generator elements are given by:

$$\tilde{a}\left((i, y_r), (j, y_p)\right) = \begin{cases} a_{ii}, & i = j, y_r = y_p \\ 0 & i = j, y_r \neq y_p \\ a_{ij} G(j, y_p) & i \neq j \end{cases}. \tag{4}$$

3. Estimation considering periodic observations

The two-dimensional process (X, Y) is fully described by matrix $\boldsymbol{A}$ and the emission distribution G, both dependent on a vector of parameters θ, i.e. $\mathbf{A} = A(\theta)$, and that $G = G_\theta$. The vector θ must be estimated from the data. More specifically, it must be estimated from a sample of observations of the process Y_t on the interval $(0, T]$ given that X_t is unobservable.

3.1 *Time discretization based on periodic observations*

Let us consider a stochastic system according to a continuous-time Markov chain $\{X_t;\ t \geq 0\}$ with state space $E = \{1, 2, \ldots, d\}$ as above; generating matrix $\boldsymbol{A}$, that is $P(t) = e^{\boldsymbol{A}t}$, for all $t > 0$, where $P(t)$ provides the transition probabilities at time *t*.

The true state of the system X_t is non observable at any time t ∈ (0, T]. However, at *n* regular pre-specified times denoted by $0 < h < 2h < \ldots < nh = T$, observations of a random process $\{Y_t;\ t \geq 0\}$ are available providing certain information about the true state of the system. Let us denote $\hat{Y}_k = Y_{kh}$ the kth observed signal, which is recorded at time kh, for $k \in \{0, 1, \ldots, n\}$.

Associated to (or embedded in) the continuous-time process we can consider the following discrete time Markov chain $\hat{X}_k = X_{kh}$, with transition matrix $\boldsymbol{Q}_h = e^{\boldsymbol{A}h}$, whose (i,j)-element is given by $Q_h(i,j) = P(X_h = j | X_0 = i)$. The discrete Markov chain $\hat{X}_k$ is called the *discrete skeleton* (at scale h) of X_t (see Kingman (1963) for ergodicity properties of Markov chains based on discrete skeletons).

Consider the sequence of times $0 < h < 2h < \ldots < nh$, then $\hat{\boldsymbol{Y}} = (\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n)$ can be seen as a realization of the embedded discrete hidden Markov model that we denote by $(\hat{X}, \hat{Y})$. We can use the results in Bickel et al. (1998) to estimate the transition matrix of this embedded hidden Markov chain $\{\hat{X}_n\}$, that is we obtain $\widehat{\boldsymbol{Q}}_h$, as well as the emission functions $G(i, dy) = P(\hat{Y}_k \in dy | \hat{X}_k = i)$, for $i \in E$.

For an interval of length *h*, we have $\boldsymbol{Q}_h = \boldsymbol{P}(h) = e^{\boldsymbol{A}h}$,; where $A = (a_{ij}\ ;\ i,j \in E)$ is the generating matrix of the hidden continuous-time Markov chain; $\boldsymbol{P}(h)$ is the corresponding transition function matrix; and, $\boldsymbol{Q}_h$ is the transition matrix of the Markov chain $\hat{X}_k$. Then, using Taylor expansion, we approximate

$$a_{ij} = \frac{1}{h}\left(Q_h(i,j) - \delta_{ij}\right) + o(1). \tag{5}$$

where $o(h)/h \to 0$ as $h \to 0$, and δ_{ij} equal to 1 if $i = j$ and 0 otherwise. Then, we define the estimator of the generating matrix as

$$\widehat{\boldsymbol{A}}_h = \frac{1}{h}\left(\widehat{\boldsymbol{Q}}_h - \boldsymbol{I}\right). \tag{6}$$

where $\widehat{\boldsymbol{Q}}_h$ is the maximum likelihood estimator of the transition matrix corresponding to the discrete-time chain $\hat{X}$, obtained as in Gamiz et al. (2023b), and $\boldsymbol{I}$ the identity matrix.

Note that since we write $\mathbf{A} = A(\theta)$ we also have that $\boldsymbol{Q}$ is a function of the parameter vector θ, that is, $\mathbf{Q} = Q(\theta)$.

3.2 *Estimation based on the EM algorithm*

Let $\hat{\theta}_n$ be the maximum-likelihood estimator of θ. For $h > 0$ fixed, let $\widehat{\boldsymbol{Q}}_{h,n} = Q(\hat{\theta}_n)$ then $\widehat{\boldsymbol{A}}_{h,n} = h^{-1}\left(Q\left(\hat{\theta}_n\right) - \boldsymbol{I}\right)$. Considering the regularity assumptions presented in Bickel et al. (1998), this estimator is a uniformly strongly consistent estimator of $\boldsymbol{A}_h$, as well as of $\mathbf{A}$, as proven in Gamiz et al. (2023b). Additionally, $\hat{G}(i,y) = G_{\hat{\theta}_n}(i,y), i \epsilon E, y \epsilon \mathcal{Y}$.

Our aim now is to estimate the vector of parameters θ, so that in turn we could estimate $\mathbf{A} = A(\theta)$ and $G(i,y)$.

Given a sample of observations $\hat{\boldsymbol{Y}} = \{\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n\}$, we estimate vector of parameters θ by the EM-algorithm. Let us consider the function $Q\left(\theta \middle| \theta^{(m)}\right) = E_{\theta^{(m)}}\ [ln\, f(\hat{X}, \hat{\boldsymbol{Y}} | \theta)]$, by means of this function and by successive iterations we obtain an approximation of the estimate of θ, where $ln\, f(\hat{X}, \hat{\boldsymbol{Y}} | \theta)$ is the log-likelihood corresponding to the complete dataset, which would be obtained from the bivariate process $\left(\hat{X}, \hat{Y}\right)$ if it was fully observed.

$$Q\left(\theta \middle| \theta^{(m)}\right) = \sum_{n=1}^{N-1} \sum_{i,j \in E} P_{\theta^{(m)}}\left(\hat{X}_n = i, \hat{X}_{n+1} = j \middle| \hat{\boldsymbol{Y}}\right) \ln P_\theta(i,j) + \sum_{n=1}^{N-1} \sum_{i,j \in E} P_{\theta^{(m)}}\left(\hat{X}_n = i \middle| \hat{\boldsymbol{Y}}\right) \ln P_\theta(i,j) \sum_{k=1}^{K} \ln g_i\left(\hat{Y}_{kn}\right). \tag{7}$$

3.2.1 *The EM algorithm*

Let us consider that the state space of the hidden MC is given by $E = \{1, 2, \ldots, d\}$ and for the observable process we have $\mathcal{Y} = \{y_1, y_2, \ldots, y_s\}$, i.e. we have a discrete state space for the observable process. Let us consider a sample of observations $\hat{\boldsymbol{Y}} = \{\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n\}$.

E-Step. For a given $\theta^{(m)}$, compute the following probabilities:

$$P_{\theta^{(m)}}\left(\hat{X}_n = i, \hat{X}_{n+1} = j \middle| \hat{\boldsymbol{Y}}\right), \mathrm{n} \in \{1, \ldots, \mathrm{N} - 1\};\ \mathrm{i,j} \in \mathrm{E}.$$

and

$$P_{\theta^{(m)}}\left(\hat{X}_n = i \middle| \hat{\boldsymbol{Y}}\right), \mathrm{n} \in \{1, \ldots, \mathrm{N}-1\};\ \mathrm{i,j} \in \mathrm{E}.$$

M-Step. Update $\theta^{(m)}$ to $\theta^{(m+1)}$. Then, maximize the first part of equation (2) to obtain

$$\hat{p}^{(m+1)}(i,j) = \frac{\sum_{n=1}^{N} P_{\theta^{(m)}}(\hat{X}_n = i, \hat{X}_{n+1} = j \middle| \hat{\boldsymbol{Y}})}{\sum_{n=1}^{N} P_{\theta^{(m)}}(\hat{X}_n = i \middle| \hat{\boldsymbol{Y}})}$$

Finally, maximization of the second half of equation (2) to derive

$$\hat{G}(i,y) = \frac{\sum_{n=1}^{N} P_{\theta^{(m)}}(\hat{X}_{n-1} = i)\mathbf{1}_{\{Y_n = y\}}}{\sum_{n=1}^{N} P_{\theta^{(m)}}(\hat{X}_{n-1} = i \middle| \hat{\boldsymbol{Y}})}$$

4. Reliability and maintenance effectiveness measures

Let us consider that the state space $E = \{1, 2, \ldots, d\}$ can be split into two subsets, that is, $E = U \cup D$ where U is the set of *working states* with $U = \{1, 2, \ldots, r\}$ and D is the set of *down states,* $D = \{r+1, \ldots, d\}$. Additionally, the observation process *states*, $\mathcal{Y}$, that is, the information we received through the observation process about the system state can be categorized into two groups of signals: subset $\mathcal{Y}_1$ that includes the signals indicating a *good* level of performance, and subset $\mathcal{Y}_2$ that comprises the *warning* signals referring to a serious problem in the system that requires the system operation interruption, and therefore its failure.

Let us define $\widetilde{U} = U \times \mathcal{Y}_1$, the set of states of the two-dimensional process that indicate a *good* level of performance of the system, and $\widetilde{D} = \widetilde{E} \backslash \widetilde{U}$, the set of states of the two-dimensional process that indicate a *system failure.*

The corresponding generating matrix of the process (X, Y) given in (2) can be written in blocks as follows

$$\widetilde{\boldsymbol{A}} = \begin{pmatrix} \widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{U}} & \widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{D}} \\ \widetilde{\boldsymbol{A}}_{\widetilde{D}\widetilde{U}} & \widetilde{\boldsymbol{A}}_{\widetilde{D}\widetilde{D}} \end{pmatrix}. \tag{8}$$

4.1 *Reliability*

Let T be the first time the system visits the set $\widetilde{D}$, $T = \inf\{t > 0; (X_t, Y_t) \in \widetilde{D}\}$. Considering this, the reliability of the system can be defined as $\tilde{R}(t) = P(T > t)$.

Conditioning on the initial state of the two-dimensional process $(i, y) \in \widetilde{U} = U \times \gamma_1$ we have

$$\tilde{R}_{(i,y)}(t) = P_i(T > t) = P_i(X_s \in U, Y_s \in \mathcal{Y}_1, 0 < s \leq t) = P_i\big((X_t, Y_t) \in \widetilde{U}, 0 < s \leq t\big). \tag{9}$$

Then, we can write

$$\tilde{R}(t) = \sum_{(i,y)\in\widetilde{U}} \tilde{R}_{(i,y)}(t) = \tilde{\alpha} \exp\big(\widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{U}} t\big)\, \mathbf{1}_{\widetilde{U}}, t \geq 0. \tag{10}$$

being $\tilde{\alpha}$ the initial probability vector of the two-dimensional process and $\mathbf{1}_{\widetilde{U}}$ a column vector of ones of appropriate dimensions.

Based on a sample of the observable process $\hat{\boldsymbol{Y}} = (\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n)$, with $\hat{\boldsymbol{Y}} = Y(t_{nh})$ for a fixed h small enough, we define the following estimator of the reliability function in continuous time, plugging-in the estimator of the generating matrix $\widetilde{\boldsymbol{A}}$, $\widetilde{\boldsymbol{A}}_h$, in (7)

$$\tilde{R}_h^*(t) = \sum_{(i,y)\in\widetilde{U}} \tilde{R}_{(i,y)}(t) = \tilde{\alpha} \exp\big(\widetilde{\boldsymbol{A}}_{h,\widetilde{U}\widetilde{U}}^* t\big)\, \mathbf{1}_{\widetilde{U}}, t \geq 0. \tag{11}$$

The proposed estimator is asymptotically Normal, as shown in Gámiz et al. (2023).

4.2 *Failure rate function and Rate of Occurrence of Failures*

We can obtain the Failure rate function as

$$\tilde{r}(t) = \sum_{(i,y)\in\widetilde{U}} \frac{\tilde{f}_{(i,y)}(t)}{\tilde{R}_{(i,y)}(t)} = \frac{\tilde{\alpha} \exp(\widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{U}} t)(-\widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{U}} \mathbf{1}_{\widetilde{U}})}{\tilde{\alpha} \exp(\widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{U}} t)\mathbf{1}_{\widetilde{U}}}, t \geq 0. \tag{12}$$

Based on the sample of the observable process $\hat{\boldsymbol{Y}} = (\hat{Y}_0, \hat{Y}_1, \ldots, \hat{Y}_n)$, with $\hat{\boldsymbol{Y}} = Y(t_{nh})$ for a fixed h small enough, and by plugging-in $\widetilde{\boldsymbol{A}}_h$, we obtain an estimator of the failure rate function $\tilde{r}_h^*(t)$

The expression for the ROCOF is

$$\rho_F(t) = \sum_{(i,y)\in\widetilde{U}} \tilde{p}_{(i,y_r),(j,y_p)}(t)\tilde{a}\left((i, y_r), (j, y_p)\right) = \tilde{p}_{\widetilde{U}}(t)\widetilde{\boldsymbol{A}}_{\widetilde{U}\widetilde{D}}. \tag{13}$$

where $\tilde{p}_{(i,y_r),(j,y_p)}(t) = P\big(X_t = j, Y_t = y_p \middle| X_0 = i, Y_0 = y_r\big)$. And by plugging-in $\widetilde{\boldsymbol{A}}_h$ in (5) we obtain an estimator, $\tilde{\rho}_{Fh}^*(t)$.

4.3 *Maintenance effectiveness*

Let us assume a maintenance action takes place in the system after some time, it could be a scheduled one, or it could be performed based on the system condition (Condition Based Maintenance (CBM)).

After the maintenance action the system is in a new and, generally, better state. The new state of the system depends on the maintenance action performed as well as on how well it was performed.

After the *mth* maintenance, the system new state will be chosen according to initial probability vector $\tilde{\alpha}_m$. This is the initial probability vector of the coupled process *(X,Y)*, being α the initial probability vector of the hidden process, and $\alpha_i = P(X_0 = i)$, $\tilde{\alpha}_{0(i,y_r)} = \alpha_i G(i, y_r)$. In fact, the initial probability vector of the hidden process after the *mth* action, α^m, is the one that captures how the system has recovered after the maintenance, the distribution of the observable process is not affected by the maintenance policy.

Additionally, although in general the maintenance action restores the system to a better condition, in some circumstances, such as an excessive delay on maintaining the system, it could not have the desire effect and even worsen its condition. All of this can be captured by vector α^m.

Finally, α^m shall depend on the number of interventions the system has undergone and the calendar age of the system. As time passes it becomes more difficult to recover the system to the desire performance state (see Clavereau (2010)). We can establish a maximum number of interventions (M) or a maximum calendar age (T).

Here we consider the mean time to failure (MTTF) to determine how well the system recovers after a maintenance action. At the beginning of the system operation, the system is new. In this case the mean time to failure (MTTF) is given by $\boldsymbol{MTTF} = \tilde{\alpha}\left(\widetilde{\boldsymbol{A}}^*_{h,\tilde{U}\tilde{U}}\right)^{-1}\mathbf{1}_{\tilde{U}}$, where $\alpha = (1,0,\dots,0)_{1\times d}$ is the initial probability vector of the hidden process.

After the *mth* maintenance action the Mean Time to Failure ($MTTF_m$) is given by:

$$\boldsymbol{MTTF_m} = \tilde{\alpha}_m\left(\widetilde{\boldsymbol{A}}_{\tilde{U}\tilde{U}}\right)^{-1}\mathbf{1}_{\tilde{U}}. \tag{14}$$

We propose to measure the effectiveness of the maintenance in terms of the meantime to failure after a maintenance intervention:

$$ef_m = \frac{\boldsymbol{MTTF_m}}{\boldsymbol{MTTF_{m-1}}}, m = 1,2,.. \tag{15}$$

where $\boldsymbol{MTTF_0} = \boldsymbol{MTTF}$, the Mean Time to Failure at the beginning of the system operation. In general, $ef_m \leq 1$, if $ef_m = 1$, the system the system has gone back to the state after the previous maintenance operation.

For a maintenance action to be As Good As New (AGAN), we will need to compare $\boldsymbol{MTTF_m}$ with $\boldsymbol{MTTF_0}$, if $ef_m^0 = \frac{\boldsymbol{MTTF_m}}{\boldsymbol{MTTF_0}} = 1$ the maintenance is AGAN.

5. Numerical application

Let us consider a system standing shocks. The system deteriorates with shocks, and the state of the system can be described by the set $E = \{1, 2, \dots, d\}$, 1 indicating that the system is new and has no received any shocks, and *d* indicating that the system has failed due to the arrival of a shock. Each time a shock arrives at the system, the system state changes from state *i* to *i+1* with probability *p*, meaning that the shock caused damage to the system, and remains in state *i* with probability *1-p*, for $i = 1, \dots, d-1$. Once the system has reached state *d* the system is replaced. The transition probability matrix for this system is given by:

$$P = \begin{pmatrix} 1-p & p & \cdots & 0 \\ 0 & 1-p & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

We do not observe the true state of the system, but we receive a signal which is produced depending on the current state of the system with a particular probability distribution. We assume that the state space for the signal process is a discrete state space $\mathcal{Y} = \{y_1, y_2, \dots, y_s\}$, therefore for the signal process we have a matrix M of dimension $d \times s$ whose element (i, y_r) is given by $G(i, y_r)$. In this example we assume that $M(d, y_s) = 1$ and for the rest of the rows in the matrix we consider the probability distribution of a Binomial law with size $s-1$ and probability p_i which decreases with the value of i.

Let us consider that a system where $d = 10$and $s = 20$. We consider that signals y_{15} to y_{20} are warning signals. Therefore, the system up states are given by $U = \{1, 2, \dots, 9\}$ and the set of good signals is given by $\mathcal{Y}_1 = \{y_1, \dots, y_{15}\}$. Finally, let $p = 0.6$.

For this system, its Reliability function is presented in Figure 1.

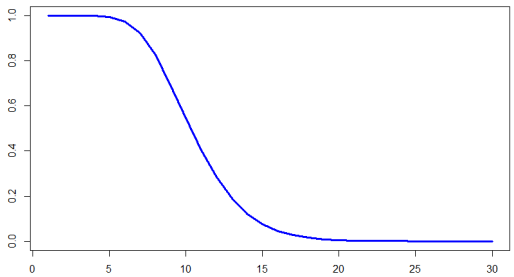


Figure 1. Reliability estimation of the shock model

In table 1 we provide some values of MTTF at the beginning of the system operation and after a maintenance action. We assume that all the maintenance actions were performed on time and that only the system age or how well the maintenance was performed can reduce the efficiency of the maintenance.

For this example we consider that at the beginning of the system´s operation $\alpha = (1,0,\dots,0)$, and then, after the *mth* maintenance intervention $\tilde{\alpha}_m = (1 - 0.2 * m, 0.2, \dots, 0.2, 0)$, where the probability $0.2 * m$ is equally distributed among the phases 2 to $m + 1$, with m=1,2,3,4. For this system the $MTTF_0 = 11.14379$.

m	$MTTF$	ef_m	ef_m^0
1	10.81048	0.97	0.97
2	10.14385	0.94	0.91
3	9.143918	0.9	0.82
4	7.811664	0.85	0.7

Table 1. Maintenance efficiency measures

6. Conclusions

In the present paper we consider a CTHMM to model the evolution in time of a system. We do not have direct observations of the state of the system, but we do receive some signals that provide some information about the true system state. We estimate the parameters of the HMM by a discretization strategy and obtain estimators that allow us to provide reliability measures of interest. We also define a measure to determine maintenance effectiveness if the system is maintained throughout its lifetime.

Acknowledgements

This work is jointly supported by the Spanish Ministry of Science and Innovation-State Research Agency through grants numbered PID2020-120217RBI00 and PID2021-123737NB-I00.

References

Bickel PJ, Ritov Y, Ryden T (1998) Asymptotic normality of the maximum-likelihood estimator for a general Hidden Markov model. *Annals of Statistics*, **26**(4), 1614-1635.

Doyen L, & Gaudoin, O (2004) Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering and System Safety* **84,** 45–56.

Duan C, Li Y, Pu H, Luo J (2022) Adaptative monitoring scheme of stochastically failing systems under hidden degradation processes. *Reliability Engineering and System Safety* **221**, 108322.

Gámiz ML, Limnios, N, Segovia-García MC (2023a) Hidden Markov models in reliability and maintenance. *European Journal of Operational Research* **304**, 1242–1255.

Gámiz ML, Limnios, N, Segovia-García MC (2023b) The continuous-time hidden Markov model based on discretization. Properties of estimators and applications. *Statistical Inference for Stochastic Processes* **26** (3), 525-550.

Kijima M (1989) Some results for repairable systems with general repair. *Journal of Applied Probability* **26**: 89–102.

Kingman JFC (1963) Ergodic properties of continuous-time Markov processes and their discrete skeletons. *Proc London Mathematical Society* **13**(1), 593-604.

Mor B, Garhwal S, Loura, A (2021) A Systematic Review of Hidden Markov Models and Their Applications. *Archives of Computational Methods in Engineering* **28**, 1429–1448.

Pham H, Wang H. (1996) Imperfect maintenance. *European Journal of Operational Research* **94**, 425-438.

Wang Y, Lee IC, Hong Y, Deng X (2022) Building degradation index with variable selection for multivariate sensory data. *Reliability Engineering and System Safety* **227**, 108704.

From theory to practice: the process of optimisation of preventive maintenance policies

Shaomin Wu

Kent Business School, University of Kent

13th IMA International Conference on Modelling in Industrial Maintenance and Reliability MIMAR2025

July 8, 2025

Outline

- Uncertainties and Challenges
- Two Models:
  - The exponential smoothing of intensity functions
  - Average of lower and upper bounds of the failure intensity function
- Dealing with Uncertainty
- Conclusions and Suggestions

Table of Contents

1. Uncertainties and Challenges
2. Two models
3. Exponential smoothing of intensity functions
4. Dealing with Uncertainty
5. Conclusions and Suggestions

Need of Implementing PM policies in practice

- Comparison between Preventive Maintenance (PM) and Predictive Maintenance (PdM)

Aspect	PM	PdM
Trigger	Time/usage-based	Condition/data-based
Data Requirement	Low	High (sensor & operational data)
Technology	Basic	Advanced (IoT, ML, analytics)
Expertise	Maintenance personnel	Multidisciplinary team
Implementation Cost	Low	High (initial)
Optimization Capability	Moderate	High

- PdM is more accurate, but demands more resources; PM is less accurate, but less resource-hungry

Need of Implementing PM policies in practice

- Comparison between Preventive Maintenance (PM) and Predictive Maintenance (PdM)

Aspect	PM	PdM
Trigger	Time/usage-based	Condition/data-based
Data Requirement	Low	High (sensor & operational data)
Technology	Basic	Advanced (IoT, ML, analytics)
Expertise	Maintenance personnel	Multidisciplinary team
Implementation Cost	Low	High (initial)
Optimization Capability	Moderate	High

- PdM is more accurate, but demands more resources; PM is less accurate, but less resource-hungry

Need of Implementing PM policies in practice

- Comparison between Preventive Maintenance (PM) and Predictive Maintenance (PdM)

Aspect	PM	PdM
Trigger	Time/usage-based	Condition/data-based
Data Requirement	Low	High (sensor & operational data)
Technology	Basic	Advanced (IoT, ML, analytics)
Expertise	Maintenance personnel	Multidisciplinary team
Implementation Cost	Low	High (initial)
Optimization Capability	Moderate	High

- PdM is more accurate, but demands more resources; PM is less accurate, but less resource-hungry

A PM Policy—Age Replacement Policy (ARP)

- Denote the replacement cost for a unit as c_r, the repair cost as c_m, and the lifetime distribution as $F(t)$. To minimise the following long-run expected cost per unit of time, one can obtain the optimal time interval T^* of an age replacement policy:

$$C_{\mathrm{ARP}}(t_0) = \frac{c_r + c_m F(t_0)}{\int_0^{t_0}(1 - F(t)dt}. \tag{1}$$

- A good decision-making needs considering uncertainty

A PM Policy—Age Replacement Policy (ARP)

- Denote the replacement cost for a unit as c_r, the repair cost as c_m, and the lifetime distribution as $F(t)$. To minimise the following long-run expected cost per unit of time, one can obtain the optimal time interval T^* of an age replacement policy:

$$C_{\mathrm{ARP}}(t_0) = \frac{c_r + c_m F(t_0)}{\int_0^{t_0}(1 - F(t)dt}. \tag{1}$$

- A good decision-making needs considering uncertainty

Uncertainties/Challenges in Development of PM policies

- System structure and operating
  - Unknown dependence between components
  - Uncertain usage intensity
  - Dynamic operating condition
- Challenges in data scarcity
  - Censored data
  - Lack of failure data
  - Scalability is an issue for larger, complex systems
- Uncertainty in parameter estimation
  - Misspecification of models, models are generic and not tailored to specific systems
  - Difficult in model (or PM policy) validation
- Administration related uncertainties

Uncertainties/Challenges in Development of PM policies

- System structure and operating
  - Unknown dependence between components
  - Uncertain usage intensity
  - Dynamic operating condition
- Challenges in data scarcity
  - Censored data
  - Lack of failure data
  - Scalability is an issue for larger, complex systems
- Uncertainty in parameter estimation
  - Misspecification of models, models are generic and not tailored to specific systems
  - Difficult in model (or PM policy) validation
- Administration related uncertainties

Uncertainties/Challenges in Development of PM policies

- System structure and operating
  - Unknown dependence between components
  - Uncertain usage intensity
  - Dynamic operating condition
- Challenges in data scarcity
  - Censored data
  - Lack of failure data
  - Scalability is an issue for larger, complex systems
- Uncertainty in parameter estimation
  - Misspecification of models, models are generic and not tailored to specific systems
  - Difficult in model (or PM policy) validation
- Administration related uncertainties

Uncertainties/Challenges in Development of PM policies

- System structure and operating
  - Unknown dependence between components
  - Uncertain usage intensity
  - Dynamic operating condition
- Challenges in data scarcity
  - Censored data
  - Lack of failure data
  - Scalability is an issue for larger, complex systems
- Uncertainty in parameter estimation
  - Misspecification of models, models are generic and not tailored to specific systems
  - Difficult in model (or PM policy) validation
- Administration related uncertainties

Some Interesting Findings

1. the use of assumption-heavy models, which might not be applicable in real-world situations with variable or uncertain conditions;
2. the problem of data scarcity for costly components, which has not been resolved beyond balancing sampling cost and error;
3. the intricacy and limited interpretability of the suggested measures, which could prevent practitioners from adopting them;
4. Result-validation rarely appears. Most papers surface only in the numerical examples with which authors validate their results
5. Most PM policies have limited take up, or are not appropriated.

