Titre: Title:	Reinforcement Learning with Data-Driven Prediction Methods for Optimal Condition-Based Maintenance Strategies
Auteur: Author:	Mina Mikhail
Date:	2022
Type:	Mémoire ou thèse / Dissertation or Thesis
Référence: Citation:	Mikhail, M. (2022). Reinforcement Learning with Data-Driven Prediction Methods for Optimal Condition-Based Maintenance Strategies [Ph.D. thesis, Polytechnique Montréal]. PolyPublie. https://publications.polymtl.ca/10244/

URL de PolyPublie: PolyPublie URL:	https://publications.polymtl.ca/10244/
Directeurs de recherche: Advisors:	Soumaya Yacout, & Mohamed-Salah Ouali
Programme: Program:	Doctorat en génie industriel

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Reinforcement Learning with Data-Driven Prediction Methods for Optimal Condition-Based Maintenance Strategies

MINA MIKHAIL

Département de mathématiques et de génie industriel

Thèse présentée en vue de l'obtention du diplôme de *Philosophiae Doctor*

Génie industriel

Mars 2022

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Cette thèse intitulée:

Reinforcement Learning with Data-Driven Prediction Methods for Optimal Condition-Based Maintenance Strategies

présentée par **Mina MIKHAIL**

en vue de l'obtention du diplôme de *Philosophiae Doctor*

a été dûment acceptée par le jury d'examen constitué de :

Luc ADJENGUE, président

Soumaya YACOUT, membre et directrice de recherche

Mohamed-Salah OUALI, membre et codirecteur

Tamer GAYED, membre

Vahid EBRAHIMIPOUR, membre externe

DEDICATION

To my wife Merna, who gave constant support and encouragement every step of the way, to my family, my father Refaat, my mother Jehan, my sister Marina and my brother Marco who have always believed that I can accomplish great things.

ACKNOWLEDGMENT

I have been blessed with wonderful friends without whom I would not have made it through my Ph.D. degree; no words or actions of gratitude can outweigh how each and every one of you supported and guided me during tough and dry seasons: Michael Kamel, Kerelous Waghen, Mina Rady, and Mina Makram. I am grateful from all my heart to every one of you.

I would not have been here without the sincere advice of my master's degree supervisors, Prof. Mustafa Arafa (at the American University in Cairo) and Prof. Hesham Hegazi (at Cairo University). I could not be more grateful for everything they taught me. Also, they were the first to encourage me to take this opportunity and accept the new challenge.

I want to thank my team at CanmetENERGY. I specially thank my manager Yves Poissant and my colleague Nayeem Ninad, who were always supportive and understanding.

I would like to express my sincere gratitude to my supervisors Prof. Soumaya Yacout and Prof. Mohamed-Salah Ouali, who have supported me throughout this research project. They consistently helped me to formulate my ideas concretely. Along my research journey, they totally and consistently supported me in focusing on realistic objectives and harnessed all obstacles to keep me aligned with my research objectives. I admire their great support, care, and patience during this fruitful research.

RÉSUMÉ

La dégradation est un phénomène naturel qui se produit dans les différents systèmes en raison de l'utilisation et de l'exposition à l'environnement. Par conséquent, une maintenance est nécessaire pour maintenir les systèmes dans des conditions de fonctionnement sûres. Au cours de la dernière décennie, la maintenance conditionnelle s'est avérée capable d'améliorer la fiabilité des systèmes et de réduire les coûts. La maintenance conditionnelle est une stratégie de maintenance qui décide des actions de maintenance en détectant les premiers signes de défaillance à partir de l'état du système. Le développement rapide des technologies de surveillance et d'acquisition de données a encouragé l'application de stratégies de maintenance conditionnelle.

Malgré l'intérêt croissant pour la maintenance conditionnelle et les différentes études proposées dans ce domaine, ce sont des lacunes existantes qui nécessitent davantage d'investigations et de recherches. La littérature s'accorde sur la distinction de la maintenance conditionnelle car elle utilise des informations collectées à partir des systèmes, néanmoins la majorité des études proposées proposent des stratégies de maintenance conditionnelle basées sur des modèles avec des paramètres supposés qui ne sont pas liés aux informations ou données collectées à partir de systèmes réels. De tels modèles ont également des difficultés à envisager des actions de maintenance plus applicables autres que les actions de remplacement. Nous abordons ces problèmes en proposant un modèle et une méthodologie capables d'utiliser des informations réelles collectées sur différents systèmes qui n'ont pas d'hypothèses et d'envisager différentes actions de maintenance. Nous trouvons également une lacune dans la recherche concernant l'utilisation de la durée de vie utile résiduelle. La durée de vie utile restante principalement utilisée dans la littérature pour optimiser les intervalles d'inspection, nous proposons un arrangement différent pour utiliser la durée de vie utile restante avec une méthode de solution basée sur les données pour atteindre une meilleure rentabilité. Enfin, pour tenter de combler l'écart consistant à ne considérer qu'un seul mode de détérioration, le modèle proposé dans ce travail a traité ce problème en utilisant un thème axé sur les données similaire proposé pour combler les deux écarts précédents.

Cette thèse propose un modèle et une méthodologie génériques basés sur les données qui combinent principalement des modèles de prédiction de détérioration et une méthode de solution sans modèle prédéfini, pour optimiser une stratégie de maintenance conditionnelle préventive à plusieurs

niveaux d'intervention. Cette méthodologie n'est pas limitée à certaines hypothèses sur le processus de détérioration et apprend le processus de détérioration directement à partir de données réelles. Il prend également en compte les actions de réparation préventive à plusieurs niveaux à côté des actions de remplacement pour résoudre le problème de maintenance de manière pratique. La méthode de solution sans modèle prédéfini, en particulier l'apprentissage par renforcement, a facilité l'applicabilité de la méthodologie car elle obtient la solution à l'aide d'un apprentissage interactif sans avoir besoin d'estimer aucun paramètre. Plus tard, le concept de durée de vie utile restante du système est adopté pour améliorer la rentabilité des stratégies obtenues. Une méthode qui adopte des techniques de survie non paramétriques ainsi qu'une approche basée sur la fiabilité est proposée pour estimer la durée de vie utile restante à utiliser dans le cadre d'une nouvelle conception pour la fonction de récompense de la méthode d'apprentissage par renforcement. Enfin, cette méthodologie est développée pour être capable de traiter le problème des modes de défaillance multi-dégradation.

Deux points méritent d'être mentionnés. Premièrement, les stratégies obtenues grâce à cette méthodologie associent l'état de dégradation du système à l'action de maintenenance appropriée. Ce qui différe des stratégies de maintenance basées sur les conditions de seuil des actions largement proposées. Deuxièmement, un ensemble de données réelles est utilisé dans cette recherche pour tester la méthodologie proposée.

La contribution de la méthodologie proposée est présentée en trois volets qui peuvent être utilisés ensemble pour résoudre des problèmes complexes. De plus, chacun des volets a sa capacité à résoudre certains problèmes d'applications différentes. Indépendamment de l'individualité de chaque volet, la méthdodologie proposée se développe progressivement de manière séquentielle.

ABSTRACT

Degradation is a natural phenomenon that occurs in different systems as a result of usage and environmental exposure. Accordingly, there is a need for maintenance and inspection to keep the systems in safe and functional condition. Over the past decade, condition-based maintenance has been found to be capable of improving the reliability of systems and reducing costs. Condition-based maintenance is a maintenance strategy that decides and optimizes maintenance actions by detecting early signs of failure from the condition of the system. The rapid development of data monitoring and acquisition technologies encouraged the application of condition-based maintenance strategies.

Despite a growing interest in condition-based maintenance and various studies that have been proposed in that domain, there are existing opportunities that call for more investigation and research. The literature agrees on the distinction of condition-based maintenance, as it uses information collected from the systems; nevertheless, the majority of the studies have proposed condition-based maintenance strategies based on models with assumed parameters that are not related to information or to data collected from real systems. Such models also suffer from challenges in considering more applicable maintenance actions, other than replacement actions. We address these issues by proposing a model and methodology that are capable of using real information collected from different systems that do not have assumptions and consider different maintenance actions and allow optimally carry out proactive maintenance. We also find a gap in the research with remaining useful life utilization. Remaining useful life is mainly used in the literature to optimize inspection intervals. So based on the estimation of the remaining useful life, some planned inspections are canceled, and others are added. We propose a different arrangement for using the remaining useful life with a data-driven solution method as prediction and reinforcement learning models to reach better cost-efficiency. Finally, in an attempt to address the gap of considering only a single deterioration mode, the proposed model in this work treats this problem using a similar data-driven theme proposed to address the two previous gaps.

This thesis proposes a generic data-driven model and methodology that mainly combines deterioration prediction models and model-free solution methods to optimize a multilevel preventive condition-based maintenance strategy. This methodology is not limited to certain

assumptions about the deterioration process and learns the deterioration process directly from real data. It also considers multi-level preventive repair actions, aside from the replacement actions, to tackle the maintenance problem in a practical way. The model-free solution method, specifically, reinforcement learning, strengthened the applicability of the methodology, as it obtains the solution using an interactive learning manner without a need to estimate any parameters. Later, the concept of the system's remaining useful life is adopted to improve the cost-effectiveness of the strategies obtained. A method that adopts survival techniques, together with a reliability-based approach, is proposed to estimate the remaining useful life that is used as a part of a new design for the reward function for the reinforcement learning method. Finally, this methodology is developed such that it is capable of addressing the multi-deterioration failure modes problem.

Two points are worth mentioning: first, the strategies obtained through this methodology map from state to action differently than the widely proposed actions' threshold condition-based maintenance strategies. Second, a real dataset is used in this research to test the proposed methodology.

The contributions of this proposed methodology are presented in threefold that can be employed together to solve complex problems. Each has its own capacity to solve certain problems in different applications. While each is an individual solution, they have been developed progressively, in a sequential manner.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
CBM	Condition-based Maintenance
CM	Corrective Maintenance
DP	Dynamic Programming
GPI	General Policy Iteration
KM	Kaplan Meier
LAD	Logical Analysis of Data
LP	Linear Programming
MDP	Markov Decision Process
ML	Machine Learning
MTTF	Mean Time to Failure
OSM	Original System Manufacturer
PM	Preventive Maintenance
RF	Random Forest
RL	Reinforcement Learning
RUL	Remaining Useful Life
SARSA	State Action Reward State Action

SVM	Support Vector Machine
TBM	Time-Based Maintenance
TD	Temporal Difference
TPM	Transition Probability Matrix
UE	User Experience

CHAPTER 1 INTRODUCTION

Systems that are used in the production of goods (i.e., production machines) or for service providers (i.e., aircraft, vehicles, roads, and railways) witness deterioration over time and from use (Wang, Hongzhou 2002). Such systems should be kept in conditions that allow them to operate safely, as the failure of these systems encompasses high costs and high risks. Thus, failure needs to be mitigated and avoided through maintenance. Maintenance costs can represent up to 70% of production or service costs (Lopes Gerum et al., 2019; Yacout, 2010). As a result, maintenance optimization is a necessity in the growing competition between manufacturers of goods and service providers.

Maintenance is defined as a set of activities that can restore a failed or a deteriorated system to operational conditions that enable it to perform its functions (Ahmad et al., 2012). A few decades ago, maintenance activities were presupposed to be performed after failure to restore systems to operating condition. Over the past decades, and with the second industrial revolution, the concept of maintenance experienced a dramatic transformation from being reactive to proactive. The era of mass production started with the second industrial revolution, and during this period, a strong need for cost minimization and reliability maximization emerged to optimize production (Sakib & Wuest, 2018). Generally, maintenance strategies are classified into two categories: corrective maintenance (CM) and preventive maintenance (PM) strategies, as Figure 1.1 depicts. CM is reactive run to failure maintenance, as it waits until failure takes place and then maintenance is performed. Contradictory PM is proactive maintenance with strategies that suggest performing maintenance before failure takes place to avoid it or reduce its rate (Alaswad & Yisha, 2017; Hongzhou, 2002). With an increasingly competitive market, different organizations have realized that maintenance is an essential business function that needs to be managed efficiently. Therefore, CM is no longer an option, as it encompasses expensive costs and leads to production losses. All of the attention has been given towards developing optimal PM strategies.

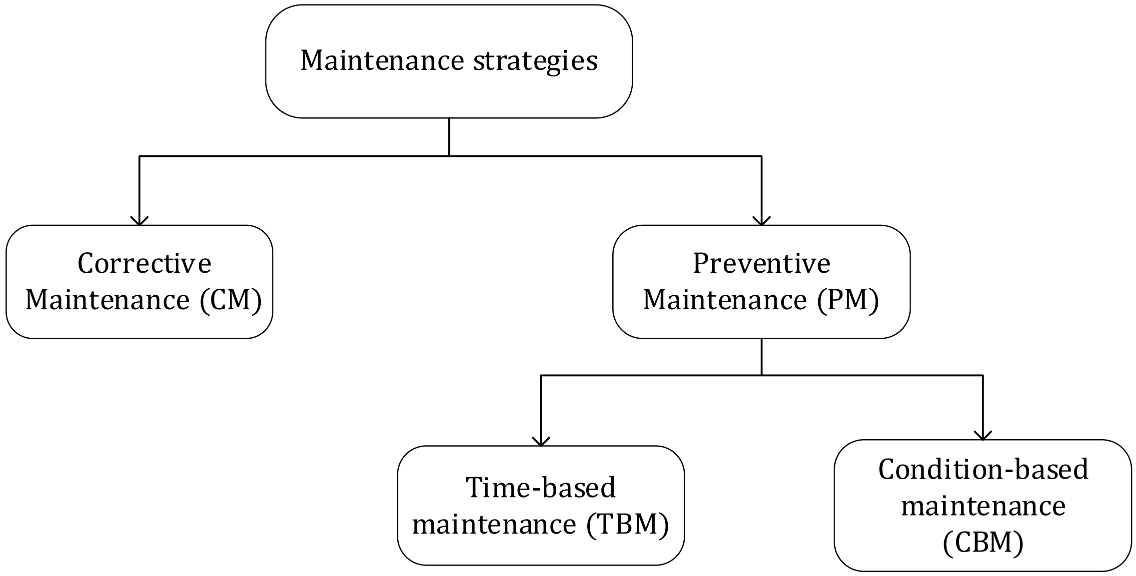


Figure 1.1. Maintenance Strategies Classification.

1.1 Preventive Maintenance

In the early era of PM, it was performed either based on the original system manufacturer (OSM) recommendations or on the user experience (UE) (Zheng & Makis, 2020). Normally, OSM recommends maintenance based on calendar age or usage time and such recommendations are founded upon a scientific approach. UE is built over time with the usage of the system. Users learn a PM strategy for the system they use as they experience different situations. In most cases, engineers or technicians are the individuals who acquire this experience. The main limitation of UE PM strategies is the dependency on experienced personnel. If an experienced person departs, problems with maintenance will arise. Later, in the late 1940s and during the 1950s, the operation research approach began to be developed for PM optimization (Ahmad et al., 2012; Boros et al., 2000; Hafez et al., 2019; Shin & Jun, 2015). Generally, PM has two main categories: time-based maintenance (TBM) and condition-based maintenance (CBM) are shown in Figure 1.1.

1.1.1 Time-Based Maintenance

TBM is a category of PM strategies based on usage time or interval. TBM strategies are traditional and easy to implement since the maintenance action is performed after a certain usage time or interval (de Jonge et al., 2017). The usage time or interval to perform the maintenance is based on the system's failure time analysis. TBM strategies assume that the failure rate follows the bathtub curves shown in Figure 1.2. According to the bathtub curve, a system failure rate passes through three stages (Bai et al., 2020). The first stage is infant mortality, with a decreasing failure rate at the beginning of the system's age or life cycle. The second stage is the random failure stage, with a low constant failure rate. Finally, the third stage is the wear-out stage, in which failures occur at an increasing rate(Ahmad et al., 2012; Bai et al., 2020). TBM mainly targets the wear-out stage to avoid failures through PM actions (Ahsan et al., 2020).

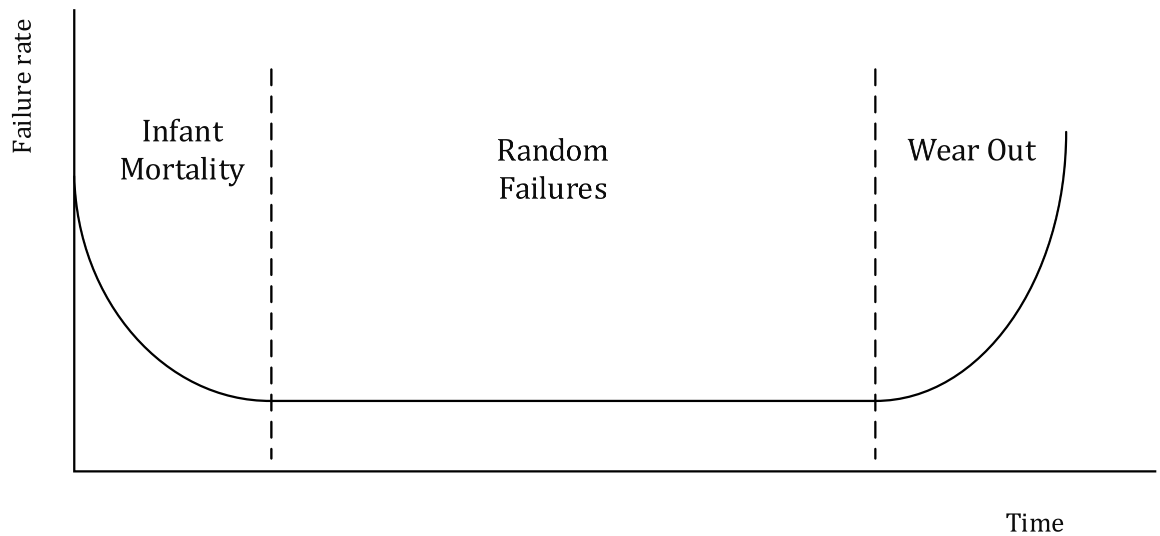


Figure 1.2. Example of the Bathtub Curve.

A summary of the general TBM steps is shown in Figure 1.3. Based on failure times, the system's reliability is modeled using different statistical distributions, e.g., the Weibull distribution, Normal distribution, Binomial distribution, and Poisson distribution. After modeling the system's reliability, TBM seeks to obtain the time or interval for performing maintenance with minimum cost. Obtaining the optimal interval for performing maintenance is a trade-off between the PM cost and CM cost if a failure takes place. The general mathematical model for TBM strategies was first proposed by Barlow and Hunter (1960). This mathematical decision model of TBM strategies includes only the replacement action, as the PM action returns the system to as-good-as new. The replacement is performed either at fixed intervals T with cost C_{PM} or at failure with cost C_{fail}. The cost associated with replacement after failure includes unplanned system downtime, loss of production, and PM replacement costs. On the other hand, the planned PM replacement includes only the maintenance cost. The unplanned system's downtime cost and the loss of production cost are high costs, even higher than the PM replacement, that fact increases the unfavorability of the failure events. The mathematical model by Barlow and Hunter (1960) aimed to obtain the optimal replacement interval that minimizes the total maintenance cost based on the planned PM replacement cost, the cost of unplanned maintenance due to failure, the cumulative failure distribution function, and the reliability function.

Other TBM strategies also consider a minimal repair action. Minimal repairs are performed after failure to retune the system to the operating conditions on which the system was just before failure. The minimal repair cost is lower than the PM replacement cost. Under such strategies, when failure takes place before the defined replacement interval T, a minimal repair action is performed. The objective of strategies with minimal repairs is to obtain the replacement interval T that minimizes the long-run expected cost per unit time. They also consider the expected number of failures that may take place until reaching this replacement interval based on the minimal repair cost C_{MR}, the expected number of failures at time T and the preventive replacement cost C_{PM}.

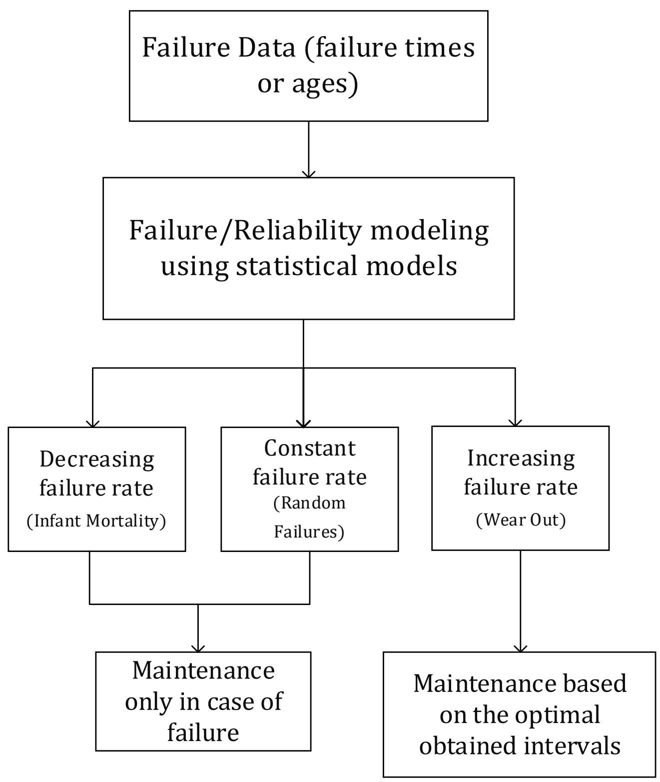


Figure 1.3. Summary of the general TBM steps.

Various TBM strategies have been proposed in the literature to cover different applications. Comprehensive literature about TBM strategies and applications can be found in (Ahmad et al., 2012; Kim et al., 2016). Regardless of the reasonable results obtained based on TBM strategies, two serious limitations arise when TBM is used. Firstly, TBM needs failure data, which may not be available for all systems. The failure of some systems can lead to catastrophic results, so no failure is allowed; thus no failure data is available. Other systems have relatively long-life cycles, which make collecting failure data a difficult process. Secondly, TBM considers only the use of time or interval to decide maintenance actions. It is clear that the use of time or interval is not sufficient to determine whether a system will witness a failure soon or after a long time. The manner of use also significantly affects the system's failure (Quatrini et al., 2020). Therefore, TBM may lead to unnecessary replacement actions or even allow failures in some cases.

1.1.2 Condition-Based Maintenance

The limitations that have arisen within TBM, together with the developments in condition monitoring through sensors that are capable of collecting and transmitting the systems' conditions, have led to an increasing interest in CBM. The beginning of CBM implementation goes back to the 1970s (Ahmad et al., 2012; Baldin, 1986). Performance of the systems is commonly related to the deterioration process, which is a natural phenomenon in which deterioration takes place with usage. CBM provides a more efficient approach than TBM, as CBM detects early signs of failure by observing the deterioration process and accordingly deciding on a maintenance action (Jardine et al., 2006). Deciding a maintenance action based on the systems' conditions is more realistic, as the manner of use significantly affects the system's conditions. Different indicators can be used to identify the deterioration process and the systems' conditions, i.e., noise, vibration level, crack size, temperature, lubrication oil viscosity, and pressure level. Over the usage, the deterioration process is naturally evolving and changes in the different indicators occur as crack size propagation and vibration magnitude increase. When the values of these measured heath conditions reach a certain level, failure takes place (Jardine et al., 2006). Therefore, the approach that CBM follows to decide maintenance actions leads to strategies that decrease the life cycle cost and avoid catastrophic failure. A variety of models and methods have been proposed in an attempt to develop optimal

CBM strategies. A summary of general CBM steps is shown in Figure 1.4, while a more detailed discussion about CBM is provided in Chapter 2.

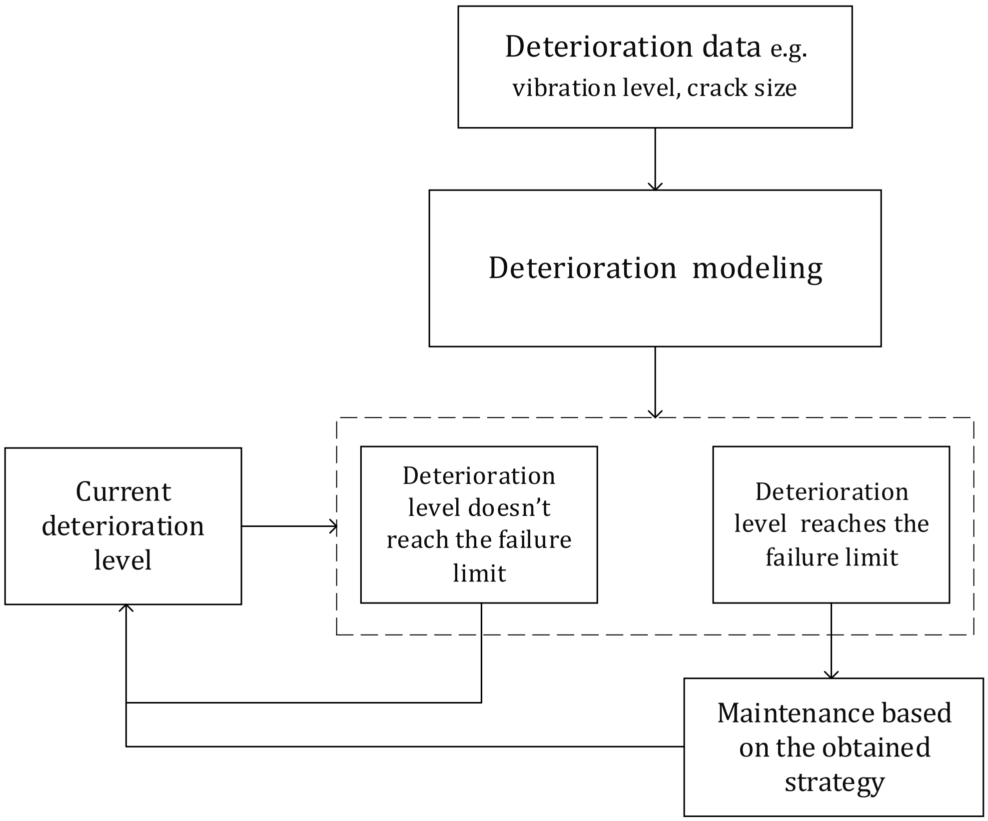


Figure 1.4. Summary of general CBM steps

1.2 Maintenance and Data

Clearly, the early shift in the concept from reactive to proactive maintenance was a need that was achieved with the help of data. At the beginning, with TBM strategies, data was the base for the models to obtain optimal replacement intervals T. Even when PM is based on UE, maintenance experts have developed this experience based on data they collect and observe during their use of the systems. Because of a lack of this failure data and the availability of other types of data, attention has been directed to CBM. It is important to acknowledge that the availability of data, in addition to the need for more optimal maintenance strategies, has contributed to the shift to CBM. CBM requires a clear understanding of a system's normal operating conditions and failure modes. This understanding can be acquired either through prior comprehensive knowledge about the system, or by using data that has been collected to monitor the system. Based on prior comprehensive knowledge of the system, fault detection can be done through model-based fault detection methods. Unfortunately, this way is not very effective since new patterns of faults may occur with time. Covering all of the fault types and patterns using these models is unfeasible. Using the data collected to monitor the system can help understand the systems' normal and faulty conditions. Relying on data to achieve successful CBM is more applicable through the following steps: data collection, data processing and analyzing, and accordingly suggesting the proper maintenance actions (Khan et al., 2021; Yacout, 2010). The data collection process is concerned with acquiring and storing different health indicators about the systems. With the advances in data sensing and storage, the data acquisition process witnesses remarkable improvements, and data availability is no longer present. The data processing and analyzing step focuses on transforming the collected data into useful forms using different techniques. Using the output of the data analysis process, proper maintenance actions can be recommended.

Various types of models and methods are proposed in the literature to use the available data in developing practical and optimal CBM (de Jonge & Scarf, 2020). Statistical-based models, such as Weibull analysis, are used extensively with the CBM problem (Aboura et al., 2014; Bracke, 2020; Khan et al., 2021; Qiuyu & Meiju, 2021; Sgarbossa et al., 2018; Zhang, S.-x. et al., 2015). Despite the wide usage of the Weibull analysis, it has a critical limitation. Weibull analysis is limited to failure data, which is unavailable in many applications. Other studies have adopted different analysis models that do not suffer from this limitation, e.g., logical analysis of data (LAD) (Ghasemi & Esmaeili, 2015; Jocelyn et al., 2020; Lo et al., 2019; Yacout, 2010), artificial neural networks (ANN) (Hafez et al., 2019; Phuc et al., 2019; Santolamazza et al., 2018), support vector machines (SVM) and random forests (RF) (Muhamad et al., 2020; Soonsung et al., 2018; Tan et al., 2019). Such models are generally classified as machine learning (ML) based methods.

1.3 Maintenance and Machine Learning

According to El Naqa and Murphy (2015) *"Machine learning is an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment."* As the amount of data collected during the systems' condition monitoring process comes to unprecedented levels, ML methods became increasingly adopted in the industrial domain. Generally, ML methods can analyze large amounts of data to help optimize industrial operation. ML methods have been used through different means to tackle various challenges related to CBM. For example, a ML pattern recognition algorithm might contribute to diagnostic capabilities of the CBM. LAD, ANN, and decision trees are adopted to help with such fault detection tasks (Apinantanapong & Nivesrangsan, 2021; Dettenborn et al., 2020; Kaparthi & Bumblauskas, 2020; Mortada et al., 2014; Ragab et al., 2019; Zhou, P. & Yin, 2019). The faults and their root causes are obtained and analyzed using such methods (Bulla & Birje, 2021; Waghen & Ouali, 2019). Failure prediction is another important point to consider for optimal CBM, as it defines a specific point of time in the future for failure. Different ML methods such as SVM, RF, ANN, k-nearest neighbor have been adopted to enable this forecasting task (Alves et al., 2020; Hamaide & Glineur, 2019; Kaparthi & Bumblauskas, 2020; Khorsheed & Beyca, 2020; Nowaczyk et al., 2013; Silva & Capretz, 2019; Xiang et al., 2018). The results obtained using such methods have shown that ML failure prediction models are effective for a variety of applications, such as wind turbines, aircraft, and production machines (Leukel et al., 2021). Optimizing the maintenance/production joint, inventory levels, and spare parts flow is also another important issue that has been tackled using ML, specifically, reinforcement learning (RL) methods. Various studies have adopted RL to tackle the aforementioned decision-making problems in maintenance (Compare et al., 2020; Huang et al., 2020; Ramírez-Hernández & Fernandez, 2007; Wang, X. et al., 2016; Wei et al., 2020; Xanthopoulos et al., 2018; Yousefi et al., 2020). Regardless of the progress made in the literature, the maintenance domain still a strong nominee domain to benefit from the growth of efficient data analysis that contributes to the decision-making field (Ingemarsdotter et al., 2021).

1.4 Objective and Contributions

The work proposed in this thesis aims to develop a genuine, data-driven approach for CBM modeling and optimization that makes use of the available data in a way that addresses various limitations exist related to: (1) the deterioration modeling assumptions; in most of the cases, certain models are used to model the deterioration. These models assume a certain shape or bath for the deterioration; moreover, these models have certain parameters, and those parameters are normally assumed and are not estimated from real data. (2) the possible maintenance actions; only replacement actions are considered for CBM. Other maintenance actions are not widely considered, regardless of their existence and benefits. (3) RUL is mainly used to optimize the inspection intervals, limiting the usage of RUL only for this purpose overlooks other benefits of using RUL directly within the optimization criterion, especially when the solution is based on methods such as RL. (4) the existence of multi deterioration failure modes has been disregarded in the proposed CBM models. (5) a threshold paradigm is widely followed in most of the proposed CBM strategies. In such strategies, the solution is a replacement threshold, and in some cases, a repair threshold when a repair action is considered.

The contributions of the work proposed in this thesis are threefold. All three aspects of this work have a common main theme that progressively grows to tackle more and more complicated challenges.

Contribution 1: A data-driven methodology for multi-level CBM strategy.

The first contribution proposes a data-driven methodology based on machine learning for CBM optimization. Machine learning models such as prediction models and reinforcement learning models are adopted and integrated to address challenges related to the modeling of, and solutions for, CBM strategies. Multi-level preventive repair actions are considered in the CBM strategy without suffering limitations related to the solutions. The obtained CBM strategy does not follow a widely used threshold paradigm, as it is a map from state to action. The proposed methodology is applied to a real case study and its performance is validated through a comparison with various other strategies.

Contribution 2: A data-driven methodology with a non-parametric survival method for optimal multi-level CBM strategies.

The second contribution aims to improve the performance of the methodology proposed in contribution 1 by combining it with a nonparametric survival technique. The nonparametric survival technique is used to estimate RUL for the system based on its condition through a reliability approach. RUL is used as a part of the optimization criteria that aim to maximize RUL while keeping a low level of maintenance cost. Based on this criterion, a new design for the RL's reward function is proposed. The new design for the reward function yields better results that minimize the average maintenance cost when compared to optimization criteria that only consider the cost of the maintenance actions.

Contribution 3: Optimal CBM strategies for systems with multi-deterioration failure modes.

The third contribution aims to develop a CBM model that addresses the maintenance problem for systems with multi-deterioration failure modes. This model is developed based on contribution 2 and it proposes certain essence modifications to allow the problem of multi-deterioration failure modes to be tackled. The model proposed is applied to a real case study and the strategy obtained is validated. The validation process is performed to examine the optimality of the obtained strategy and to test the main points considered in the proposed CBM model.

1.5 Organization of the Thesis

This thesis includes seven chapters. Next chapter 2 presents recent studies that address the CBM problem; various aspects related to the modeling, optimization criteria, and solution method are discussed. The problem statement, research objective, and the contribution of this thesis are presented also at the end of this chapter. Chapter 3 provides an overview of the proposed methodology and discusses how this methodology achieves the research objectives. Chapter 4 presents an example that tests a simple methodology the uses RL as a solution method against traditional maintenance optimization methods. Chapter 5 presents the first contribution of this thesis: a data-driven methodology to obtain CBM strategy is proposed and validated. Chapter 6 presents the second contribution of this thesis, in which RUL is employed with RL to reach the most cost-efficient CBM strategy. The third contribution of this thesis is presented in Chapter 7, in which the effect of considering multiple deterioration failure modes and their dependency on the CBM strategies is studied. Chapter 8 concludes the thesis and possible future research topics.

CHAPTER 2 LITERATURE REVIEW

This literature review presents a review of CBM that aims to help understand the fundamentals and recent state of the art of CBM modeling approaches and solution methods, and the challenges that still need to be addressed. Most existing literature is focused on single-component systems. Few studies consider multi-components, and the main target of this literature is studying dependencies between components (Hong, H. P. et al., 2014). Our interest is mainly in single-component systems and their challenges. CBM literature can be discussed from different aspects; this literature is built to consider mainly the different aspects of CBM models as optimization criteria, deterioration models, and the related solution methods. Further in this chapter, remarks and conclusions about CBM in general and the revised studies will be provided. Finally, the motivation for this research is discussed at the end of this chapter.

2.1 CBM and Optimization Criteria

Optimality is the key aspect of a successful maintenance strategy. Different optimization criteria are considered in CBM for minimizing cost, minimizing downtime, and maximizing availability (Alaswad & Yisha, 2017; Lin, S. et al., 2020). The cost of maintenance includes several parameters such as inspection cost, downtime cost, PM replacement cost, and corrective replacement cost. Minimizing a CBM strategy cost involves obtaining a threshold for PM replacement, and in some cases, obtaining inspection frequency to minimize the total maintenance cost over a finite or infinite horizon based on the problem (Huynh et al., 2011). Generally, CBM models follow a framework of three consecutive steps: (1) define the deterioration condition, (2) decide the maintenance action (do nothing, preventively replace, or correct preventively), (3) determine the time for the next inspection as Figure 2.1. This framework can be found in different publications, such as (Diyin, Makis, et al., 2015; Fouladirad, Mitra & Grall, 2015; Xuejing et al., 2010; Yang et al., 2017).

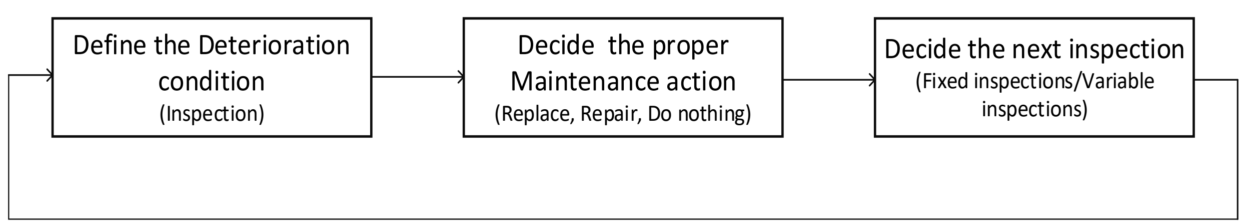


Figure 2.1 CBM general steps.

The inspection step encompasses certain costs related to the process itself and the system downtime cost; therefore, avoiding unnecessary inspections can reduce the maintenance costs. This fact encourages developing CBM strategies that optimize maintenance inspection intervals (Deloux et al., 2015; Fouladirad, M. & Grall, 2014; Golmakani, 2020). An interesting approach followed in such strategies is using the remaining useful life (RUL) to decide whether the inspection is needed (Lin, L. et al., 2019; Salih, 2020; Vu et al., 2016). In such cases, the inspection intervals are flexible and could be increased or decreased based on the RUL. At inspection, based on the deterioration condition the RUL is estimated, and one of two of the following scenarios is followed: (1) if the RUL is longer than the scheduled inspection interval, then this interval increases. (2) if the RUL is shorter than the scheduled inspection interval, a failure will take place before the next inspection, so the inspection interval is shortened to avoid failure. Different approaches are proposed to estimate the RUL and employ this scheme (Hamidi et al., 2016; Mei & Jie, 2012; Tangbin et al., 2017; Zhen et al., 2019).

Availability of a system is another applicable measure for the CBM optimality, especially in cases when different maintenance cost parameters are not available. In such cases, the useful information is the time needed to perform PM replacement and the time needed for corrective replacement. Maintenance time and inspection time are considered as downtime and the production time is the uptime. The objective is to maximize the availability by minimizing the ratio of uptime/downtime by obtaining a threshold for PM replacement, and in some cases, the frequency of inspection (Barde et al., 2019; Klutke & Yoonjung, 2002; Lin, L. et al., 2018; Zhao et al., 2021). This optimization criterion is not as popular as cost minimization and was considered more with TBM. Examples of CBM strategies with an availability maximization objective can be found in (Ait-Kadi & Chelbi, 2010; Elsayed et al., 2006; Klutke & Yoonjung, 2002; Lin et al., 2018; Qingan et al., 2017; Zhao et al., 2021).

2.2 CBM, Deterioration Modeling, and Solution Methods

Different models have been proposed to address the deterioration modeling for CBM problems. Gamma process, Inverse Gaussian process, Wiener process, and Markov decision process (MDP) are the most adopted models in CBM problems (Alaswad & Yisha, 2017; Braga & Andrade, 2019).

Gamma process and Inverse Gaussian process are both appropriate for describing continuous monotonic deterioration processes. Cholette et al. (2019) proposed a CBM model for a heat boiler exchanger using the Gamma process to model the deterioration. The objective was to obtain an optimized CBM renewal strategy that can adopt different loading scenarios for the boiler. Phuc et al. (2015) created a CBM policy with deterioration following a Gamma process model that considers perfect and imperfect maintenance actions. The objective of the study is to obtain the optimal policy and to study the effect of the imperfect maintenance actions on the deterioration process. Examples for other studies that benefit from the gamma process for deterioration modeling can be found in (Castanier et al., 2005; Wang, Han et al., 2021; Zhang, C. & Tee, 2019; Zheng & Makis, 2020). The inverse Gaussian process is similar to the Gamma process with more flexibility. Nan et al. (2015) proposed a CBM replacement policy with an optimal inspection interval when the deterioration follows an inverse Gaussian process. The study defined an optimal inspection interval and replacement policy that minimized total operation costs. Comparable work was proposed by Renqing et al. (2017), in which the deterioration was assumed to follow the inverse Gaussian process. This work aimed to obtain an optimal CBM that maximizes availability while respecting cost constraints. With the constraint of maintenance budget an optimal CBM strategy is that maximizes availability of the product. Huynh (2021) proposed a CBM model that conforms to the deterioration as an inverse Gaussian process, where the previous maintenance actions effect was represented using the random effect of an inverse Gaussian process. Various numerical assessments confirmed that the proposed model is flexible in molding the deterioration and cost-effectiveness. Other studies that considered the inverse Gaussian process as a deterioration model can be found in (Lu et al., 2019; Wu, Z. et al., 2020; Zhang, Xinsheng et al., 2017; Zhenyu et al., 2014). A Wiener process can describe a continuous non-monotonic deterioration that takes place incrementally over time, as proposed in (Elwany et al., 2011; Mimi et al., 2014). Guo, C. et al. (2013) proposed a CBM that assumes a gradual degradation modeled by a Wiener process for mission-oriented systems. This model intended to obtain an optimized PM threshold that minimizes the maintenance cost and while considering missions related constraints. A numerical example is presented to demonstrate the proposed model. Comparable work can be found in (Hong, P. et al., 2018; Minghui et al., 2020).

The previously discussed studies considered a single deterioration failure mode; this is the case in most studies (Alaswad & Yisha, 2017). Other studies considered another failure mode; in such cases, this failure mode is sudden. The sudden failure modes are generally modeled using statistical distributions as non-homogenous Poisson and Weibull distribution (Diyin, Makis, et al., 2015; Jian et al., 2016; Li, X. et al., 2019). Models based on the Gamma process, inverse Gaussian process, and Wiener process are solved using methods like Monte Carlo simulation and genetic algorithm (Alaswad & Yisha, 2017; Quatrini et al., 2020).

Regardless of the advances accomplished through the Gamma process, inverse Gaussian process, and Wiener process model, there are still certain common limitations that appear with such models. Firstly, these models assume a certain shape or bath for the deterioration that is applicable for certain cases e.g. monotonic deterioration path that is irreversible. Secondly, another strong assumption that exists in most cases is that the model parameters are known; however, these parameters are usually unknown in practice and should be estimated (Huynh, 2021). The estimation process for the parameter from the deterioration data is not a straightforward process and needs data that has certain statistics and properties based on the model. Finally, using continuous state space is not always practical from an engineering point of view as there is no need - or it is not applicable – to consider every value

Distinct from the aforementioned models, the MDP is another model for the deterioration process that assumes discrete state space. MDPs are found to be appropriate to model discrete deterioration without assuming a certain bath for it. MDPs assume discrete deterioration, which is practical from the engineering point of view, since in many cases there is no need - or it is not applicable - to consider every value (Alaswad & Yisha, 2017; Zheng & Makis, 2020). In general, MDPs are the framework for decision-making problems under uncertainties. It is important to mention that MDPs can be used to model both the deterioration and the maintenance problem from the beginning, or a different deterioration model can be aggregated into discrete states and can form the CBM problem. These facts make MDPs interesting and useful models; more detail about MDP and the solution method are discussed in the following sections.

2.3 CBM and MDP Modeling

CBM models that assume discrete-state deterioration are generally modeled using MDP or its extensions as a Semi-Markov decision process and a Hidden Markov decision process. Braga and Andrade (2019) proposed a CBM model for railway wheelsets that model the deterioration as a Markov process. An optimal maintenance policy that minimizes the maintenance cost is obtained considering three possible actions: do nothing, renewal, and turning action. A similar study that addressed the problem of the high-speed train wheels is proposed by Mingcheng et al. (2018). In this case, the deterioration was modeled as a Semi-Markov process and the CBM strategy obtained was found to minimize the long-term expected cost per time unit. Other studies that assume similar deterioration modeling can be found in (Byon & Ding, 2010; Farnoosh & Makis, 2015; Kurt & Kharoufeh, 2010; Liang et al., 2019). The previously mentioned studies estimated the transition probability matrix (TPM) for the MDP and solved the problem using methods such as dynamic programming (DP) and linear programming (LP). An essential assumption for DP and LP is the availability of a perfect MDP model. This assumption of a perfect, complete model is hard to ensure in many cases, as estimating TPM is a challenging process (Braga & Andrade, 2019; Mandiartha et al., 2017; Van Otterlo & Wiering, 2012). Besides, solving the MDPs using DP requires multiple complete sweeps through the state space, which is computationally expensive. LP is less is computationally expensive than DP. It forms an optimization problem for each state and tries to find the optimal policy that maximizes or minimizes - based on the problem - the total return, while respecting constraints about the possible action at each state. Using branching and bound algorithms, LP can efficiently obtain the optimal solution for large MDP (Lopes Gerum et al., 2019; Malek et al., 2014; Sanner & Boutilier, 2012). However, as with DP, LP is still limited by an assumption of a perfect MDP model with a defined TPM. This assumption led the CBM, based on MDP, to be limited to certain applications. The TPM is unknown and it either takes assumed values or has to be estimated from the deterioration data. As mentioned, the estimation process of the TPM has been found to be a challenging and subjective process (Braga & Andrade, 2019; Liang et al., 2019; Mandiartha, Duffield, Razelan, et al., 2017). The process is subjective, as it uses different methods and steps based on the available data and the application. Then, moreover, the difficulty of such models cannot be generalized.

Another category of methods that can be used to solve MDP based models is the Model-Free method, namely, RL. RL, in contrast to DP and LP, can obtain an optimal solution for MDP without the assumption of a complete MDP model (Buoniu et al., 2018). RL obtains the solutions directly from the data. It requires only episodes of data. The episodes are sequenced tuples of state, action, and reward. It has been found that, in many cases, it is easy to get the data episodes, while it is infeasible to obtain the complete probability distribution of all the possible transitions (Sutton & Barto, 2018). Moreover, RL does not experience challenges with large size problems. RL has been used with CBM to overcome the limitations related to TPM estimation and, in some cases, the limitation of high computational cost (Rabbanian et al., 2021).

2.4 Maintenance and RL Decision-Making

This section proposes a general review of the applications of RL in maintenance strategies. Adsule et al. (2020) proposed an optimized CBM strategy based on RL. This study assumed that the deterioration can be assessed using health index HI, $HI(t) = HI_{initial} + m*t$. in the equation that models the HI, m is a deterioration rate obtained from Gaussian distribution and t is the time. Besides the replacement action, a minor repair action is possible. The minor repair action is assumed to reduce the deterioration rate. A numerical example is proposed in the study for illustrative purposes. Unfortunately, no real cases were addressed in this study. Zhang, N. and Si (2020) proposed a CBM model for a multi-component system using RL. The deterioration of the components was assumed to follow either a compound Poisson process or a Gamma process and the maintenance model for the whole system is modeled as MDP. The purpose of using RL is to obtain a CBM strategy that considers all the components and their dependencies. Two numerical studies were carried out to prove the optimality of the obtained CBM strategy through the proposed model. The two studies address systems with a different number of components to test the scalability of the model. It is worth mentioning that in the two numerical studies, the parameters of the deterioration models were assumed, and not obtained, from real data. Other studies adopted RL for obtaining an optimal TBM policy for a multi component system, as proposed by (Barde et al., 2016; Barde et al., 2019).

On the level of maintenance/production control, Liu, Y. et al. (2019) proposed a dynamic selective maintenance approach based on RL. The proposed approach addressed the problem of production/maintenance joint. The objective of the proposed approach is to optimize the maintenance to allow the execution of multiple consecutive missions over a finite horizon. The failure time of each component was assumed to conform to an arbitrary distribution. Kuhnle et al. (2019) addressed the problem of maintenance scheduling for parallel working machines based on RL. The objective of the proposed maintenance schedule is to maximize the number of completed jobs by opportunistically maintaining production machines during breaks. The failure times of the production machines are assumed to follow Weibull distributions. The obtained maintenance schedule leads to a reduction in downtime, increases in the production output, and a reduction in the maintenance costs compared to existing maintenance schedules. Comparable studies are proposed in (Hu, Y. et al., 2021; Ling et al., 2018; Wang, X. et al., 2016; Xanthopoulos et al., 2018). Other studies profited from RL application in maintenance to manage the flow of the parts, as proposed in (Compare et al., 2020; Rocchetta et al., 2019).

2.5 Discussion

The discussed studies demonstrate the variety of CBM strategies and models proposed to tackle various maintenance challenges, yet these studies still endure certain limitations. The Gamma process, Inverse Gaussian process, and Wiener process succeeded in providing deterioration models for various CBM strategies and applications; however, the parameters of such models are, in most of the cases, assumed (Huang et al., 2020; Huynh, 2021). Using assumed parameters means that no real data is involved in the model. Another point to consider is that using these models limits the study to certain applications, since each model can represent certain deterioration shapes or bathes. Markov processes are not limited to certain deterioration shapes if TPM is available. As discussed, TPM can be either assumed or estimated directly from the data. The estimation process is challenging and cannot be generalized (Braga & Andrade, 2019; Mandiartha, Duffield, Thompson, et al., 2017). As MDPs are good models for decision-making problems in general, CBM profits from this point by merging deterioration models as Gamma process with MDP and used RL to solve the problem without the need for the TPM (Huang et al., 2020). Such cases overcame the limitations related to TPM estimation; nevertheless, limitations related to models based on the Gamma process arose again. Also, RL is adopted excessively in the case of maintenance/production control problems where the deterioration is treated based on time-based models and not on condition-based methods, which led to consider the maintenance problem to the TBM one (Kuhnle et al., 2019; Liu, Y. et al., 2019).