Some Interesting Findings

1. the use of assumption-heavy models, which might not be applicable in real-world situations with variable or uncertain conditions;
2. the problem of data scarcity for costly components, which has not been resolved beyond balancing sampling cost and error;
3. the intricacy and limited interpretability of the suggested measures, which could prevent practitioners from adopting them;
4. Result-validation rarely appears. Most papers surface only in the numerical examples with which authors validate their results
5. Most PM policies have limited take up, or are not appropriated.

Some Interesting Findings

1. the use of assumption-heavy models, which might not be applicable in real-world situations with variable or uncertain conditions;
2. the problem of data scarcity for costly components, which has not been resolved beyond balancing sampling cost and error;
3. the intricacy and limited interpretability of the suggested measures, which could prevent practitioners from adopting them;
4. Result-validation rarely appears. Most papers surface only in the numerical examples with which authors validate their results
5. Most PM policies have limited take up, or are not appropriated.

Some Interesting Findings

1. the use of assumption-heavy models, which might not be applicable in real-world situations with variable or uncertain conditions;
2. the problem of data scarcity for costly components, which has not been resolved beyond balancing sampling cost and error;
3. the intricacy and limited interpretability of the suggested measures, which could prevent practitioners from adopting them;
4. Result-validation rarely appears. Most papers surface only in the numerical examples with which authors validate their results
5. Most PM policies have limited take up, or are not appropriated.

Some Interesting Findings

1. the use of assumption-heavy models, which might not be applicable in real-world situations with variable or uncertain conditions;
2. the problem of data scarcity for costly components, which has not been resolved beyond balancing sampling cost and error;
3. the intricacy and limited interpretability of the suggested measures, which could prevent practitioners from adopting them;
4. Result-validation rarely appears. Most papers surface only in the numerical examples with which authors validate their results
5. Most PM policies have limited take up, or are not appropriated.

Dealing with Scalability and Uncertainty

- To deal with data scarcity $\rightarrow$ to use failure data from a subsystem
- To mitigate the uncertainty PM policy optimisation $\rightarrow$ to collectively account for uncertainties in maintenance policy optimisation

Dealing with Scalability and Uncertainty

- To deal with data scarcity $\rightarrow$ to use failure data from a subsystem
- To mitigate the uncertainty PM policy optimisation $\rightarrow$ to collectively account for uncertainties in maintenance policy optimisation

Table of Contents

1. Uncertainties and Challenges
2. Two models
3. Exponential smoothing of intensity functions
4. Dealing with Uncertainty
5. Conclusions and Suggestions

Recurrent event data analysis: practical needs

- Modelling of time between failures of technical systems (software or hardware).
- Modelling of insurance claims/warranty claims: times between claims.

Recurrent event data analysis: practical needs

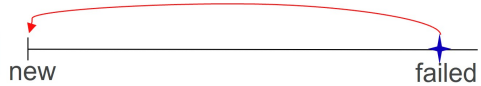
- Modelling of time between failures of technical systems (software or hardware).
- Modelling of insurance claims/warranty claims: times between claims.

Recurrent event data analysis: practical needs

- Modelling of time between failures of technical systems (software or hardware).
- Modelling of insurance claims/warranty claims: times between claims.

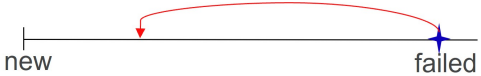
Effectiveness of Repair

Scenario 1: Perfect repair, i.e., as good as new (AGAN)



Scenario 2: Minimal repair, i.e., as bad as old (ABAO)

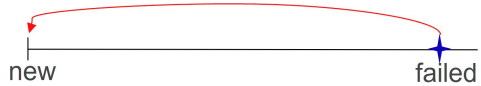
Scenario 3: Imperfect repair, between AGAN and ABAO



- Perfect repair: Renewal processes
- Minimal repair Non-homogeneous Poisson processes
- Imperfect repair: Many failure process models have been developed

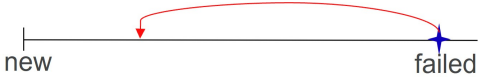
Effectiveness of Repair

Scenario 1: Perfect repair, i.e., as good as new (AGAN)



Scenario 2: Minimal repair, i.e., as bad as old (ABAO)

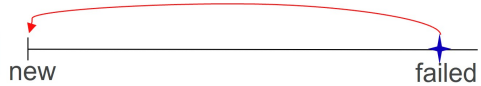
Scenario 3: Imperfect repair, between AGAN and ABAO



- Perfect repair: Renewal processes
- Minimal repair Non-homogeneous Poisson processes
- Imperfect repair: Many failure process models have been developed

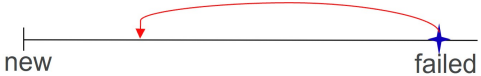
Effectiveness of Repair

Scenario 1: Perfect repair, i.e., as good as new (AGAN)



Scenario 2: Minimal repair, i.e., as bad as old (ABAO)

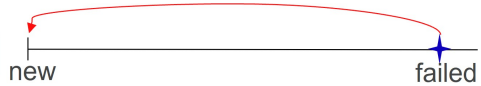
Scenario 3: Imperfect repair, between AGAN and ABAO



- Perfect repair: Renewal processes
- Minimal repair Non-homogeneous Poisson processes
- Imperfect repair: Many failure process models have been developed

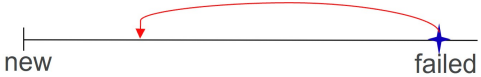
Effectiveness of Repair

Scenario 1: Perfect repair, i.e., as good as new (AGAN)



Scenario 2: Minimal repair, i.e., as bad as old (ABAO)

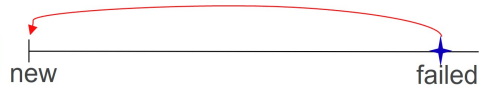
Scenario 3: Imperfect repair, between AGAN and ABAO



- Perfect repair: Renewal processes
- Minimal repair Non-homogeneous Poisson processes
- Imperfect repair: Many failure process models have been developed

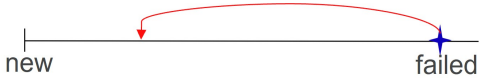
Effectiveness of Repair

Scenario 1: Perfect repair, i.e., as good as new (AGAN)



Scenario 2: Minimal repair, i.e., as bad as old (ABAO)

Scenario 3: Imperfect repair, between AGAN and ABAO



- Perfect repair: Renewal processes
- Minimal repair Non-homogeneous Poisson processes
- Imperfect repair: Many failure process models have been developed

Development in Failure Processes: Models

- Imperfect repair: Many failure process models have been developed
  - ARA (arithmetic reduction of age) (Doyen & Gaudoin, 2004)
  - ARI (arithmetic reduction of intensity) model (Doyen & Gaudoin, 2004),
  - hybrid: the geometric process (Lam, 1988)

- Models can be developed based on the failure intensity function

$$\lambda_i^{(N_{i,t})}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t \downarrow 0} \frac{P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \qquad (2)$$

where $P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t+\Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$.

Development in Failure Processes: Models

- Imperfect repair: Many failure process models have been developed
  - ARA (arithmetic reduction of age) (Doyen & Gaudoin, 2004)
  - ARI (arithmetic reduction of intensity) model (Doyen & Gaudoin, 2004),
  - hybrid: the geometric process (Lam, 1988)

- Models can be developed based on the failure intensity function

$$\lambda_i^{(N_{i,t})}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t \downarrow 0} \frac{P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \qquad (2)$$

where $P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t+\Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$.

Development in Failure Processes: Models

- Imperfect repair: Many failure process models have been developed
  - ARA (arithmetic reduction of age) (Doyen & Gaudoin, 2004)
  - ARI (arithmetic reduction of intensity) model (Doyen & Gaudoin, 2004),
  - hybrid: the geometric process (Lam, 1988)

- Models can be developed based on the failure intensity function

$$\lambda_i^{(N_{i,t})}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t \downarrow 0} \frac{P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \qquad (2)$$

where $P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t+\Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$.

Development in Failure Processes: Models

- Imperfect repair: Many failure process models have been developed
  - ARA (arithmetic reduction of age) (Doyen & Gaudoin, 2004)
  - ARI (arithmetic reduction of intensity) model (Doyen & Gaudoin, 2004),
  - hybrid: the geometric process (Lam, 1988)

- Models can be developed based on the failure intensity function

$$\lambda_i^{(N_{i,t})}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t \downarrow 0} \frac{P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \qquad (2)$$

where $P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t+\Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$.

Development in Failure Processes: Models

- Imperfect repair: Many failure process models have been developed
  - ARA (arithmetic reduction of age) (Doyen & Gaudoin, 2004)
  - ARI (arithmetic reduction of intensity) model (Doyen & Gaudoin, 2004),
  - hybrid: the geometric process (Lam, 1988)

- Models can be developed based on the failure intensity function

$$\lambda_i^{(N_{i,t})}(t|\mathscr{H}_{i,t-}) = \lim_{\Delta t \downarrow 0} \frac{P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}}{\Delta t}, \qquad (2)$$

where $P\{N_i(t+\Delta t) - N_i(t) = 1|\mathscr{H}_{i,t-}\}$ is the probability that component i fails within the interval $(t, t+\Delta t)$, given the history of failures of the component up to time t, $\mathscr{H}_{i,t-}$.

Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

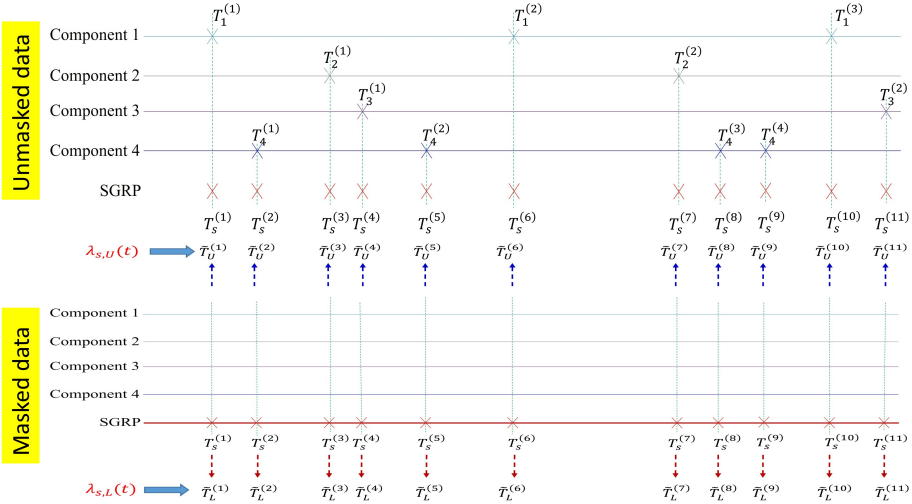
Masked Failure Data

- A common challenge in reliability engineering: a lack of failures, especially at the component level
- Masked failure data (Non-repairable vs repairable)
  - Types of masked failure data

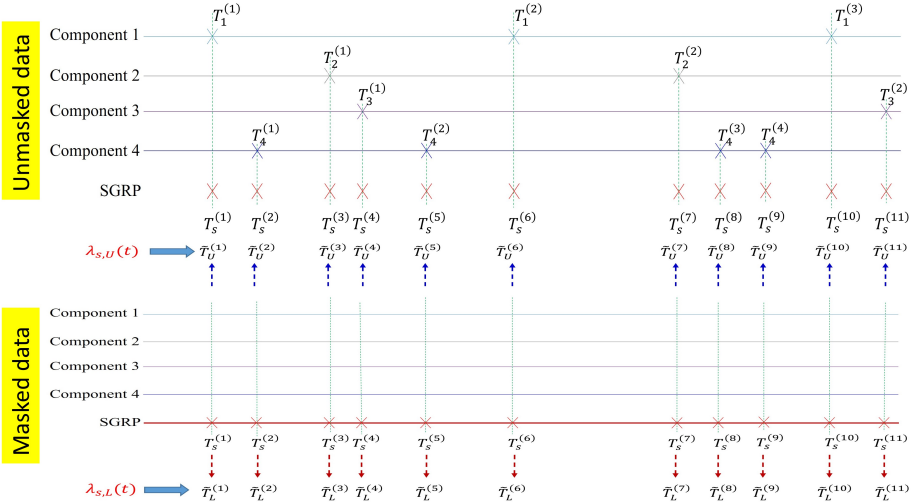
known failure time & known cause	known failure time & unknown cause
unknown failure time & known cause	unknown failure time & unknown cause

  - Examples: no-fault-found (NFF): A report shows that NFF observations can be high up to 85% in avionics in Boeing 777 electronics and can costs up to 90% of total maintenance cost (Singh, Prakash & Samad, 2021)
- In this presentation, we focus on the masked failure data with
  - Known: Times between failures are known
  - Unknown: Causes of failures are unknown

Masked Failure Data & Superimposed Renewal Processes (SRP)



Masked Failure Data & Superimposed Renewal Processes (SRP)



Prior Work on SRP

- *When time $t \to +\infty$ and the number of components in the system is infinite $m \to \infty$*
  - the observed failure times of an SRP tend to be distributed with a homogeneous Poisson process (HPP) (Cox & Smith, 1954; Drenick, 1960)
  - Lawrance (1973) obtains the survival function of the system for the case of the superposition of a finite number of stationary point processes

$$R_s(t) = \lim_{k\to\infty} P(T_s^{(k)} \geq t) = \left(\sum_{i_0=1}^{m} \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(0) \right)^{-1} \sum_{i_0=1}^{m} \left(R_{i_0}(t) \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(t) \right), \tag{3}$$

where m is the number of renewal processes, $R_{i_0}(t)$ is the survival function of the i_0th renewal process, $\Psi_{i_1}(t)$ is defined as $\Psi_{i_1}(t) = \int_t^\infty R_{i_1} k(u) du$.

Prior Work on SRP

- *When time $t \to +\infty$ and the number of components in the system is infinite $m \to \infty$*
  - the observed failure times of an SRP tend to be distributed with a homogeneous Poisson process (HPP) (Cox & Smith, 1954; Drenick, 1960)
  - Lawrance (1973) obtains the survival function of the system for the case of the superposition of a finite number of stationary point processes

$$R_s(t) = \lim_{k\to\infty} P(T_s^{(k)} \geq t) = \left(\sum_{i_0=1}^{m} \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(0) \right)^{-1} \sum_{i_0=1}^{m} \left(R_{i_0}(t) \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(t) \right), \tag{3}$$

where m is the number of renewal processes, $R_{i_0}(t)$ is the survival function of the i_0th renewal process, $\Psi_{i_1}(t)$ is defined as $\Psi_{i_1}(t) = \int_t^\infty R_{i_1} k(u) du$.

Prior Work on SRP

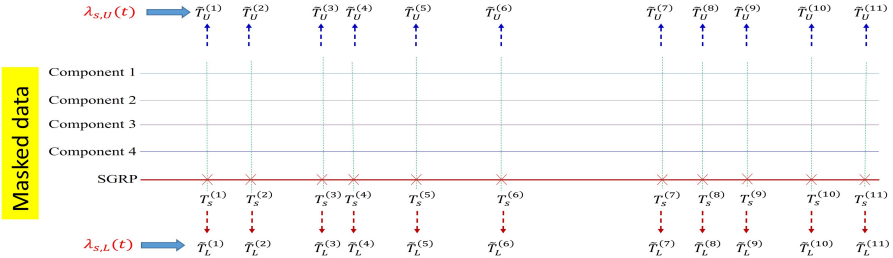
- *When time $t \to +\infty$ and the number of components in the system is infinite $m \to \infty$*
  - the observed failure times of an SRP tend to be distributed with a homogeneous Poisson process (HPP) (Cox & Smith, 1954; Drenick, 1960)
  - Lawrance (1973) obtains the survival function of the system for the case of the superposition of a finite number of stationary point processes

$$R_s(t) = \lim_{k\to\infty} P(T_s^{(k)} \geq t) = \left(\sum_{i_0=1}^{m} \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(0) \right)^{-1} \sum_{i_0=1}^{m} \left(R_{i_0}(t) \prod_{\substack{i_1=1,\ldots,m \\ i_1 \neq i_0}} \Psi_{i_1}(t) \right), \tag{3}$$

where m is the number of renewal processes, $R_{i_0}(t)$ is the survival function of the i_0th renewal process, $\Psi_{i_1}(t)$ is defined as $\Psi_{i_1}(t) = \int_t^\infty R_{i_1} k(u) du$.

Knowledge Gaps & Challenges

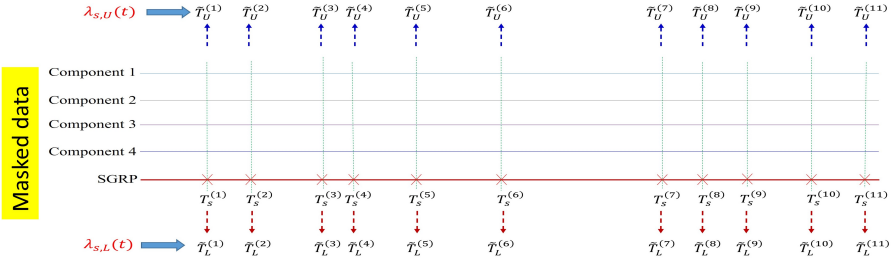
- With a limited number of failures from a component
- The failure process of a m-component subsystem follows the superposition of m-thinned stochastic processes
- Challenge 1 Masked failure data exist;



- Challenge 2 How to develop approximate approaches to estimating the superimposed failure process of a multi-component system?

Knowledge Gaps & Challenges

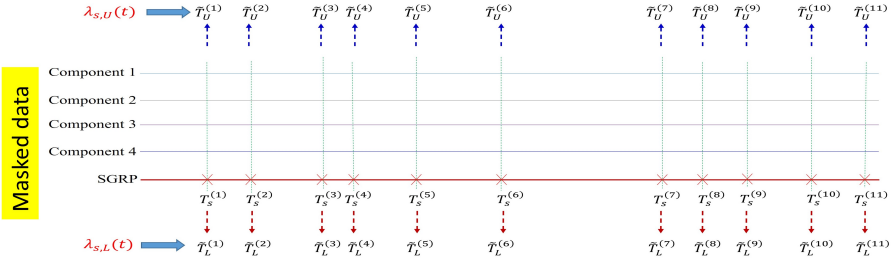
- With a limited number of failures from a component
- The failure process of a m-component subsystem follows the superposition of m-thinned stochastic processes
- Challenge 1 Masked failure data exist;



- Challenge 2 How to develop approximate approaches to estimating the superimposed failure process of a multi-component system?

Knowledge Gaps & Challenges

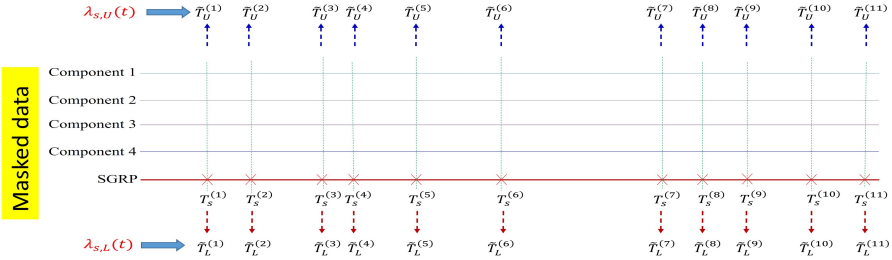
- With a limited number of failures from a component
- The failure process of a m-component subsystem follows the superposition of m-thinned stochastic processes
- Challenge 1 Masked failure data exist;



- Challenge 2 How to develop approximate approaches to estimating the superimposed failure process of a multi-component system?

Knowledge Gaps & Challenges

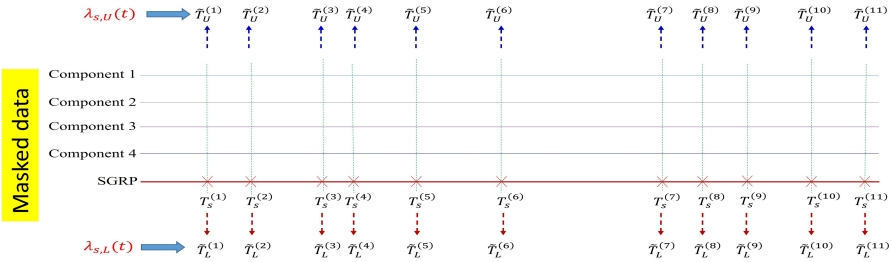
- With a limited number of failures from a component
- The failure process of a m-component subsystem follows the superposition of m-thinned stochastic processes
- Challenge 1 Masked failure data exist;



- Challenge 2 How to develop approximate approaches to estimating the superimposed failure process of a multi-component system?

Knowledge Gaps & Challenges

- With a limited number of failures from a component
- The failure process of a m-component subsystem follows the superposition of m-thinned stochastic processes
- Challenge 1 Masked failure data exist;

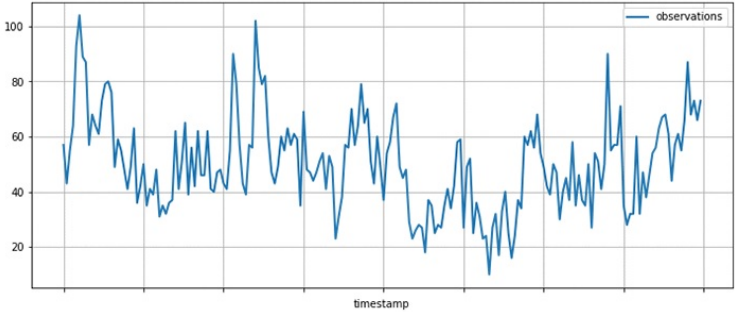


- Challenge 2 How to develop approximate approaches to estimating the superimposed failure process of a multi-component system?

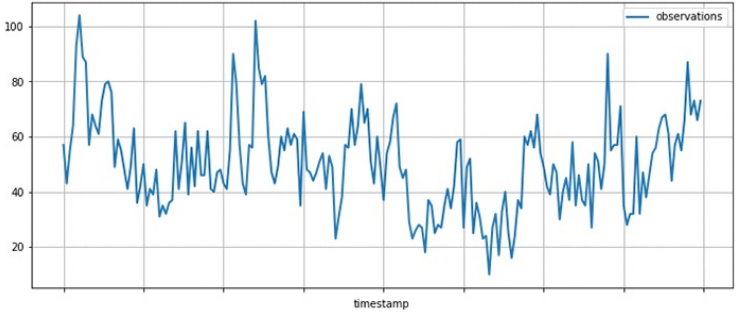
Table of Contents

1. Uncertainties and Challenges
2. Two models
3. Exponential smoothing of intensity functions
4. Dealing with Uncertainty
5. Conclusions and Suggestions

Model I: Recurrent Event Data and Time Series data



Model I: Recurrent Event Data and Time Series data



Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Recurrent Event Data and Time Series data

- Time series models: Exponential smoothing methods, ARIMA, & LSTM
- Recurrent event models:
  - Non-parametric methods for the mean cumulative intensity function
  - Intensity function methods: Prentice-Williams-Peterson model, Wei-Lin-Weissfeld model, etc
- Use a time series model in recurrent event data analysis?
  - Parsimonious model: the number of parameters should be as small as possible
  - Model selection: Which type of model fits for the purpose?