The majority of the discussed studies and the existing CBM models are generally limited to maintenance actions that return the system to as-good-as-new condition; however, other maintenance actions exist at the practical level (Alaswad & Yisha, 2017; Zhang, N. & Si, 2020). Certain studies consider minimal repair actions that return the system to as-bad-as-old condition (Fu et al., 2020; Xanthopoulos et al., 2018), and some others consider maintenance actions called general repair or imperfect repair actions; such actions were assumed to improve the deterioration condition by a random amount (Adsule et al., 2020; Ling et al., 2018; Liu, Y. et al., 2019). The type or degree of maintenance actions considered is also related to the failure modes considered in the models. Few studies discussed multi-failure modes for the same system and the studies that consider more than one failure mode assume one hard failure mode that occurs randomly or suddenly without any warning, and another soft failure mode that is described as deterioration failure mode, as proposed in (Diyin, Jinsong, et al., 2015; Jian et al., 2016; Li, X. et al., 2019; Rui & Makis, 2020). In such cases, the hard failure mode is corrected only by replacement, and the soft failure mode can be corrected by replacement, minimal repair, or general repair actions. In real cases, systems experience multiple soft or deterioration failures, not only one; there is a chasm in considering this fact.

2.6 Research Motivations

CBM has attained remarkable achievements in terms of efficient maintenance that minimizes cost and increases reliability; yet, there are still a number of open-ended research points, as explained in the discussion section. This research is motivated by the following main issues:

1- Developments in condition monitoring, data acquisition, and data storing technologies have increased the availability of the data that can be used in maintenance planning as deterioration data. There is a need for CBM models that use the available data in a generally applicable way that is not limited to a specific application or limited by specific assumptions that limit the applicability of these CBM models.

2- CBM takes maintenance actions based on the system's conditions. These conditions have a direct relation to the system's RUL and some studies utilize the RUL to obtain better results. An interesting open question for research is to explore how RUL can be employed with data-driven solution approaches as RL to obtain more cost-efficient maintenance strategies.

3- Multiple deterioration failure modes are involved in the deterioration process of the same system, most of the existing CBM models focus on a single failure deterioration mode. There is an unanswered question about how considering multiple deterioration failure modes in the same CBM model and hoe it may affect the cost-efficiency of the obtained CBM strategy.

CHAPTER 3 REASEACH METHODOLOGY AND WORK SYNTHESIS

This research provides a model and methodology to obtain optimal multi-level CBM strategies. Multi-level terminology refers to the possible maintenance action that can be performed to improve the system's deterioration condition by different levels. Multi-level preventive repair actions are considered, in addition to the no maintenance or do-nothing action and the preventive or corrective replacement actions. The proposed model and methodology are generic and data-driven. They are generic since they are applied to a wide range of applications without subjective steps or assumptions. Subjective steps or assumptions are steps or assumptions that are related to certain applications or cases that cannot be generalized to other applications. They are data-driven since both the model and the methodology, including the solution method, are purely based on the deterioration and maintenance data.

3.1 Basic Elements of the Proposed Methodology

MDP is the first, basic element for our proposal. MDPs are a good model for decision-making problems. Moreover, they do not assume a certain bath or shape for the deterioration. However, MDPs witness two main limitations, which we have addressed in our proposal. Firstly, the estimation of the TPM of the MDP is a challenging and subjective process. A prediction deterioration model is proposed, adopted and integrated with the MDP. This prediction model learns the deterioration process directly from almost any available deterioration data without assuming a certain shape or bath for the deterioration. Secondly, solving MDP problems using traditional methods as DP or LP needs a perfect MDP model; this means a model with a TPM. To overcome this challenge, the RL method is adopted to our problem. RL methods are, in general, sensitive to the design of the reward function. Generally, when RL was used with maintenance, the optimization criterion is to minimize the average maintenance cost and the reward function used is negative the maintenance cost. This reward is used in the first contribution, chapter 5, with the other elements of the model and the methodology. Later, a different reward function design is proposed in this work. The new design incorporates the system's RUL in addition to the maintenance cost. RUL is used in optimizing the inspection frequency and it yields praiseworthy results, but we propose a different way to utilize the RUL in maintenance decision optimization. The proposed reward function design is more appropriate and enables RL to obtain the optimal solution. Reward functions that consider only cost could be misleading and influence the solution obtained by the RL. The proposed RUL's estimation method is also data-driven and adopts and integrates a nonparametric and semi-parametric survival function methods and reliability-based approach. The previously described model and methodology that is developed considers a single deterioration failure mode. Deterioration failure modes are failure modes that are described through the deterioration process and take place progressively. To address the overlooked problem over multi-deterioration failure modes, we develop a methodology that has the same elements and theme as the one previously discussed. In this methodology, the deterioration prediction models are developed to consider multi-failure modes and their interactive effect on each other. Multi-level preventive repair actions are enabled for the different failure modes, where the actions are capable of improving the different failure modes at different levels. The method for RUL estimation is also developed using a reliability-based approach, taking into account multi-failure modes. Figure 3.1 presents the scheme for the proposed the model and methodology.

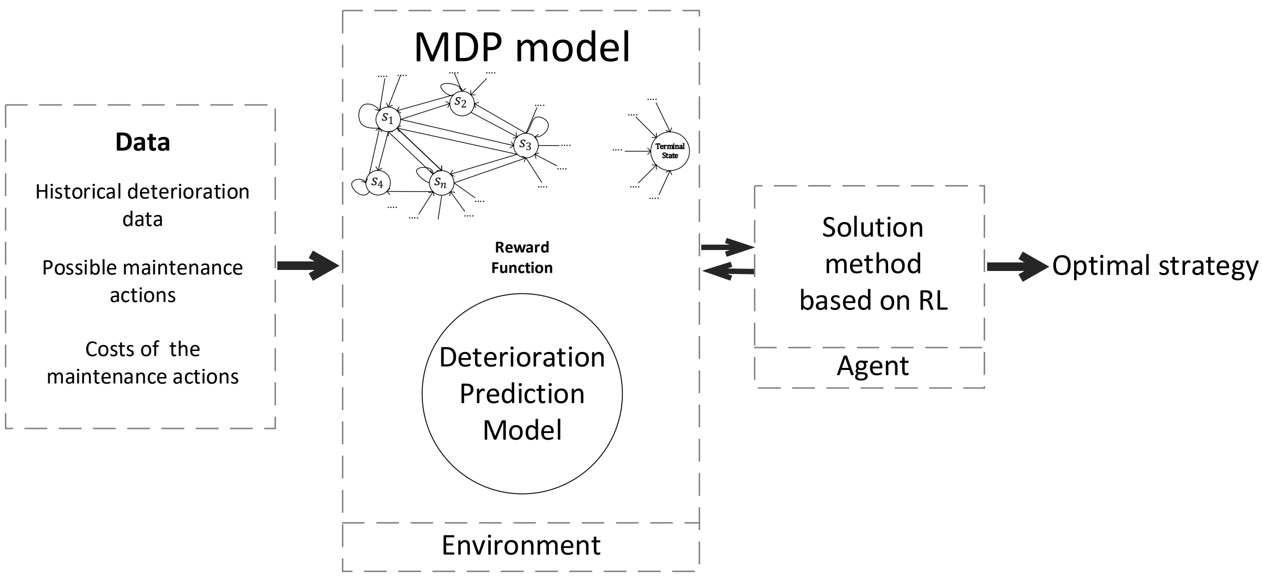


Figure 3.1 General scheme for proposed model and methodology

This research addresses four main challenges related to the assumptions of deterioration modeling, the type of the considered actions and their relation to the deterioration modeling and the solution approach, the employment of RUL in the decision-making process and finally the multi-deterioration failure modes and their effects on the obtained CBM strategies. These challenges are addressed in three contributions that are presented in detail in chapters 5 through 7. The elements of the model and methodology proposed in the thesis are individual scientific contributions that can be applied separately to solve different maintenance problems. However, the progressive fashion of their development is a central point. It is worth mentioning that each of the three contributions is tested and validated using a real case study.

3.2 Case Study

The case study proposed in this research is based on real data for identical sheet metal culverts. The deterioration conditions of the culvert are assessed regularly through inspection. Different health indicators or failure modes are evaluated to estimate the culvert general deterioration condition. According to the experts, the most dominant failure modes are defects in material and cracking and assembly defects. The general deterioration condition of the culvert is assessed on a 0-100 scale, with 100 meaning as-good-as new condition and 0 meaning very deteriorated condition. This 0-100 scale is discretized by the experts into five levels of deterioration: A, B, C, D, and E, where E is the failure level, as table 3.1 shows.

Table 3.1 Levels of deterioration condition.

Deterioration levels	Description	Value on 0-100 scale
A	Perfect	85 or more.
B	Very good	71 – 84
C	Good	56 – 70
D	Acceptable	41 – 55
E	Bad (Failure)	40 or less.

The value of the deterioration condition on the 0-100 scale is based on the health indicators, or failure mode levels. The defects in material failure mode are assessed using a five-level scale; 5 is the best and 1 is the worst, and the cracking and assembly defect failure mode is assessed using a three-level scale level: 3 is the best and 1 is the worst.

This data is used based on two perspectives. The first perspective observes the condition of the culvert based on its general deterioration level. The second perspective observe more details and use the different deterioration failure mode to describe the culvert condition. In the first and second contributions, chapters 5 and 6 respectively, the deterioration process is described by the general deterioration condition and its five levels A, B, C, D, and E without using details about the different failure modes and their effects. In the third contribution, chapter 7, the deterioration process description required more detail, so the failure mode levels, are used to describe the process.

This case study, with its real data, is used to test the different contributions proposed and show their scientific and practical advantages. This case study is selected to be representative and general enough that if the methodology is applied to other cases, it should work properly.

The structure of the rest of this thesis is organised into 5 chapters. Chapter 4 presents a practical example to test a simple methodology that uses RL as a solution against traditional maintenance optimization methods. This example is a simple first attempt towards the contributions proposed. It strengthened our ideas about adopting RL for more complex maintenance problems. Also, it confirms the importance of the reward function for the RL and opened the door to propose a new design for the reward function. Chapters 5, 6, and 7 present the three main contributions with developments of the model and the methodology. Chapter 8 is a general discussion about the results and the limitations of this research. Finally, chapter 9 provides the conclusions and recommendations.

CHAPTER 4 ARTICLE 1: OPTIMAL PREVENTIVE MAINTENANCE STRATEGY USING REINFORCEMENT LEARNING

Paper presented at the 4th North American International Conference on Industrial Engineering and Operations Management, Toronto, ON

http://ieomsociety.org/toronto2019/papers/44.pdf

Authors: Mina Mikhail, Soumaya Yacout, Mohamed-Salah Ouali

Abstract

Taking optimal maintenance decisions is a challenging process as different maintenance actions have different effects on the system. Maintenance is defined as a set of associated techniques, tools and management actions that aim to maintain or restore the functioning state of the system. Maintenance excellence is the balance between performance and risk, therefore this helps improve the sustainability of production. Traditionally, maintenance decisions are taken based on human experience and on the basic information known about the system. With the availability of data collected during the system's life cycle; machine learning approaches can help develop optimal strategies for maintenance actions. This paper proposes and depicts an optimization model imported from the machine learning field and developed to find optimal preventive maintenance strategies. The main objective of this developed optimization model is to minimize the downtime and allow the system to take autonomous decisions. In this work, the maintenance strategy is modeled as a Markov Decision Process (MDP). MDP is a classical forming of sequential decision-making problem. Reinforcement learning (RL) model is then developed to solve the problem interactively. RL uses the MDP to define the interaction between the learning agent and the environment. The final output from this method in an optimal policy allows providing optimal actions in different situations.

Keywords

Preventive Maintenance, Systems Reliability, Reinforcement Learning, Markov Decision Process.

4.1 Introduction

4.1.1 Maintenance Problem

The growing competitive environment in the manufacturing field forces different organizations to reduce their costs (Barde et al., 2016). The two main expenditure sources that can be reduced, without any loss of the quality level, are the energy consumption cost and the maintenance cost. Proper maintenance can also lead to optimizing energy consumption. The main objective of maintenance excellence is to ensure the maximum reliability and availability of the system with the minimum cost and without affecting the production quality. Maintenance can be classified into corrective maintenance (CM) and preventive maintenance (PM). CM is the maintenance actions that take place when the failure of the system occurs, thus these actions are the result of failure and they aim to restore the system to specific conditions. PM is performed while the system is still operating; and considered as set of activities performed to retain the system in specific conditions. Several steps are needed to achieve PM, which are inspection, detection and prevention of anticipated failure (Hongzhou, 2002).

In this paper, our main concern is obtaining optimal PM strategy to minimize the downtime of the system. For this type of PM, the component is replaced after a specific time T or after failure, depending on which occurs first (Shey-Huei et al., 1995). Classically the optimal replacement time T^* is obtained by solving an optimization problem. An example that implements this concept is proposed in (AbdelHaleem & Yacout, 1998). It has been noted that the solution given by this method may lead to local optimal time to replace for the different components of the system (Barde et al., 2016). To overcome this limitation, RL has been used to provide data-driven optimized solutions that could outperform the classical optimization techniques.

4.1.2 Reinforcement Learning

Reinforcement learning (RL) is an area of machine learning algorithm that is concerned with the decision-making process. A software agent learns to take actions by interacting with a dynamic environment. The agent knowledge is enhanced by using scaler value feedback, which is related to a reward function. The agent learns how to take actions that lead to maximizing this reward function (Wiering & Van Otterlo, 2012). Mapping from the situation (state) to an optimal action is the main output of the RL and it is in the form of optimal policy to be followed in different situations. RL is solving problems modeled as Markov decision processes (MDPs). Unlike dynamic programming, RL does not need the probability transitions matrix and does not perform full backups to solve problems (Sutton & Barto, 2011). RL algorithms are model-free; they use exploration and exploitation techniques, and interaction with the environment to provide the optimal actions in different situations.

Game theory and robotics are the most popular domains for RL (Sutton & Barto, 2011), the industrial field is one of the domains impacted by RL. Xanthopoulos et al. (2018) proposed an approach to obtain near-optimal control policy for production-maintenance joint based on reinforcement learning. The solution proposed in this paper aimed to minimize the sum of two conflicting objective functions: the average inventory level and the average number of backorders. Kuhnle et al. (2019) addressed the optimization of an opportunistic maintenance schedule for parallel working machines. The aim of this paper was to reduce downtime and increase production output. Xiao et al. (2016) investigated the problem of scheduling maintenance for two different series machines to sustain a certain buffer level between the two machines. Mattila and Virtanen (2011) proposed a maintenance scheduling for a fleet of aircrafts. The main objective of this paper is how to select maintenance times for different aircraft to keep a high level of readiness of the aircrafts fleet. This situation takes place when the activities of the fleet are not planned a priori. Liu, Q. et al. (2019) proposed an optimized maintenance plan that considers restrictions related to the resources as the spare parts that are available in stock. Also, Compare et al. (2018) treated the problems of gas turbine parts flow management by considering a preventive maintenance plan and stochastic failures of gas turbines. All the previous literature is related to assembly line scheduling, optimal inventory level and optimal maintenance production joint schedules. Some limited literature refers to the use of RL to optimize maintenance plans only. Barde et al. (2019) proposed optimized preventive maintenance strategies for a fleet of military trucks. Three different preventive maintenance strategies were optimized using Monte Carlo reinforcement learning methods (MRCL). Also Barde et al. (2016) proposed another solution to optimize opportunistic preventive maintenance for a multi-component system with a hierarchical structure. The optimized strategy was obtained using temporal difference reinforcement learning algorithm SARSA(λ). Based on the literature review a variety of work has been proposed in the field of using RL with maintenance. Most of this work is directed towards scheduling production and maintenance times to reach a certain required level of inventory. A smaller portion of the proposed work focuses its efforts to optimize the preventive maintenance plans using RL techniques.

This paper presents a model to optimize PM strategies using RL. The main goal of this model is to minimize the downtime of the system through autonomous decisions based on data-driven machine learning techniques. The reward function is constructed in terms of system reliability. The reliability of the system is obtained using the Kaplan-Meier estimate, which is a nonparametric survival function. This model avoids the limitations related to the parameters needed as inputs for traditional optimization methods like the renewal reward theory. The rest of this paper is organized as follow: section 2 provides a description of the problem and the proposed solution. Section 3 contains a numerical example. Conclusions and future work are presented in section 4.

4.2 Model Description

4.2.1 Problem Description

Maintenance plans that aim to maintain the equipment in a functioning state are looking for maintenance excellence. As mentioned, balancing maintenance excellence is a challenging task, since to ensure high performance or reliability levels for systems, maintenance actions should be performed regularly and within short periods. These frequent maintenance activities have high costs related to spare parts cost, labor cost and cost due to loss of availability. To solve this problem a compromise is needed. Most of the PM plans addressed this compromise by using the renewal reward theory. This theory proposes an explicit form to solve this compromise (Blischke & Murthy, 2003). AbdelHaleem and Yacout (1998) proposed a preventive maintenance strategy for a fleet of military trucks based on the renewal theory. The replacement time for each component is obtained by the renewal reward theory. The objective of this work was to obtain the optimal replacement time that minimizes the downtime of the system. The main limitation of this method is that many parameters needed as input to the model. In addition, the obtained solution is dependent on the inputs and any small change in the values of these parameters leads to different solutions. As an alternative, such limitations could be overcome by using the data collected during the equipment's life cycle to propose solutions based on machine learning. As maintenance is a decision-making problem, RL methods can be used to optimize maintenance strategies by using the data collected during the life cycle.

4.2.2 Reinforcement Learning Model

4.2.2.1 Model Description

In the proposed solution, a Markov decision process (MDP) is used to model the problem of searching for an optimal maintenance strategy. MDP is used since it is the classical forming of sequential decision-making problem and RL uses it to define the interaction between the learning agent and the environment.

MDP has mainly the following four elements:

1. Discrete state-space S in which, at every time step, a new state $s_t \in \mathcal{S}$ takes place.
2. A set of actions is available in action space A, in which, at every time step, an action $a_t \in$ A is taken.
3. Transition probabilities between the states $P(s_{t+1}|s_t, a_t)$, which is the probability of being in state s_{t+1} given that the system was in state s_t and action a_t is performed.
4. The reward function $r(s_t, a_t)$, which is the reward of performing an action a_t at state s_t.

RL is capable of solving this MDP problem without the need for the transition probability matrix. A principal point in this modeling process is how the reward function is designed. The objective of this work is to minimize the system's downtime. That could be achieved through eliminating the failures by PM actions without exaggerating in the frequency of PM actions. Therefore, finding the optimal time for the PM actions is the solution to this problem. To fulfill these requirements without being restricted to use certain parameters, nonparametric estimation of the system $R(t)$ could be used for designing the reward function. A full description of the reward function is provided in the next section.

4.2.2.2 Model Formulation

As discussed, RL is a model-free concept, so there is no need for the transition probabilities matrix. The main elements that need to be defined are the states, the actions, and the reward function. In the proposed model the system state is defined by the age G and W, which denotes the system status either normal W = *1* or failure W = 0. Then the following vector defines the equipment's state at any time:

$s_t = (G_t, W_t)$	Eq. 4.1

The action at every time step can be either a PM action $a_t = 1$ or it may be "do-nothing" $a_t = 0$. The last element to be defined is the reward function $r(s_t, a_t)$. As discussed, the reward function design is proposed in a nonparametric form. In addition, it should represent the objective of the model which is minimizing the downtime without introducing very frequent unnecessary PM actions. To achieve this compromise the following design for the reward function is proposed:

$r(s_t, a_t) = \begin{cases} -R(t), & if\ a_t = 1 \\ -1, & if\ a_t = 0\ and\ W_{t+1} = 0 \\ R(t), & otherwise \end{cases}$	Eq. 4.2 Eq. 4.3 Eq. 4.4

Where R(t) denotes the reliability function. Equation (4.2) represents PM action before failure takes place. In this case, the reward that the agent receives has a small penalty with a negative value of reliability. This value is associated with the time when PM action is taken. The purpose of introducing this penalty aims to avoid any needless PM actions and encourage the agent to pick the right time for PM actions. Equation (4.3) represents the failure of the system before the PM action is taken. In case, the agent is penalized by a relatively large value, as the action that was selected by the agent leads to failure. Equation (4.4), R(t) represents a positive value reward that the agent receives if no failure takes place and while no PM action is taken. As mentioned, the value of R(t) is obtained by the Kaplan-Meier estimate which is a nonparametric survival function.

4.2.2.3 Solution Description

The objective of the RL is to find the best policy π that can maximize the expected reward r^* over an infinite time horizon as shown in Equation (4.5) (Sutton & Barto, 2018):

$r^* = {}^{max}_{\pi}\mathbb{E}[\sum_{t=0}^{\infty} \gamma^t . r(s_{t,}\pi(a_t))\|s_o]$	Eq. 4.5

Where γ is the discount factor, $\gamma \in (0,1)$, and s_o is the initial state of the equipment.

The reward r depends on the actions taken in the different states under a certain policy π. So, all the RL algorithms start with an arbitrary policy π and evaluate the state-action value function Q(s, a) under π, then they keep improving the policy until reaching the optimal policy π^*. SARSA (λ) (State-Action-Reward-State-Action) algorithm is suitable in our case since it does not need to wait until the end of an episode to update the value of Q(s, a). It just needs one step forward. Moreover, it provides fast convergence and it is not computationally expensive (Sutton & Barto, 2011). SRASA (λ) estimates the value function Q(s, a) by using temporal-difference methods that are combined with eligibility traces. The value function Q(s, a) is updated every transition from one state-action pair to another. In this way, the value function is continuously updated. At the same time, policy π is keeping updating towards π^* by using the greedy approach (Sutton & Barto, 2018). The update for the value function is done as shown in Equation (4.6):

$Q(s_t, \mathrm{a}_t) = Q(s_t, \mathrm{a}_t) + \gamma\ [\alpha . Q(s_{t+1}, \mathrm{a}_{t+1}) - Q(s_t, \mathrm{a}_t)]$	Eq. 4.6

Where γ is the discount factor, γ ∈ (0,1) and α is the learning rate, α ∈ (0,1).

The value function Q(s, a) estimate provides the optimal policy π^*. Then π^* provides the optimal action to take at each state. Therefore, optimal time $T^*$for taking replacement action is obtained through this optimal policy π^*. The next section, section3, presents an evaluation of the proposed model through a numerical example.

4.3 Model Evaluation

AbdelHaleem and Yacout (1998) proposed an optimized PM plan for a fleet of military trucks. The objective of the proposed plan is to minimize the downtime of the system. The optimal replacement time for each component was obtained by solving the optimization Equation (4.7) that is based on the renewal theory.

$$argmin_{T^*} D = \frac{t_p \, . \, (1 - F(T) + t_f \, . \, F(T))}{(T + t_P) \, . \, (1 - F(T) + \left[\left(t_f + E[t \mid t \, \leq \, T] \right) \right] . \, F(T) \,)}$$	Eq. 4.7

Where T^* is the optimal replacement time to minimize the downtime D, t_p is the time to perform preventive maintenance action, t_f is the time for replacing the component in case of failure, F(T) is the failure probability of component at time T and $E(t|t \leq T)$ is the expected time to failure given that failure happened before T. Failure data were gathered for all the components of the trucks and the failure probability distribution for each component was modeled by a Weibull cumulative density function then T^* was obtained for each component.

Two components, the brake, and the coupling are selected to evaluate the proposed model. The optimal replacement times T^* for the two components are obtained by the proposed model.

4.3.1 Finding the Optimal Replacement Time using the Proposed Model

In order to obtain the optimal replacement time T^* by the proposed model, MDP is used to model the problem of obtaining the optimal maintenance strategy. The proposed RL model solves the MDP without the need for the transition probability matrix. The Input for the proposed model is data in form of episodes; each episode consists of tuples of state, action, and reward. The time step between every two tuples is 10 hrs, thus the age is defined for the state every 10 h also the decision is taken every 10 hrs. The state is defined by the age G and W which denotes the system status either normal $W = 1$ or failure $W = 0$. The actions at each state are either a PM action $a_t = 1$ or it may be "do-nothing" $a_t = 0$. The reward function is defined by the reliability R(t). The reliability is obtained using the Kaplan-Meier estimate.