Model I: Exponential Smoothing of Intensity (ESI) Model

- Simple exponential smoothing model: Given a time series $\{x_1, x_2, \ldots\}$, the simple exponential smoothing model can be expressed by

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t$$

- Let a sub system be composed of m components in series.
- Before the first failure of the system, the components have intensity functions $\frac{1}{m}\rho^{m-1}\lambda_0(t), \ldots, \frac{1}{m}\rho\lambda_0(t), \frac{1}{m}\lambda_0(t)$, respectively, where $\rho > 0$.

Model I: Exponential Smoothing of Intensity (ESI) Model

- Simple exponential smoothing model: Given a time series $\{x_1, x_2, \ldots\}$, the simple exponential smoothing model can be expressed by

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t$$

- Let a sub system be composed of m components in series.
- Before the first failure of the system, the components have intensity functions $\frac{1}{m}\rho^{m-1}\lambda_0(t), \ldots, \frac{1}{m}\rho\lambda_0(t), \frac{1}{m}\lambda_0(t)$, respectively, where $\rho > 0$.

Model I: Exponential Smoothing of Intensity (ESI) Model

- Simple exponential smoothing model: Given a time series $\{x_1, x_2, \ldots\}$, the simple exponential smoothing model can be expressed by

$$\hat{x}_{t+1} = \alpha x_t + (1 - \alpha)\hat{x}_t$$

- Let a sub system be composed of m components in series.
- Before the first failure of the system, the components have intensity functions $\frac{1}{m}\rho^{m-1}\lambda_0(t), \ldots, \frac{1}{m}\rho\lambda_0(t), \frac{1}{m}\lambda_0(t)$, respectively, where $\rho > 0$.

Model I: The ESI(m) Model

- The failure intensity of the system after the N_t-th failure is given by

$$\lambda(t|\mathscr{H}_{t-}) = \begin{cases} \dfrac{1}{m}\displaystyle\sum_{k=0}^{m-1} \rho^{m-k-1}\lambda_0(t), & \text{if } N_t = 0, \\ \dfrac{1}{m}\left(\displaystyle\sum_{k=0}^{N_t-1} \rho^{m-k-1}\lambda_0(t - T_{N_t-k}) + \displaystyle\sum_{k=N_t}^{m-1} \rho^{m-k-1}\lambda_0(t)\right), & \text{if } 1 \leq N_t < m, \\ \dfrac{1}{m}\displaystyle\sum_{k=0}^{m-1} \rho^{m-k-1}\lambda_0(t - T_{N_t-k}), & \text{if } N_t \geq m. \end{cases} \tag{4}$$

where $\rho > 0$. t is the time since the system starts, $t \in (0, +\infty)$. We refer to the count process with intensity $\lambda(t|\mathscr{H}_{t-})$ as an ESI model, or more precisely, an ESI(m) model.

- If $\rho = 1$, the model is referred to as the mean average of intensity model (MAI)

Model I: The ESI(m) Model

- The failure intensity of the system after the N_t-th failure is given by

$$\lambda(t|\mathscr{H}_{t-}) = \begin{cases} \dfrac{1}{m}\displaystyle\sum_{k=0}^{m-1} \rho^{m-k-1}\lambda_0(t), & \text{if } N_t = 0, \\ \dfrac{1}{m}\left(\displaystyle\sum_{k=0}^{N_t-1} \rho^{m-k-1}\lambda_0(t - T_{N_t-k}) + \displaystyle\sum_{k=N_t}^{m-1} \rho^{m-k-1}\lambda_0(t)\right), & \text{if } 1 \leq N_t < m, \\ \dfrac{1}{m}\displaystyle\sum_{k=0}^{m-1} \rho^{m-k-1}\lambda_0(t - T_{N_t-k}), & \text{if } N_t \geq m. \end{cases} \tag{4}$$

  where $\rho > 0$. t is the time since the system starts, $t \in (0, +\infty)$. We refer to the count process with intensity $\lambda(t|\mathscr{H}_{t-})$ as an ESI model, or more precisely, an ESI(m) model.
- If $\rho = 1$, the model is referred to as the mean average of intensity model (MAI)

Model I: Some Properties

- Proposition 1: For a given m (the number of components in the system), if $\int_0^\infty \lambda_0(u)dG_k(u) < \infty$, then $E[\lambda(t|\mathscr{H}_{t-})] < \infty$ for $t \to \infty$, where $G_k(u)$ is the probability distribution of time to the k-th failure.

- Proposition 2: $\lambda(t|\mathscr{H}_{t-})$ may show monotonicity under some conditions, as shown below.
  - If $\lambda_0(t)$ is an increasing function in t, then $\lambda(t|\mathscr{H}_{t-})$ is an increasing function for $t \in (T^+_{N_t}, T^-_{N_t+1})$;
  - If $\lambda_0(t)$ is a decreasing function in t, then $\lambda(t|\mathscr{H}_{t-})$ is a decreasing function for $t \in (T^+_{N_t}, T^-_{N_t+1})$.

Model I: Some Properties

- Proposition 1: For a given m (the number of components in the system), if $\int_0^\infty \lambda_0(u)dG_k(u) < \infty$, then $E[\lambda(t|\mathscr{H}_{t-})] < \infty$ for $t \to \infty$, where $G_k(u)$ is the probability distribution of time to the k-th failure.

- Proposition 2: $\lambda(t|\mathscr{H}_{t-})$ may show monotonicity under some conditions, as shown below.
  - If $\lambda_0(t)$ is an increasing function in t, then $\lambda(t|\mathscr{H}_{t-})$ is an increasing function for $t \in (T^+_{N_t}, T^-_{N_t+1})$;
  - If $\lambda_0(t)$ is a decreasing function in t, then $\lambda(t|\mathscr{H}_{t-})$ is a decreasing function for $t \in (T^+_{N_t}, T^-_{N_t+1})$.

Model I: Simulation Results

The means and standard deviations of $(-\log(L))$ from 30 repetitions.

n	m	Estimated values of $(-\log(L))$ Nine existing models									New models	
		$q=2$		$q=3$					$q=4$		$q=3$	$q=2$
		RP	NHPP-PL	GP	Kijima I	Kijima II	ARI_m	ARA_m	BBIP	Model II	ESI	MAI
	5	45.33	42.40	43.08	41.65	41.17	40.82	41.20	42.18	40.60	40.82	41.00
		(5.07)	(4.47)	(4.21)	(4.25)	(4.35)	(4.21)	(4.29)	(4.21)	(4.23)	(4.25)	(4.28)
15	15	28.73	24.79	26.72	24.18	23.96	24.15	24.25	24.32	23.14	23.84	24.02
		(4.21)	(3.38)	(3.23)	(3.57)	(3.68)	(3.72)	(3.72)	(3.23)	(3.65)	(3.77)	(3.67)
	25	23.16	19.35	21.62	18.32	18.33	18.36	18.31	18.96	17.53	18.01	18.24
		(3.95)	(3.58)	(3.04)	(3.84)	(3.74)	(3.83)	(3.80)	(3.47)	(3.96)	(3.89)	(3.87)
	5	91.80	86.67	88.15	86.02	85.46	85.30	85.51	86.34	85.09	85.22	85.33
		(9.10)	(8.13)	(8.10)	(8.08)	(7.99)	(8.33)	(8.16)	(8.10)	(8.10)	(8.06)	(8.17)
30	15	57.69	50.53	53.60	49.96	49.34	50.02	49.92	49.98	48.42	49.06	49.23
		(7.70)	(6.03)	(5.99)	(6.19)	(6.14)	(6.15)	(6.12)	(5.98)	(6.23)	(6.13)	(6.18)
	25	45.03	36.95	41.76	36.46	35.97	36.26	36.26	36.52	35.22	35.67	35.84
		(7.32)	(5.57)	(5.45)	(5.66)	(5.81)	(5.92)	(5.79)	(5.44)	(5.83)	(5.93)	(5.95)
	5	136.12	127.52	130.17	126.61	125.76	125.60	125.71	127.20	125.15	125.41	125.56
		(13.97)	(11.82)	(12.19)	(11.69)	(11.63)	(11.75)	(11.66)	(11.84)	(11.31)	(11.68)	(11.62)
45	15	84.94	74.17	79.01	73.65	73.12	73.99	73.52	73.49	72.15	72.75	72.96
		(11.66)	(7.87)	(8.77)	(7.95)	(8.75)	(8.24)	(7.93)	(7.86)	(7.73)	(8.06)	(8.02)
	25	64.92	52.24	59.85	51.72	51.31	51.69	51.62	51.68	50.57	51.09	51.30
		(12.03)	(6.77)	(9.13)	(6.94)	(7.07)	(7.20)	(7.11)	(6.71)	(7.00)	(7.02)	(7.09)

Model I: Comparison on the Simulated Data

- Comparing the AIC values of the 11 models, then the ESI has the smallest AIC value in 10 out of the 15 cases. That is, the MAI outperforms the other 10 models in term of AIC.

- Comparing the values of the $\mathrm{AIC_c}$ and BIC of the 11 models, then the MAI has the smallest values of the $\mathrm{AIC_c}$ and BIC in 11 out of the 15 cases. That is, the MAI outperforms the other models in terms of both $\mathrm{AIC_c}$ and BIC, as shown in the fifth column.

Model I: Comparison on the Simulated Data

- Comparing the AIC values of the 11 models, then the ESI has the smallest AIC value in 10 out of the 15 cases. That is, the MAI outperforms the other 10 models in term of AIC.

- Comparing the values of the $\mathrm{AIC_c}$ and BIC of the 11 models, then the MAI has the smallest values of the $\mathrm{AIC_c}$ and BIC in 11 out of the 15 cases. That is, the MAI outperforms the other models in terms of both $\mathrm{AIC_c}$ and BIC, as shown in the fifth column.

Model I: 15 Real-world Datasets for Model Validation

No.	Dataset	n	Data source	Model
1	Hydraulic system (LHD 1)	23	Kumar and Klefsjö (1992)	NHPP-PL
2	Hydraulic system (LHD 3)	25	Kumar and Klefsjö (1992)	NHPP-PL
3	Hydraulic system (LHD 9)	27	Kumar and Klefsjö (1992)	NHPP-PL
4	Hydraulic system (LHD 11)	28	Kumar and Klefsjö (1992)	NHPP-PL
5	Hydraulic system (LHD 17)	26	Kumar and Klefsjö (1992)	NHPP-PL
6	Hydraulic system (LHD 20)	23	Kumar and Klefsjö (1992)	NHPP-PL
7	Air conditioner (TBF 7909)	24	Proschan (1963)	HPP
8	Air conditioner (TBF 7912)	30	Proschan (1963)	HPP
9	Air conditioner (TBF 7913)	27	Proschan (1963)	HPP
10	Air conditioner (TBF 7914)	23	Proschan (1963)	HPP
11	Compressor	24	Yanez et al. (2002)	Kijima model I
12	Main propulsion motor	24	Yanez et al. (2002)	Kijima I
13	Powertrain System 510	55	Guida and Pulcini (2009)	BBIP
14	Powertrain System 514	35	Guida and Pulcini (2009)	BBIP
15*	Diesel engine	56	Lee (1980)	NHPP-WLL

Model I: Results of Comparison with Other 9 Models on the 15 Datasets

No.	ESI		MAI	
	$-\log(L)$	$-\log(L)$ of the "$q = 3$" models	AIC	$\mathrm{AIC_c}$ & BIC
1	Model II	ESI	NHPP-PL	NHPP-PL
2	ESI	ESI	MAI	MAI
3	ESI	ESI	NHPP-PL	NHPP-PL
4	ESI	ESI	MAI	MAI
5	$\mathrm{ARA_m}$	$\mathrm{ARA_m}$	MAI	MAI
6	Model II	ESI	MAI	MAI
7	GP	GP	MAI	MAI
8	Model II	ESI	NHPP-PL	NHPP-PL
9	$\mathrm{ARI_m}$	$\mathrm{ARI_m}$	MAI	MAI
10	Model II	ESI	MAI	MAI
11	$\mathrm{ARA_m}$	$\mathrm{ARA_m}$	MAI	MAI
12	$\mathrm{ARI_m}$	$\mathrm{ARI_m}$	NHPP-PL	NHPP-PL
13	Kijima II	Kijima II	MAI	MAI
14	BBIP	ESI	ESI	MAI
15	$\mathrm{ARI_m}$	$\mathrm{ARI_m}$	MAI	MAI
Frequency	4×Model II	8×ESI	10×MAI	11×MAI

Model I: Comparison based on the 15-Real-Datasets

- When comparing the AIC values of the 11 models, we find that the ESI has the smallest AIC value in 10 out of the 15 cases. That is, the MAI outperforms the other 10 models in term of AIC, as shown in the fourth column.

- When comparing the values of the AICc and BIC of the 11 models, we can find that the MAI has the smallest values of the AICc and BIC in 11 out of the 15 cases. That is, the MAI outperforms the other models in terms of both AICc and BIC, as shown in the fifth column.

Model I: Comparison based on the 15-Real-Datasets

- When comparing the AIC values of the 11 models, we find that the ESI has the smallest AIC value in 10 out of the 15 cases. That is, the MAI outperforms the other 10 models in term of AIC, as shown in the fourth column.

- When comparing the values of the AICc and BIC of the 11 models, we can find that the MAI has the smallest values of the AICc and BIC in 11 out of the 15 cases. That is, the MAI outperforms the other models in terms of both AICc and BIC, as shown in the fifth column.

Model II: Bounds of the Intensity Function of a SGRP

- **Proposition 1:** If replacement is performed on each failed component, or the failure process of the system is the SGRP, then

$$\sum_{i=0}^{n-1} \lambda(t - T_s^{(N_{s,t}-i)}) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda(t - T_s^{(N_{s,t})}), \tag{5}$$

where $T_s^{(k)} = 0$ if $k \leq 0$ and $N_{s,t} \geq 0$, $T_{s_i}^{(k)}$ is the time when the kth failure of system i occurs.

- **Proposition 2:** Under some assumptions, the failure intensity $\lambda_s(t)$ of the system after the $N_{s,t}$-th repair satisfies

$$\sum_{i=0}^{n-1} \lambda_i^{(N_{s,t}-i)}(t) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda^{(N_{s,t})}(t), \tag{6}$$

where $\lambda_i^{(N_{s,t}-i)}(t) = \lambda(t)$ for $N_{s,t} - i \leq 0$.

Model II: Bounds of the Intensity Function of a SGRP

- **Proposition 1:** If replacement is performed on each failed component, or the failure process of the system is the SGRP, then

$$\sum_{i=0}^{n-1} \lambda(t - T_s^{(N_{s,t}-i)}) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda(t - T_s^{(N_{s,t})}), \tag{5}$$

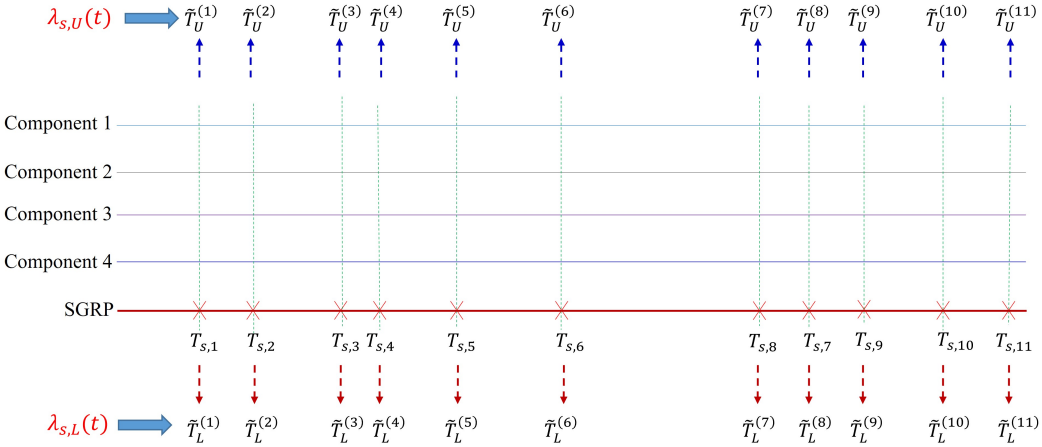
where $T_s^{(k)} = 0$ if $k \leq 0$ and $N_{s,t} \geq 0$, $T_{s_i}^{(k)}$ is the time when the kth failure of system i occurs.

- **Proposition 2:** Under some assumptions, the failure intensity $\lambda_s(t)$ of the system after the $N_{s,t}$-th repair satisfies

$$\sum_{i=0}^{n-1} \lambda_i^{(N_{s,t}-i)}(t) \leq \lambda_s(t) \leq (n-1)\lambda(t) + \lambda^{(N_{s,t})}(t), \tag{6}$$

where $\lambda_i^{(N_{s,t}-i)}(t) = \lambda(t)$ for $N_{s,t} - i \leq 0$.

Model II: Lower and Upper Bounds



Model II: An Weighted Average Model

Based on Proposition 2, we can use the following model, which is the weighted average of the lower bound and the upper bound of the Inequality (6), respectively, to approximate the rate of the SGRP of the n components in a series system for $N_{s,t} \geq 1$.

$$\widetilde{\lambda}_s(t) = \delta \sum_{i=0}^{n-1} \lambda^{(N_{s,t}-i)}(t) + (1-\delta)\left((n-1)\lambda(t) + \lambda^{(N_{s,t})}(t)\right),$$

$$= (n-1)(1-\delta)\lambda(t) + (1-\delta)\lambda^{(N_{s,t})}(t) + \delta \sum_{i=0}^{n-1} \lambda^{(N_{s,t}-i)}(t) \qquad (7)$$

where $\delta \in [0,1]$.

Model II: Lower and Upper Bounds

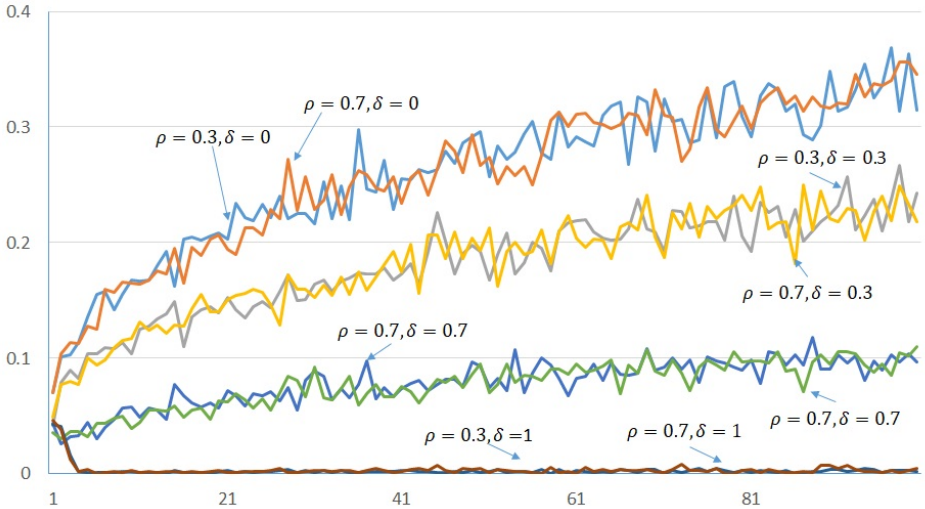


Table of Contents

1. Uncertainties and Challenges
2. Two models
3. Exponential smoothing of intensity functions
4. Dealing with Uncertainty
5. Conclusions and Suggestions

Accounting for Uncertainty with Modern Portfolio Theory

- A good decision-making needs considering uncertainty
- Modern portfolio theory (MPT), or mean-variance analysis, is a mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk
- Modern portfolio theory for PM policy optimisation
- other methods: Value at Risk (VaR), Conditional VaR (CVaR)

Accounting for Uncertainty with Modern Portfolio Theory

- A good decision-making needs considering uncertainty
- Modern portfolio theory (MPT), or mean-variance analysis, is a mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk
- Modern portfolio theory for PM policy optimisation
- other methods: Value at Risk (VaR), Conditional VaR (CVaR)

Accounting for Uncertainty with Modern Portfolio Theory

- A good decision-making needs considering uncertainty
- Modern portfolio theory (MPT), or mean-variance analysis, is a mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk
- Modern portfolio theory for PM policy optimisation
- other methods: Value at Risk (VaR), Conditional VaR (CVaR)

Accounting for Uncertainty with Modern Portfolio Theory

- A good decision-making needs considering uncertainty
- Modern portfolio theory (MPT), or mean-variance analysis, is a mathematical framework for assembling a portfolio of assets such that the expected return is maximized for a given level of risk
- Modern portfolio theory for PM policy optimisation
- other methods: Value at Risk (VaR), Conditional VaR (CVaR)

Dealing with Uncertainty

- **Age replacement**: A unit is replaced with a new identical unit either at a pre-determined age T or upon failure if it occurs earlier. The expected cost per unit of time is

$$C_a(T|\boldsymbol{\theta}) = \frac{c_r F(T|\boldsymbol{\theta}) + c_m \bar{F}(T|\boldsymbol{\theta})}{\int_0^T \bar{F}(t|\boldsymbol{\theta})dt}. \tag{8}$$

- **Block replacement** A unit is replaced with a new identical unit upon failure or at fixed times $kT(k = 1, 2, ...,)$. The expected cost per unit of time is

$$C_b(T|\boldsymbol{\theta}) = \frac{c_r M(T|\boldsymbol{\theta}) + c_m}{T}. \tag{9}$$

Dealing with Uncertainty

- **Age replacement**: A unit is replaced with a new identical unit either at a pre-determined age T or upon failure if it occurs earlier. The expected cost per unit of time is

$$C_a(T|\boldsymbol{\theta}) = \frac{c_r F(T|\boldsymbol{\theta}) + c_m \bar{F}(T|\boldsymbol{\theta})}{\int_0^T \bar{F}(t|\boldsymbol{\theta})dt}. \tag{8}$$

- **Block replacement** A unit is replaced with a new identical unit upon failure or at fixed times $kT(k = 1, 2, ...,)$. The expected cost per unit of time is

$$C_b(T|\boldsymbol{\theta}) = \frac{c_r M(T|\boldsymbol{\theta}) + c_m}{T}. \tag{9}$$

MPT for PM Policy Optimisation

- Then, the expected total cost of maintenance of the n systems is given by

$$\mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta}) = \sum_{i=1}^{n} C_i(T_i|\boldsymbol{\theta}_i), \tag{10}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ and $\boldsymbol{T} = (T_1, \ldots, T_n)$.

- One may select the decision $\boldsymbol{T}$ by

$$\min_{\boldsymbol{T} \in \mathbb{R}} \mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta}), \tag{11}$$
$$\text{subject to}: \mathrm{V}_{\boldsymbol{\theta}}(\mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta})) \leq \nu_0.$$

where ν_0 reflects decision maker's risk attitude.

- Other methods

MPT for PM Policy Optimisation

- Then, the expected total cost of maintenance of the n systems is given by

$$\mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta}) = \sum_{i=1}^{n} C_i(T_i|\boldsymbol{\theta}_i), \tag{10}$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_n)$ and $\boldsymbol{T} = (T_1, \ldots, T_n)$.

- One may select the decision $\boldsymbol{T}$ by

$$\min_{\boldsymbol{T} \in \mathbb{R}} \mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta}), \tag{11}$$
$$\text{subject to}: \mathrm{V}_{\boldsymbol{\theta}}(\mathrm{C}(\boldsymbol{T}|\boldsymbol{\theta})) \leq \nu_0.$$

where ν_0 reflects decision maker's risk attitude.