For the SARSA (λ) algorithm that solves the proposed model, the learning rate is selected to be 0.001. This learning rate is small enough to eliminate rough fluctuations if any noise appears on the data. The discounting factor is selected to be 0.6 to introduce an acceptable level of uncertainty about the actions in the future. Decaying exploration rate ϵ_n is used, $\epsilon_n = {}^{1}/_{(1+n)}$ where n is the nth episode. The decaying exploration rate is used to ensure the convergence to the optimal solution (Sutton & Barto 2011). To define the size of the data sample needed to train the SARSA (λ) algorithm, the average change in the value function Q(s, a) is measured against the number of episodes. The required sample size can be defined as the number of episodes that lead to a low and stable average change of Q(s, a) (Sutton & Barto, 2011). As shown in figure 1, when using 10000 episodes or more the average change in Q(s, a) is stable at a value less than 0.05. Therefore, a sample size of 10000 episodes is sufficient to be used for training SARSA (λ) algorithm to solve the model.

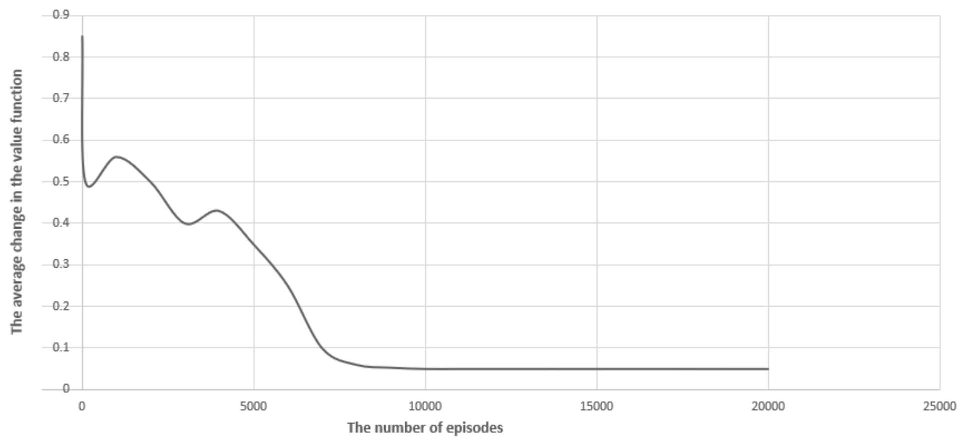


Figure 4.1 The average change in the value function versus the number of episodes.

4.3.2 Discrete Event Simulation

After obtaining T^* for the brake and the coupling, these values are compared with the values provided by AbdelHaleem and Yacout (1998). To complete the comparison a discrete event simulation for 100,000 hrs is performed. The simulation is performed using the optimal replacement time obtained by the two models to compare their performance. The simulation inputs are:

1. The optimal replacement time which defined by the proposed model and by the model of (AbdelHaleem & Yacout, 1998).
2. The time needed for PM action t_p and the time needed to correct a failure t_f are respectively 0.7 h and 3.5 h for the brake and .857 h and 6 h for the coupling.

3. The failure distribution function Weibull distribution P(t; λ, k), for the brake λ= 3933.12, k = 143.60 and for the coupling λ= 1406.84, k = 115.21. All the parameters t_f, t_p, λ, and k are obtained from (AbdelHaleem & Yacout, 1998) and(Barde et al., 2019).

At each decision time, a random failure time (FT) is generated from the failure distribution function then:

If $T^* >$ component age (G), then check if FT > G, then we move to the next step and the age of the component is increased by the time interval, while if FT < G, then the component is replaced, its age is reset to 0, the downtime is increased by t_f and 1 is added to the failure counter.

If $T^* < G$, then the component is replaced, its age is reset to 0, the downtime is increased by t_p, and 1 is added to the replacement counter.

The outputs from this simulation are the total downtime of the equipment due to PM action or failure, the number of PM actions, and the number of failures in 100000 hrs. Comparison and results are shown in table 1, table 2, and figure 2.

Table 4.1 Comparison between the results of the two models for the Brake.

	Brake			
	Optimal replacement	Number of PM actions	Number of Failures	Total Downtime hrs.
(AbdelHaleem & Yacout, 1998) Model	2250	444	0	310.8
Proposed Model	3800	261	2	189.7

Table 4.2 Comparison between the results of the two models for the Coupling

	Coupling			
	Optimal replacement	Number of PM actions	Number of Failures	Total Downtime hrs.
(AbdelHaleem & Yacout, 1998) Model	2160	0	715	4290
Proposed Model	1320	757	0	648.75

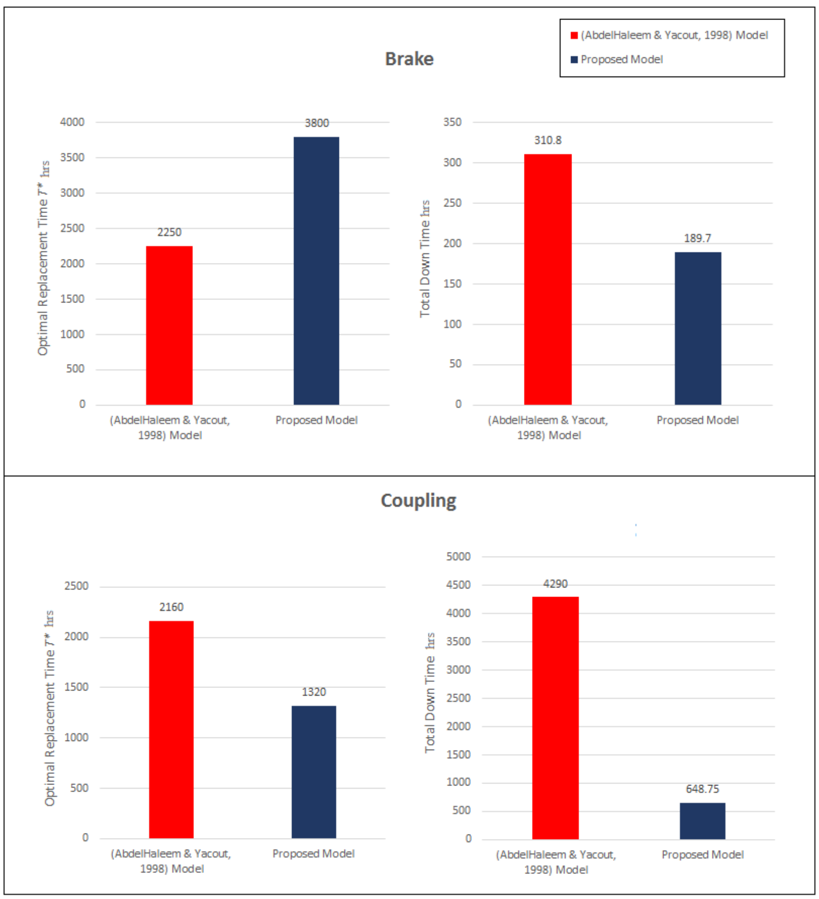


Figure 4.2 Simulation Results for the model by (AbdelHaleem & Yacout, 1998) and the proposed model.

4.4 Discussion of the Results

Figure 2, Table 1, and Table 2 show a comparison between the proposed model and the model by (AbdelHaleem & Yacout 1998). In terms of the optimal replacement time T^* the two models provide different results. For the brake T^*= 2250 hrs using the original model while by the proposed model T^*= 3800 hrs. For the coupling T^*= 2160 hrs by the original model while by the proposed model T^*= 1320 hrs. It is notable that in the brake's case T^* using the proposed model is bigger than T^* by the original model, while in the case of the coupling T^* by the proposed model is smaller than this by the original model. To investigate which of the two models having a better solution the discrete event simulation was performed using each solution. The results of the simulation show that the proposed model was able to outperform the original model by yielding lower downtime in both of the cases. For the brake, the downtime due to PM actions or failures correction is reduced by 39%. For the coupling a great reduction for the downtime by 84% took place. The interpretations of these results can be concluded as follows:

1. In the case of the brake, the original model performed the PM actions with a high frequency which is not needed. These unneeded actions lead to an increase the downtime. The proposed model found the value of T^* that eliminated the unneeded PM actions. A limitation appeared in the solution by the proposed model; that it allows two failures.
2. In the case of the coupling, the original model did not take any PM actions, it just corrected the failures. The proposed model obtained T^* that eliminated the failures to minimize the downtime.

From the previous interpretations, we could conclude the proposed model provides better performance than the original model. In addition, the proposed model is not dependent on any parameters.

4.5 Conclusion

The objective of this work aimed to find an optimal PM strategy, using the RL area imported from machine learning. This maintenance strategy minimizes the downtime of the systems. The proposed solution in this work to this problem is to develop an RL model that could be solved by SARSA (λ) algorithm. First, the challenges related to the maintenance problem are described with the limitations of the available models. Mainly these limitations are both the need and the high dependency of the available models on the input parameters and their values. The proposed model overcomes this limitation. In the proposed solution the problem was modeled as MDP, the developed model was based on the reliability of the system. The Reliability was obtained using the Kaplan-Meier estimate. Then an optimal solution to the problem was founded by the SARSA(λ) algorithm. The problem was solved without the need for the transition probabilities matrix. In addition, the proposed model is not dependent on any parameters as the failure distribution function, t_p, or t_f. These capabilities of RL as a nonparametric data-driven model enable it to solve the real-time problem of autonomous decision-making.

To evaluate the performance of the proposed model, a numerical comparison between the original model and the proposed model through discrete event simulation was performed. The proposed model outperformed the original model in two different cases, as the proposed model provides T^* that yield less downtime when used to perform PM actions.

Areas for further research are: i. solve the limitation related to failure allowance as in the case of the brake. ii. extend the model for the mult-component systems, so all the components are considered together in the same model. iii. extend the model to include minimizing the maintenance cost explicitly in the objective.

CHAPTER 5 ARTICLE 2: A DATA-DRIVEN METHODOLOGY FOR OPTIMAL MULTI-LEVEL CONDITION-BASED MAINTENANCE DECISION-MAKING

Submitted at Computers & Industrial Engineering journal

Manuscript Reference No.: CAIE-D-21-03003

Authors: Mina Mikhail, Mohamed-Salah Ouali, Soumaya Yacout

Abstract

This paper proposes a data-driven methodology that combines a deterioration prediction model and reinforcement learning method to optimize a multi-level preventive condition-based maintenance strategy. The strategy consists of ordinary replacements and do-nothing actions in addition to multi-level repair actions that temporarily mitigate the deterioration of a system. The optimization problem is formulated using a finite Markov decision process with the objective of minimizing the total average maintenance cost over a finite horizon. The data-driven prediction model learns the deterioration process without any assumptions about its shape or bath. Then, the prediction model interacts with a reinforcement learning agent to obtain an optimal sequence of maintenance actions without the need for the Markov decision process's transition probability matrix. The strategy obtained is a map from the state to the action, which differs from the widely presented action threshold condition-based maintenance strategies. This proposed methodology is applied to a case study of sheet metal culverts system and its results are compared to different maintenance strategies, starting from new and degraded states.

Key Words*:* Reinforcement learning, Deterioration prediction, Condition-based maintenance, Multi-level preventive repair.

5.1 Introduction

Following optimal maintenance strategies is necessary since as maintenance has a significant impact on the operational efficiency of various systems and assets. A maintenance strategy (or policy) can be defined as a set of actions performed on a system based on certain decision criteria to sustain or restore its safely functioning state (Ahmad et al., 2012). In the literature, the terms strategy and policy are used interchangeably. In this work, we are using the term "strategy" to describe our model, or when we refer to literature that uses that term. The term "policy" is used when referring to literature that uses that term. In general, maintenance strategies can be classified as corrective maintenance (CM) that is performed after failure, or preventive maintenance (PM) that is done to improve a system's conditions while it is still operating (de Jonge et al., 2017). CM actions have high costs of failure and require excessive amounts of time; as a result, they are mostly undesirable. PM actions are supposed to lead to lower costs, and they are carried out during planned withdrawals, so they do not affect the availability of the system.

Time-based preventive maintenance (TBM) and condition-based preventive maintenance (CBM) are the main classes of PM (Hu, J. & Zhang, 2014; Yang et al., 2017). For TBM, statistical models such as Weibull, Normal, and Poisson distributions are used to perform a lifetime analysis and to determine the intervals of PM maintenance actions (Peng et al., 2010; Shafiee & Finkelstein, 2015; Zhang, N. & Yang, 2015). TBM models have two main limitations: 1) the need for failure data and 2) a disregard for a system's conditions and effects of usage. As a result, TBM may lead to failure or unnecessary maintenance.

CBM is capable of overcoming these limitations by recommending maintenance actions based on the information collected about the system's conditions to avoid both failures and unnecessary maintenance (Chen, N. et al., 2015; Jardine et al., 2006; Jingjing et al., 2019). It has been found that CBM can effectively improve a system's reliability and reduce maintenance costs (Alaswad & Yisha, 2017; He et al., 2017; Zhou, Z.-J. et al., 2012). The distinction of CBM over TBM increases interest in molding and optimization methods of CBM strategies. Markov decision process (MDP) is one of the most widely used methods to model CBM strategies (Braga & Andrade, 2019; Khaleghei & Makis, 2015; Li, X. et al., 2018). MDPs are useful to model this kind of problem since they are able to address a system's deterioration process (Zhang, Xueqing & Gao, 2012). Moreover, MDPs are appropriate models for sequential decision-making under uncertainties in which the actions affect subsequent situations and not only the immediate situation (Puterman, 2014). Methods such as linear programming (LP), dynamic programming (DP), and reinforcement learning (RL) are used to obtain the optimal solution for an MDP optimization problem (Lopes Gerum et al., 2019; Malek et al., 2014; Sutton & Barto, 2018). While estimating the transition probability matrix (TPM) for the MDP is necessary for LP or DP solution methods, the RL method does not need TPM. RL is a data-driven method that solves the problem in an interactive manner (Sutton & Barto, 2018).

Another issue of great concern is the effect of maintenance actions on a system's conditions. Most PM strategies consider replacement actions that return a system's conditions to as good as new and a do-nothing action, which leaves the system's conditions as they are. The maintenance strategy in such cases has a threshold for performing the PM action. In many applications, other maintenance actions are possible, such as general and minimal repairs. While general repair actions may improve a system's conditions by different levels, minimal repair actions restore the system to a condition that is as bad as the old system's condition before failure (Liu et al., 2016; Wu, F. et al., 2015; Zheng & Makis, 2020). Examples for implementing such actions in CBM can be found in (Adsule et al., 2020; Braga & Andrade, 2019; Liang et al., 2019; Ling et al., 2018). In the following two sub-sections, the relative work and limitations are discussed.

5.1.1 MDP and Maintenance

Braga and Andrade (2019) proposed a maintenance optimization approach for railway wheelsets based on the MDP. The MDP was adopted to model wear and damage and to offer possible actions. The states of the wheelsets were defined by the diameter change, the damage that occurred, and the mileage since the last turning action. Three different actions were proposed: renewal, turning, and do nothing. An empirical approach was used to obtain the (TPM) for the MDP under each action. Policy iteration and value iteration approaches based on DP are used to find the optimal strategy with regards to cost over a life cycle. Mandiartha, Duffield, Razelan, et al. (2017) proposed an optimization model for pavement using MDPs. The optimization model aimed to minimize maintenance costs. The state space of the model has eleven states and the action space has three actions: routine action, periodic action, and rehabilitation action. The TPM was obtained from another publication that discussed the same application (Mandiartha, Duffield, Thompson, et al., 2017). The estimation of the TPM in (Mandiartha, Duffield, Thompson, et al., 2017) used a mathematical method that is based on assumptions related to the pavement structure and the traffic volume. The problem was solved using a dual LP formulation. The obtained maintenance strategy was explained and interpreted. Liang et al. (2019) proposed an optimized CBM strategy for long-life assets. A Semi-Markov decision process (SMDP) was used to model the maintenance policies and to obtain the optimal policy. The proposed approach was used to optimize the thresholds for both major and minor maintenance actions to minimize the time-averaged operation cost. Comparable work could also be found in (Fan et al., 2019; Huang et al., 2020; Huy et al., 2020).

The above-mentioned papers have demonstrated that MDP based models that consider a system's condition achieve cost optimal maintenance strategies. However, they still experience limitations related to obtaining the MDP's TPM, which is essential for the DP and LP techniques to solve the problem. DP solution techniques have two significant limitations: they are computationally expensive and need TPM to be capable of solving the problem. DP performs multiple sweeps through the state space during the solution process to obtain the result, which leads to high computational cost. The recent advances in the computational capabilities of computers allow us to overcome this limitation. In addition, LP proposed methods to solve the MDP problem with less computational cost (Malek et al., 2014; Mandiartha, Duffield, Razelan, et al., 2017). Such methods could be better than DP in terms of computational cost, but as with DP, they need TPM to solve the problems. According to the authors, obtaining the TPM is a challenging limitation. It contains particular steps related to the application and cannot be generalized or transferable to other applications (Braga & Andrade, 2019; Mandiartha, Duffield, Razelan, et al., 2017).

In attempt to overcome these limitations, RL is adopted to solve PM models based on the MDPs. RL is capable of solving such problems without the need for TPM. It needs episodes of data sequences that are tuples of (state, action, reward) to obtain an optimal solution. Generally, the state is an observation that describes the system's conditions. Based on the state, an action is selected by the RL agent. Finally, the reward is a return that is provided to the RL agent, depending on both the state and the action that is taken. The availability of data helps with the application of such approaches.

5.1.2 RL and Maintenance

Barde et al. (2019) formulated three replacement PM strategies for multi-component military trucks as MDPs. For the first strategy, each component is replaced at failure and at replacement intervals. For the second, an overhaul action for the system replacement, at a fixed interval, is added to the first strategy. The third is a group-based strategy that allows opportunistic maintenance actions, so when a component is replaced due to failure, or there is a PM replacement, the components that are in a neighborhood with that component are also replaced. These replacement PM strategies were optimized using the Monte Carlo RL (MCRL) method, where the objective is to minimize the system's downtime by obtaining the optimal replacement interval for each component and the optimal overhaul time for the system replacement. The results show that the MCRL method results outperformed the results obtained by the optimization problem for downtime minimization solved using simulation. Barde et al. (2016) developed two opportunistic PM replacement maintenance policies for a system that exhibits a hierarchical structure. In the first PM replacement policy, each component is replaced upon failure and at the replacement interval. The second policy considers the system hierarchy. When the system experiences downtime due to either a component failure replacement or PM replacement, the first policy is applied, and other components could be replaced if their remaining time to PM is less than a certain predetermined value. The problem is modeled as MDP and the solution was obtained using the temporal difference RL algorithm, named SARSA(λ). The problem was also solved using the renewal reward theory. A simulation was performed to compare and measure the average cost per unit time for the policies obtained using the two solving approaches. A solution based on SARSA(λ) for the second policy was found to conclude with the minimum cost per unit time.

In the previous works, MDPs and RL were applied to optimize the maintenance policy, and RL was shown to outperform the ordinary optimization techniques. However, the proposed maintenance policies are mainly TBM and do not consider the system's condition, which may lead to a failure or unnecessary maintenance, as discussed earlier. Also, they only consider the replacement and do-nothing maintenance action and aim to obtain the replacement intervals based on the system's age.

Liu, Y. et al. (2019) proposed a dynamic selective maintenance approach based on RL. For selective maintenance, aged components are maintained during the system's breaks between the different production tasks. In this work, the joint between production and maintenance is modeled as MDP. The states of the components were defined by their effective age, and either they have failed or not using a Weibull density function. RL was used to solve the problem to overcome the "curse of dimensionality". The objective of the proposed work is to optimize a selective maintenance strategy that maximizes the expected number of successful future multiple consecutive tasks. A relative work was proposed by Kuhnle et al. (2019) to optimize an opportunistic maintenance schedule for parallel working machines by using the Deep-Q-Networks RL algorithm. The breakdown times of the machines were determined using a combination of two Weibull density functions. The opportunity to perform PM for the parallel working machines had been investigated through the proposed work. The maintenance schedule obtained leads to a reduction in downtime, an increase in the production output, and a reduction in the maintenance costs compared to the TBM or CM policies.

Xanthopoulos et al. (2018) used RL to optimize a parametric production-control policy for deteriorating manufacturing systems. The objective of this policy is to minimize the sum of two conflicting objective functions: the average inventory level and the average number of backorders. The deterioration of the system was defined to follow assumed deterioration failure rates related to time. Preventive maintenance action was considered. Its threshold was part of the optimization problem and a corrective maintenance action was applied in the case of failure. The optimal joint production/maintenance control policy was obtained and compared with different parametric production and maintenance policies. The obtained policy based on RL was able to outperform the ordinary policies based on ad hoc. Ling et al. (2018) considered a two-machine flow line system with a buffer zone in between. The failure time of the two machines was taken to follow a Gamma density distribution that decreases as the machine ages. The actions considered are preventive repair, corrective repair, and minimal repair. The optimal maintenance strategy for the whole system is obtained by considering maintenance resource constraints, and the RL algorithm is used to obtain the optimal solution that minimizes the total system average cost rate.

Adsule et al. (2020) proposed a CBM policy with a health index employed to define the deterioration level. The proposed policy considered minor maintenance action in addition to the replacement and do-nothing actions. The health index was a function of time and a deterioration rate that followed a Gaussian distribution, and a RL solution technique was suggested to obtain an optimal maintenance strategy that minimizes the long-run average cost. An illustrative example is proposed to elaborate on the idea without an application to a real problem. A comparable work that was applied to a multi-component system with dependent competing risks was proposed by Zhang, N. and Si (2020).

The maintenance strategies and policies developed in the discussed literature based on RL have shown some advancement compared to other various solutions. Nevertheless, these efforts are oriented towards the maintenance/production control problem and do not focus on the maintenance problem. Also, the systems' deterioration and failures were defined based on either statistical density functions or assumed failure and deterioration rates related to age or time. Only replacement or minimal repair actions were considered and CBM strategies aimed to identify a preventive replacement threshold.

To this end, we can conclude that the CBM demonstrated real advances in optimal maintenance decision-making. In addition, adopting RL helps overcome the need for TPM, as its estimation process is challenging and can not be generalized for different applications. The first limitation that still exists is the assumptions that have been made to model the deterioration. The deterioration is either modeled directly using age, with a model that mainly depends on age, or the deterioration is assumed to follow a certain statistical model. Treating the deterioration in such a way leads to either TBM or to CBM that is limited due to deterioration modeling assumptions. The second limitation is related to maintenance actions that have been considered. Mainly, replacement, do-nothing, and minimal repair actions are considered. However, multi-level preventive repair actions that improve a system's deterioration in different levels are not as extensively considered, even though they have great potential to lead to more cost-effective maintenance strategies. In addition, they are practiced in different applications.

To overcome the aforementioned limitations, we propose a generic data-driven modeling and optimization methodology for a multi-level CBM strategy. The methodology combines a prediction model and the RL method. The data-driven prediction model is proposed to learn a system's deterioration from the available deterioration historical data. This prediction model is then merged into the environment that represents the system to be maintained, with which RL interacts to obtain the optimal maintenance strategy. The originality of this work is based on its ability to tackle real issues in maintenance problems that are normally treated with assumptions. These assumptions are typically related to the shape of the deterioration, its modeling parameters, and possible maintenance actions. Such assumptions limit the applicability on a practical level. The contribution of the proposed methodology and the multi-level CBM strategy is threefold: 1) it addresses the deterioration through a data-driven method that is widely applicable without limiting assumptions. 2) It is different from the traditional CBM strategy that defines a threshold for a preventive replacement action; instead, the proposed multi-level CBM strategy provides mapping from a state to action, as it suggests an action for each state. 3) It considers the potential cost savings that could be achieved by including multi-level preventive repair actions that are practiced in different applications.

The rest of this article is organized as follows. Section 2 provides the proposed methodology for the multi-level CBM strategy. It contains descriptions for the MDP model, the deterioration prediction model, RL methods and their integration. Section 3 presents a case study using real data. Moreover, validation of the results obtained is proposed in section 3. Section 4 concludes the paper and presents future research.

5.2 Proposed Methodology for Multi-level CBM Strategy Optimization

We propose a generic methodology for a multi-level CBM strategy that tackles the limitations related to TPM estimation, the deterioration modeling assumptions, and the limited maintenance actions. The proposed methodology employs an MDP to model the CBM strategy. MDPs are a relevant method for sequential decision-making under uncertainties, where the actions affect subsequent situations and not only the immediate situation. The proposed strategy includes multi-level preventive repair actions that improve the system’s deterioration by different levels, in addition to the ordinary replacement and do-nothing actions. The deterioration process is learned from historical data through a data-driven deterioration prediction model that is capable of learning any type of deterioration without the need for assumptions. The RL is integrated with the deterioration prediction model by merging the prediction model in the environment of the RL, as Figure 5.1 shows.