- Other methods

Numerical Examples

- Collectively optimising three maintenance policies
  - Age replacement policy
  - Block replacement policy
  - Periodic replacement with minimal repair

- Results

Method	Separately
Age replacement	$T_a = 1.03,\ C_a(T\|\theta) = 21.05,\ Var(C_a(T\|\theta)) = 139.5$
Block replacement	$T_a = 1.03,\ C_a(T\|\theta) = 21.05,\ Var(C_a(T\|\theta)) = 139.5$
Periodic replacement	$T_p = 0.49,\ C_p(T\|\theta) = 41.04,\ Var(C_a(T\|\theta)) = 2.13$
Collectively	$\nu_0 = 1.23 \times 10^7.\ T_a = 1.01,\ T_b = 0.54\ \&\ T_p = 0.59$

Numerical Examples

- Collectively optimising three maintenance policies
  - Age replacement policy
  - Block replacement policy
  - Periodic replacement with minimal repair

- Results

Method	Separately
Age replacement	$T_a = 1.03,\ C_a(T\|\boldsymbol{\theta}) = 21.05,\ Var(C_a(T\|\boldsymbol{\theta})) = 139.5$
Block replacement	$T_a = 1.03,\ C_a(T\|\boldsymbol{\theta}) = 21.05,\ Var(C_a(T\|\boldsymbol{\theta})) = 139.5$
Periodic replacement	$T_p = 0.49,\ C_p(T\|\boldsymbol{\theta}) = 41.04,\ Var(C_a(T\|\boldsymbol{\theta})) = 2.13$
Collectively	$\nu_0 = 1.23 \times 10^7.\ \ T_a = 1.01,\ T_b = 0.54\ \&\ T_p = 0.59$

Table of Contents

1. Uncertainties and Challenges
2. Two models
3. Exponential smoothing of intensity functions
4. Dealing with Uncertainty
5. Conclusions and Suggestions

Conclusions

- Limitations of the proposed models
  - The proposed models are recursive models
  - how can we define the level of granularity?
- Future work
  - To improve the existing models and research on validation methods
  - How can we deal with uncertainty better?

Conclusions

- Limitations of the proposed models
  - The proposed models are recursive models
  - how can we define the level of granularity?
- Future work
  - To improve the existing models and research on validation methods
  - How can we deal with uncertainty better?

Conclusions

- Limitations of the proposed models
  - The proposed models are recursive models
  - how can we define the level of granularity?
- Future work
  - To improve the existing models and research on validation methods
  - How can we deal with uncertainty better?

Conclusions

- Limitations of the proposed models
  - The proposed models are recursive models
  - how can we define the level of granularity?
- Future work
  - To improve the existing models and research on validation methods
  - How can we deal with uncertainty better?

Conclusions

- Limitations of the proposed models
  - The proposed models are recursive models
  - how can we define the level of granularity?
- Future work
  - To improve the existing models and research on validation methods
  - How can we deal with uncertainty better?

Suggestions

- Individualise systems, use all possible information, including ambient data, auxiliary variables
- Expert experience: using any information with predictive value that the algorithm does not consider
- Large Language Models for reliability assessment

Suggestions

- Individualise systems, use all possible information, including ambient data, auxiliary variables
- Expert experience: using any information with predictive value that the algorithm does not consider
- Large Language Models for reliability assessment

Suggestions

- Individualise systems, use all possible information, including ambient data, auxiliary variables
- Expert experience: using any information with predictive value that the algorithm does not consider
- Large Language Models for reliability assessment

Thank you all!

Many thanks to my co-authors: Prof Phil Scarf, Prof Frank Coolen, Prof Majid Asadi, Dr Bin Liu

Many thanks to Rasa and Phuc for their invitation

Questions?

Resources and Acknowledgement

- This presentation is based on the following papers:
  - Wu, S., & Asadi, M. (2024). A preventive maintenance policy and a method to approximate the failure process for multi-component systems. *European Journal of Operational Research*, 318(3), 825-835.
  - Wu, S. (2019). A failure process model with the exponential smoothing of intensity functions. *European Journal of Operational Research*, 275(2), 502-513.
  - Wu, S., & Scarf, P. (2017). Two new stochastic models of the failure process of a series system. *European Journal of Operational Research*, 257(3), 763-772.
  - Wu, S., Coolen, F. P., & Liu, B. (2017). Optimization of maintenance policy under parameter uncertainty using portfolio theory. IISE Transactions, 49(7), 711-721.

References

- Jovicic, E., Primorac, D., Cupic, M., & Jovic, A. (2023). Publicly available datasets for predictive maintenance in the energy sector: A review. *IEEE access*, 11, 73505-73520.
- Singh, V. P., Ganguly, K., & Samad, T. A. (2022). No fault found: a systematic literature review and future research agenda. *International Journal of Quality & Reliability Management*, 39(5), 1281-1306.
- Cox, D. R., & Smith, W. L. (1954). On the superposition of renewal processes. *Biometrika*, 41(1-2), 91-99.
- Drenick, R. F. (1960). The failure law of complex equipment. *Journal of the Society for Industrial and Applied Mathematics*, 8(4), 680-690.
- Whitt, W. (1982). Approximating a point process by a renewal process, I: Two basic methods. *Operations Research*, 30(1), 125-147.

References (cont.)

- Albin, S. L. (1984). Approximating a point process by a renewal process, II: Superposition arrival processes to queues. *Operations Research*, 32(5), 1133-1162.
- Lawrance, A. J. (1973). Dependency of intervals between events in superposition processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 35(2), 306-315.
- Doyen, L., & Gaudoin, O. (2004). Classes of imperfect repair models based on reduction of failure intensity or virtual age. *Reliability Engineering & System Safety*, 84(1), 45-56.
- Lam, Y. (1988). Geometric processes and replacement problem. *Acta Mathematicae Applicatae Sinica*, 4, 366-377.

Asset management and reliability theory for infrastructure managers

Some key challenges in moving from science to industry practice: open problems, knowledge transfer and standardization

Dr. Thomas GUILLON

R&D Senior Expert at RTE, Vice-chair of IEC TC 123

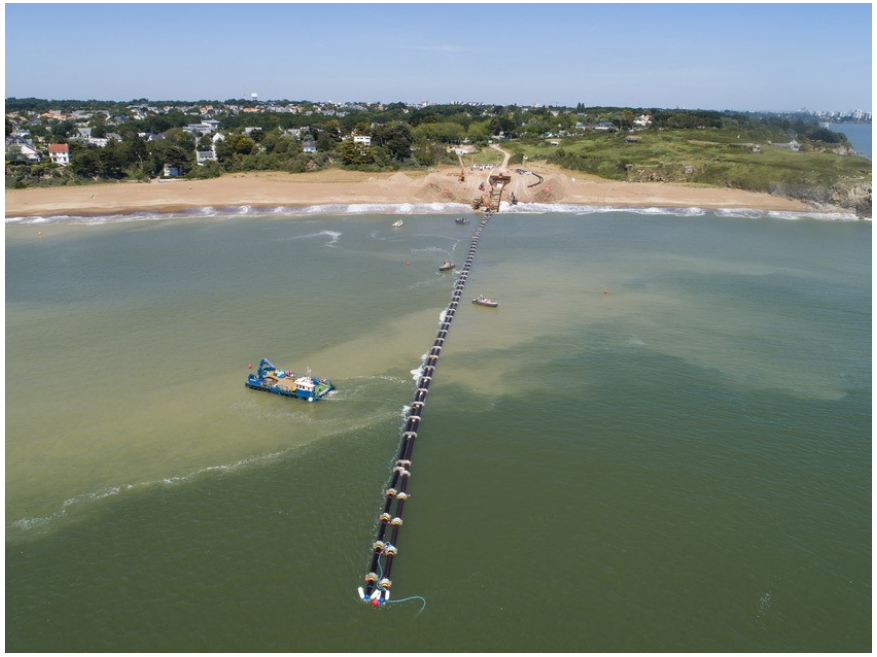
Key challenges facing power system networks

Power system networks are facing unprecedented challenges as they adapt to a rapidly changing energy and industrial landscape.

- **Asset Renewal**: Maintain and renew ageing infrastructures to ensure reliability and continuity of current network.
- **Network Development**: Support the energy transition and electrification of end-uses through the expansion and adaptation of the grid.
- **New Grid Connections**: Integrate new renewable energy sources, data centers, and reindustrialization projects.

Power systems need also to deal with emerging External Constraints:

- Climate change
- Increasing tensions in global supply
- Stricter environmental regulations impacting asset design.

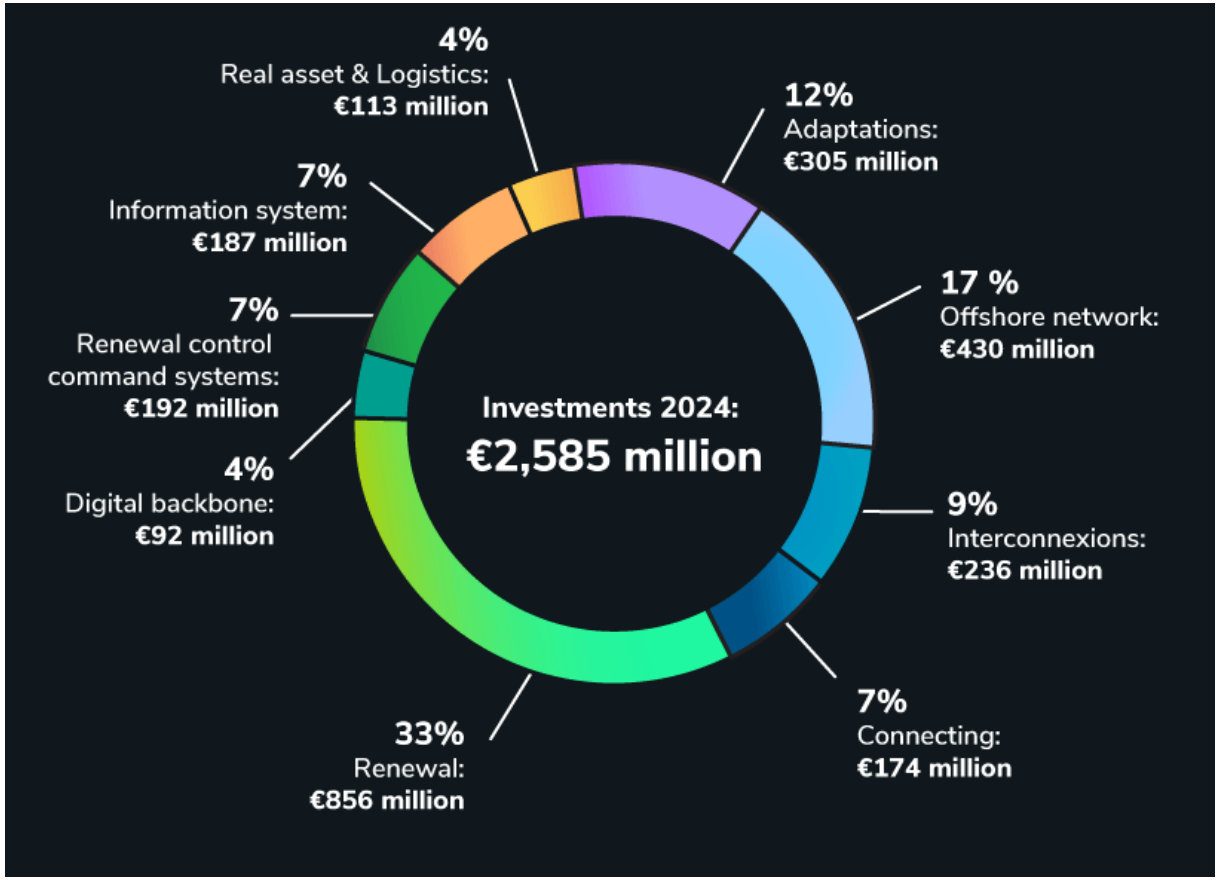


Some figures about RTE

RTE is the French Transmission System Operator
The largest in size in Europe:

- 98 210 km of overhead lines
- 7 800 km of underground lines
- 540 km of submarine lines
- 2 842 substations

- 10 400 employees
- 5 559 M€ of revenue in 2024
- 2 585 M€ of investments in 2024



Asset Management Under Pressure: Investment and Uncertainty

The growing challenges increases the demands on resources pressuring the **justification of renewal investments** in a context of limited budgets.

Regulators ask for numbers, but:

- Lack of reliable data on failures and lifetimes.
- Wide variety of approaches/heuristics providing numbers: reliability theory, health indices, condition monitoring, predictive maintenance & AI, risk matrices.
- Resource constraints often delay renewal plans, what are the effects on the objectives?

The risk decision-making context might be unclear.
How can science and standardization help justify or guide renewal decisions?

Scientific Foundations for Asset Management Decisions

Asset management provides a structured framework to ensure that assets deliver value (ISO 55000 series).

Value is primarily defined in terms of failure risk mitigation
When to replace? — and what else should be considered

We rely on 3 scientific pillars:

1. **Risk Science** – Which risk management strategies should be considered?

2. **Reliability Theory** – Survival analysis, renewal theory, and maintenance policies.

3. **Socio-Economic Evaluation** – Societal costs, Discounting, economic indicators, time horizons.

Age-Based Replacement Policy: The Foundational Example

Age-based replacement policies are very common — at least driven by accounting depreciation periods.

Reliability theory allows us to define an **optimal replacement age** that minimizes the asymptotic cost per unit of time:

$$C(a) = \frac{c_f\, F(a) + c_p\, S(a)}{\int_0^a S(x)\,\mathrm{d}x} + C_m$$

Where:

- $C(a)$ the cost function to minimize
- F cumulative distribution function with $S = 1 - F$
- c_f replacement cost upon failure
- c_p preventive replacement cost
- C_m fixed routine maintenance cost per unit time

Justifying a replacement age requires a significant amount of **knowledge**.

1 Risk science

What is a risk?

Asset renewal is often driven by the risk of failure — but how can that risk be characterized?

Risk: Combination *(C,U)* of undesirable consequences *C* of an event (or activity) and associated uncertainties *U*.

Uncertainty: Imperfect or incomplete information/knowledge about a hypothesis, a quantity or the occurrence of an event:

- Aleatory uncertainty – Inherent variability within a population, typically modeled using probabilities.
- Epistemic uncertainty – Incomplete knowledge that limits understanding or modeling of the system
- Normative ambiguity – Differing views on values, priorities and objectives among stakeholders.
  (ex. Constraints on resources, conflicts between operations rules, ambiguity in roles and responsibilities, etc.)

SRA (2018) Aven, T., & *al.* Society for risk analysis glossary.
Aven, T., & Thekdi, S. (2021). Risk science: An introduction. Routledge.
Aven, T., & Renn, O. (2020). Some foundational issues related to risk governance and different types of risks. *Journal of Risk Research, 23*(9), 1121-1134.

The three risk management strategies depending on uncertainty

Strategy	Context	Objective
Risk analysis	Strong knowledge for representing: • uncertainty using probabilities • quantified (monetized) consequences	Calibrate and compare risk mitigation options based on the strength of knowledge.
Cautionary principle	Significant epistemic uncertainty (lack of knowledge, data or understanding of the underlying phenomena).	Improve knowledge and collect data. Implement temporary measures: • Warning • Limitation of activities • Resilience • Protection
Discursive strategy	Differing views among stakeholders on values, priorities, or objectives regarding risk representation and controls.	Share knowledge and objectives. Seek consensus on objectives and risk controls.

Aven, T., & Renn, O. (2020). Some foundational issues related to risk governance and different types of risks. *Journal of Risk Research, 23*(9), 1121-1134.

IEC 63223-2 Risk-Informed Decision-Making process

IEC 63223-2
Management of network assets in power systems
Risk-informed Decision-Making process

42 Experts from 14 Countries in WG4 of TC 123

Seek a consensus based on risk science to:

- Characterize the nature of uncertainties (aleatory, epistemic, and normative ambiguity)
- Highlight the most relevant risk management strategies
- Provide decision-makers with a clear framework to assess and compare risk mitigation options
- Ensure transparency, consistency, and traceability in decision-making

Expectations of the project:

- Possibility of certification for power system operators
- Independent from, but compatible with ISO 55001
- Designed for easy adoption, supports current heuristics and fosters continuous improvement

The risk management strategies

6. Risk management strategies

6.1 Risk analysis

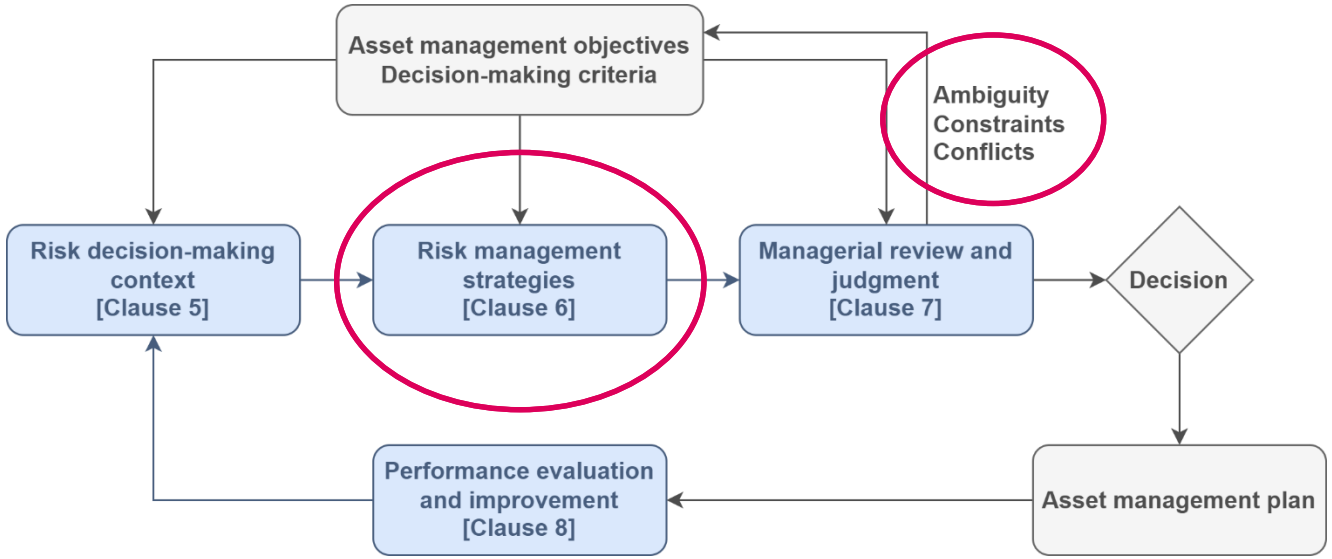
6.1.1 Risk mitigation options identification
6.1.2 Probability models
6.1.3 Consequences evaluation
6.1.4 Risk mitigation options evaluation
6.1.5 Strength of Knowledge

6.2 Cautionary principle

6.2.1 Actions to improve knowledge
6.2.2 Identifying cautionary measures
6.2.3 Impact of the cautionary measures on the stakeholders

6.3 Discursive strategy

6.3.1 Sharing knowledge and objectives
6.3.2 Alignment on values and objectives
6.3.3 Alignment on risk controls



Vote planned for summer 2025 (CDV Stage)
Final publication expected early 2026

Example of clause structure and wording

6.1.1 Risk mitigation options identification

Introduction

Risk analysis, in conjunction with a quantitative approach, is intended to compare risk mitigation options by considering investment and operational costs, valued consequences, and measures of uncertainty, in alignment with the decision-making criteria.

Requirement

The organization shall identify and select the risk mitigation options to be evaluated consistently with the risk decision-making context and risk identification outlined in 5.1.

Illustration

EXAMPLE: Some usual preventive risk mitigation options are opportunistic replacement, preventive replacement by age, inspection to detect failure, conditional replacement by inspection of deterioration, conditional replacement triggered by an alarm from online monitoring, patch deployment, etc. (see IEC 60300-3-11:2009 [18] for maintenance types).

NOTE 1: Run-to-failure replacement can be considered as a risk mitigation option serving as a reference for comparing other options, particularly to assess the added value of other risk mitigation solutions.

Recommendation

The organization should ensure that the knowledge characterized in 5.2 is sufficient to conduct a risk analysis and evaluate the selected risk mitigation options. In the absence of such knowledge, the organization should rely on the cautionary principle outlined in section 6.2 or the discursive strategy described in section 6.3.

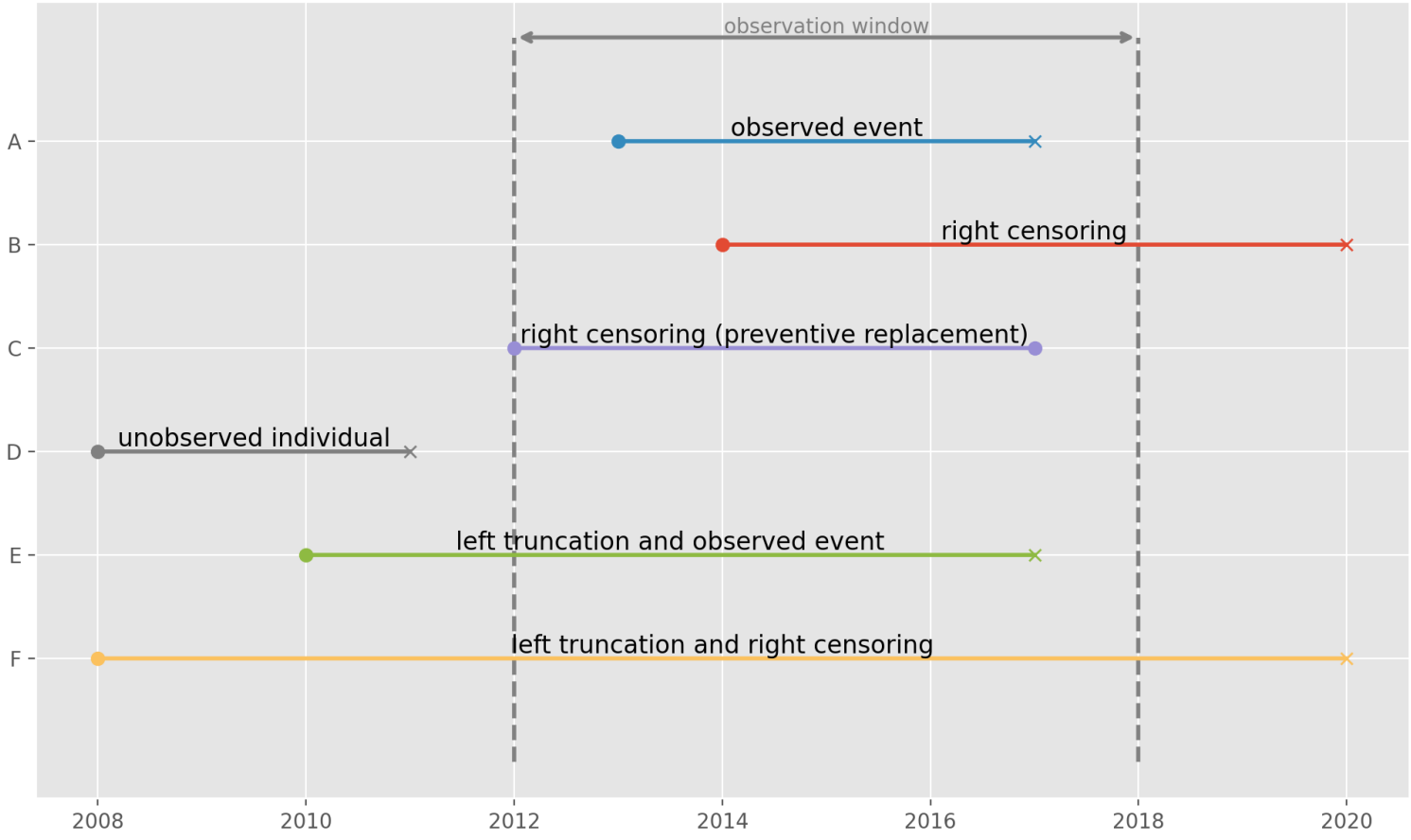
NOTE 2: Clauses 6.1.2, 6.1.3, 6.1.4 and 6.1.5 only apply when a risk analysis is effectively conducted.