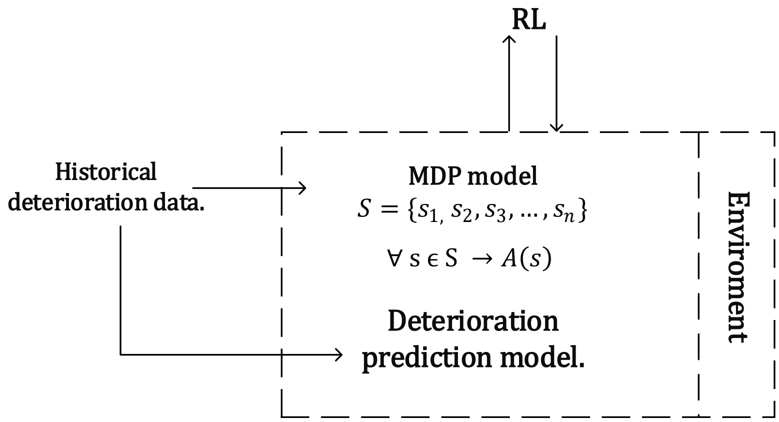


Figure 5.1 Schematic representation for the proposed methodology.

5.3 Maintenance Strategy Description and Model

The proposed multi-level CBM strategy is modeled as MDP and concerns deteriorating systems with an identified and finite state space that are subject to multi-level preventive repairs, preventive and corrective replacements, and do-nothing actions. Each of the multi-level preventive repair actions has a specific effect on the deterioration level and encounters a certain cost. The following points summarize the MDP assumptions and model expressions that are given in the present paper.

- Finite-state space describes the system's states $S = \{s_{1,} s_2, s_3, \dots, s_n\}$ the state s that identifies the system state could be defined by one or more variables, e.g., the system's deterioration level and the system's age.
- The deterioration of the systems is continuous in time, but it could be discretized into multiple levels, as shown in Table 5.1. This discretization process is practical from an engineering point of view, as in a lot of the cases there is no need - or it is not applicable – to consider every value (Alaswad & Yisha, 2017; Zheng & Makis, 2020). Then, at any time t the system's deterioration level falls into any of the categories. The age is continuous, but it could be discretized to a finite state space.
- The deterioration of the system is Markovian, which means that the conditional probability of moving from state i at time t to state j at time $t + 1$ depends only on the current state i; this condition can be defined as:
  $p_{ij} = pr\,\{s_{t+1} = j | s_t = i\} = pr\,\{s_{t+1} = j | s_t = i, s_{t-1,} \dots, s_0\}$.
- The system deterioration level is observed through inspection. The worst state of the system is assumed to be a functional failure observed during the inspection. That means that when the system reaches this failure state it will function inefficiently, which also encompasses a high risk in terms of safety. This is considered to be a functional failure.

- At each state $s \epsilon S$ a set of different actions $A(s)$ could be performed. Multi-level preventive repairs, preventive and corrective replacement, and do-nothing actions can be performed in the proposed CBM strategy.
- Each action has a cost $C(s, a)$. The cost is related to both the action and the initial state, as Table 5.2 shows.
- The last component of the MDP is the transition probability matrix. The elements of the matrix can be defined as $p_{ij} = pr\{s_{t+1} = j | s_t = i, a_t = a\}$. The transition probability is based on both the state and the action performed. This transition probability matrix is not needed in this work, since it adopts the RL method to solve the problem.

Table 5.1 Discretization of the system's deterioration.

Index (s)	**Deterioration level**	**Description**
A	Perfect	New state.
B	Very good	Damage initiation.
C	Good	Minor damage.
D	Acceptable	Moderate damage.
E	Bad	Major damage affects safety and functionality (functional failure)

Table 5.2 Different possible actions with the associated costs.

Action (a)	**Description of actions' effects**	**Actions' cost function $C(s, a)$**
Do-nothing	Do-nothing action, the system will continue to deteriorate.	0
Preventive repair 1	Improve the deterioration by one level.	CR_a^d is the repair cost for $a^{th} = 1,2,3$ level improvement at $deterioration\ level\ d = \{A, .., D$
Preventive repair 2	Improve the deterioration by two levels.	
Preventive repair 3	Improve the deterioration by three levels.	
Preventive repair 4	Improve the deterioration by four levels.	
Preventive replace	Preventive replacement action.	C_r is the preventive replacement cost.
Corrective replace	Corrective replacement after failure	C_c is the corrective replacement cost.

5.3.1 Deterioration Prediction and RL

Solving by interaction is the main characteristic of the RL method. The interaction occurs between the RL agent, which is based on the SRARA (State-Action-Reward-State-Action) algorithm that takes a maintenance action based on a given system's state and the environment returns the predicted new system state and its reward in terms of the maintenance cost, as illustrated in Figure 5.2.

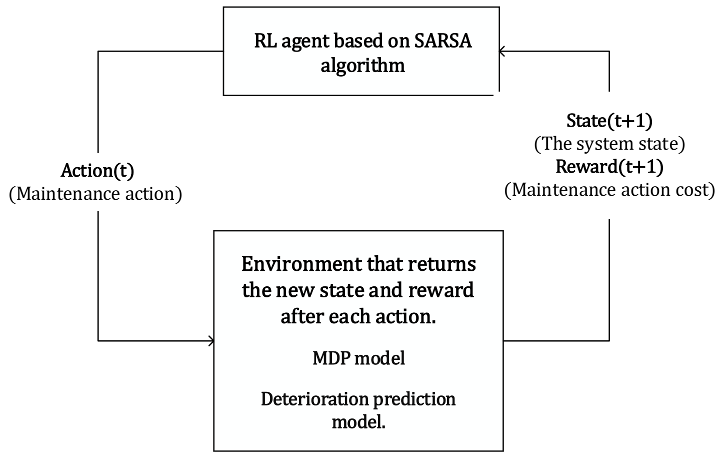


Figure 5.2 RL agent learning process.

Specifically, the environment represents the system to be maintained. The regulations that regulate the environment dynamics rely on the system's deterioration and its improvements over time. They take place as a result of the effects of the different maintenance actions as shown in Table 5.2. The system deterioration results from natural aging and from leaving the system without maintenance. The deterioration of the same system follows different paths and shapes depending on the conditions of use, age, and previous maintenance actions. Defining the progress of the system's degradation under different conditions is a challenging issue. To address this issue, the data-driven prediction model is used to predict the system deterioration level. The system's age is part of the information provided to the prediction model to ensure that it is capable of capturing the increase in failure risk that could take place over time as its age increases. This prediction model is merged in the environment to represent the dynamics of the system's deterioration. Figure 5.3 shows a schematic representation of the proposed deterioration prediction model. Different data-driven prediction models are convenient with the proposed methodology. The random forests technique is adopted in the proposed case study in section 5.4 to predict the state of the system over the age, as it is capable of performing linear and nonlinear prediction tasks with high accuracy and moderate amounts of data (Carvalho et al. 2019; Falamarzi et al. 2019).

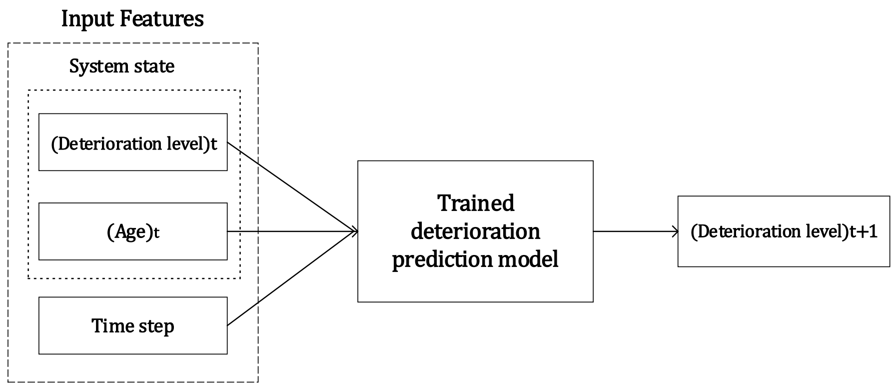


Figure 5.3 Deterioration prediction model.

This work adopts the temporal-difference (TD) RL agent, specifically based on the SARSA algorithm. SARSA is a TD RL algorithm that provides fast convergence without being computationally expensive (Sewak, 2019). Figure 5.4 shows the adopted agent based on the SARSA algorithm, where $Q(s,a)$ is the value function for a state action pair, $r(s,a)$ is the reward obtained after choosing action a in state s, α is the learning rate, γ is the discounting factor, and ε is the threshold for following a greedy policy. The optimal maintenance strategy is obtained through the interaction of the RL agent with its environment.

RL agent based on SARSA algorithm

Initialize arbitrary values $Q(S,A)\ \forall\ s = (\text{Deterioration Level}, age)\ \in\ S, a\ \in\ A(s)$
while $n \leq number\ of$ episods
Choose starting state $s = (\text{Deterioration Level}, age)\ \in\ S$
Choose $action\ a\ \in\ A(s)$ following ε-greedy approach.
Get the reward $r_{t+1} = r(s,a) = action\ cost$ and observe the next state $s^{'}$
Update the state-action pairs value function:
$Q(s,a) = Q(s,a) +\ \alpha\ [r_{t+1} +\ \gamma Q(s^{'},a^{'}) - Q(s,a)]$
Update: $s = s^{'}, a = a^{'}$
Continue until termination takes place either by failure or preventive replacement
$n = n + 1$

α: learning rate. γ: discounting factor. ε: threshold for acting greedy.

Figure 5.4 RL agent adopted to the maintenance problem.

5.4 Case Study

The purpose of this section is to apply the proposed methodology to a real case study to prove its applicability and to validate the solution obtained with it. This case study deals with sheet metal culverts' maintenance problem and uses real data about the deterioration and maintenance of the sheet metal culverts. The optimal multi-level CBM strategy for the sheet metal culverts that minimizes the maintenance costs over a finite time horizon is derived and validated through a comparison with other strategies; among them are well-known corrective and preventive maintenance strategies.

5.4.1 MDP States, Actions, and Rewards

In this case study, the culvert deterioration level is defined using a performance index that combines different failure mode measurements obtained through a regulatory inspection process. The performance index takes a continuous value between 0, which depicts the worst condition of deterioration, and 100, which denotes the best condition of the deterioration, meaning no deterioration has taken place. This continuous value is discretized into five levels by the domain experts, as per the example that was shown in Table 5.1. The deterioration levels, and the age of the culvert, are used to identify the culvert's state for the MDP model. From the available data, culverts can have a maximum age of 60 years, and then the age span to be considered is between 0 and 60 years. At the age of 0, the culvert is new, which means it does not need any type of maintenance. The culvert state can be any of the deterioration levels at any age between 10 and 60, with increments of 10 years, which is a reasonable step size to use to update the maintenance strategy for such applications. The different combinations between the deterioration level and age result in 30 different states. The failure is defined by either exceeding the age of 60 years or reaching the deterioration level E. Given that the age affects the deterioration process, the choice of maintenance actions is affected by the age of the systems. For systems with the same deterioration level and different ages, different maintenance actions are optimal. For each state, a set of actions can be performed. There are 84 state-action pairs, as shown in Figure 5.5. Figure 5.5 also shows different maintenance action costs. The cost of multi-level preventive repairs per unit of length ($/u.L) is a function of improvement level, deterioration level, and age. The corrective replacement maintenance action after failure cost is three times the preventive replacement. The high cost of a corrective replacement penalizes failures.

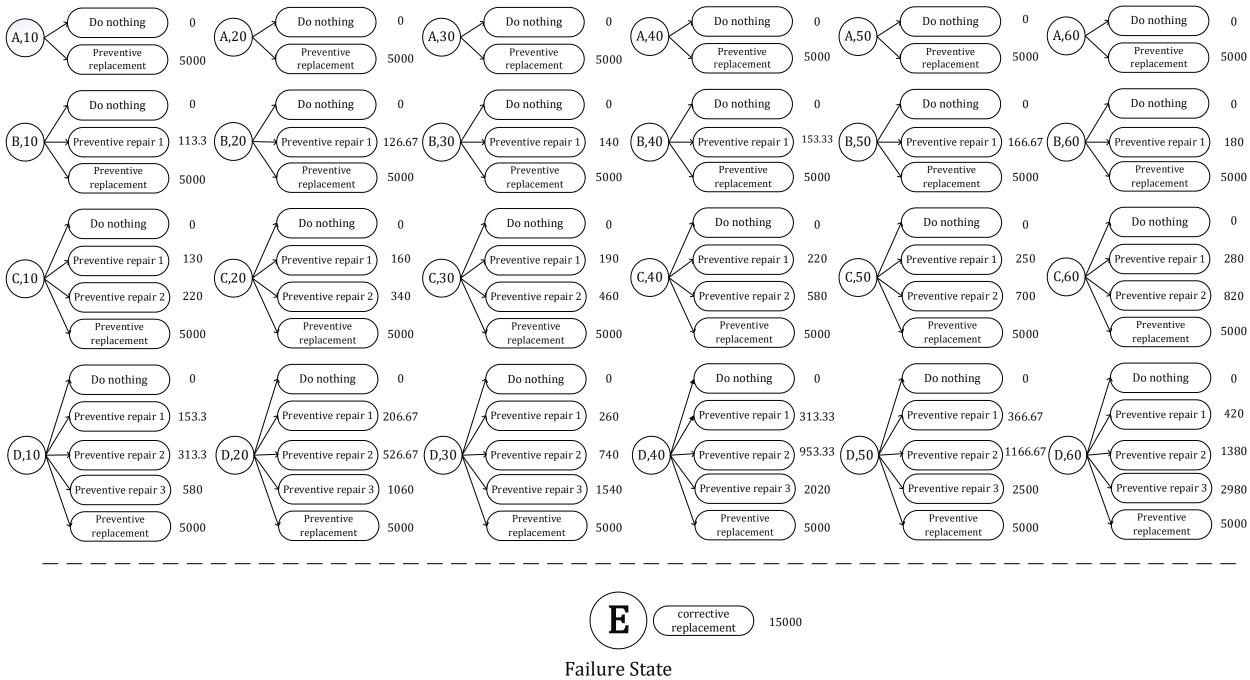


Figure 5.5 State action pairs according to deterioration level and age.

To summarize, the state is defined by deterioration level and age, while the reward of an action is the maintenance action cost. Then, the effects of the maintenance actions and the system deterioration process need to be defined. The effects of the maintenance actions are defined in Table 5.2 and a deterioration prediction model is trained using the historical deterioration data.

5.4.2 Data-driven Deterioration Prediction Model

The random forest algorithm is adopted for the deterioration prediction model. Two random forest prediction models are built. In the first model, the categorical values of the deterioration index are used. In the second model, the continuous values of the deterioration index are used. The two models are built to compare their performance and to assess the information that may be lost as a result of this categorization.

The random forest has hyperparameters that need to be tuned to maximize the accuracy of the model. These parameters are the number of trees, the criterion to measure the quality of the split, the number of features to consider in order to obtain the best split, and the maximum depth of the tree. The number of trees defines how many trees are used to build the random forest model. The splitting criterion is a function based on it the decision tree algorithm decides how to make a split on certain input features to create a node. The feature is selected each time using a comparison criterion for the different features based on the split. This comparison is not done between all of the input features, but is done between a subset of the input features. The subset size is a parameter that needs to be optimized. The maximum depth of the tree is the number of layers on the longest bath between the tree's root and a leaf. More details about the random forest hyperparameters can be found in (Lior, 2014) (Pedregosa et al., 2011). Grid search is used to obtain the hyperparameters that maximize the accuracy of the model. It is an exhaustive concept, as it tries every combination. In our case, this type of search is possible, since the data size of 2665 observations is adequate for this type of search. To obtain the optimized hyperparameters, the random forest prediction model is trained, and the grid search is applied.

Tables 5.3 and 5.4 show the selected hyperparameters for discrete and continuous deterioration levels, respectively. Using the hyperparameters defined by the grid search, the two prediction models are trained. For the discrete prediction model, the training accuracy is 82.7%, and the testing accuracy is 74%. For the continuous prediction model, training accuracy is 91.3% and testing accuracy is 90.4%. Notably, the performance of the continuous deterioration model is better than the performance of the discrete deterioration model by 15.6%. This performance is expected in the discrete model, deterioration can only be five levels, which leads to the loss of some of the information that is found in the continuous model. This difference in performance is reasonable because discretization can lead to the loss of much more information. The discrete model is used in this study since the discretization is done by domain experts and the performance of the discrete model is still acceptable.

Table 5.3 Selected hyperparameters for a discrete deterioration prediction model.

Parameter	**Available values**	**Selected values for Discrete deterioration model**
Number of trees	Integer: optional, can be any value. The defined search space is [100, 250, 500, 750, 1000]	500
Criterion	"Gini" impurity factor or "Entropy" for the information gained	Gini
Max depth	Integer: optional, can be any value. The defined search space is [3, 5, 7, 9, 11, 13]	9
Number of features	Square root (Number of features), Log2(Number of features), or Number of features	Number of features

Table 5.4 Selected hyperparameters for the continuous deterioration prediction model.

Parameter	**Available values**	**Selected values for Continuous deterioration model.**
Number of trees	Integer: optional can take any value. The defined search space is [100, 250, 500, 750, 1000]	250
Criterion	"mse" mean square error or "mae" mean absolute error.	mse
Max depth	Integer: optional can take any value. The defined search space is [3, 5, 7, 9, 11, 13]	9
Number of features	Square root (Number of features), Log2(Number of features), or Number of features	Square root (Number of features)

5.4.3 RL Agent

The ε-greedy method is followed to enhance the exploration during the training process of the RL agent. Additionally, the value function $Q(s, a)$ has an arbitrary optimistic starting value of 10 to encourage the exploration process. Based on the ε-greedy approach at each state s, an action a that appears to be optimal is chosen by a probability of $1 - \varepsilon + \frac{\varepsilon}{|A(s)|}$ and each of the other actions is chosen with a probability of $\frac{\varepsilon}{|A(s)|}$. The selection of exploration rate ε, learning rate α, and discounting factor γ is a crucial point to consider. The values of these parameters are selected by the researcher in most cases based on the problem type and previous experience. To ensure an optimal selection of values, two approaches are followed in our work. The first approach is applied to the discounting factor γ. The optimal selection of the discounting factor is based on the fact that in episodic tasks, γ is proven to be effective when it takes a value near 1 (Pitis, 2019). Based on this conclusion, γ is chosen to be 0.9 in the present application. Since for ε and α there is no clear recommendation in the literature, a systematic search is done, as shown in Figure 5.6. Both ε and α take values between 0 and 1 but a very large value for ε risks convergence. Fig. 6 shows a change in the average cost per unit length when α takes different values between 0 and 1, and ε = 0.01, 0.05, 0.1, 0.15, 0.2, 0.3, and 0.4, respectively. The minimum average cost is obtained at ε=0.05 and α=0.9. The optimal maintenance strategy is shown in Table 5.5.

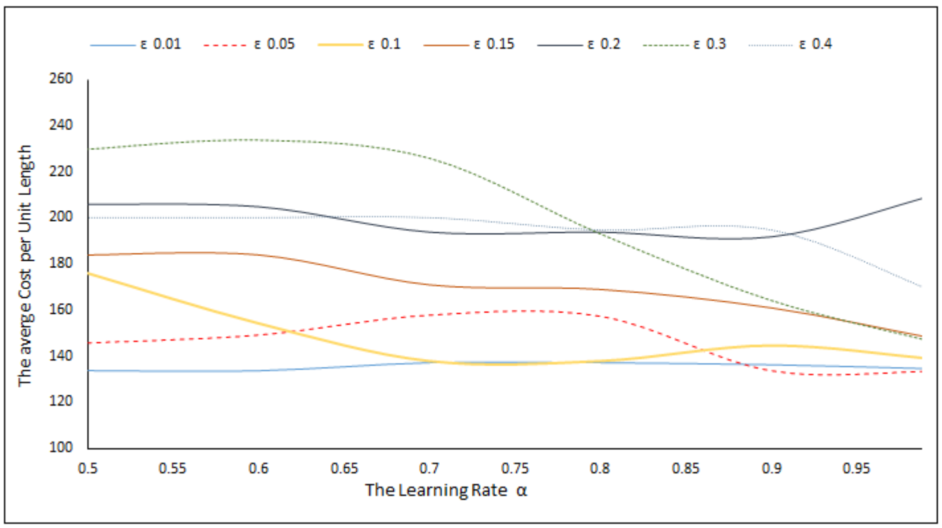


Figure 5.6 Optimal Values for Learning and Exploration Rates. The SARSA algorithm is trained using 5000 episodes. The average cost per unit of length is obtained by averaging the maintenance cost over its age, starting from all of the available states.

Table 5.5 Optimal maintenance strategy.

State	Action	State	Action	State	Action	State	Action	State	Action	State	Action
(A,10)	Do nothing	(A,20)	Do nothing	(A,30)	Do nothing	(A,40)	Do nothing	(A,50)	Do nothing	(A,60)	Preventive replacement
(B,10)	Do nothing	(B,20)	Do nothing	(B,30)	Repair 1	(B,40)	Repair 1	(B,50)	Do nothing	(B,60)	Preventive replacement
(C,10)	Repair 1	(C,20)	Do nothing	(C,30)	Repair 1	(C,40)	Repair 1	(C,50)	Repair 1	(C,60)	Preventive replacement
(D,10)	Repair 1	(D,20)	Repair 1	(D,30)	Repair 1	(D,40)	Repair 1	(D,50)	Repair 1	(D,60)	Preventive replacement
E: Corrective replacement											

The results obtained demonstrate that the optimal actions of the strategy are not always the ones with the lowest cost. This is due to the immediate and delayed reward trade-off. For the immediate reward, some actions may appear to be less costly at present but performing these actions will lead to higher costs in the future. RL treats this trade-off to find an optimal strategy. Moreover, the results confirm that age affects the optimal selection of an action. The optimality of this strategy is investigated in the following section.

5.4.4 Validation Approach

The form of the obtained solution is a multi-level CBM strategy that defines the optimal action for each state of the system. The obtained strategy is supposed to minimize the average maintenance cost over its age. Validating the optimality of the obtained solution through the proposed methodology is an important issue to consider. Validation is performed by comparing the average maintenance cost over age for different strategies. This validation is conducted using a simulation, as shown in Figure 5.7.

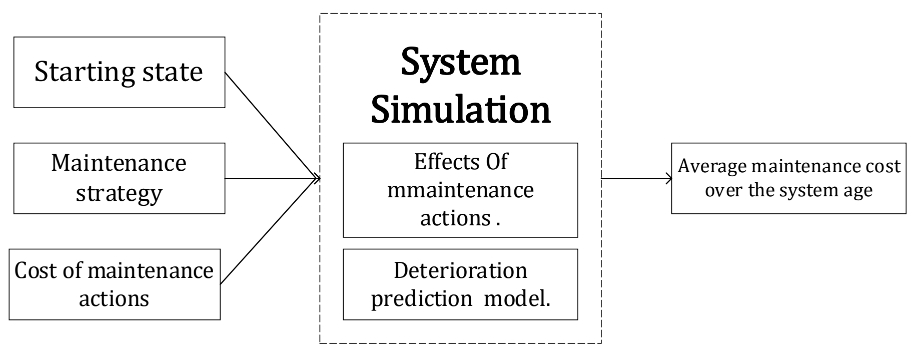


Figure 5.7 System simulation that compares different maintenance strategies.

The inputs for the simulation are the maintenance strategy and the cost of the actions. The simulation of the system includes the system deterioration prediction model and the effects of the maintenance actions. The output is the average maintenance cost over the system's age. The simulation is done by selecting a starting state; this starting state can be any of the system's states. Starting from this state, a maintenance strategy is followed until the system fails and corrective replacement is needed, or until a preventive replacement action takes place. The simulation keeps track of the costs of the actions performed, then the average cost over the system's age is calculated.

To prove the optimality of the obtained strategy, it is compared against two well-known strategies, in addition to evaluating all other possible strategies. The two well-known strategies are: 1) the CM strategy, which is comprised of corrective replacement and do-nothing actions only, and 2) the ordinary PM strategy, which considers both preventive and corrective replacements and a do-nothing action. 3) Finally, all of the strategies combine multi-level preventive repair actions, preventive and corrective replacements, and a do-nothing action. The first two strategies included in the comparison CM, and ordinary PM, are obtained by training our model with actions that are considered in each strategy. A brute force method is adopted to obtain all of the possible strategies that also consider multi-level preventive repair actions, preventive and corrective replacements, and a do-nothing action. The role of the brute force method is to try all of the possible combinations of actions that exist. In general, the brute force method is not efficient. In this case, it will confirm the optimality of the solution obtained by comparing it against all of the existing solutions. Furthermore, the proposed methodology holds for larger state spaces, in which the application of the brute force approach is difficult.

5.4.5 Validation Results

Two different states from the state space are selected as starting states to perform the comparison. The two starting states are selected randomly. Any state can be used to perform the validation, as the obtained solution is supposed to be optimal starting from any state. The two starting states selected to perform the comparison are (new state: deterioration level is A and the age is 0) and (degraded state: deterioration level is B and age is 30). Any other states can be selected to perform the test; these two points have been selected as an example.