2 Reliability Theory

Information Systems and Observation Biases

Infrastructure operators are mostly focused on real-time operations:

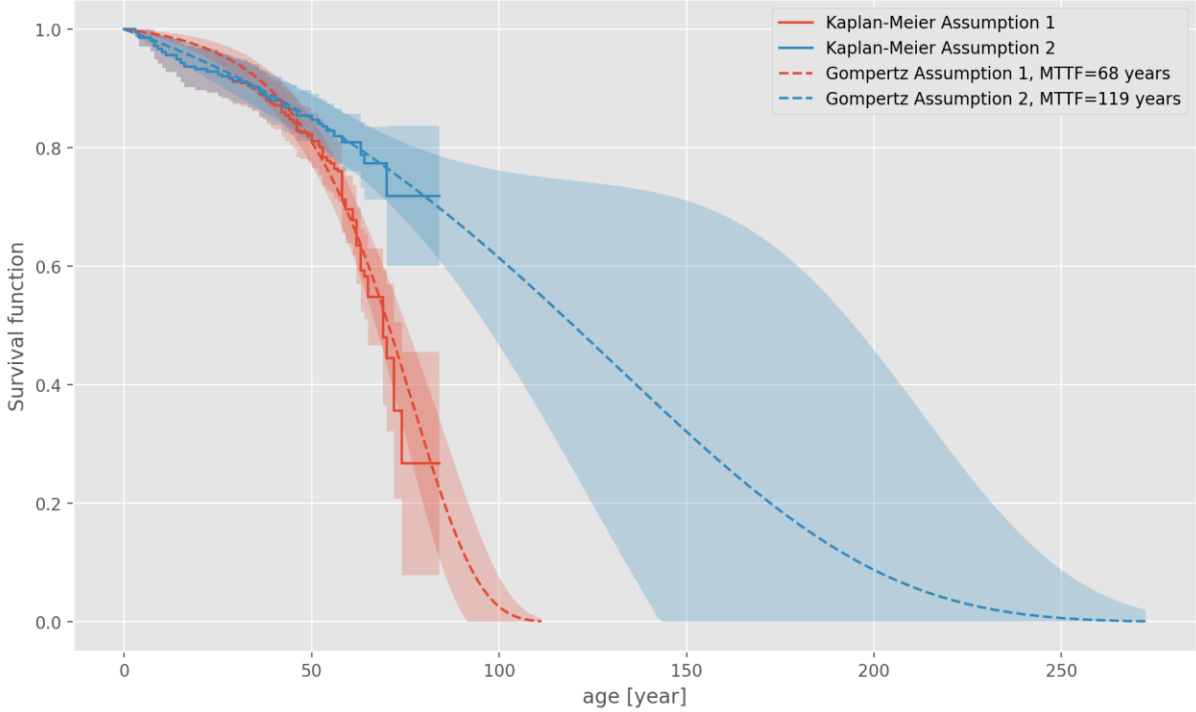
- Asset registries are often not linked to event or maintenance operation databases
- Information systems only began accurately recording events after a certain year



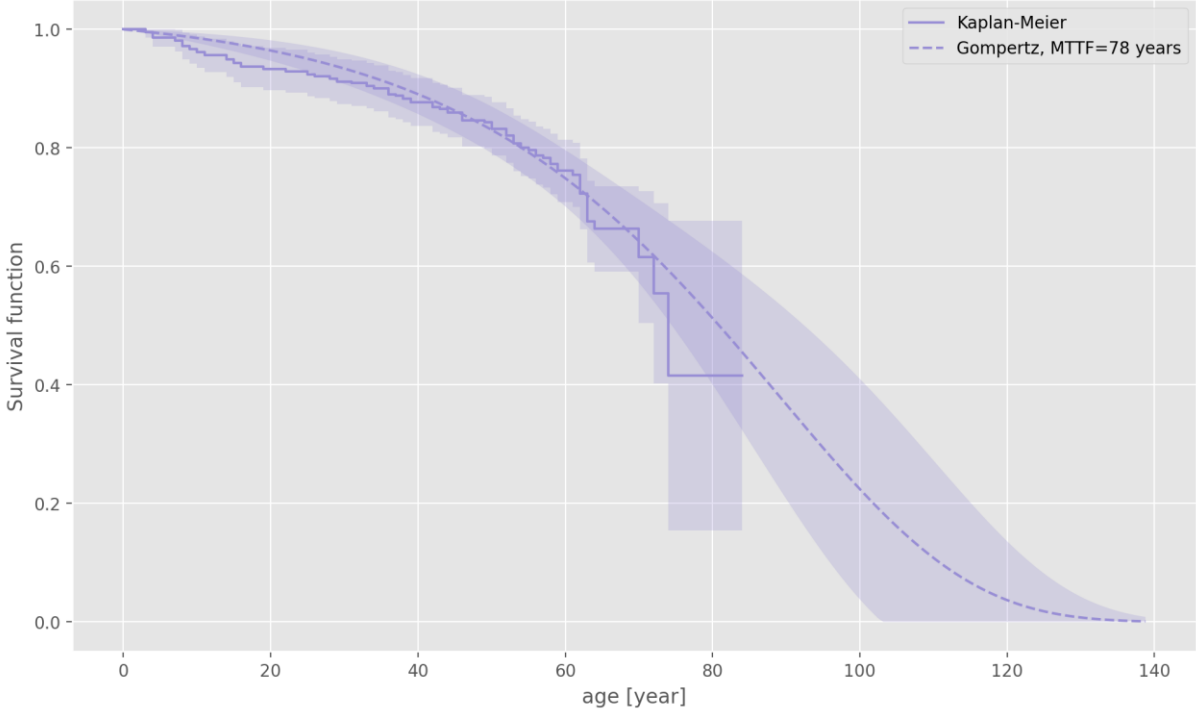
Example of Ambiguity in Asset Replacement Records

Power transformer data set with 1,436 units. For 49 of them, the reason for removal is unclear: was it a grid reconfiguration (right-censoring) or a failure (event)?

106 or 57 events before investigation



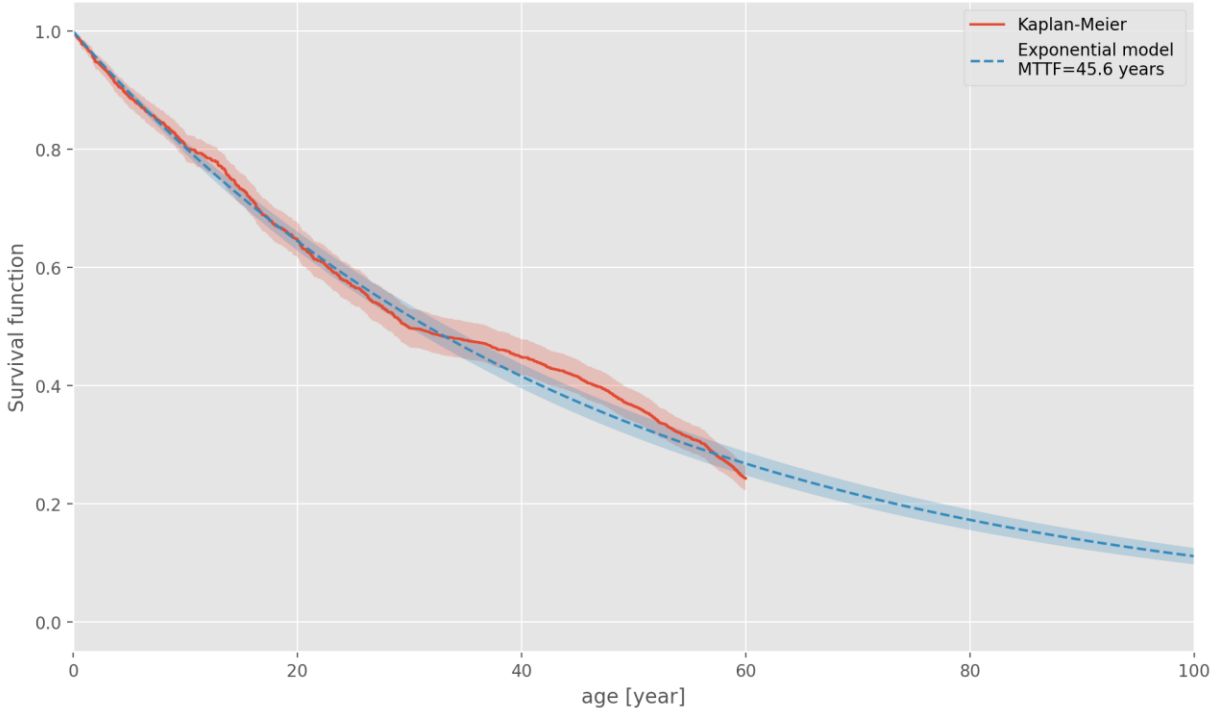
80 events after investigation



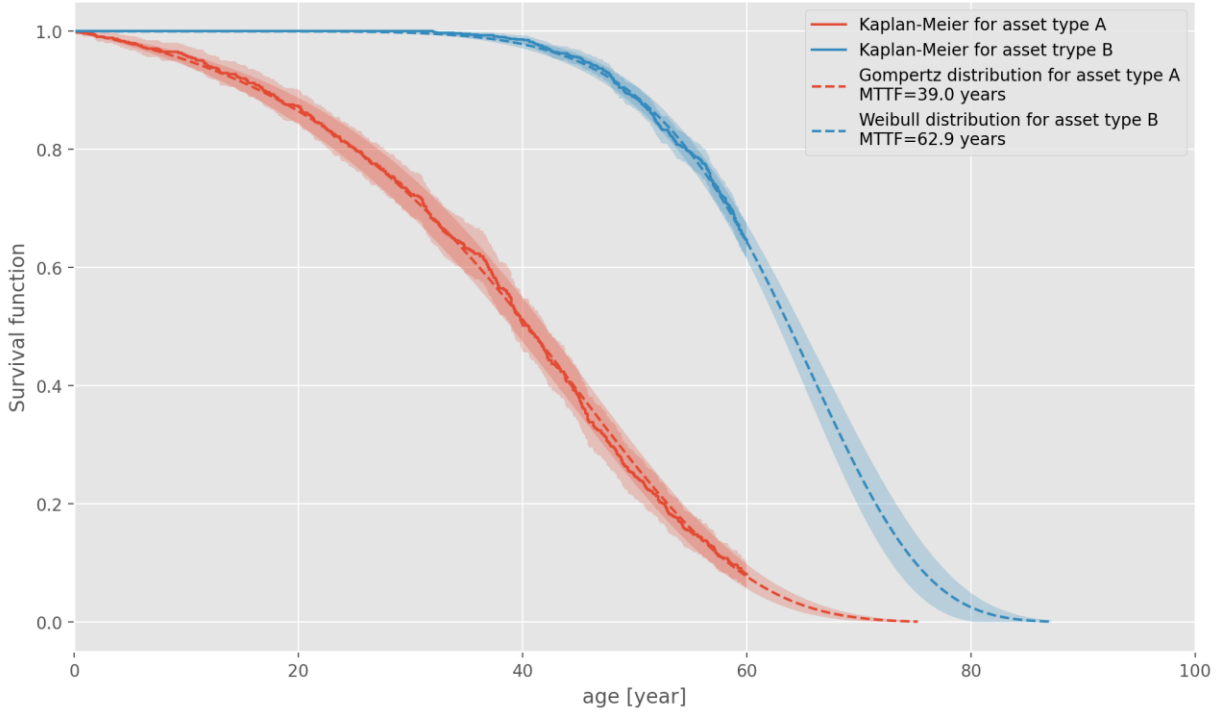
Typical Example Merging Asset and Event Data

Survival analysis using raw data, followed by extensive cleaning and sample separation.

Before cleaning



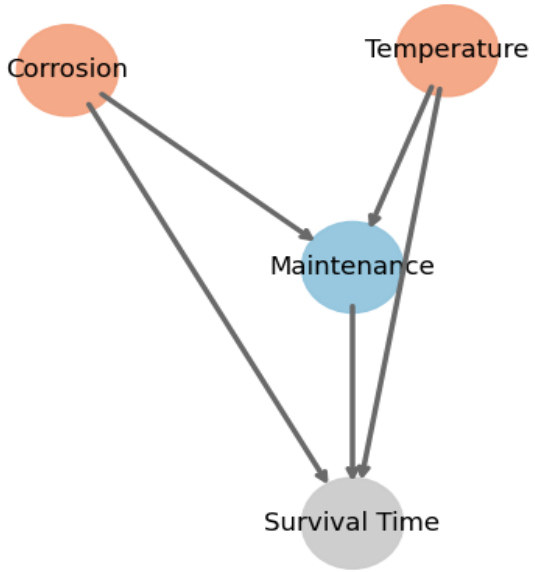
After cleaning



Figueroa, Guillon & Oueslati (2024). Industrial Data and Survival Analysis: How to Avoid Mistakes? *Congrès Lambda Mu 2024*, 14 au 17 octobre, Bourges.

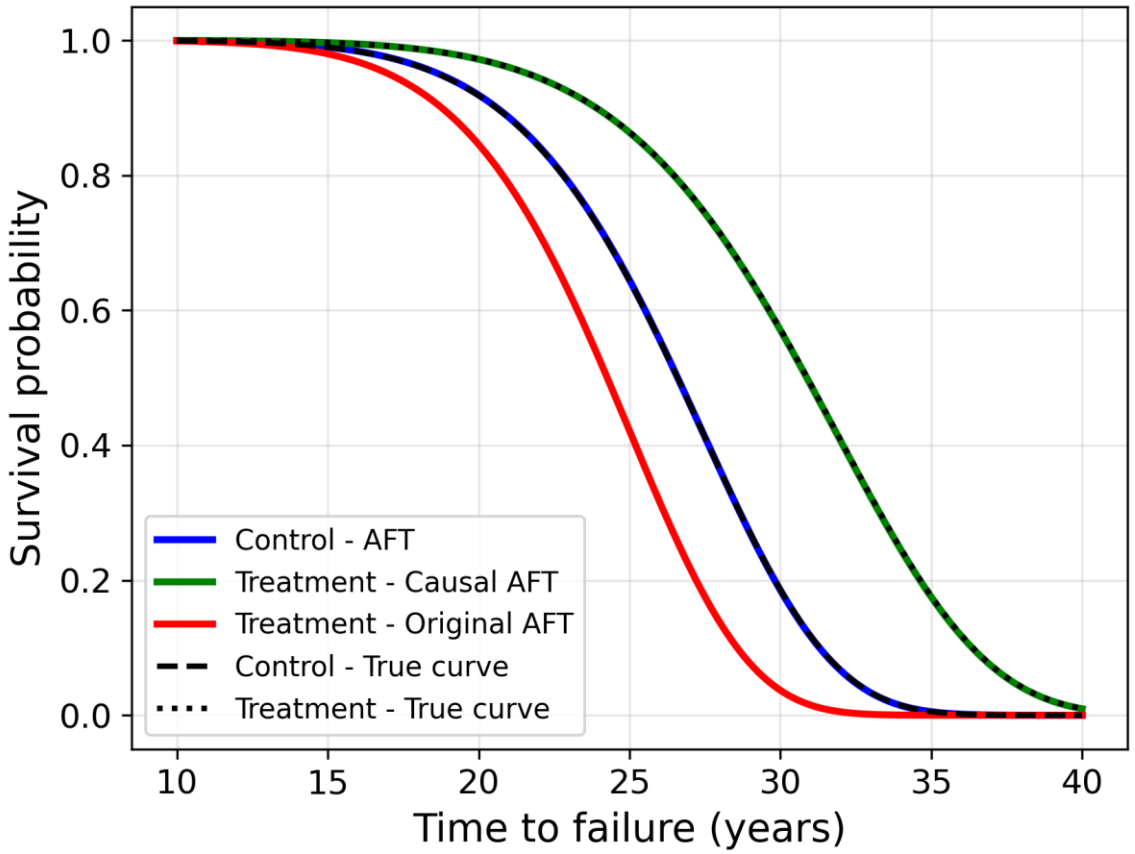
Causal Effect of Routine Maintenance

How does routine maintenance affect the lifetime distribution of assets — and can it be optimized?
There is a sampling bias, as routine maintenance tends to be more frequent when covariates are unfavorable.



Optimizing the age of replacement a, along with the routine maintenance effort ξ:

$$C(a,\xi) = \frac{c_f\, F(a;z,\xi) + c_p\, S(a;z,\xi)}{\int_0^a S(x;z,\xi)\,\mathrm{d}x} + C_m(\xi)$$



Work in progress, Younsi, Zeng, Guillon & Fremond (2025)

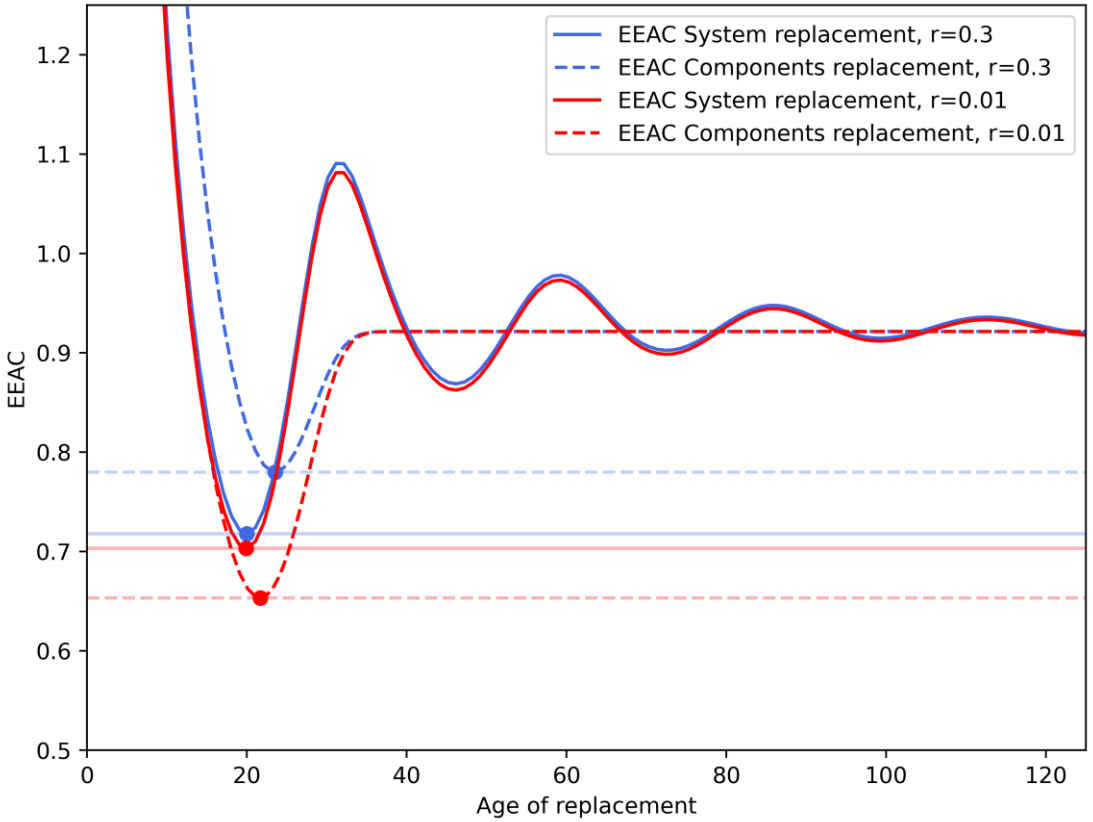
Preventive Replacement: Entire Asset or Component by Component?

Example: Three-phase disconnectors with six engines/arms.

Planned outage costs can be shared across multiple interventions: $c_p = c'_p + c_I$

Does imperfect repair represent well this context to decide of the replacement of the entire asset?

Work in progress, Herzhaft, Guillon & Fremond (2025)



3 Socio-economic evaluation

Accounting for Negative Externalities

To support decisions that enhance network performance, we distinguish between direct expenses and societal costs representing what we are willing to spend to avoid consequences:

- Value of a Statistical Life: VSL = 3M$€_{2010}$/fatal accident
- Shadow price of carbon: VCO_2 = 256$€_{2023}$/tCO2e in 2025
- Value of Lost Load: VoLL = 34k$€_{2022}$/MWh

Replacement cost upon failure could be:

$$c_f = c_f' + q_{SL}VSL + q_{CO2}VCO_2 + q_L VoLL$$

"Life has no price", "These values can't be fixed". Yet they are always implicit in the decision.
Example: if you replace at age *a*, what VoLL would make that decision optimal?

$$VoLL = \frac{1}{q_L}\left(\frac{c_p}{h(a)\int_0^a S(x)\mathrm{d}x\ - F(a)} - c_f' + c_p\right)$$

Viscusi, W. (2018). *Pricing lives: Guideposts for a safer society*. Princeton University Press.

Cost Functions and Discounting

Risk reduction options must be compared over long investment horizons, using **discounting** to reflect the time value of money.

Different **cost functions** can be considered depending on the time horizon, if there is only one cycle or multiple replacements, discounting or not in continuous or discrete time.

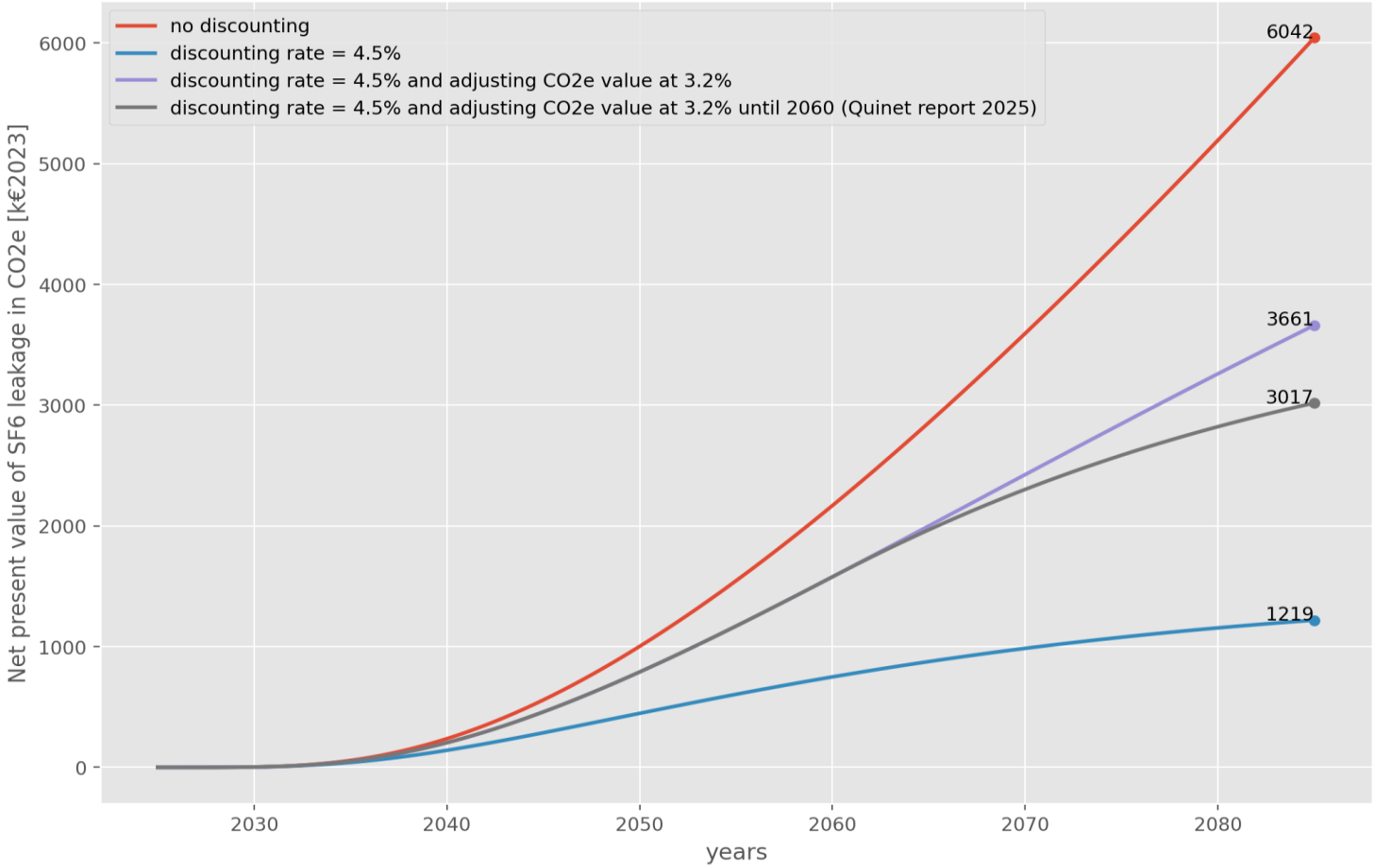
For an age replacement policy, infinite time horizon with discounting in continuous time and multiple replacements, the asymptotic expected equivalent annual is:

$$EEAC_{\infty}(a) = \frac{c_f \int_0^a e^{-\delta x} f(x)\,\mathrm{d}x + c_p e^{-\delta a} S(a)}{\delta \int_0^a e^{-\delta x} S(x)\,\mathrm{d}x} + C_m$$

But what does it mean for societal costs ?

How Discounting Affects Societal Costs and Decision-Making

Example: Gas Insulated Substation operated for 60 years (finite time horizon).
Non-Homogeneous Poisson process modelling the SF6 leakage with ABAO repairs.



Investement cost ~ 4,5 M€$_{2023}$

Typical Heuristics for Asset Management Decisions in Power System Network Organizations

An Almost Kaplan-Meier Estimator — That Doesn't Work

Operational teams lack standards on survival analysis and age-based replacement.
They rely on technical brochures compiling international practices containing many errors.

$$S(t) = \frac{N_s(t)}{N_F(t) + N_S(t)}$$

$$\text{Expected failures} = \sum_{i=1}^{n} \frac{F(t_i + k) - F(t_i)}{1 - F(t_i)}$$

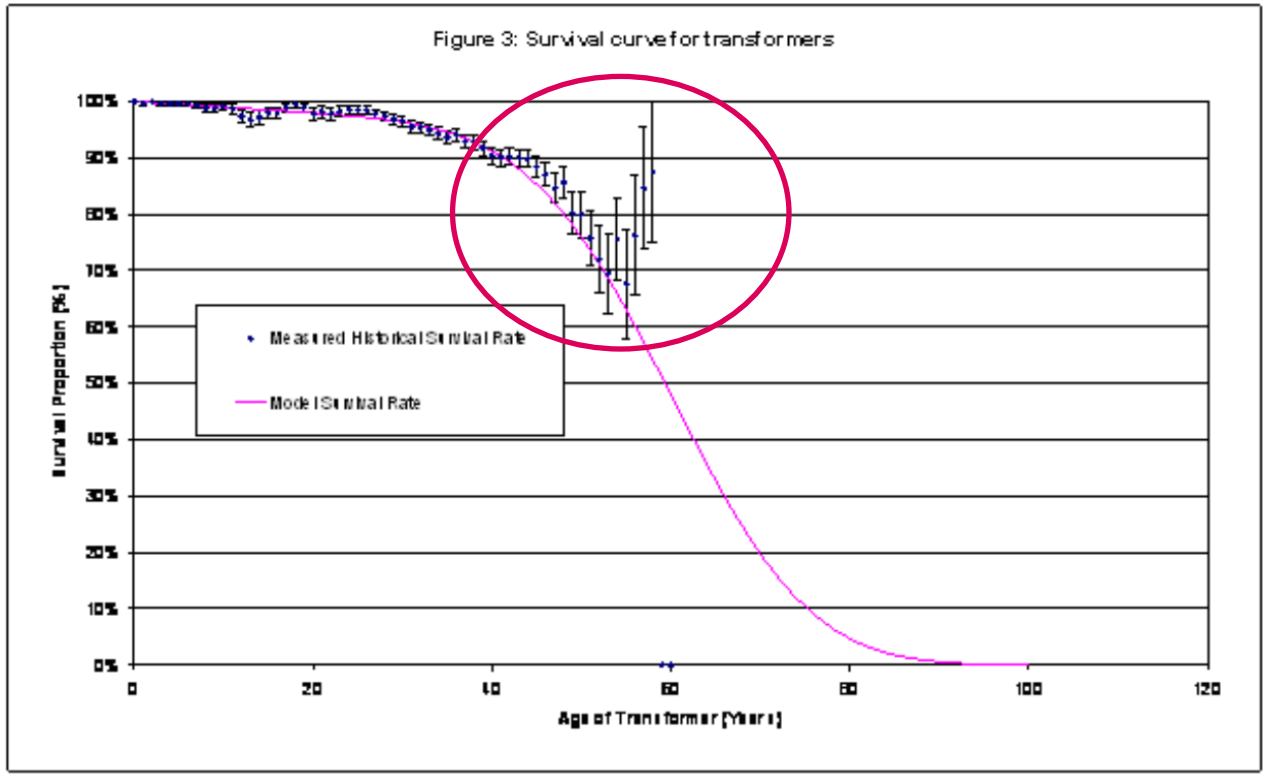


FIGURE 44 - SURVIVAL CURVE FOR TRANSFORMERS

CIGRE TB 706 Guidelines for the Use of Statistics and Statistical Tools on Life Data (2017), e.g., p 44 & p81-82

The Curse of Asset Management: HI, PoF & Risk Matrix

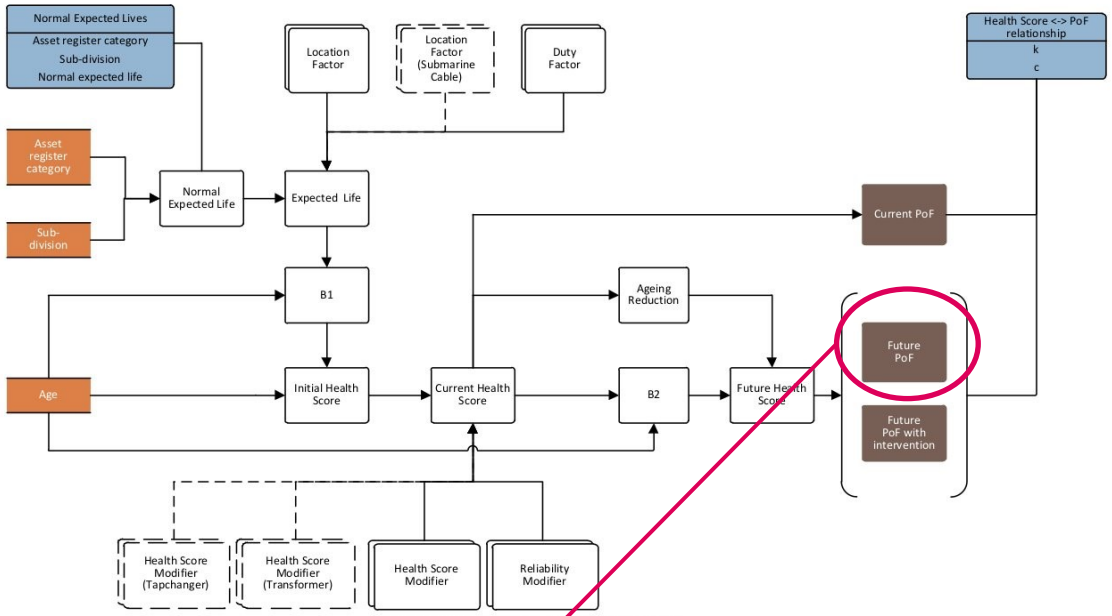
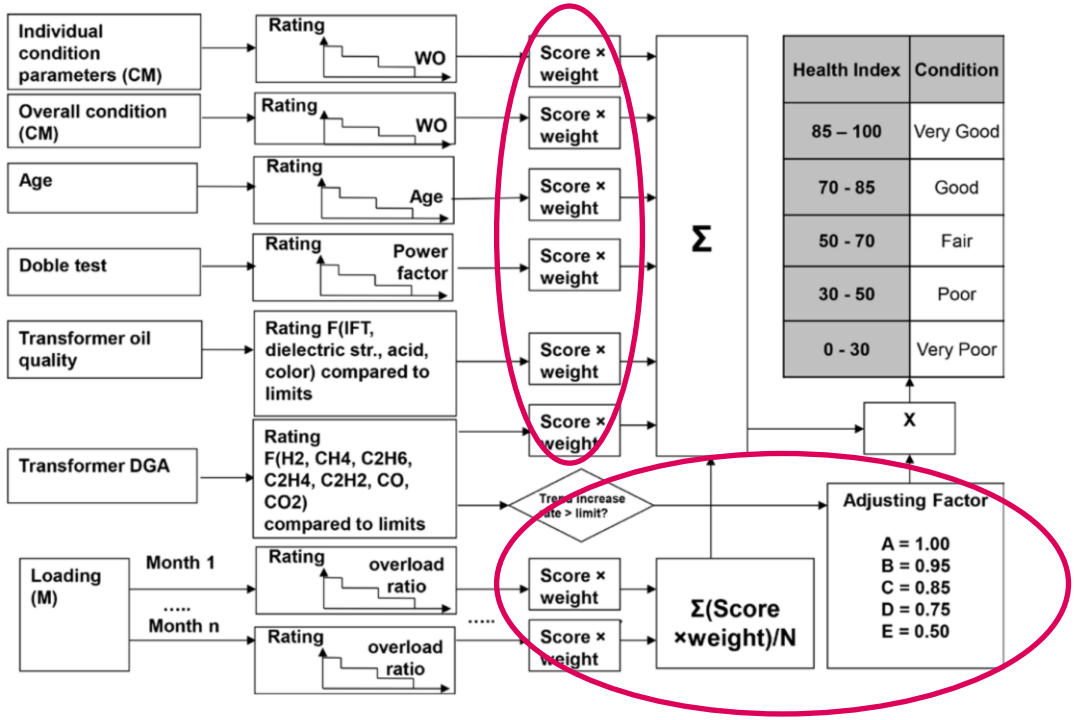


FIGURE 4: PROBABILITY OF FAILURE

$$\mathbf{PoF} = \mathbf{K} \times \left[\mathbf{1} + (\mathbf{C} \times \mathbf{H}) + \frac{(\mathbf{C} \times \mathbf{H})^2}{\mathbf{2!}} + \frac{(\mathbf{C} \times \mathbf{H})^3}{\mathbf{3!}} \right]$$

$$\textit{Present value of future risk}_{0-n} = \left[\sum_{i=0}^{n} (\mathbf{PoF}_i \times (\mathbf{1} + \mathbf{r})^{-i}) \right] \times \mathbf{CoF}$$

DNO Common Network Asset Indices Methodology Version 2.1. (2021)



IEC Strategic asset management of power networks, White Paper (2015)

What is PoF (Probability of Failure)?

PoF appears to be an annual probability. What does it refer to?

1. $PoF_1(a) = P(T \leq a+1 | T > a) = 1 - \frac{S(a+1)}{S(a)}$
2. $PoF_2(a) = P(a < T \leq a+1) = F(a+1) - F(a)$
3. $PoF_3(a) = \lim_{\Delta a \to 0} \frac{P(a < T \leq a + \Delta a | T > t)}{\Delta a} = h(a)$
4. $PoF_4(a) = P(T \leq a) = F(a)$

PoF is estimated as if it was PoF1, mistakenly interpreted as PoF3 (not a probability), used as if it was PoF2 to calculate expected NP

How to Bring Reliability Theory into Asset management Practices

- Reframe decision problems, avoiding prioritization
- Develop standards in reliability (survival analysis, renewal theory, maintenance policies)
- Build skills both internally and externally (including consulting firms)
- Define an organizational structure suited to implement reliability practices
- Develop software tools and libraries to support end users
  https://github.com/rte-france/relife

Key Takeaways

🔧 **Reliability Models: A Strategic Tool**

- Essential for planning infrastructure renewal
- Anticipate failures, manage spares, secure supply chains

🧭 **Models Guide Decisions — Know How and Where**

- Reliability models support decisions, not just analysis
- Uncover the optimization problem behind the model request
- Ask how it may impact existing decision processes

📊 **Data is Always Possible**

- Cross-reference databases
- Sample, measure, install sensors
- Collaborate with peers and suppliers

🎓 **Bridging the Knowledge Gap**

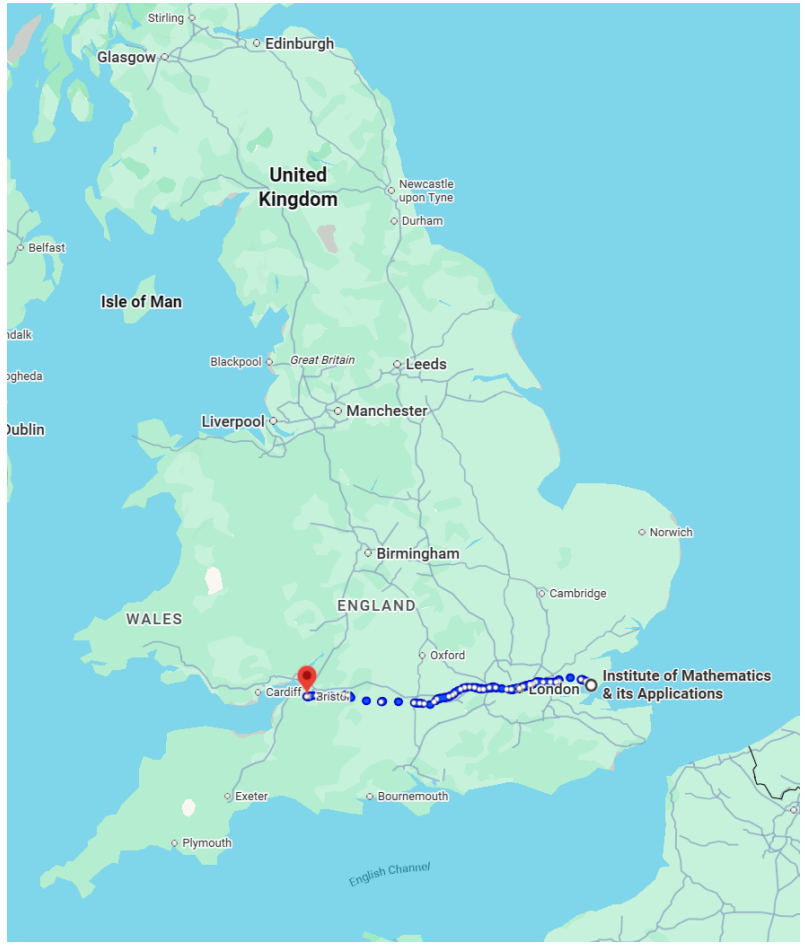
- Reliability theory is still underused by operators and regulators
- Let’s change that — standards like IEC 63223-2 pave the way for more

A Path Linking Reliability Theory and Asset Management

Asset management has evolved without drawing on the foundations of reliability theory.

The Institute of Asset Management

https://theiam.org/



INSTITUTE OF MATHEMATICS & ITS APPLICATIONS

https://ima.org.uk/

Thank you for your work!

thomas.guillon@rte-france.com

Stochastic Approaches for Optimizing Maintenance of Continuously Degrading Systems in the Era of AI: Are they still Relevant?

Antoine Grall
Université de Technologie de Troyes, France
antoine.grall@utt.fr

MIMAR 2025 - Nancy, France

Outline

- Context
- Framework for maintenance optimization
  - Problem and main options
  - Stochastic Markov-renewal approach
- Examples of frugal models
  - Unknown degradation parameters
  - Imperfect maintenance
  - Latent degradation process
- Concluding remarks

General context

General context

"Optimizing" Maintenance decision-making

- Integration of monitoring and system reconfiguration actions
- Economic stand point (direct maintenance costs)
- Environmental criteria
- Reliability / availability

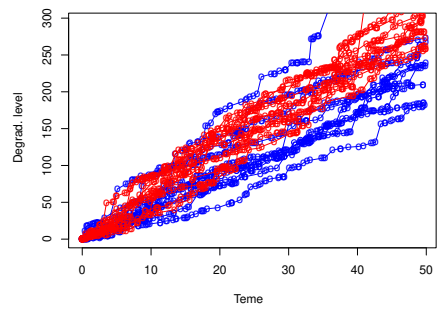
Possible strategies

- Condition-based maintenance
- Predictive maintenance
- Proactive maintenance
- Prescriptive maintenance

... or others!

Continuous Deterioration leading to failure

- Continuous state space
- Time continuous
- Monitoring: Inspections / specific sequences
- Hard or soft Failure = limit degradation level



General context

Many possible approaches for modeling and solving

- Between Knowledge-based and data-driven models: a continuum

- **Decision-making**
  - based on specific parametric rules or directly derived from optimization problem;
  - explicitly derived from lifetime prognosis (post-prognosis approach) or implicitly obtained from degradation behavior;
  - Long term cost rate or discounted cost, expectation on finite time horizon.
- **Mathematical framework**
  - Markov-Decision Processes and dynamic programming;
  - Renewal and Markov renewal theory;
  - Stochastic simulation methods (Monte Carlo);
  - ...

General context

Many possible approaches for modeling and solving

- Between Knowledge-based and data-driven models: a continuum

- **Decision-making**
  - based on specific parametric rules or directly derived from optimization problem;
  - explicitly derived from lifetime prognosis (post-prognosis approach) or implicitly obtained from degradation behavior;
  - Long term cost rate or discounted cost, expectation on finite time horizon.

- **Mathematical framework**
  - Markov-Decision Processes and dynamic programming;
  - Renewal and Markov renewal theory;
  - Stochastic simulation methods (Monte Carlo);
  - ...

General context

Common discourse in the context of digital transformation:
that can be heard about maintenance x.0

- Development of technologies and methodologies for monitoring
  - New autonomous, communicating sensors
  - Increasingly powerful computing resources (advanced data processing methods are becoming affordable, natural language interaction with AIs is available)

- Masses of data are available (in real time) on the behavior of the object of study and its environment
  - Monitoring and analysis of operating mode
  - Monitoring (and modeling) of degradation phenomena
  - Many powerful AI-based approaches e.g. for PHM (especially for clustering and classification)

General context

- **What I can experience** (potentially biased point of view)
  - In general, (very) complex and poorly formulated problems from “real world”
  - Search for solutions with decisions that are easy to understand and simple to implement in practice (“Why replace a component that’s not totally degraded? ” may still be an issue)
  - Still frequently very small usable data sets (or I haven’t met the right people nor the right companies)

- Use of mathematical modeling as a bridge between “real” problems and practical solutions
- The importance of not neglecting “classical” but frugal models

General context

- **System**: Passive component (structure) or dynamic (controlled) system subject to a degradation mechanism
  (Cracking (fatigue, stress corrosion), railways track geometry, degradation of large gap power electronics components (MOSFET-SiC, threshold voltage evolution), evolution of mechanical properties of metal pressure vessel, turbofan engine, wind turbine blade angle control system, suspension systems in heavy-duty trucks, cooling system of an optical sensor, ...)
- **Monitoring information**: Successive inspections or diagnostic sequences potentially costly, not necessarily performed periodically, few or no breakdown or failure data - **small data**
- **Deterioration evolution**: direct measurement of degradation level or synthesis of an indicator

General context

Optimizing Maintenance of Continuously Degrading Systems: General problem formulation

$$\min_{A}\left(\min_{\Theta} \; F \; (A,\Theta)\right)$$

General context

Optimizing Maintenance of Continuously Degrading Systems: General problem formulation

includes degradation model, assessment criterion, decision structure

$$\min_{A}\left(\min_{\Theta} F(A,\Theta)\right)$$

decision variables or actions

model parameters (estimated from data)

☞ **Correlation between the number of parameters and the amount of information** (biais, sensitivity, underfitting and overfitting ...)

General context

Optimizing Maintenance of Continuously Degrading Systems: General problem formulation

☒ Focus on cases with small data sets (few measurements from few inspections)

⇒ Models with a limited number of parameters

☞ Gamma, Inverse Gaussian, Wiener stochastic processes and extensions

$$\Delta X \sim f_\Theta\left(\Delta t\right) \quad \text{and } f_\Theta(\Delta t_1) + f_\Theta(\Delta t_2) = f_\Theta(\Delta t_1 + \Delta t_2)$$

Based on increments; Degradation trajectories = samples paths of the process

☒ Willingness to preserve the continuous nature of the degradation

☒ Simple configurations highlighting specific questions from real life

Framework for maintenance optimization

Maintenance problem

✗ Continuously deteriorating system, non continuously monitored, with actions[1] based on its condition, enabling “control” on its life cycle

✗ Question: what and when?

☞ How to respond the question in an “optimal” way.

☞ Economic point of view, dependability issues (unavailability, probability of failure ...), environmental impact (carbon footprint ...)

-> maintenance cost

☞ Stochastic aspect lead to consider expected (mean) costs

[1] inspection or assessment process, usage modification, repair or replacement

Maintenance problem

- **Focus on one possible approach** (among others) given by
  - Framework of Renewal theory
    - ✗ Cumulated cost per time unit on infinite time horizon
    - ✗ Continuous time and continuous state space
    - ✗ Requires the choice of a parametric maintenance decision rule
    - ✗ Few decision variables - easy to explain, easy to deal with non periodic inspections without combinatorial explosion

Maintenance problem

- **Focus on one possible approach** (among others) given by
  - Framework of Renewal theory
    - ✗ Cumulated cost per time unit on infinite time horizon
    - ✗ Continuous time and continuous state space
    - ✗ Requires the choice of a parametric maintenance decision rule
    - ✗ Few decision variables - easy to explain, easy to deal with non periodic inspections without combinatorial explosion
- **Other (and popular) approach**
  - Framework of Markov Decision Processes
    - ✗ Expected cost on a finite time horizon
    - ✗ Discounted cost on infinite time horizon
    - ✗ Discrete time problem (and finite state space)
    - ✗ Optimal policy
    - ✗ ... gateway to Reinforcement Learning

Semi-Regenerative Approach

Cost evaluation and optimization:

- Ex. of decision rule: inter-inspections time $\Delta T = m(x)$ and preventive maintenance threshold M
- Long run expected cost per time unit

$$C_\infty(A) = \lim_{t\to\infty} \frac{\mathbb{E}\left[C(t \mid A)\right]}{t}$$

where A= Decision variables and $C(t \mid A)$ = Cumulated cost up to time t

- **Step 1: Evaluation of $C_\infty(A)$ using renewal theory**

$$\text{classically } C_\infty(A) = \frac{\mathbb{E}_0\left[C(S \mid A)\right]}{\mathbb{E}_0[S]} \text{ with } S \text{ renewal cycle}$$

- **Step 2: Optimization (classical gradient-based algorithm)**

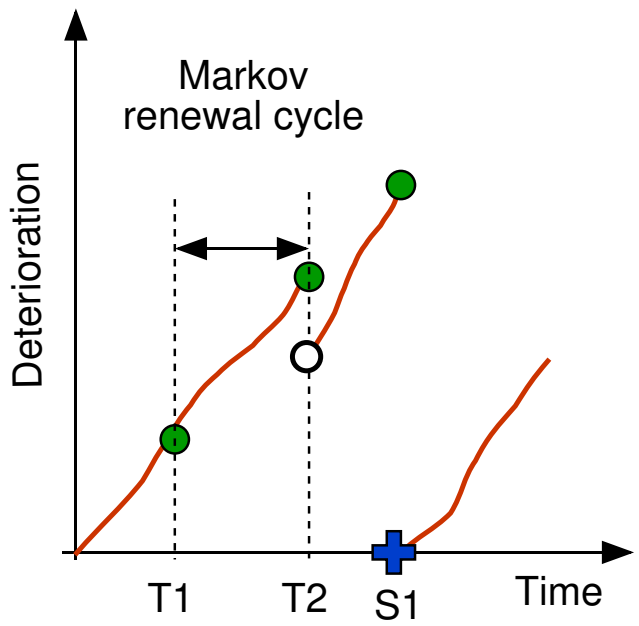
Semi-Regenerative Approach

Markov renewal process

- Perfect inspection + Markov property : the knowledge of the condition allows to summarize all the past history ⇒ **Markov renewal process** (inspection times)
- Result for Markov renewal process (embedded in a renewal process)

$$C_\infty(A) = \frac{\mathbb{E}_{\pi}\left[C(T \mid A)\right]}{\mathbb{E}_{\pi}\left[T\right]}$$

- Stationary law of the degradation level at inspection times
- Time between two successive inspections (length of a semi-regenerative cycle)



- Shorter cycle ⇒ Simplified analysis, but requires the evaluation of π

Semi-Regenerative Approach

Stationary law π of the degradation level at inspection times?

1. Look at the system state at (after) maintenance time: $Y_n = X_{T_n}$.
   $(Y_n)_{n\in\mathbb{N}}$ is a Markov chain whose transition probability function can be obtained by listing the scenarios after an inspection, given the current state x

$$\text{Example: } \mathbf{P}(dy|x) = \overline{F}_{\alpha\cdot m(x),\beta}(M-x)\,\delta_0(dy) + f_{\alpha\cdot m(x),\beta}(y-x)\,\mathbb{1}_{\{x\le y<M\}}\,dy$$

"new" after maintenance — degradation without repair

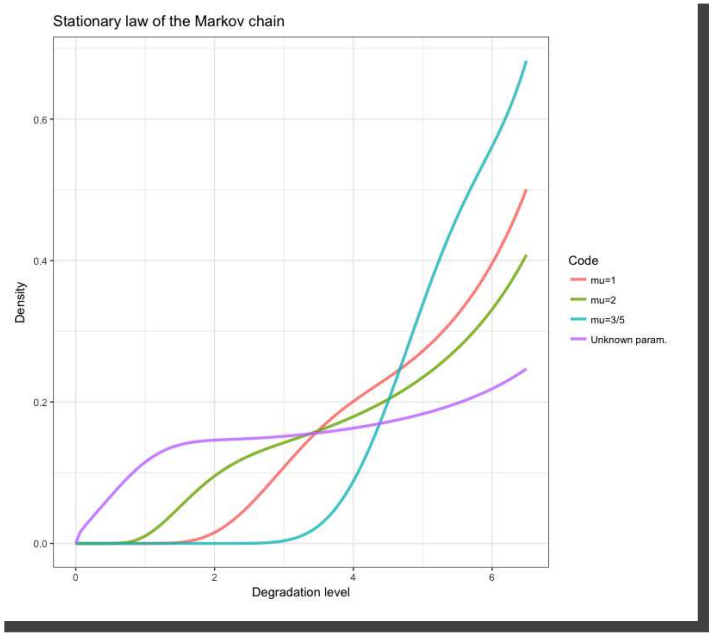
2. Probability distribution π:

- solution of the stationary equation

$$\pi(x) = \int_{[0;M)} \mathbf{P}(.|x)\pi(dy)$$

- with a general form identical to $\mathbf{P}(dy|x)$ i.e.

$$\pi(dy) = a\delta_0(dy) + (1-a)b(y)dy$$



Unknown degradation parameters [a]

[a] From E. Mosayebi Omshi, A. Grall, and S. Shemehsavar. A dynamic auto-adaptive predictive maintenance policy for degradation with unknown parameters. European Journal of Operational Research, 282(1):81–92, 2020.
DOI:10.1016/j.ejor.2019.08.050

Problem

✗ Unknown degradation parameters:
Can inspections information be useful in case of a maintenance policy with non-periodic inspections i.e. with 1 to 3 inspections per cycle? How to proceed? How to dynamically improve decision-making?