For the first starting state (new state: deterioration level is A and the age is 0): 1) CM strategy takes a replacement action when failure is observed. 2) Ordinary PM strategy is defined as the optimal threshold for preventive replacement actions to be at the state, as defined by deterioration level C and an age of 40 years. 3) The brute force method results in 64 different maintenance strategies - other than the one obtained - that can be followed starting from this point. The CM strategy, the ordinary PM strategy, and the 64 other strategies are introduced into the simulation and the average costs are calculated. Table 5.6 shows the results.

Table 5.6 Simulation results that compare the average cost of different maintenance strategies starting from the new state.

Strategy	**Average cost per unit of length over the age**
Obtained by the model	89.06
CM strategy	300
Ordinary PM strategy	125
Best strategy of 64 strategies	89.11
Worst strategy of 64 strategies	500

From Table 5.6, it is clear that the strategy obtained through the proposed methodology is the optimal one. Also, it is noteworthy that a sub-optimal strategy does exist in strategies obtained through the brute force approach, but the obtained strategy is the optimal one.

For the second starting state (degraded state: deterioration level is B and the age is 30): 1) the CM strategy makes a replacement action when a failure is observed. 2) An ordinary PM strategy is defined the optimal threshold for preventive replacement action, which is the same as the one defined earlier. 3) The brute force approach results in 30 different maintenance strategies, other than the one obtained, that can be followed starting from this point. The CM strategy, the ordinary PM strategy, and the 30 other strategies are introduced to the simulation and the average costs are calculated. Table 5.7 illustrates the results. In this case, also, the proposed methodology is capable of obtaining the optimal strategy.

Table 5.7 Simulation results comparing the average cost of different maintenance strategies starting from a degraded state.

Strategy	**Average cost per unit of length over the age**
Obtained by the model	89.83
CM strategy	300
Ordinary PM strategy	125
Best strategy of 30 strategies	92.72
Worst strategy of 30 strategies	375

Figure 5.8 presents a comparison between the optimal, sub-optimal, and worst maintenance strategies. The comparison is performed by running the simulation based on each strategy. The simulation should be run for at least a complete life cycle, so that maintenance strategies reach a steady state. The run time for the simulation is selected as three times the maximum culvert age, with a step size of 10 years for demonstrative reasons. Figure 5.8 (A) shows the results for the first starting state (the new state: the deterioration level is A and the age is 0) and Figure 5.8 (B) shows the results for the second starting state (degraded state: deterioration level is B and the age is 30).

The results presented in Table 5.6, Table 5.7, and Figure 5.8 confirm two main points: 1) the obtained multi-level CBM through the proposed methodology is the optimal one and outperforms all other strategies. 2) Considering the multi-level preventive repair actions, aside from the ordinary replacement do-nothing actions in the proposed strategy, leads to minimizing the average maintenance cost more than considering only the replacement and do-nothing actions.

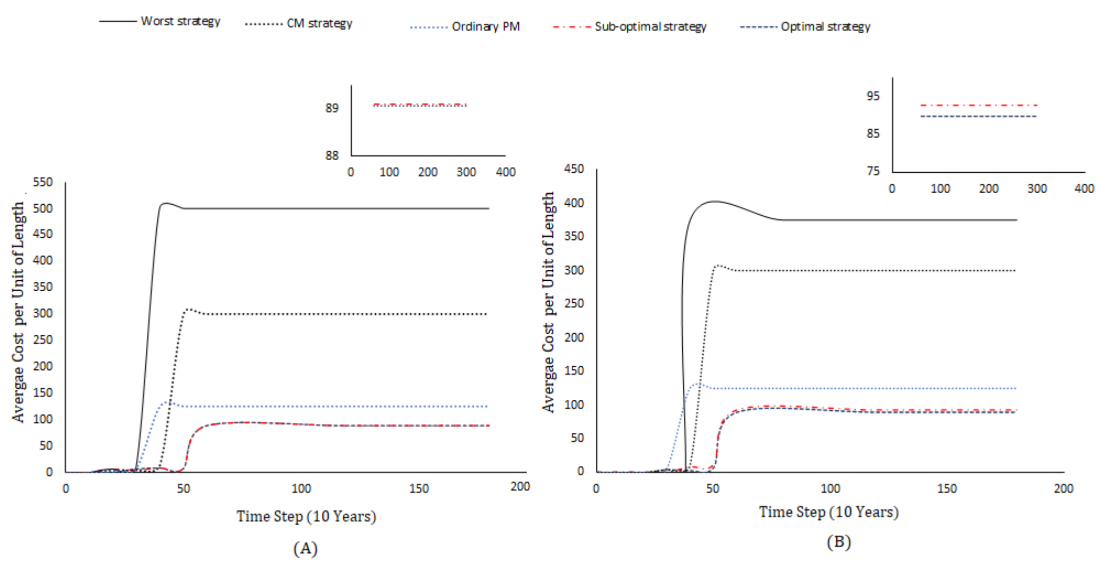


Figure 5.8 Comparison between the optimal, sub-optimal, and worst maintenance strategies when applied for 180 years. The comparison is in terms of the average cost per unit of length over age, starting from a new state (A) and degraded state (B).

5.5 Conclusion

Condition-based Maintenance (CBM) strategies based on Markov Decision Process (MDP) show advances in the optimization of the maintenance problem, yet they still experience challenging limitations. When Dynamic Programming (DP) or Linear Programming (LP) is adopted to solve these optimization problems, estimating the Transition Probability Matrix (TPM) and the high computational cost are the main limitations. In an attempt to overcome these limitations, Reinforcement Learning (RL) is employed to solve such problems. However, in this case, the main issue was the maintenance/production control problem. In addition, the deterioration is addressed based on either statistical density functions or assumed failure and deterioration rates related to age or time. Addressing the deterioration in such a way either returns the problem to being a TBM, does not provide sufficient representation for the deterioration, or limits the ability to transfer it to other cases. Besides, in most of the CBM strategies, multi-level preventive repair actions have not been explored enough to examine their efficiency.

This paper presents a generic data-driven methodology to optimize multi-level CBM strategies that tackle these challenges. The proposed methodology integrates a RL method and a prediction model. This methodology does not suffer the discussed limitations of developing or solving MDP based models. The deterioration of the system is modeled using a data-driven prediction model. The data-driven deterioration prediction model can learn the real deterioration process from the historical deterioration data. It can learn and model any deterioration process without any assumptions. This prediction model is merged in the environment that represents the system. Besides the widely considered replacement maintenance and do-nothing actions, multi-level preventive repair actions are considered in the CBM strategy. RL is adopted to solve the optimization problem by interacting with the environment, which comprises the data-driven deterioration prediction model.

A real case study on sheet metal culverts is proposed to verify the optimality and applicability of the proposed methodology. For the culvert system, the optimal maintenance strategy is obtained based on the proposed methodology. Then, it is compared to various other strategies to examine its optimality. The obtained strategy is compared against 1) CM strategy, which takes a replacement action when a failure is noticed, 2) an ordinary PM strategy that is optimized to obtain the preventive replacement threshold, and 3) all of the possible strategies that consider multi-level preventive repair actions, preventive, corrective replacement, and do-nothing actions. A brute force approach is used to obtain the last set of strategies. The comparison is done based on average maintenance cost over age. It has been proven that the solution obtained is optimal and leads to the minimum average maintenance cost over age. Also, it is worth noting that considering multi-level preventive repair actions leads to minimal average maintenance costs. It is important to emphasize that the obtained solution is optimal from any starting state in the system state space. Thus, if the system is left without maintenance for certain interval or was maintained inefficiently and has reached a degraded state, this approach is still applicable.

This proposed methodology applies to any system or application as long as historical data and maintenance information are available. The availability of the deterioration data is not a limitation, as this type of data is widely available due to recent advances in data acquisition and storage. Furthermore, the state size does not cause any limitations. The case study proposes a moderate state space size to allow for validation and interpretation of results. For further research, it is of interest to examine the effect of using different data-driven prediction algorithms to address deterioration. Moreover, the effect of imperfect maintenance and the interaction between the different failure modes of the system should be studied.

CHAPTER 6 ARTICLE 3: A DATA-DRIVEN METHODOLOGY WITH A NONPARAMETRIC RELIABILITY METHOD FOR OPTIMAL CONDITION-BASED MAINTENANCE STRATEGIES

Submitted at Reliability Engineering & System Safety journal

Manuscript Reference No.: JRESS-S-22-00435

Authors: Mina Mikhail, Soumaya Yacout, Mohamed-Salah Ouali

Abstrt

Condition-based maintenance (CBM) strategies have been receiving increasing attention due to their advances in terms of avoiding failures and decreasing maintenance costs. This paper proposes a CBM optimization methodology that adopts the remaining useful life (RUL) as a part of its optimization criterion. The proposed methodology combines data-driven prediction model and reinforcement learning (RL) method with a nonparametric survival model, called Kaplan-Meier (KM) product limit. The prediction model learns the system's deterioration process from historical data. The RL model integrates the prediction model with a customized reward function to fulfill the objective of the CBM strategy that simultaneously maximizes the system's RUL and minimizes the cost of maintenance through selecting the proper maintenance actions. The RUL is estimated based on the appropriate KM survival curve, which represents the time at which the system reaches a predetermined deterioration level. Both the prediction model and the KM model are generic and are not based on certain assumptions. The CBM decision-making problem that enables preventive and corrective replacements besides multi-level preventive repairs at periodic intervals is modeled using a Markov decision process (MDP). Each of the considered maintenance actions restores the system's condition to a certain deterioration level. The RL is used to obtain the optimal maintenance strategy without the need for the transition probability matrix (TPM) of the MDP. Contrary to the widely proposed threshold based traditional CBM strategies, the proposed methodology provides an optimal CBM strategy that maps each system state to the appropriate maintenance action that maximizes the RUL over the maintenance cost. A case study dealing with sheet metal culverts systems is used to demonstrate the relevance and the cost efficiency of the obtained CBM strategy. The obtained CBM strategy through the proposed methodology is found to be more cost-efficient when it is compared to widely proposed baseline strategies. An important

quality to highlight is the generic character of this methodology as it applies to a wide range of problems without depending on certain statistical assumptions.

6.1 Introduction

A maintenance strategy is a decision-making criterion that maintains or restores a system at a given performance level related to its reliability, availability, maintainability, and safety according to a maintenance budget and expected services (Han et al., 2020; Jardine et al., 2006). Cost, downtime, and reliability, or a combination of the three, are mostly considered to be optimization criteria of maintenance strategies (Hongzhou 2002; Fan et al. 2019). Age-based maintenance (ABM) and condition-based maintenance (CBM) are the two main preventive maintenance strategies (Zhang and Yang 2015; Yang et al. 2017). In the case of ABM, maintenance decisions are based on the usage interval or age, regardless of the deterioration level or the performance level of the system (Shafiee and Finkelstein 2015). On the other hand, CBM selects a maintenance action based on the system's condition as defined by the deterioration level or performance level. Among the different maintenance categories, CBM has obtained significant attention because it considers both safety and cost by avoiding unnecessary maintenance activities that could be introduced by ABM (Bousdekis et al. 2018; Shi et al. 2020).

A large variety of CBM strategies employ Markov decision processes (MDPs) to model the maintenance decision-making problem. According to Braga and Andrade (2019), MDPs are ideal structures for CBM models since they do not assume certain deterioration paths or shapes and they involve the process of sequential decision-making under uncertainty. Mandiartha et al. (2017) proposed a CBM optimization model for pavement. The problem is modeled as an MDP and solved using dual linear programming (LP) formulation to obtain a maintenance strategy that minimizes the total average cost. Braga and Andrade (2019) proposed a CBM decision model for railway wheelsets using MDP that was solved using dynamic programming (DP) to minimize the total maintenance cost over the life cycle. Sancho et al. (2021) adopted a similar approach for rail tracks and the proposed CBM planning aimed to minimize the cost over an infinite time horizon. Similar CBM strategies are proposed for different applications such as wind turbines, machining, and manufacturing machines (Li et al. 2017; J. Wang et al. 2019; Rui and Makis 2020). An extensive literature review is proposed by Alaswad and Yisha (2017) for interested readers. Despite the encouraging results obtained in the reported literature, it is still based on a strong limitation of the need for the transition probability matrix (TPM) to solve the proposed models using LP or DP. Various authors have reported that estimation of the TPM has been found to be challenging and contains steps that cannot be generalized (Mandiartha et al. 2017; Braga and Andrade 2019; Sancho et al. 2021). Also, it is worth noting that most of the optimization criteria are related to minimizing costs.

Reinforcement learning (RL) methods are adopted to address challenges related to the estimation of the TPM for maintenance problems. RL is capable of solving problems modeled as MDP without the need for the TPM (Wiering and Van Otterlo (Eds.) 2012; Sutton and Barto 2018); RL only needs episodes of data. The episodes are sequential tuples of state, action, and reward. Adsule et al. (2020) proposed a CBM strategy that minimizes the long-run average cost. The deterioration is defined using a health index that is modeled as linear growth. The solution approach based on RL is explained using a small illustrative example with assumed parameters, but not applied to a case study. A CBM strategy for a multi-component system with competing risks and a cost minimization objective is proposed by Zhang and Si (2020). The deterioration of the system components is represented using either the compound Poisson process or the gamma process. Two numerical examples are used to demonstrate and verify the optimality of the obtained solution. The RL method has also shown promising results for optimizing maintenance/production control policy with either CBM or TBM (Huang et al., 2020; Kuhnle et al., 2019; Ling et al., 2018; Wang, X. et al., 2016). In such work, the maintenance problem is constrained by certain production requirements.

From the literature discussed in the previous paragraphs, the major similarity between most existing CBM research is the aim to minimize the maintenance cost by considering a system's conditions to eliminate both unneeded maintenance and failure. The remaining useful life (RUL) is a key point that helps maintenance managers achieve more cost-effective maintenance strategies. Predicting or estimating the RUL based on degradation data could be accomplished through statistical data-based or machine learning methods (J. Guo et al. 2020). These methods are found to be accurate and effective (Si et al. 2017). A review of recent studies discusses such methods can be found in (Bousdekis et al. 2018; J. Guo et al. 2020).

Zhen et al. (2019) proposed RUL prediction method using a hidden Markov model with auto-correlation to minimize the cost per unit time of a CBM strategy. At each inspection time, the RUL is predicted to make sure that the system could survive until the next inspection takes place. Based on the RUL prediction, a PM threshold is obtained. It yields a lower average maintenance cost than the one obtained without considering the RUL estimation. (Du et al. 2020) proposed a CBM strategy that optimizes the inspection interval based on the RUL prediction. The strategy considered two types of degradation: natural degradation and random shock. RUL estimation allowed the inspection intervals and the PM threshold that minimize the long-term expected maintenance cost to be obtained. Huynh et al. (2019) proposed a methodology for a predictive maintenance decision framework for a deteriorating system subject to inspection and replacement. In the proposed methodology, the RUL estimation is used to decide whether an irregular/extra inspection is needed, in addition to planned inspections, to avoid potential failure before the next planned inspection. The proposed methodology with a flexible inspection option demonstrated improvements in terms of cost reduction over the traditional fixed inspections. A comparable approach that optimizes both the inspection interval and the PM threshold is proposed by Huynh (2021). Other examples can be found in (Deutsch, J. & He, 2018; Deutsch, Jason et al., 2017; Elsheikh et al., 2019; Guo, J. et al., 2018; Huynh et al., 2014; Huynh et al., 2019; Salih, 2020). From the literature discussed, it is remarkable that most of the existing literature has used RUL to update the inspection intervals or to obtain an optimized preventive replacement threshold.

This paper proposes a novel methodology to optimize CBM strategies. The methodology merges a data-driven deterioration prediction model, RL method, and nonparametric survival techniques. The CBM strategy is modeled using MDP and includes multi-level preventive repair actions that improve the deterioration level by different levels in addition to preventive replacement, corrective replacement, and do-nothing. To overcome the limitations related to the TPM estimation for the MDP, the RL method is adopted. The deterioration is learned through the adopted data-driven prediction model, then it is integrated into the RL environment. The reward function for the RL is designed to incorporate both the RUL and the maintenance cost simultaneously. The RUL estimation is based on a reliability approach that employs Kaplan-Meier (KM) as a nonparametric survival technique.

The contribution of this work is threefold:

- The deterioration phenomenon is addressed through a generic practical model that is not limited to certain applications and does not assume certain parameters or a certain shape for the deterioration, as is the case in most literature.
- A new design for the reward function that incorporates RUL and the cost is proposed. This reward function provides the RL with better feedback about the effects of the maintenance actions on the system. That helps in obtaining more cost-effective strategies. The proposed method for RUL estimation uses a nonparametric model with a reliability-based approach. The nonparametric model namely KM is selected as it does not rely on certain assumptions.
- The multi-level preventive repair actions that are mostly overlooked are considered within the CBM strategy. Such actions restore the system's condition to a previous deterioration level and they lead to a more cost-efficient CBM strategy.

Based on the actions considered and the proposed methodology, the CBM strategy that is obtained provides a map from a condition or state to action. This means that at each deterioration condition, an optimal maintenance action is provided and not only a threshold for preventive replacement.

The remainder of this article is structured as follows. Section 6.2 covers the details of the proposed methodology, the MDP model for the CBM strategy and the combination of the RL, the nonparametric survival techniques KM method for RUL estimation, and the data-driven prediction model. Section 6.3 presents a case study based on real data for identical sheet metal culverts. The proposed methodology is applied to the data of the culverts to obtain an optimal CBM strategy. The results, interpretations, and validation of the results are also discussed in section 6.3. Section 6.4 summarizes the contribution, the findings, and introduces areas of research for future consideration.

6.2 Proposed Methodology

A schematic showing the main procedure of the proposed methodology is presented in Figure 6.1. Using the historical data the states and actions of the MDP are developed. The deterioration is learned by the data-driven deterioration prediction model that is based on Random Forest (RF) algorithm. The survival curves corresponding to the deterioration levels are constructed using the

KM method. To construct the KM curves, different levels of deterioration are defined and the times to reach each level are used to construct the corresponding KM curve. The MDP, deterioration prediction model, and the KM curves are integrated into the environment that stands for the system to be maintained. The RL agent adapted to our problem then interacts with the environment and the final output is optimal CBM strategy inform of optimal action a^* for each state $s_i \in S$ that is defined through both the deterioration level and the system age. The details are presented in the following two sections 6.2.1 and 6.2.2.

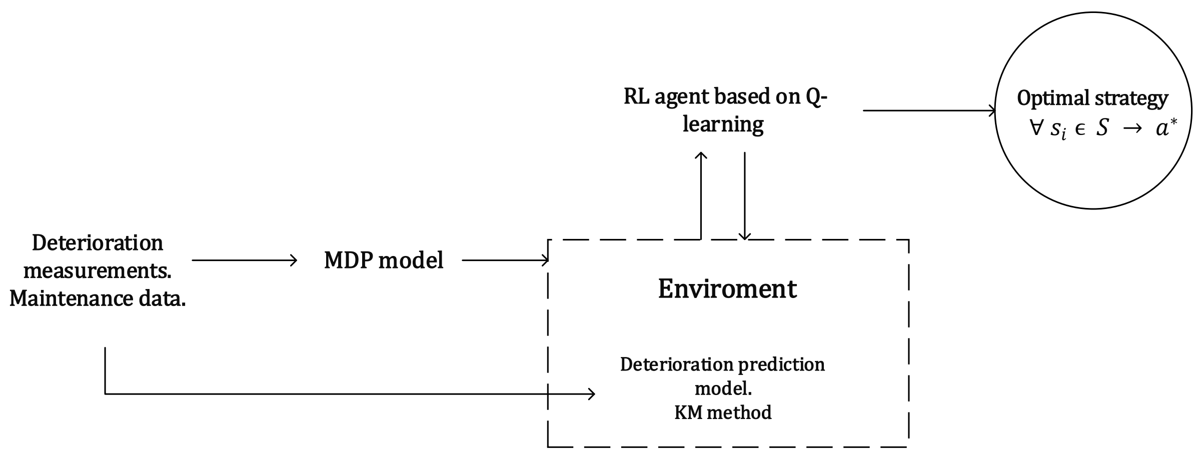


Figure 6.1 Main procedure for the methodology.

6.2.1 MDP Model

The MDP model generally has four main elements to be defined: the state space, the action space, the rewards, and finally, the TPM between the states. The states of the system to be maintained are defined by the deterioration level and age. The system's continuous deterioration condition is discretized into N number of deterioration levels (DLs) $\{D_1, D_2, D_3, \dots, D_N\}$ based on the application. Then, the state-space S has n discrete finite number of states $S = \{s_{1,} s_2, s_i, \dots, s_n\}$, each state $s_i = (\text{DL}, age)$. The failure of the system is defined by a certain DL; when the system reaches this DL, it is considered to have failed. The action space consists of m different actions for each state $A(s_i) = \{a_0, a_1, a_j, \dots, a_m\}$. The action space includes preventive and corrective replacements, do-nothing, and multi-level preventive repair actions. The multi-level preventive repair actions improve the system's DL. The possibility of the multi-level preventive repair actions is based on the DL; e.g., if the DL is "as good as new," no repair actions are possible. The execution of each action results in a certain reward. The reward function is designed to meet the CBM strategy's objective of maximizing the system's RUL while keeping the maintenance cost at a low level. The reward function $r_t(s_i, a_j) = \mathrm{RUL}_t / C(s_i, a_j)$, where RUL is estimated based on the system's deterioration level after the maintenance activity at age t and $C(s_i, a_j)$ is the maintenance action cost. The reason for this reward function design is to provide the RL with enough feedback about the effect of the different actions on the RUL together with their associated costs. Both the RUL and the maintenance cost are important factors to consider while deciding which maintenance action to perform. This reward function provides this kind of information in a relative way so the RL became capable of selecting the maintenance actions that minimize the cost over the system's age. The proposed estimation technique for RUL is presented in section 6.2.2.

As has been mentioned, the TPM that defines the transition between the different states is not needed. Since the solution approach is based on the RL method, the transition probability distribution is not needed. RL needs episodes of data that consist of tuples of state, action, and reward. The dynamics of the episodes are defined through the effect of the maintenance actions or the deterioration process. The effects of the maintenance actions and the deterioration are defined by the maintenance data and the data-driven deterioration prediction model, respectively. Figure 6.2 shows how the deterioration prediction model is constructed. It also depicts that the deterioration prediction model considers the age besides the deterioration level as the age affects the deterioration process and considering it preserves the intrinsic character of the deterioration process. We can expect that at older ages systems tend to deteriorate with higher rates.

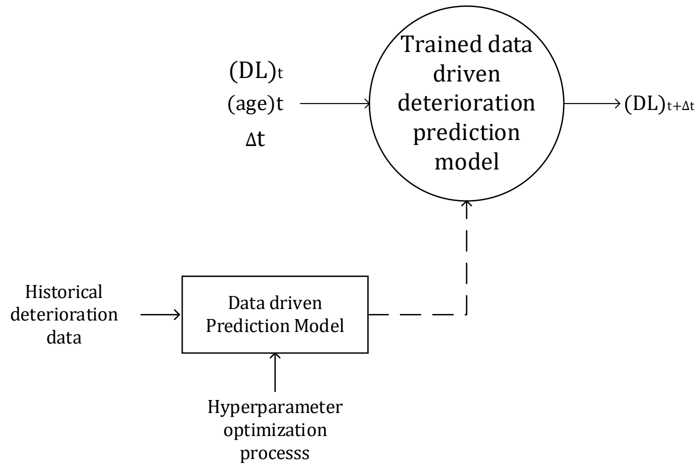


Figure 6.2 Deterioration prediction model.

6.2.2 RUL Estimation Method

The proposed reward function $r_t(s_i, a_j) = \mathrm{RUL}_t / C(s_i, a_j)$ consists of two elements, the remaining useful life RUL_t of the system at age t just after the maintenance action, and the maintenance action cost $C(s_i, a_j)$ The action costs are obtained from historical maintenance data. The estimation technique of the RUL proposed is based on a reliability approach. The formula used to estimate RUL Eq. 6.1 is adopted from (Elsayed, Elsayed A, 2003; Ragab et al., 2019), where RUL_t is the remaining useful life at a certain age t, $R(t_l)$, and $R(t_{l-1})$ are the survival probabilities at different times and the difference between t_{l-1} and t_l is a time step.

$$\mathrm{RUL}_t = \frac{\sum_{t_l=t}^{\infty} t_l\big(R(t_{l-1}) - R(t_l)\big)}{R(t_{l-1})} - t$$	Eq. 6.1

According to the previous formula, the estimation of the RUL requires a reliability or survival function. The KM estimate for survival function as shown in Eq. 2 is a nonparametric survival technique that does not rely on certain statistical assumptions (Klein & Moeschberger, 2006) and is adopted in the proposed methodology to estimate RUL. The choice of the KM model as a non-parametric model to estimate the RUL is motivated by the concern to keep the proposed methodology generic and does not suffer from assumptions that limit its applicability, in the same way as the data-driven deterioration prediction model and RL are selected. In Eq. 6.2 $R(t)$ is the survival function at time t, t_i is a time when at least one failure took place, d_i is the number of systems that have failed up to the time t_i, and n_i is the number of systems that have survived up to time t_i.

$$R(t) = \prod_{i:t_i \leq t} (1 - \frac{d_i}{n_i})$$	Eq. 6.2

In the proposed RUL estimation method, the failure is defined by reaching a certain DL, then the times or ages to reach this certain DL are the failure times that are needed. During its life, the system undergoes different DLs until it reaches failure. Based on lifetime samples, we can construct a KM curve for each DL using the times at which the different systems reached each DL, as such the failure event is not only the catastrophic failure that is traditionally used in the literature but also failure can be defined by reaching certain deterioration level. Figure 6.3(a) shows an example for different deterioration curves, and it shows four deterioration levels (D1, D2, D3, and D4) using the times at which the system reached the different predetermined DL a KM curve corresponding to this DL can be construed. The KM curves that are shown in Figure 6.3(b) are constructed for each DL using the data from the different deterioration curves where the deterioration reaches a certain level.

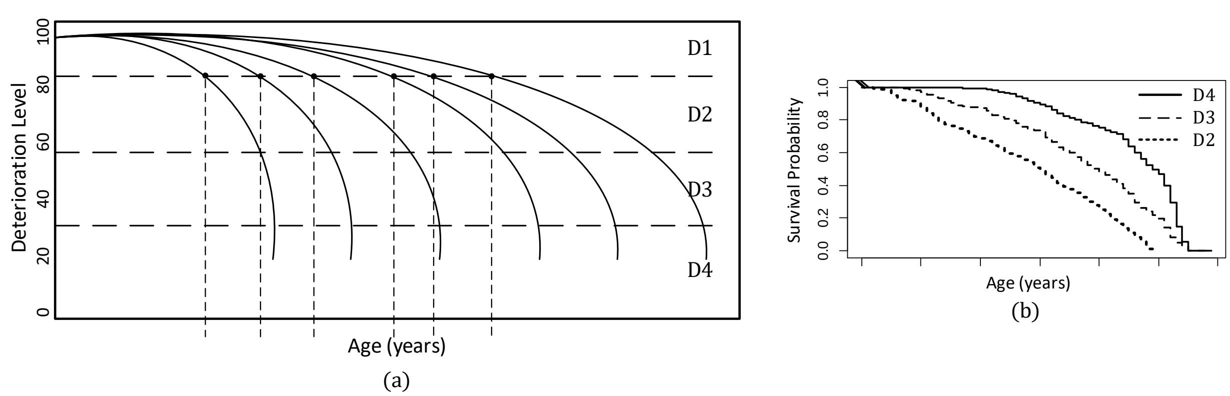


Figure 6.3 Example of the KM survival curves characterizing three DLs.

Using Eq. 6.1, the RUL is estimated based on the KM survival curve of the corresponding DL. Doing so, the estimation of the RUL considers not only the system's age but also its DL and the survival curve that it corresponds to. Two scenarios for estimating the RUL will be followed and their results will be compared. The difference between the two scenarios is the state that is based on which the RUL is estimated, as Figure 6.4 shows. In the first scenario, the RUL is estimated directly after the maintenance action is performed. While the first scenario provides a shortsighted estimation for the RUL, the second scenario provides a farsighted estimation. In other words, the second scenario provides more information about how the deterioration process will affect the system's RUL in the future. Such information could help a RL agent to take better actions. The strategies that can be obtained are based on the two scenarios will be compared in order to investigate the effect of exchanging the RUL estimation for the reward function on the obtained CBM strategy.

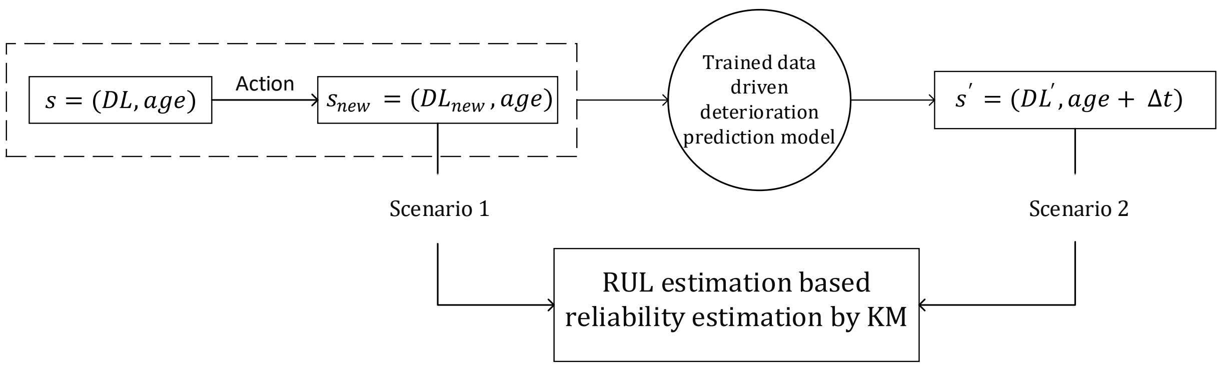


Figure 6.4 Two scenarios to estimate the RUL.

6.2.3 RL Based Solution

RL is a machine learning branch that concerns the decision-making problem. The RL's agent learns and provides optimal actions by interacting with the environment. The environment that represents the problem should be structured as MDP. The RL's agent observes the state, takes an action, and it receives a reward. The agent aims to maximize the return G_t by taking the optimal actions at each time step t over a certain time horizon T based on the problem as shown in Eq. 6.3. By definition, the return is the total discounted reward r.

$G_t = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots + \gamma^{T-1} r_T = \sum_{l=0}^{T-1} \gamma^l r_{t+l+1}$	Eq. 6.3

The discounting factor $\gamma \in [0,1)$ avoids the infinite return values for continuous problems and also helps determine a present value for future rewards.

Different algorithms are proposed in the domain of RL. Q-learning is an off-policy algorithm based on the temporal difference (TD) approach (Sutton and Barto 2018). Q-learning learns the optimal policy by evaluating the value function $Q(s,a)$ of the state action pairs, as shown in Eq. 6.4.

$Q(s_t, a_t) = Q(s_t, a_t) + \alpha\,[r_{t+1} + \gamma max\, Q(s_{t+1}, a) - Q(s_t, a_t)]$	Eq. 6.4

The TD approach makes updates after each time step and does not wait until the termination of the task or the episode to update $Q(s,a)$. Generally, RL agents could choose the actions either on-policy or off-policy. An off-policy update for the value function is based on the optimal action-value function of the next state, regardless of the policy followed (Sutton & Barto, 2018). Figure 6.5 shows a schematic representation of the proposed solution technique and basic, simplified steps of the Q-learning algorithm, adapted to our problem.

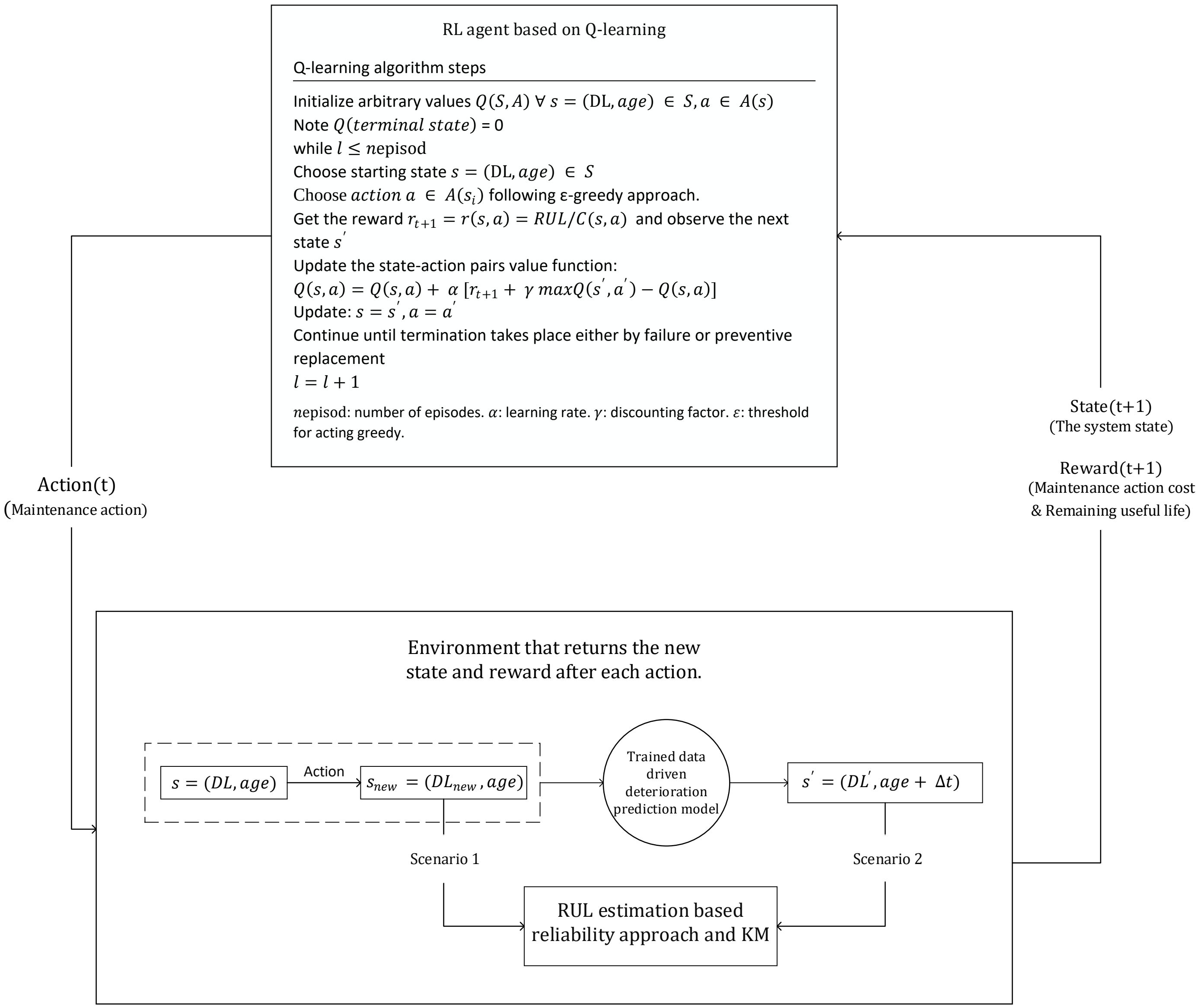


Figure 6.5 Schematic representation of the proposed model-free RL solution technique.

6.3 Case Study

The objective of this section is to illustrate and examine the applicability of the proposed methodology by applying it to solve a real CBM problem concerning identical sheet metal culverts. The culverts are inspected regularly, and their deterioration is assessed using different measures that are concluded and represented in a single deterioration condition by domain experts. The deterioration condition has a 0-100 range scale, this range is discretized – by the domain experts - into five DLs, as Table 6.1 shows. The DL for a culvert could be improved by multi-level preventive repair actions proposed in Table 6.2. The other possible actions are a do-nothing action, preventive replacement action, and corrective replacement action performed in case of failure.

Table 6.1 Discrete deterioration levels.

DL	DL description	Deterioration condition range
A	Perfect	85 or more.
B	Very good	71 – 84
C	Good	56 – 70
D	Acceptable	41 – 55
E	Bad (Failure)	40 or less.

Table 6.2 The actions and their description.

Action	Description
a_0	Do-nothing action allows the system to deteriorate.
a_1	Repair action improves the system's DL by 1 level.
a_2	Repair action improves the system's DL by 2 levels.
a_3	Repair action improves the system's DL by 3 levels.
P_r	Preventive replacement action returns the system to as good as new conditions.
C_r	Replacement action is performed in case of failure, and it returns the system to as good as new

6.3.1 CBM Strategy Model

As discussed in section 6.2, MDP is the base model, so for this culverts system, the elements of the MDP have to be defined, except the transition probability matrix, which is not needed.

The state-space $\boldsymbol{S}$ The states will be represented by both the culvert's DL and the culvert's age $s = (DL, age)$. Based on the available deterioration data, the maximum age that a culvert could be reached is 80 years. A time step of 10 years is used; this time step is appropriate to update the DL for this application. The combination of the discrete DL and the age results in 32 states, excluding the new state when the age is zero, and the failure states, which are defined by reaching DL E at any age.

The action space $\boldsymbol{A(s)}$: Do-nothing and preventive replace actions are possible for all of the states. The possibility for multi-level preventive repair actions depends on the DL of the state. For DL equal to A, no repair actions are possible, for DL equal B a_1 action is possible, for DL equal C a_1, and a_2 are possible and for DL equal D a_1, a_2, and a_3 actions are possible. If the DL reaches E, only the corrective replacement action is possible. The size of the state-action space is 120 and considers only the age up to 80 years, but if the age exceeds 80 years during learning, the new state is added to the model and the optimal action for it is obtained.

The reward $\boldsymbol{r(s,a)}$ In the proposed model, the reward is related to the RUL of the system after performing the maintenance action and the performed action cost $r_t(s_i, a_j) = \mathrm{RUL}_t / C(s_i, a_j)$ year/ dollars per unit length of the culvert (year/\$/u.L)., The preventive replacement action cost is 5100 \$/u.L, the corrective replacement cost is 15100 \$/u.L, the cost for a do-nothing action is 100 \$/u.L, and the costs of the rest of the actions differ based on the DL, the age as shown in Table 6.3. The costs of the multi-level preventive repairs shown in Table 6.3 depends on the system's state and not only the action, i.e. the same action has different costs when performed at different system states. Also, the corrective replacement cost is almost three times the preventive replacement that exhibits disfavor of the failure event. The RUL is estimated as discussed in section 6.2.2 following two scenarios, as Figure 6.4 shows.

Table 6.3 Costs of the actions at different states (\$/u.L).

(s,a)	$C(s,a)$	(s,a)	$C(s,a)$	(s,a)	$C(s,a)$
$((B,10),a_1)$	213.3	$((C,10),a_1)$	230	$((C,10),a_2)$	320
$((B,20),a_1)$	226.67	$((C,20),a_1)$	260	$((C,20),a_2)$	440
$((B,30),a_1)$	240	$((C,30),a_1)$	290	$((C,30),a_2)$	560
$((B,40),a_1)$	253.3	$((C,40),a_1)$	320	$((C,40),a_2)$	680
$((B,50),a_1)$	266.67	$((C,50),a_1)$	350	$((C,50),a_2)$	800
$((B,60),a_1)$	280	$((C,60),a_1)$	380	$((C,60),a_2)$	920
$((B,70),a_1)$	293.3	$((C,70),a_1)$	410	$((C,70),a_2)$	1040
$((B,80),a_1)$	306.67	$((C,80),a_1)$	440	$((C,80),a_2)$	1160
$((D,10),a_1)$	253.3	$((D,10),a_2)$	413.3	$((D,10),a_3)$	680
$((D,20),a_1)$	306.67	$((D,20),a_2)$	626.67	$((D,20),a_3)$	1160
$((D,30),a_1)$	360	$((D,30),a_2)$	840	$((D,30),a_3)$	1640
$((D,40),a_1)$	413.3	$((D,40),a_2)$	1053.3	$((D,40),a_3)$	2120
$((D,50),a_1)$	466.67	$((D,50),a_2)$	1266.67	$((D,50),a_3)$	2600
$((D,60),a_1)$	520	$((D,60),a_2)$	1480	$((D,60),a_3)$	3080
$((D,70),a_1)$	573.3	$((D,70),a_2)$	1693.3	$((D,70),a_3)$	3560
$((D,age \geq 80),a_1)$	626.67	$C((D,age \geq 80),a_2)$	1906.67	$((D,age \geq 80),a_3)$	4040

Using the historical deterioration data, a prediction model is trained using the Random Forest (RF) algorithm. RF is a tree-based algorithm that shows advances in the prediction tasks with reasonable amounts of data (Falamarzi et al., 2019; James et al., 2013). The model is trained to predict the system's DL after any time step Δt based on its current age and DL. The model's hyperparameters are tuned through grid search to obtain the prediction model with the highest possible accuracy. The final optimized model training accuracy is 82.7%, and the testing accuracy is 74%. The hyperparameters of the final optimized model are: the number of trees in the forest is 500 trees, the maximum depth of the trees is 9, the splitting criteria is Gini, and the number of features to split on is the total number of features. The model training and testing accuracies are acceptable as the model is based on the categorical values for DL, not the continuous ones.

6.3.2 The Results

The described model is merged into the environment that represents the culvert. The Q-learning RL agent is trained to obtain the solution by interacting with the environment. The interactive training of the RL agent is trained using 2500 episodes. The episodes are terminated if failure takes place or if preventive replacement action is performed. The chosen value for the discounting factor γ is 0.9 and for the learning rate α is 0.1. As the task we have is episodical, which means it has explicit termination, the discounting factor should be selected to be near 1 (Pitis, 2019) and with the number of episodes taken to be 2500, a learning rate of 0.1 is suitable. The exploration rate ε value decays over time to allow a high rate of exploration at the beginning of the training and a low rate towards the latest episodes. The starting value of ε is 1 and it keeps decays by small increments of $(c_1 * \varepsilon)/(c_2 * current\ episode\ number)$ after each episode. The values for the two constants c_1 and c_2 are taken to be 0.25 and 0.5; such values allow a moderate decrease for ε. Table 6.4 summarises the case study parameters and Table 6.5 shows the solution obtained by the agent for the two scenarios of RUL estimation.

Table 6.4 Summary for the case study parameters.

MDP model parameters	
Number of discrete deterioration levels	5
State-space S size	34+
State action pairs space size	120+
Deterioration prediction model	Random forest
Q-learning based agent parameters	
Number of episodes	2500
Discounting factor γ	0.9
Learning rate α	0.1
Exploration rate ε	Decaying: 1 to 0.05

Table 6.5 The optimal strategy obtained by the RL agent.

State A	Action		State B	Action	
	Scenario1	Scenario2		Scenario1	Scenario2
$(A, 10)$	a_0	a_0	$(B, 10)$	a_1	a_1
$(A, 20)$	a_0	a_0	$(B, 20)$	a_1	a_1
$(A, 30)$	a_0	a_0	$(B, 30)$	a_1	a_1
$(A, 40)$	a_0	a_0	$(B, 40)$	a_1	a_1
$(A, 50)$	a_0	a_0	$(B, 50)$	a_1	a_1
$(A, 60)$	a_0	a_0	$(B, 60)$	a_0	a_0
$(A, age \geq 70)$	P_r	P_r	$(B, age \geq 70)$	P_r	P_r
State C	**Action**		**State D**	**Action**	
	Scenario1	Scenario2		Scenario1	Scenario2
$(C, 10)$	a_2	a_2	$(D, 10)$	a_3	a_3
$(C, 20)$	a_2	a_2	$(D, 20)$	a_3	a_3
$(C, 30)$	a_2	a_2	$(D, 30)$	a_3	a_3
$(C, 40)$	a_2	a_2	$(D, 40)$	a_3	a_3
$(C, 50)$	a_2	a_1	$(D, 50)$	a_3	a_3
$(C, 60)$	a_1	a_1	$(D, 60)$	a_3	a_2
$(C, age \geq 70)$	P_r	P_r	$(D, age \geq 70)$	P_r	P_r
$State\ E, any\ age, C_r$					

Table 6.5 indicates that the two scenarios for estimating the RUL in the reward function yield similar strategies, except for two states $(C, 50)$ and $(D, 60)$. For the state $(C, 50)$ a_2 repair action is selected based on scenario 1 versus a_1 repair action selected based on scenario 2. For the state $(D, 60)$ a_3 repair action is selected based on scenario 1 versus a_2 repair action selected based on

scenario 2. The obtained actions based on scenario 2 are repair actions with less improvement effect on the DL for the two states. It can be observed that the two states have high DLs and old ages.

6.3.3 Validation and Interpretation

To validate the proposed methodology, the obtained results are compared with three baseline strategies. The comparison will be based on the average cost per unit of age. The three baseline strategies are:

- ***Baseline 1***. A CBM strategy that does not consider the RUL, and the reward function is based on the maintenance action cost, only to minimize the maintenance cost.
- ***Baseline 2***. A CBM strategy that uses the same reward function design as the one proposed with preventive repair, corrective repair, and do-nothing. It does not consider multi-level repair actions.
- ***Baseline 3.*** A corrective maintenance strategy that only performs corrective replacement after failure.

To obtain the baseline strategies, the Q-learning-based agent is trained based on the reward function and available maintenance actions for each strategy. Tables 6.6 and 6.7 show the first two baseline strategies, respectively. The third strategy is run to failure strategy. To estimate the cost per unit of age, a simulation that follows the algorithm shown in Figure 6.6 is used.

Table 6.6 Optimal results of Baseline 1 strategy.

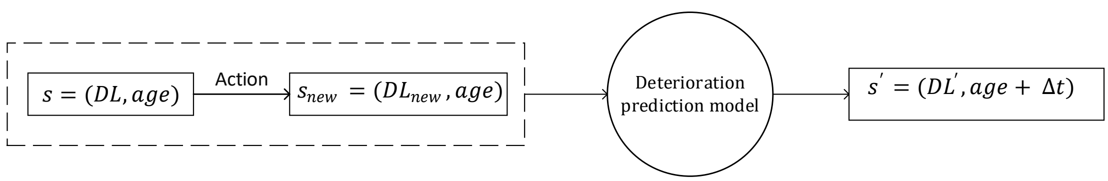
State	Actio	State	Actio	State	Actio	State	Actio	
$(A, 10)$	a_0	$(B, 10)$	a_1	$(C, 10)$	a_2	$(D, 10)$	a_2	
$(A, 20)$	a_0	$(B, 20)$	a_0	$(C, 20)$	a_1	$(D, 20)$	a_1	
$(A, 30)$	a_0	$(B, 30)$	a_1	$(C, 30)$	a_0	$(D, 30)$	a_1	$(E, any\ age)$
$(A, 40)$	a_0	$(B, 40)$	a_0	$(C, 40)$	a_1	$(D, 40)$	P_r	
$(A, 50)$	a_0	$(B, 50)$	a_0	$(C, 50)$	a_1	$(D, 50)$	P_r	C_r
$(A, 60)$	a_0	$(B, 60)$	a_0	$(C, 60)$	a_1	$(D, 60)$	P_r	
$(A, age \geq 70)$	P_r	$(B, age \geq 70)$	P_r	$(C, age \geq 70)$	P_r	$(D, age \geq 70)$	P_r	

Table 6.7 Optimal results of Baseline 2 strategy.

State	Actio	State	Actio	State	Actio	State	Actio	
$(A, 10)$	a_0	$(B, 10)$	a_1	$(C, 10)$	a_2	$(D, 10)$	a_2	
$(A, 20)$	a_0	$(B, 20)$	a_0	$(C, 20)$	a_1	$(D, 20)$	a_1	$(E, any\ age)$
$(A, 30)$	a_0	$(B, 30)$	a_1	$(C, 30)$	a_0	$(D, 30)$	a_1	
$(A, 40)$	a_0	$(B, 40)$	a_0	$(C, 40)$	a_1	$(D, 40)$	P_r	
$(A, 50)$	a_0	$(B, 50)$	a_0	$(C, 50)$	a_1	$(D, 50)$	P_r	C_r
$(A, 60)$	a_0	$(B, 60)$	a_0	$(C, 60)$	a_1	$(D, 60)$	P_r	
$(A, age \geq 70)$	P_r	$(B, age \geq 70)$	P_r	$(C, age \geq 70)$	P_r	$(D, age \geq 70)$	P_r	

Inputs	The maintenance strategy (mapping from state to action) The cost of maintenance actions.
Initiate	cost=0, age=0

follow the maintenance strategy to determine the action.



$s = s'$

cost= cost+ performed maintenance action cost

age=age+Δt

Figure 6.6 Algorithm for obtaining cost per unit age to compare maintenance strategies.

Table 6.8 Average maintenance cost for all strategies considered.

Maintenance strategy	Average maintenance cost ($/u.L/year)
Proposed strategy, RUL based on scenario 1	116.7
Proposed strategy, RUL based on scenario 2	108.9
Baseline 1 strategy	137.5
Baseline 2 strategy	137.5
Baseline 3 strategy	310

The results for an average maintenance cost shown in Table 6.8 depict that the proposed reward function design that uses the RUL as part of optimization criteria leads to more cost-effective maintenance strategies. The two proposed CBM strategies (scenarios 1 and 2) outperform the three baseline strategies. It is important to note that Baseline 1 and Baseline 2 strategies are optimal according to their own optimization criteria, which minimizes the maintenance cost without considering the RUL. This means the proposed optimization criterion that is based on the RUL demonstrates better performance over the ordinary optimization criteria, based only on cost.