- "Failure" when degradation level X_t crosses the critical level L
- **Process parameters are not known -> Bayesian approach**
- **Example**: degradation process = IG process with parameters $\delta = 1/\mu$ and λ
  - independent, non-negative increments
  - for each $t > s \geq 0$: $Y = X_t - X_s \sim IG(\mu(t-s), \lambda(t-s)^2)$ with $\mu, \lambda > 0$

$$\text{Increment pdf: } f(y) = \sqrt{\frac{\lambda(t-s)^2}{2\pi y^3}} \exp\left\{ \frac{-\lambda\big(y - \mu(t-s)\big)^2}{2\mu^2 y} \right\}$$

  - mean and variance of X_t are respectively μt and $\mu^3 t/\lambda$.

Bayesian framework

Unknown degradation parameters - choice of conjugate priors

- Parameter λ: gamma distribution with parameters α and β

$$f(\lambda) = \frac{\lambda^{\alpha-1}}{\Gamma(\alpha)\beta^{\alpha}} \exp\{-\lambda/\beta\},$$

- Parameter δ: conditional normal distribution with mean ξ and variance σ^2/λ

$$f(\delta|\lambda) = \sqrt{\frac{\lambda}{2\pi\sigma^2}} \exp\left\{-\frac{\lambda(\delta-\xi)^2}{2\sigma^2}\right\}.$$

Additional assumption: $P(\delta \leq 0)$ is negligible

☞ Joint distribution of (δ, λ): $f(\delta, \lambda) = f(\delta|\lambda) f(\lambda)$ with 4 hyper-parameters α, β, ξ and σ

Unknown degradation parameters - choice of conjugate priors

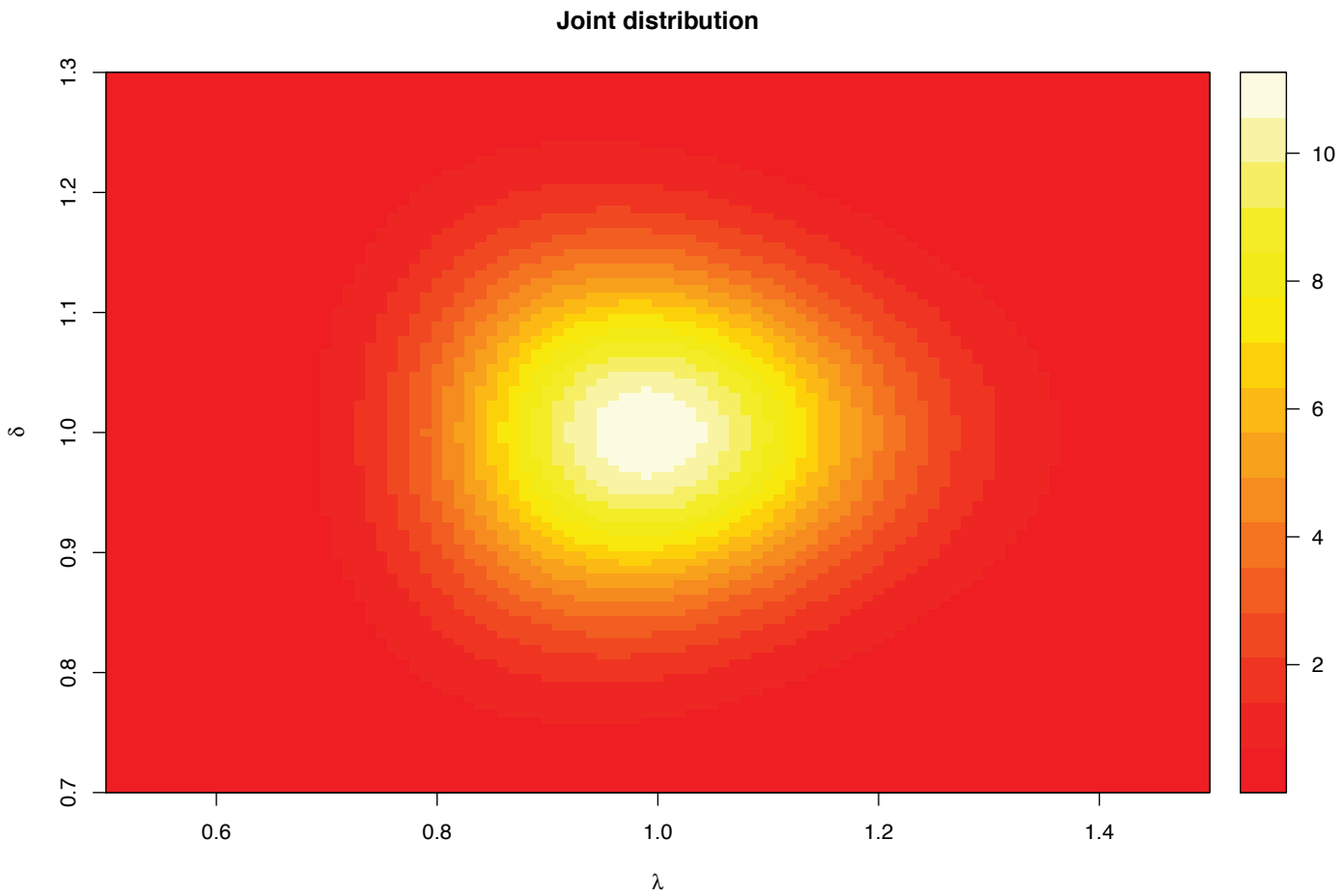


Fig.1: PDF of $(\boldsymbol{\delta}, \boldsymbol{\lambda})$ for $\boldsymbol{\alpha} = \mathbf{50}$, $\boldsymbol{\beta} = \mathbf{0.02}$, $\boldsymbol{\xi} = \mathbf{1}$, and $\boldsymbol{\sigma} = \mathbf{0.1}$

Maintenance decision rule

☒ **Preventive maintenance based on degradation level**:
at inspection time T_n

- If $X_{T_n^-} \geq M$: repair (corrective or preventive);
- If $X_{T_n^-} < M$: system left as it is.

☒ **Non-periodic inspection/maintenance planning based on RUL**

$$T_{n+1} = T_n + \tau_p(x_{T_n}, T_n),$$

where $\tau_p(x,t) = \left\{\Delta t : \Pr(X_{t+\Delta t} \geq L | X_t = x) = p\right\}$ p-quantile of RUL distribution.

☞ **Decision variables: p and M**

Stationary law

- $\{X_t\}_{t\geq 0}$ semi-regenerative process with semi-regeneration times T_k (inspec. times)
- $\{Y_n = X_{T_n}\}_{n\in\mathbb{N}}$ Markov chain (taking values in $[0, M)$) with transition probability law (conditioned on the current state)

$$\Pr(\mathrm{d}y|x) = \int_0^{\infty}\int_{-\infty}^{+\infty} \Pr(\mathrm{d}y|x,\delta,\lambda) f(\delta|\lambda) f(\lambda)\mathrm{d}\delta\mathrm{d}\lambda$$

Transition law with known parameters

$$\Pr(\mathrm{d}y|x) = \delta_0(\mathrm{d}y).\int_{M-x}^{\infty} g_{\tau_p(x)}(u)du + g_{\tau_p(x)}(y-x)I_{\{x\leq y<M\}}\mathrm{d}y$$

where

$$g_{\tau_p(x)}(z) = \sqrt{\frac{\beta}{2\pi}}\frac{\Gamma(\alpha+1/2)\tau_p(x)}{\Gamma(\alpha)} z^{-3/2}(\sigma^2 z+1)^{-1/2}\left(1+\frac{\beta(\xi z-\tau_p(x))^2}{2z(\sigma^2 z+1)}\right)^{-(\alpha+1/2)}$$

Adaptive maintenance policy

Integration of information monitoring after each replacement (heuristics)

☞ After integration of new information (bayesian update)

- Update of degradation model parameters through hyper-parameters

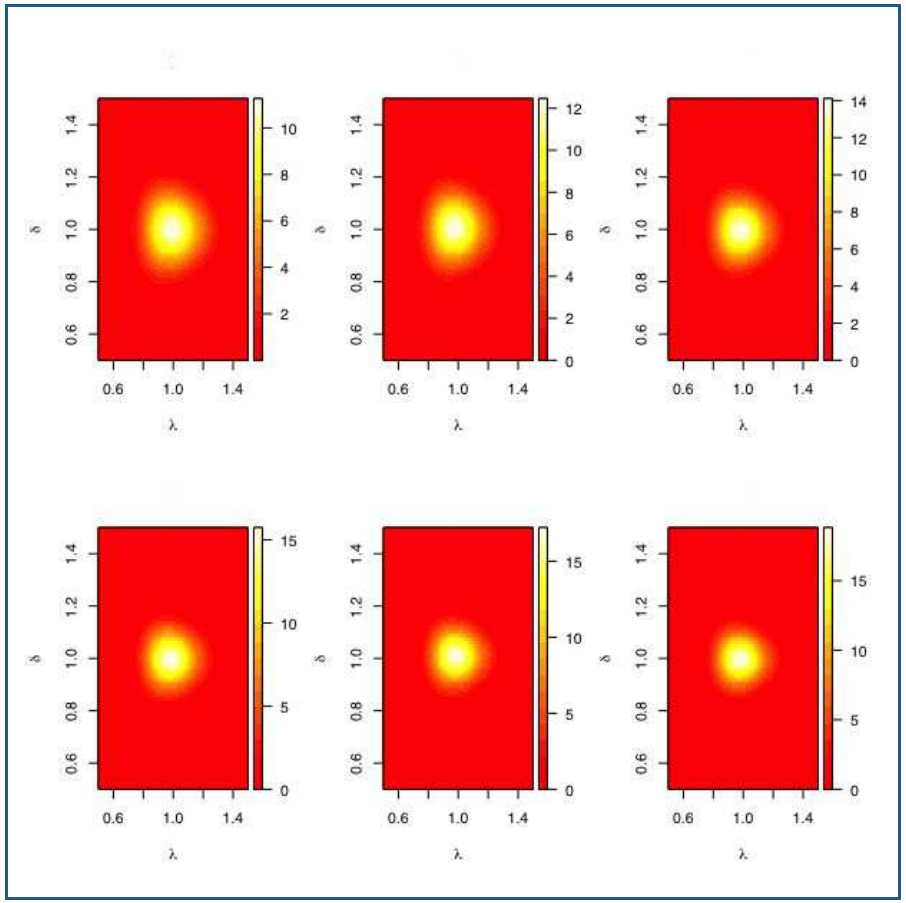
  ⇒ Update of RUL distribution
  ⇒ Impact on the p-quantile of RUL distribution
  ⇒ modification of the time between inspections

- New stationary law and associated cost function

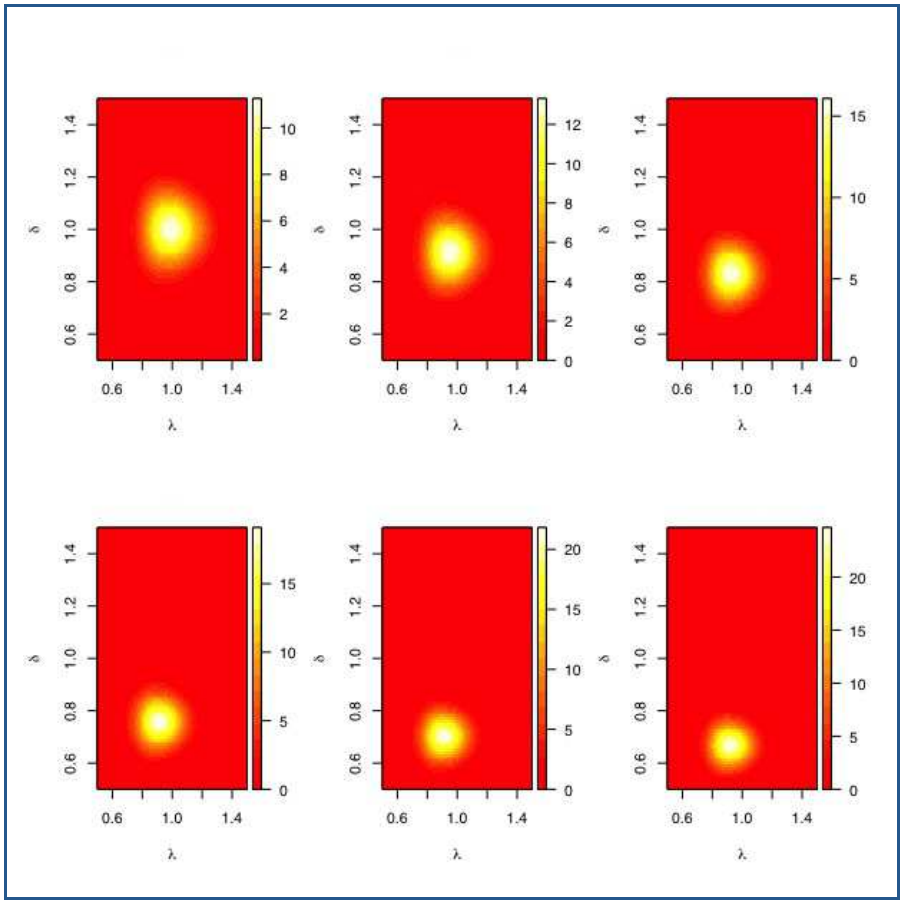
- New optimization and new decision variables $\boldsymbol{p}$ and $\boldsymbol{M}$

Numerical illustration

- Evolution of joint pdf of (δ, λ)



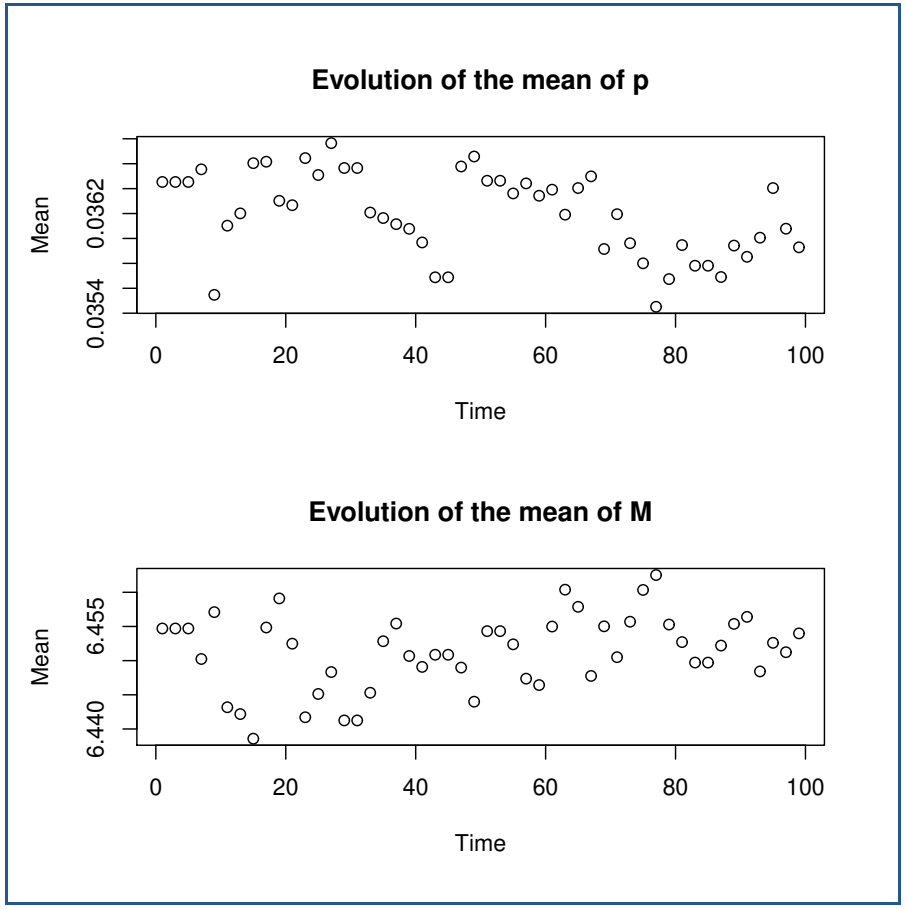
Same degradation for each cycle



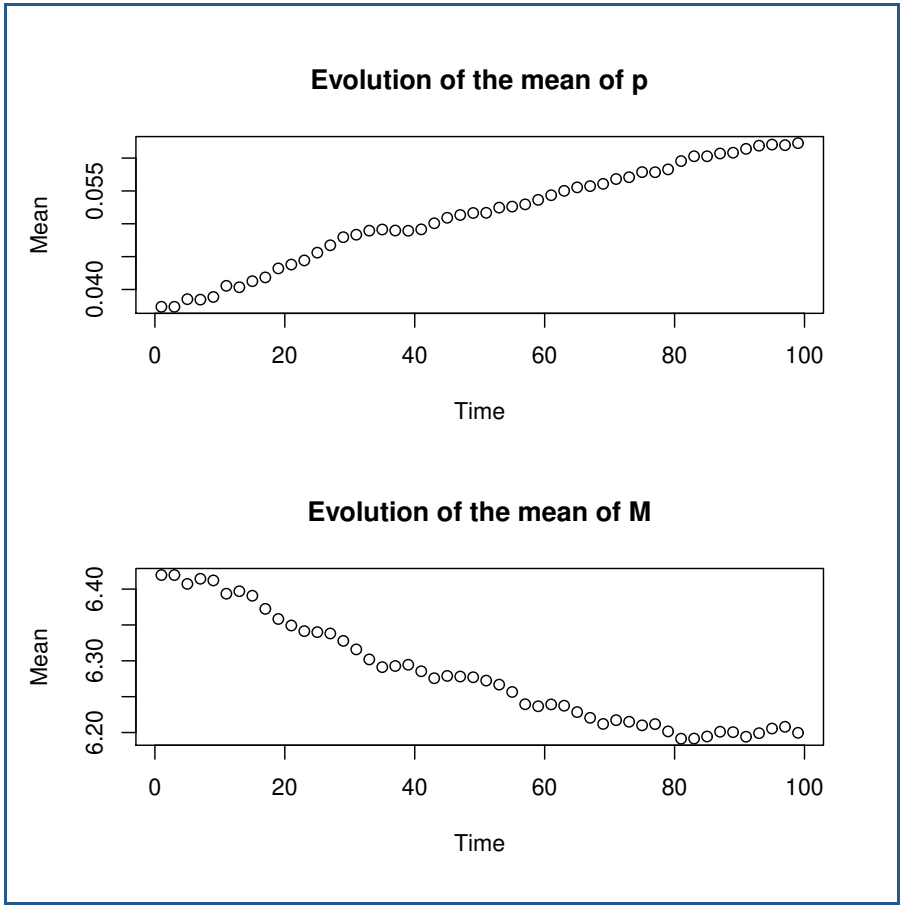
Degradation acceleration

Numerical illustration

✗ Decision variables evolution



Same degradation for each cycle



Degradation acceleration

Example of imperfect maintenance [a]

[a] From E. Mosayebi Omshi and A. Grall. Replacement and imperfect repair of deteriorating system: Study of a CBM policy and impact of repair efficiency. *Reliability Engineering & System Safety*, 215, 2021.
DOI:10.1016/j.ress.2021.107905

Problem

- Choice between 2 maintenance actions:
  - **Replacement** = perfect maintenance: $X_t = 0$ after maintenance
  - **Repair** = imperfect maintenance: $X_t \geq 0$ after maintenance
    - ☞ Effect of imperfect repair = **improvement factor model**

$$X_{T_n} = \delta . X_{T_n^-}$$

degradation level after and before repair

- ☞ δ = random variable (random effect of maintenance), in $(0, 1)$

- **Scarce on-line information on system state with non periodic inspections based on condition-based inspection planning**
- **Is it interesting to have information in advance about the performance characteristics of the maintenance team that will be coming in?**

Maintenance policy

When? And What?

- **Inspection times**: RUL-based inspection planning
  - System RUL with current degradation level equal to x (first passage time of the limit threshold)

$$\Upsilon_x = \inf\{r > 0 : X_{t+r} \geq L | X_t = x\}$$

  - Inspection scheduling

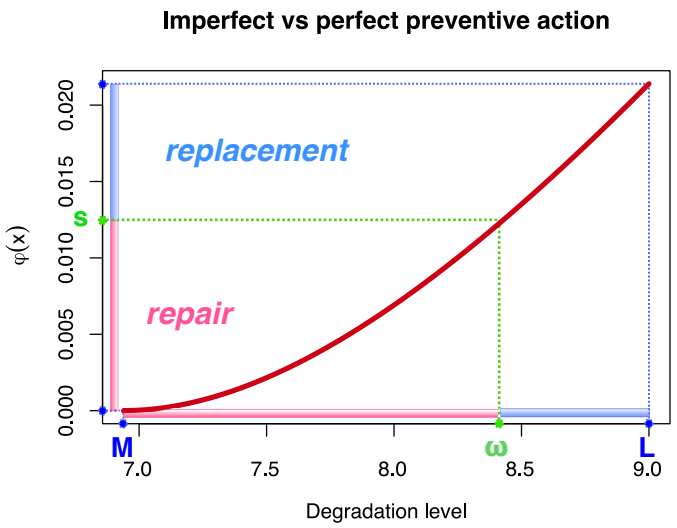
$$T_{n+1} = T_n + \tau_p(X_{T_n})$$

with $\tau_p(X_{T_n}) = \{\triangle t : G_\Upsilon(\triangle t | X_{T_n}) = p\}$ (p-quantile of RUL distribution).

Maintenance policy

☒ Maintenance decision rule

- If $X_{T_n^-} \geq L$: corrective perfect replacement ($X_{T_n} = 0$)
- If $M \leq X_{T_n^-} < L$: preventive maintenance - choice between replacement and repair based on $\varphi(x) = \mathbb{P}\left(\delta.x \geq M\right)$
  - if $\varphi(X_{T_n^-}) \geq s$ (poor repair effectiveness) then the system is replaced and $X_{T_n} = 0$;
  - if $\varphi(X_{T_n^-}) < s$ then a repair is chosen and $X_{T_n} = \delta.X_{T_n^-}$. If after imperfect maintenance $X_{T_n} \geq M$ then a perfect action is performed immediately such that $X_{T_n} = 0$.



☒ Decision variables : p for inspection, M for preventive action and s (or ω) for balance between replacement and repair

Stationary law

- **Stationary law** of degradation level at inspection times

$$\pi(y) = \int_0^M \mathrm{f}(y|x)\pi(\mathrm{d}x) \text{ with } \pi(\mathrm{d}x) = a\delta_0(\mathrm{d}x) + (1-a)b(x)\mathrm{d}x$$

$f(y|x)$ = transition probability distribution from x to y.

with

$$\mathrm{f}(y|x) = A_1(x)\delta_0 + \left(A_2(x,y)\mathbb{1}_{\{y<M\}} + f_{\tau_p(x)}(y-x)\mathbb{1}_{\{x\leq y<M\}}\right)$$

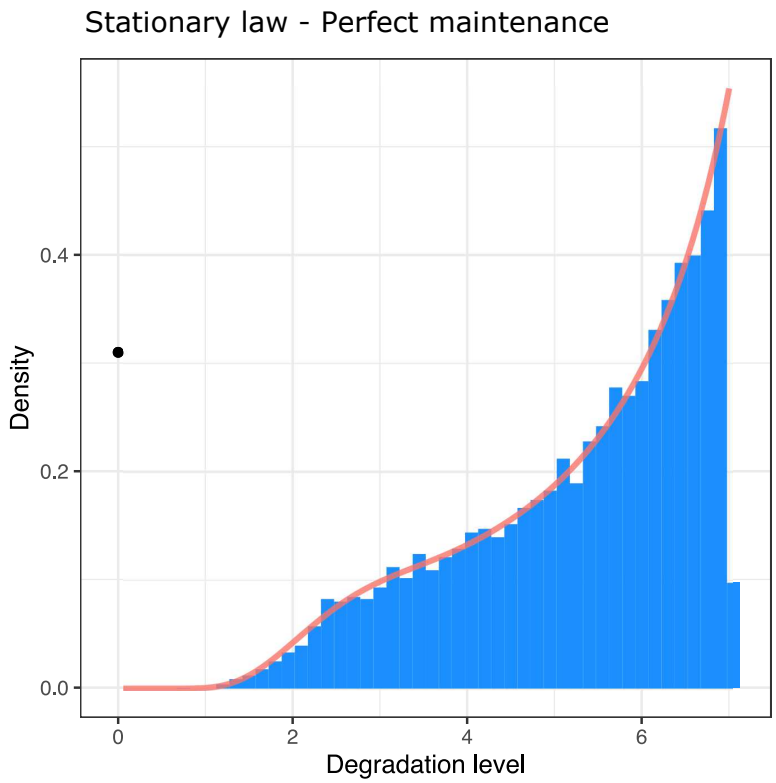
and the integral equation can be solved numerically with an iterative method based on a trapezoidal scheme

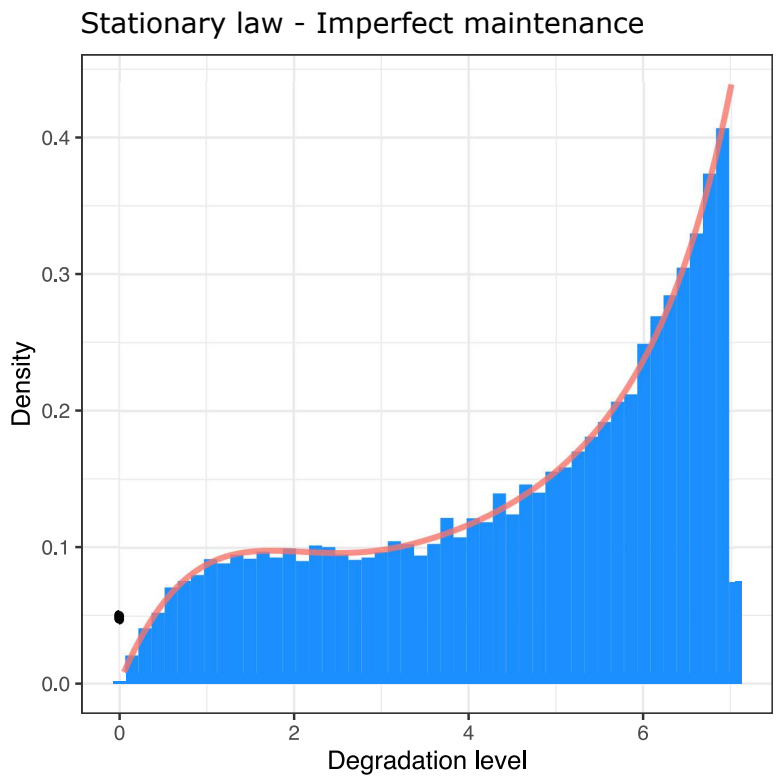
Stationary law

- **Stationary law** of degradation level at inspection times

$$\pi(y) = \int_0^M \mathrm{f}(y|x)\pi(\mathrm{d}x) \text{ with } \pi(\mathrm{d}x) = a\delta_0(\mathrm{d}x) + (1-a)b(x)\mathrm{d}x$$

$f(y|x)$ = transition probability distribution from x to y.





Example of numerical results

- **Degradation process = inverse Gaussian process**

- **Imperfect repair effect:** Beta distribution with shape parameters $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ for improvement factor

$$\mathbb{E}(\boldsymbol{\delta}) = \frac{\boldsymbol{\alpha}}{\boldsymbol{\alpha}+\boldsymbol{\beta}} \text{ and } \mathrm{Var}(\boldsymbol{\delta}) = \frac{\boldsymbol{\alpha}\boldsymbol{\beta}}{(\boldsymbol{\alpha}+\boldsymbol{\beta})^2(\boldsymbol{\alpha}+\boldsymbol{\beta}+1)}$$

- **2 different maintenance teams**

  - **Team 1: efficient maintenance team** - $\boldsymbol{\alpha} = \mathbf{0.5}$ and $\boldsymbol{\beta} = \mathbf{5}$ (right-skewed distribution).

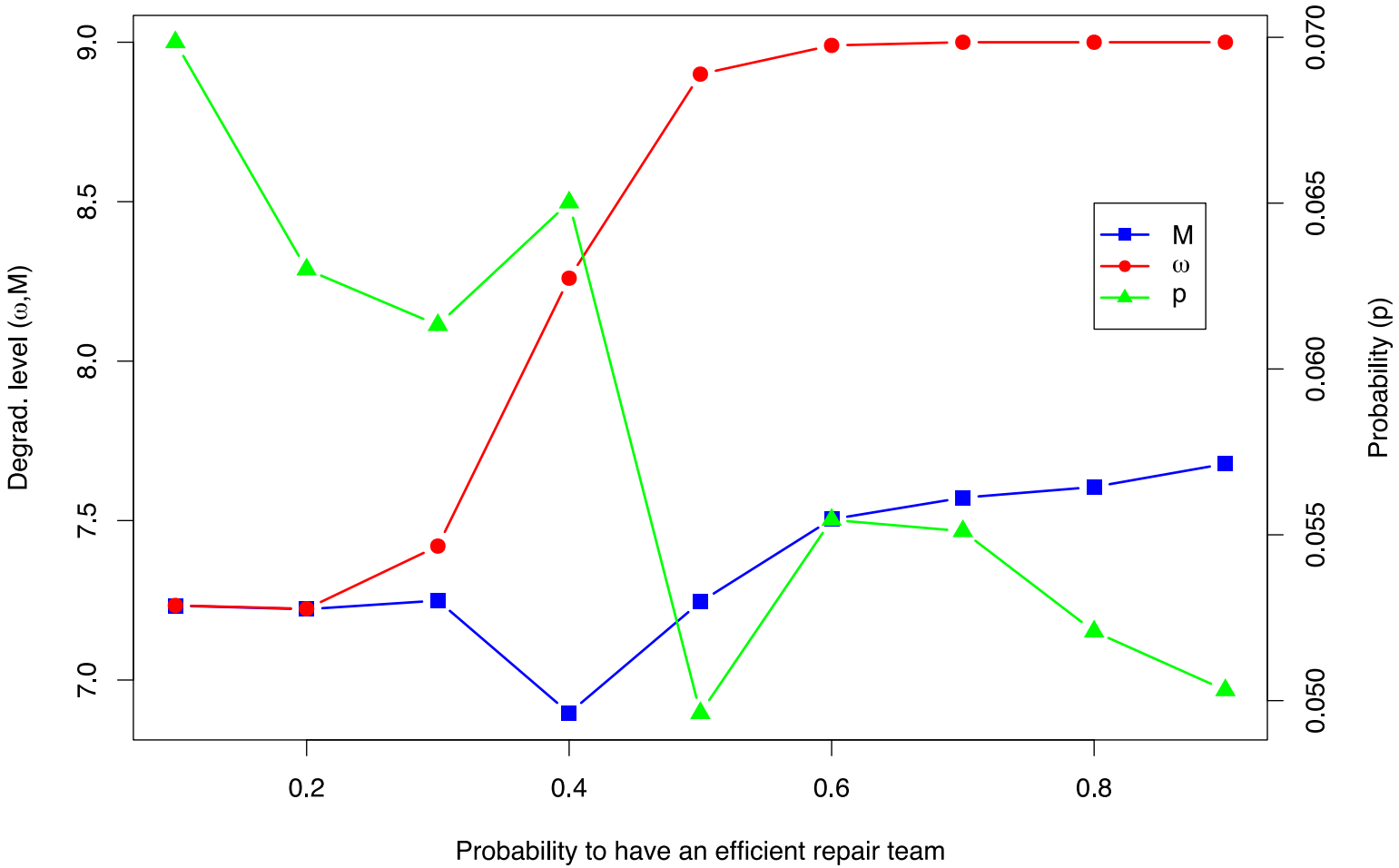
$$\mathbb{E}(\boldsymbol{\delta}) \simeq \mathbf{0.09} \text{ and } \mathrm{Var}(\boldsymbol{\delta}) \simeq \mathbf{0.113^2}$$

  - **Team 2: non efficient maintenance team** - $\boldsymbol{\alpha} = \mathbf{7}$ and $\boldsymbol{\beta} = \mathbf{5}$ (slightly left-skewed distribution).

$$\mathbb{E}(\boldsymbol{\delta}) \simeq \mathbf{0.58} \text{ and } \mathrm{Var}(\boldsymbol{\delta}) \simeq \mathbf{0.137^2}$$

Example of numerical results

- Decision variables with mixture of repair performances



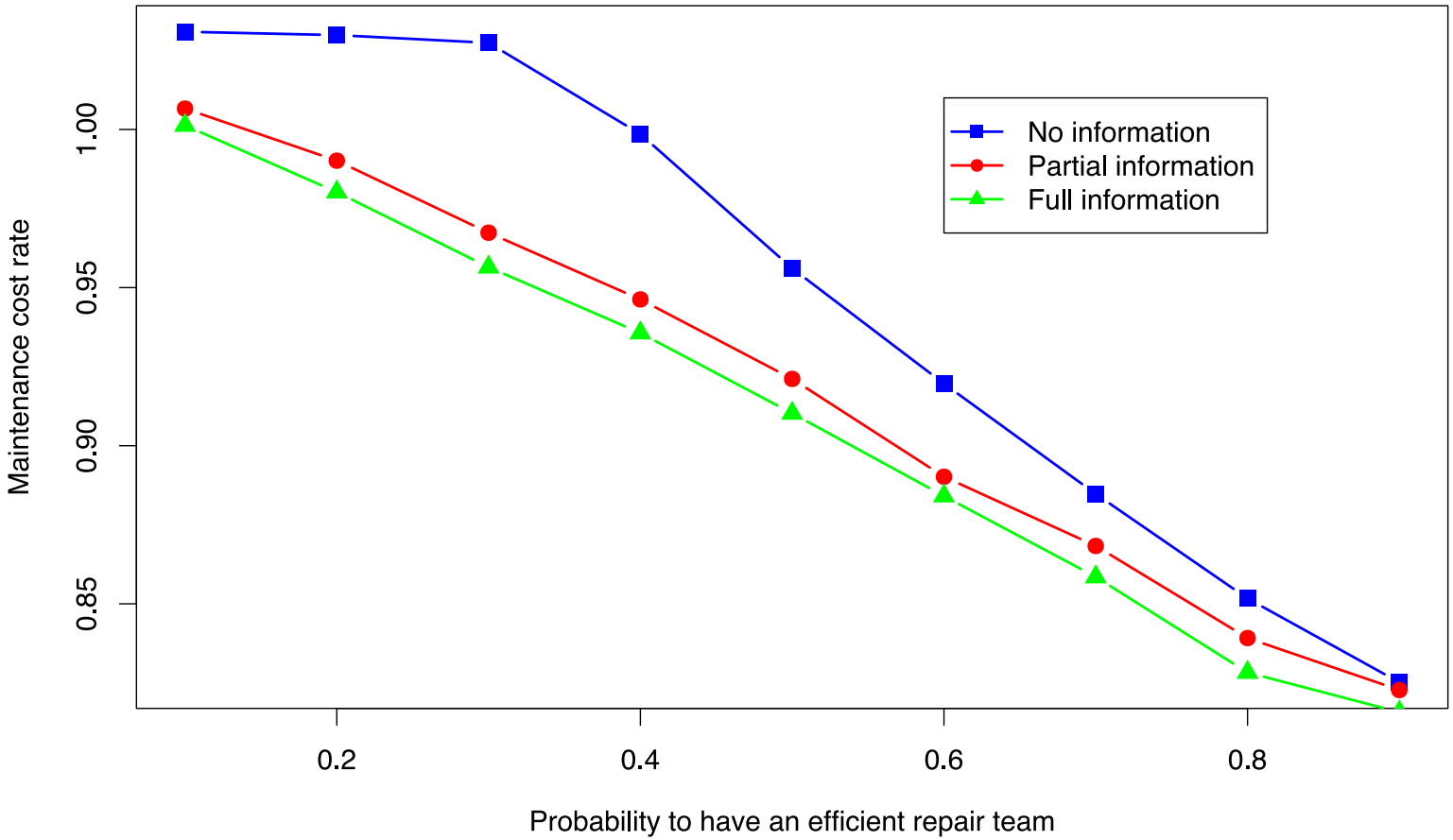
Example of numerical results

☒ **Three different maintenance strategies** corresponding to three levels of on-line available information

- **No on-line information about the mobilized maintenance team**: only the mean proportion of maintenance actions realized by each team is known Decision parameters are optimized considering a mixture of the performance characteristics associated with each team
- **Late on-line information about the mobilized maintenance team**: knowledge of the available team just at maintenance times ⇒ decision parameters $\boldsymbol{s}$ is updated to the optimal value obtained for the current active team but $\boldsymbol{p}$ and $\boldsymbol{M}$ are the ones from previous strategy
- **Early on-line information about the mobilized maintenance team**: knowledge of the available team at the beginning of the current cycle ⇒ decision parameters $\boldsymbol{s}$, $\boldsymbol{M}$ and $\boldsymbol{p}$ are updated at the beginning of each cycle to the optimal values obtained for the current active team

Example of numerical results

- Comparison of three maintenance strategies (i) no on-line information, (ii) info. at maintenance time and (iii) info. at beginning of the cycle



System with a latent degradation mode [a]

[a] E. Mosayebi Omshi, S. Shemehsavar, and A. Grall. An intelligent maintenance policy for a latent degradation system. Reliability Engineering & System Safety, 242, 2023.

DOI:10.1016/j.ress.2023.109739

Context

- ✗ Degradation process that cannot be measured or observed directly and accurately = latent process
- ✗ Indices related to the performance characteristics can be measured = markers
- ✗ Markers are statistically related to the latent degradation state − > valuable information about the unobservable degradation process

- **How to take maintenance decision from markers?**
  - Is it interesting to consider a marker directly?
  - Is it interesting to exploit the marker information to draw a predictive inference on the level of latent degradation?
  - Does the answer depend on the correlation between marker and latent degradation processes?

Marker/Latent degradation process

Predictive inference for a system surviving at time t

- Prediction of the latent degradation level X_t, given the marker level Y_t;
- **Prediction of the system failure time S given that the marker level $Y_t = y$.**

– >The second one is considered.

Example with two processes having similar behavior

- ✗ Non decreasing gradual homogeneous processes - time and space continuous
- ✗ Choice of a bivariate gamma process (Kibble's bivariate gamma process)
- ✗ Correlation coefficient between marker and latent processes as an explicit parameter

Marker/Latent degradation process

Kibble's bivariate gamma process $(X_t, Y_t)_{t\geq 0}$

- Shape parameter α, rate parameters λ_1, λ_2, and correlation ρ
- Joint distribution

$$f(x,y|\alpha,\rho) = \frac{(\lambda_1\lambda_2)^\alpha}{(1-\rho)\Gamma(\alpha)}\left(\frac{xy}{\rho\lambda_1\lambda_2}\right)^{\frac{\alpha-1}{2}} \exp\left(-\frac{\lambda_1 x+\lambda_2 y}{1-\rho}\right) I_{\alpha-1}\left(\frac{2\sqrt{\rho\lambda_1\lambda_2 xy}}{1-\rho}\right)$$

where $x, y > 0, 0 < \rho < 1, \alpha, \lambda_1, \lambda_2 > 0$ and $I_\alpha(\cdot)$ is the modified Bessel function of the first kind of order α defined as:

$$I_\alpha(z) = \sum_{k=0}^{\infty} \frac{(\frac{z}{2})^{2k+\alpha}}{\Gamma(k+\alpha+1)k!}, \qquad z > 0.$$

- Marginal distributions of X_t and Y_t are gamma with the same shape parameter α and rate parameters λ_1 and λ_2, respectively

Maintenance policies

☒ **Periodic inspections (every T units of time)**

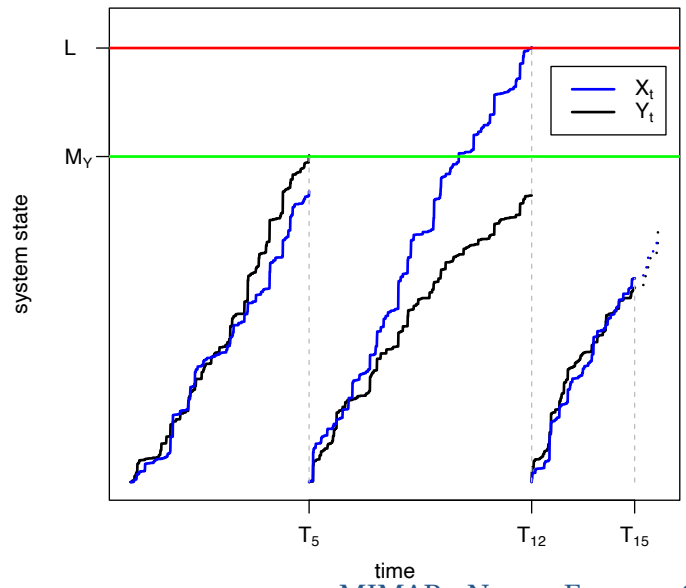
☒ **Two maintenance decision rules**

1. CMP: Fixed limit threshold M_Y on the observable marker process (maintenance action if $Y_{T_k} > M_Y$) - Two decision variables;

2. IMP: Decision based on the conditional distribution of the latent variable X given the observed marker variable Y (maintenance action if $P(X_{T_k} > M_X | Y_{T_k} = y_{T_k}, S > T_k) > p$) - Three decision variables.

• Maintenance actions are supposed to be perfect

• Failures are self announced and in that case the system is immediately replaced



Example of output: Policies comparison

- **Assessement and decision variables optimisation**
  - Long-run cumulated cost rate with $C(t) = c_i N_i(t) + c_p N_p(t) + c_c N_c(t)$
  - Analytical expression obtained by renewal theorem (renewal cycle)
  - For IMP: mathematical analysis suggests an adaptive threshold that dynamically evolves as time passes
- **Illustrative numerical example**
  - Parameters of the bivariate gamma process: $\alpha = 0.1, \lambda_1 = 0.7, \lambda_2 = 0.7$
  - Failure threshold: $L = 60$
  - Unit costs: $c_i = 5$, $c_p = 140$, $c_c = 240$.

Example of output: Policies comparison

- **Reference configuration**: observable degradation level
  - Optimal cost = **0.4590998**
  - Optimal time interval between inspections = **110.6426**
  - Optimal preventive threshold = **43.63159**.

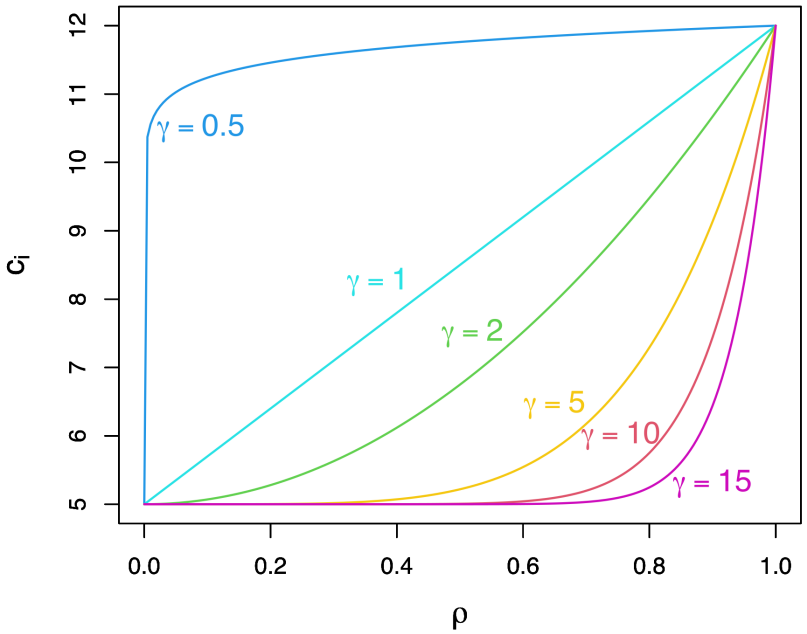
- **Relative increase for optimal cost compared to reference for different values of ρ -> benefit of considering information contained in the model (perfect here)**:

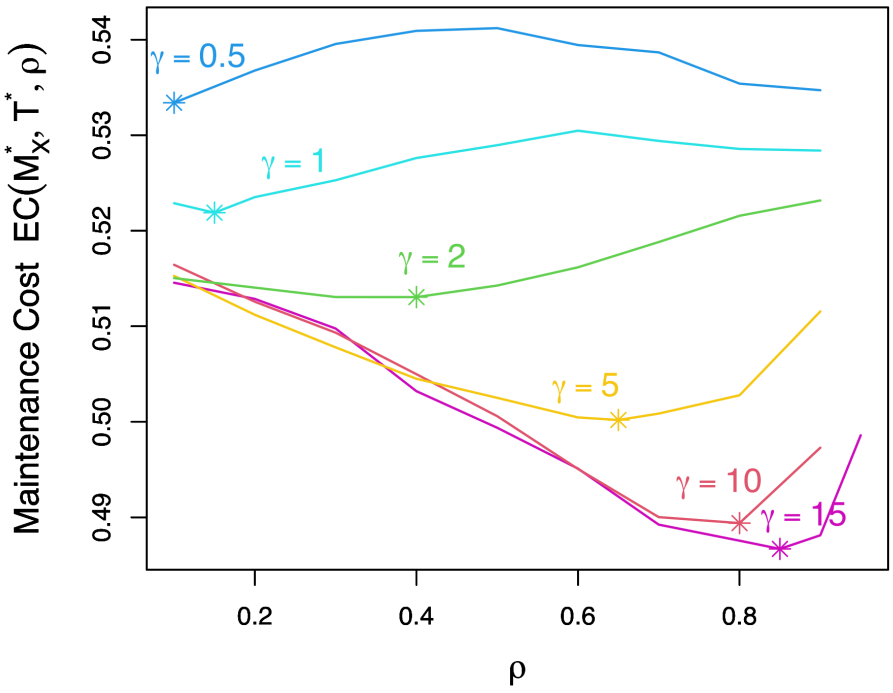
	for CMP	for IMP
$\rho = 0.9$	**3.37%**	**1.94%**
$\rho = 0.7$	**6.11%**	**3.16%**
$\rho = 0.5$	**8.35%**	**3.45%**
$\rho = 0.3$	**9.96%**	**3.58%**
$\rho = 0.1$	**11.25%**	**3.68%**

Correlation as a decision variable

Cost of monitoring performance : Inspection cost and correlation are linked

☞ For illustrative purpose: $c_i = a + b\rho^{\gamma}$
i.e. $\gamma \uparrow$ means increasing correlation becomes less costly)





Correlation as a decision variable

Cost of monitoring performance : Inspection cost and correlation are linked

☞ For illustrative purpose: $c_i = a + b\rho^{\gamma}$
i.e. $\gamma \uparrow$ means increasing correlation becomes less costly)

	EC^* CMP	ρ^* CMP	EC^* IMP	ρ^* IMP
$\gamma = 0.5$	0.533	0.10	0.497	0.11
$\gamma = 1$	0.521	0.15	0.491	0.19
$\gamma = 2$	0.511	0.40	0.478	0.22
$\gamma = 5$	0.498	0.65	0.476	0.39
$\gamma = 10$	0.489	0.78	0.475	0.51
$\gamma = 15$	0.485	0.85	0.475	0.59

Table: Optimal cost and correlation parameter for different values of γ

Correlation as a decision variable

Optimal cost and correlation parameter for different values of γ:

	EC^* CMP	ρ^* CMP	EC^* IMP	ρ^* IMP
$\gamma = 0.5$	0.533	0.10	0.497	0.11
$\gamma = 1$	0.521	0.15	0.491	0.19
$\gamma = 2$	0.511	0.40	0.478	0.22
$\gamma = 5$	0.498	0.65	0.476	0.39
$\gamma = 10$	0.489	0.78	0.475	0.51
$\gamma = 15$	0.485	0.85	0.475	0.59

($\gamma \uparrow$ means increasing correlation becomes less costly)

Concluding remarks and open issues

Conclusion

- Importance of simple decision-making processes
- Need to assess the added value of monitoring information. Academic work can help!
- Dynamic use of optimized decision variables, with updates based on information available online
- Importance of models to complement available data (and vice versa)
- Artificial intelligence is very popular and efficient (e.g. for classification and clustering) but requires a lot of data... statistical and probabilistic approaches can still help
- If lifetime prognosis is useful to maintenance decision-making, problems have to be considered jointly and integrated into a comprehensive framework

Open Issues

Many open challenges at the interface of PHM, Reliability and Maintenance

- Imperfect/Indirect monitoring
- Covariates: Integration of environmental characteristics and future operating conditions into maintenance decision models
- Multi-component systems prognosis & maintenance : scalability issues
- Influence of time horizon on decision-making
- Feedback from operation on the RUL and reconfiguration as a maintenance decisions (e.g. derating in case of controlled systems)

Thank you