In addition, it is worth noting that the two scenarios for selecting the state based on which the RUL is estimated yield different actions in some states, hence different average costs. Table 6.8 shows that scenario 2 outperforms scenario 1 and yields a lower average cost. As mentioned previously, the strategy based on scenario 2 chooses repair actions with less improvement effects for states with high deterioration levels and old ages. These actions have lower costs. The interpretation for this difference in the selection of the actions is related to the deterioration process. After the system is maintained in high deterioration levels and old age states, it witnesses a quick deterioration over the next 10 years of its age. Then, based on scenario 2, estimating the RUL based on the predicted DL state after 10 years, that gives the RL agent a better idea about what will happen in the future in terms of the system's deterioration. In other words, scenario 1 overestimates the RUL as it only has the information based on the current system state and does not consider the deterioration. Considering current DL in states that would witness fast deterioration after maintenance leads to a selection of repair actions that have high improvement levels. These actions may not be the most cost-effective for such states. We can conclude that scenario 2 provides a more farsighted reward for the RL agent, which allows the agent to select more cost-effective actions for such highly deteriorated states.

6.4 Conclusion

In this work, a novel CBM optimization methodology that combines data-driven deterioration prediction, RL method, and nonparametric survival techniques is developed. The objective is to maximize the system's remaining useful life (RUL) while keeping the maintenance cost at a low level. The deterioration process of the system is learned by a data-driven prediction model and introduced to the RL environment together with the MDP model and the RUL estimation method.

The proposed action space includes the usual preventive replacement, corrective replacement, and do-nothing actions besides multi-level preventive repair actions. The multi-level preventive repair actions improve the DL by certain amounts based on the action. The MDP transition probability distribution is not needed, as RL solution method needs only episodes of data that consist of sequential tuples of state, action, and reward. The reward function is designed to meet the CBM strategy objective using both RUL and maintenance actions cost. The proposed RUL estimation method is developed based on the Kaplan-Meier (KM) reliability model, which is a nonparametric survival technique that does not rely on any statistical assumptions. It is employed to estimate RUL in a way that considers the system's condition defined by its DL. Two scenarios for estimating the RUL are considered. The difference between the two scenarios is the way in which the system's deterioration process is considered. The first scenario is more concerned about the current deterioration level and estimates RUL based on that, and the second scenario considers the future effects of the deterioration process on the RUL estimation.

The proposed methodology is applied to a case study. The case study contains a deterioration data set collected from identical sheet metal culverts and the possible maintenance actions and their costs. Through the proposed methodology, two optimal CBM strategies are developed and solved. Each of the two strategies is optimal according to its own scenario for the RUL estimation. The two strategies are identical, except for only two states of high DL and old age. Validation is carried out by comparing the average maintenance cost for the obtained strategies with three baseline strategies: CBM strategy without the RUL, CBM strategy without the RUL and the multilevel repair actions, and the corrective replacement after failure. The comparison shows that considering the RUL as an optimization criterion leads to more cost-effective maintenance strategies.

The comparison also shows that the strategy based on scenario 2 is more cost-effective than the one based on scenario 1. The interpretation of the difference between the two scenarios is related to the deterioration process. In the highly deteriorated and old age states, quick deterioration takes place, even after maintenance. Scenario 1 estimates the RUL based on the state just after maintenance, which does not provide much information about the future. Scenario 2 estimation provides the RL agent with a better idea about what will happen in the future by allowing the deterioration over a time step, then it estimates the RUL. The actions obtained based on scenario 1 for the highly deteriorated and old age states incorporate high costs, and the system does not benefit from them over a long time as expected. In conclusion, scenario 2 provides a more farsighted reward for the RL agent, which allows the agent to select more cost-effective actions. The actions obtained based on scenario 1 for the highly deteriorated and old age states incorporate high costs, and the system does not benefit from them over a long time. In the light of this interpretation, scenario 2 is more favorable than scenario 1. Furthermore, a key finding is that the appropriate estimation of the RUL minimizes the maintenance strategy costs.

We would like to also confirm that a significant characteristic of the proposed methodology is that it does not depend on any prior assumptions other than the Markovian property for the MDP; this makes the model generic and applicable to a wide range of problems, as it only requires deterioration data.

Future research may investigate two studies: the effect of using different RUL estimation approaches other than the one that is adopted, the additional use of RUL to optimize the inspection interval aside from using it in the reward function, and the effect of using different deterioration prediction models on the optimal CBM strategy.

CHAPTER 7 ARTICLE 4: OPTIMAL CONDITION-BASED MAINTENANCE STRATEGY FOR MULTIPLE FAILURE MODES BASED ON DATA-DRIVEN MODELING AND SOLUTION METHOD

Submitted at Expert Systems with Applications journal

Manuscript Reference No.: ESWA-D-22-00134

Authors: Mina Mikhail, Mohamed-Salah Ouali, Soumaya Yacout

Abstract

This paper proposes a data-driven modeling and solution method for an optimal Condition-Based Maintenance strategy that considers multi failure modes subject to multi-level preventive repairs, in addition to the ordinary preventive and corrective replacements. The strategy is modeled as a Markov decision process and combines a data-driven deterioration prediction model, Cox proportional hazard model, and reinforcement learning method to minimize the average maintenance cost. The deterioration process related to each failure mode is learned from deterioration data using a data-driven model that is based on machine learning and then used as a state predictive model to realize the environment of the reinforcement learning model. A reward function that comprises the remaining useful life of the system and maintenance costs is designed in order to meet the optimal maintenance actions over the system deterioration age. The system's remaining useful life is estimated based on the undertaken maintenance action at a certain age and the deterioration conditions, using the Cox proportional hazard model that represents the survival curves under the competitive failure modes. A case study based on real data is used to illustrate the proposed maintenance strategy. The optimal solution takes the form of mapping from state-to-action and not just a condition-based replacement threshold related to each failure mode, as is often widely suggested. The optimal solution obtained is compared with other benchmark strategies to demonstrate its effectiveness and cost-savings abilities over time.

7.1 Introduction

A Preventive Maintenance (PM) strategy (or policy) is defined as a set of actions carried on a system based on certain decision criteria to sustain or restore its safely functioning state (Ahmad et al., 2012). An appropriate PM strategy must satisfy the increasing demand to run the systems at a high reliability level and at a low cost. PM has two main categories, time-based maintenance (TBM) and Condition-Based Maintenance (CBM) (Peng et al., 2010; Shin & Jun, 2015; Tran & Yang, 2012). TBM is periodic preventive maintenance that is carried out in fixed time intervals recommended by the systems' manufacturer or based on maintenance experts' experience (Ahmad et al., 2012; de Jonge et al., 2017). Differently, CBM is carried out based on the systems' deterioration condition (Hu, J. & Zhang, 2014; Jardine et al., 2006). The improvements in data acquisition, storage, and analytics assist the CBM strategies in being more cost-efficient than TBM ones (Bousdekis et al., 2018; Sarazin et al., 2021). While the deterioration of an industrial system implies complex processes related to several failure modes, most of the existing CBM strategies consider a single deterioration failure mode (Braga & Andrade, 2019; Mandiartha, Duffield, Razelan, et al., 2017; Sancho et al., 2021; Wang, J. et al., 2019). According to soft or deterioration failure modes, which are deterioration processes that progressively affect the system's performance, failure takes place when the deterioration level exceeds a certain predetermined threshold. In such cases, the system's deterioration is modeled using a gamma process, Winner process, Markov decision process (MDP), or semi-Markov decision process (SMDP) (Cholette et al., 2019; Fan et al., 2019; Kumar et al., 2018; Liu et al., 2021). Other literature has considered sudden failures besides the deterioration failure modes, which also refer to hard failure modes. Hard failure modes are more serious and lead to considerable costs and downtime when they occur (Sharifi, M. & Taghipour, 2020).

Li, Y. et al. (2017) proposed a CBM model for a system with two failure modes i.e., a deterioration failure mode and a shock failure mode. The deterioration failure mode is represented using a Winner process and the shocks are assumed to follow a non-homogenous Poisson process. The system state can either be normal, defective or fail. In case of a failure state, a corrective replacement is performed and two PM thresholds are determined to perform the PM actions in both the normal and defective states. Through a numerical example, the CBM strategy has proven that the solution obtained for the normal and defective states yields a lower average maintenance cost. The same problem is addressed by Li, X. et al. (2019) by proposing an optimal maintenance policy for gearbox subjected to two failure modes, namely a deterioration failure and a catastrophic failure. The maintenance optimization problem is modeled and solved using a semi-Markov decision process (SMDP) framework with three states (i.e. "good", "warning" and "failure" states).

The preventive replacement is performed when the condition of the gearbox exceeds the warning state threshold, and the corrective replacement is undertaken when the failure state is reached. The obtained results have demonstrated that taking into consideration the two failure modes helps in the early detection of failure and improves the availability when it was compared with the traditional TBM strategy. Gao et al. (2020) proposed a comparable model to jointly optimize the maintenance and production for a system with soft and hard failure modes. While soft failure follows an exponential distribution and occurs after warning signs, hard failure occurs suddenly according to a Weibull distribution with no warning. Preventive maintenance, complete replacement, and minimal repair actions are considered in the maintenance strategy. Using two numerical examples based on the proposed model, a sensitivity analysis showed that improving the system reliability and avoiding failures are effective ways to increase profit. Comparable studies are proposed in (Diyin, Jinsong, et al., 2015; Jian et al., 2016; Rahimikelarijani et al., 2020; Rui & Makis, 2020; Sharifi, Mani & Taghipour, 2021) for the benefit of readers.

Regardless of the improvements to the CBM achieved by the above-mentioned literature, they all are limited by 1. Assuming that only one soft failure mode exists in the systems, which is not the case in practice. 2. The deterioration model may depend on certain simplifying assumptions about the shape or the progression of the deterioration process. 3. Only preventive replacement action is considered, with limited exceptions, where minimal repair action is considered. 4. The obtained solution is defined as the threshold for preventive replacement. Most CBM models have overlooked the existence of multiple soft failure modes and multi-level preventive repair actions that improve the deterioration failure modes. These actions return the system deterioration conditions to somewhere between as-good-as new and as-bad-as old, aside from the preventive replacement action that improves the system to as-good-as new condition.

This paper proposes a generic CBM strategy that considers multi soft failure modes, subject to multi-level preventive repair actions based on a data-driven modeling and solution method. The model combines MDP, a data-driven deterioration model that is based on machine learning, the Cox proportional hazard model (PHM), and reinforcement learning (RL) as a solution method to minimize the average maintenance cost over time. The MDP acts as a general framework for the maintenance optimization problem since it proposes an appropriate model for sequential decision-making under uncertainties where the maintenance actions affect not only the immediate situation but its future as well (Puterman, 2014). The deterioration of the system that incorporates multi soft failure modes is represented using a machine learning prediction model, which is normally generic and does not have any assumptions related to the nature of the deterioration shape or progression. The Cox PHM is a reliability-based model that is used in many applications of CBM strategies to describe the system's survival based on its age and deterioration conditions (Chen, C. et al., 2020; Zeyang et al., 2014). It is a regression model that defines a hazard-based on a baseline hazard function and a parametric effect of the deterioration conditions. In doing so, the Cox model is an appropriate reliability-based approach to enable remaining useful life (RUL) estimation, which will be used later in the reward function of the RL method. The RL method is developed to obtain the optimal maintenance strategy in the form of mapping from a state to action. Therefore, unlike the widely proposed CBM models that provide a threshold for preventive replacement, for each state that represents a certain deterioration condition, the optimal action is determined. Literature that adopted RL as the solution method for maintenance problems has similar limitations as the CBM models disused earlier (Adsule et al., 2020; Kuhnle et al., 2019; Ling et al., 2018; Wang, X. et al., 2016). Also, most of those models consider only the preventive replacement action and use the negative value of the maintenance cost in the RL reward function to minimize the average maintenance cost. In the proposed CBM model, a meaningful reward function is designed considering the systems' RUL to obtain a more cost-effective maintenance strategy.

Therefore, the proposed CBM strategy addresses the overlooked CBM problem of the system with multi soft failure modes through a generic data-driven model and solution method. The rest of this paper is arranged as follows. Section 7.2 provides a detailed description of the CBM strategy developed, i.e., the MDP model, the deterioration prediction model and RUL estimation, and RL method modeling. Section 7.3 applies the proposed CBM strategy to a case study based on real data. The objective of the case study is to examine the optimality of the CBM strategy by comparing the obtained solution to other benchmark strategies to demonstrate its effectiveness and cost savings over time. Section 7.4 highlights the main findings of the proposed CBM model and future research challenges.

7.2 CBM Model Development

We propose a CBM model for systems that witness dependent soft multi failure modes. This model differs from the above-discussed models in three main ways: (1) It considers dependent soft multi failure modes and their effects on the systems' deterioration condition through a data-driven prediction model that is based on machine learning. (2) It employs the RUL of the system in a reward function that serves the CBM strategy of maximizing the RUL and minimizing the maintenance cost. (3) It takes into account the possible multilevel preventive repair actions that are overlooked by most of the existing CBM strategies.

7.2.1 MDP Model

The proposed CBM strategy utilizes MDP as a general model for the optimization problem. The MDP model is defined using a tuple of five elements (S, A, P, r, γ). S is the state space that consists of n finite number of states. The state $s_i \in S$ defines the system deterioration condition using different elements, in the proposed model $\forall s_i \in S = \big(FML_1, FML_2, \ldots, FML_j, \ldots, FML_k, D,$ $Age\big)$, where FML_j denotes failure mode j level, k is the total number of failure modes, D is the system's deterioration level, Age is the system's calendar age. The system's deterioration level D is obtained based on the different failure mode levels $D = f(FML_1, FML_2, \ldots, FML_k)$, and the system's failure is observed when D reaches a certain threshold. A denotes the action space. There is a set of possible maintenance actions $A(s_i) = \{a_0, a_1, \ldots.\ a_m\}$ $\forall s_i \in S$, the possible number of actions m depends on the state s_i. P is the state transition probability matrix (TPM) that is not needed in our case thanks to RL. r represents the reward function that divides the remaining useful life at a certain state s_i (RUL_{s_i}) by the maintenance cost $C(s_i, a_l)$ as follows $r(s_i, a_l) = {\mathrm{RUL}_{s_i}}/{C(s_i, a_l)}$. This customized reward function aims to yield a more cost-effective maintenance strategy. Finally, $\gamma \in [0,1)$ denotes the reward discounting factor used to decide the importance of future rewards.

7.2.2 Deterioration Prediction and RUL Estimation

Figure 7.1 depicts the general work frame for deterioration prediction using a data-driven model that is based on machine learning. The Random Forest (RF) algorithm is adopted to model the system deterioration, taking into consideration the effect and dependency of various failure models. RF is a tree-based algorithm that shows advances in the prediction tasks with reasonable amounts of data (Falamarzi et al., 2019; James et al., 2013). Each of the failure modes is modeled using a prediction model that predicts the failure mode level $FML_{jt+\Delta t}$ at $Age_{t+\Delta t}$ based on its level, the levels of all the other failure modes at t, the Age_t, and the time step Δt that represents the difference in age. This prediction model preserves the dependency between the failure modes by including all failure modes as independent variables for the prediction. Therefore, the predicted level for any of the failure modes is affected by the levels of the other failure modes. The same approach is applied to model the system's deterioration level D; it is defined based on the levels of the different failure modes and it is obtained using a prediction model that uses the levels of failure modes and age, as shown in Figure 7.2.

$$FML_{1_t}, FML_{2_t}, \dots, FML_{k_t}, Age_t, \Delta t \longrightarrow \begin{array}{c}\text{FML}_\text{j}\\ \text{Prediction}\\ \text{Model}\end{array} \longrightarrow FML_{j_{t+\Delta t}}$$

Figure 7.1 Machine learning prediction model for failure mode $\mathrm{FML}_{\mathrm{j}_{\mathrm{t+\Delta t}}}$.

$$FML_{1_{t+\Delta t}}, FML_{2_{t+\Delta t}}, \dots, FML_{k_{t+\Delta t}}, Age_{t+\Delta t} \longrightarrow \begin{array}{c}\text{D}\\ \text{Prediction}\\ \text{Model}\end{array} \longrightarrow D_{t+\Delta t}$$

Figure 7.2 Machine learning prediction model for system's deterioration level $\mathrm{D}_{\mathrm{t+\Delta t}}$.

A reliability-based method to estimate the RUL is adopted from (Elsayed, Elsayed A, 2003) and customized to our proposed model. Eq. 7.1 defines an estimation for the system's RUL, where $R(t)$ is the reliability function at time *t* and $f(t)$ is a probability-density-function.

$\mathrm{RUL} = \dfrac{1}{R(t)} \displaystyle\int_t^{\infty} \tau f(\tau) d\tau - t$	Eq. 7.1

As previously discussed, the Cox model assesses the system's reliability considering its deterioration condition, which is based on multiple failure modes. The general form for Cox PHM is shown in Eq. 7.2 (Cox, 1972; Jardine et al., 2006; Lin, D et al., 2006) where $h_o(t)$ is the baseline hazard function that describes the risk over time when the covariates have no effect, $\psi = (\psi_1, \psi_2, \psi_i, \dots, \psi_p)$ are p covariates' coefficients and $X = (x_1, x_2, x_i, \dots, x_p)$ are p covariates. In this work, the covariates are the failure mode levels $(FML_1, FML_2, \dots, FML_j, \dots, FML_k)$ and all the possible interactions between them $(FML_1 * FML_2, FML_1 * FML_3, \dots, FML_{k-1} * FML_k)$.

$\mathrm{h(t,x)} = h_o(t)\exp\left(\sum_{i=1}^{p} \psi_i\, x_i\right) = h_o(t)\exp(\psi X)$	Eq. 7.2

Using Eq. 7.2 and based on Cox PHM, Eq. 1 is reformulated as shown in Eq. 7.3. It provides an estimation for the RUL (t, X) at certain time t and certain covariate values X, which are defined based on the current system state $s_i = \left(FM_1, FM_2, \dots, FM_j, \dots, FM_k, D,\ Age\right) \in S$. Eq. 7.3 is demonstrated in the appendix.

$\mathrm{RUL(t,X)} = \frac{1}{R_o(t)^{\exp(\psi X)}} \int_t^{\infty} R_o(\tau)^{\exp(\psi X)}\, d\tau$ where $R_o(t) = \exp\left(-\int_0^t h_o(\tau) d\tau\right)$	Eq. 7.3

In Eq. 7.3, there is a need to estimate $R_o(t)$ using $h_o(t)$. Different methods can be used to estimate $h_o(t)$. It can be modeled using different distributions such as Exponential, Weibull, or normal distribution. Also, $h_o(t)$, can be estimated based on the Breslow estimator, which is a nonparametric estimator (Lin, DY 2007; Xia et al., 2018). In this work, the estimation of $h_o(t)$ is based on the nonparametric Breslow estimator that is shown in Eq. 7.4 and Eq. 7.5 (Breslow, N., 1972; Breslow, N. E. & Wellner, 2007; Xia et al., 2018). The Breslow estimator is selected to avoid any assumptions about $h_o(t)$ that may limit the applicability of the RUL estimation to certain applications. In Eq. 7.5 d_j is the number of failures at time y_j, $\mathcal{R}(y_j)$ is the set of k systems that are still at risk at time y_j, and X_k are the corresponding covariates to $\mathcal{R}(y_j)$.

$R_o(t) = \exp\left(-\int_0^t h_o(\tau) d\tau\right) = \exp -\Lambda_o(t)$	Eq. 7.4

$\Lambda_o(t) = \sum_{j:y_j \le t}\left(\frac{d_j}{\sum_{k\in\mathcal{R}(y_j)} \exp{(\psi X_k)}}\right)$	Eq. 7.5

The RUL in this case is an optimistic estimation since it assumes that the system will stay at the same deterioration condition until failure takes place, which is not true. Over time, the system deterioration condition gets worse until failure takes place.

To have a more realistic estimation for the RUL_{s_i} to be used in the proposed reward function, an average estimation for the RUL is obtained. To clarify this, let us assume a simple example of a system with a single failure mode that has four levels (1, 2, 3,4) where 4 is the best level of the failure mode and 1 is the most deteriorated level. For each of the levels, a survival curve can be constructed using the Cox model. Suppose that at age *t*, the system is at level 3. This system will keep deteriorating until failure takes place, and it will witness the other levels 2 and 1 before failure. Through Eq. 1, the RUL estimation will assume that the system remains in level 3 until failure takes place, which is not true. To overcome this problem, we average the RUL estimations given by each failure mode level 3, 2, and 1 from time *t* until failure, as Figure 7.3 shows.

Starting from any state $s_i = (FML_{1_t}, FML_{2_t}, FML_{j_t}, D_t, \ Age_t)$, get the failure modes levels and the age t.

While no failure:

Step1: Use $(FML_{1_t}, FML_{2_t}, FML_{j_t}, \ldots, FML_{k_t}, \ FML_{1_t} * FML_{2_t}, FML_{1_t} * FML_{3_t}, \ldots, FML_{k-1_t} * FML_{k_t})$ as input covariates to the Cox PHM to get the survival probabilities based on these covariates' values.

Step2: Estimate RUL as $\mathrm{RUL}_{s_i} = \mathrm{RUL(t, X)} = \frac{1}{R_o(t)^{\exp{(\psi X)}}} \int_t^{\infty} R_o(\tau)^{\exp{(\psi X)}} \, d\tau$

Step3: Keep a record of RUL estimated.

Step4: Get failure modes levels $(FML_{1_{t+\Delta t}}, FML_{2_{t+\Delta t}}, \ldots, FML_{j_{t+\Delta t}}, \ldots, FML_{k_{t+\Delta t}})$ after Δt using the prediction models.

If no failure go to Step1

else: terminate and go to step 6

Step6: RUL_{s_i} = Average value of the obtained RUL estimations until failure, where s_i is the current system state in which the deterioration condition is defined, and the age is t.

Figure 7.3 $\mathrm{RUL}_{\mathrm{s_i}}$ Estimation method.

7.2.3 RL Modeling

A RL model has two principal elements: the environment and the agent. The environment represents the problem's model with which the agent interacts to progressively learn the optimal solution. The environment in our CBM optimization problem comprises the defined MDP, the deterioration prediction, and the Cox model for the RUL estimation. The objective of the RL agent is to learn the optimal strategy π that maximizes the return for a problem modeled as MDP without using the MDP's TPM. The total return is shown in Eq. 7.6, which calculates the total discounted reward following the strategy π.

$G_t = (r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \gamma^3 r_{t+4} + \cdots + \gamma^{T-1} r_T)\|\pi = \sum_{i=0}^{T-1} \gamma^l\, r_{t+i+1}\|\pi$	Eq. 7.6

The interaction between the environment and the agent is as follows. The agent observes the environment's state $s_i \in S$, then it takes an action $a \in A(s_i)$ and receives a reward r. A typical learning process for a RL agent customized to our CBM model is based on the Q-learning algorithm, described in Figure 7.4.

Q-learning algorithm steps

Initialize arbitrary values $Q(S, A)\ \forall\ s_i = (FML_1, FML_2, \ldots, FML_j \ldots, FML_k, D,\ Age) \epsilon\ S,\ a\ \epsilon\ A(s_i)$

Note $Q(terminal\ state) = 0$

while $m \leq nepisod$

- Choose starting state $s_i = (FML_1, FML_2, \ldots, FML_j \ldots, FML_k, D,\ Age) \epsilon\ S$
- $Choose\ action\ a\ \epsilon\ A(s_i)$following ε-greedy approach.
- Get the reward $r_{t+1} = r(s_i, a_l) = {}^{\mathrm{RUL}_{s_i}}\!/_{C(s_i, a_l)}$ and observe the next state s_{i+1}
- Update the state-action pairs value function:
- $Q(s, a) = Q(s, a) + \alpha\ [r_{t+1} + \gamma max\ Q(s', a') - Q(s, a)]$
- Update: $s = s_{i+1}$;
- Continue until termination takes place either by failure or preventive replacement.
- $m = m + 1$

FML_j is the failure mode j level, k is the total number of failure modes, D is the system's deterioration level, and Age is the system calendar age.

nepisode: number of episodes to train. α: learning rate. γ: discounting factor. ε: threshold for acting greedy.

Figure 7.4 Basic steps of Q-learning algorithm customized to the proposed CBM model.

The Q-learning is an off-policy algorithm that is based on the temporal difference (TD) technique (Sutton & Barto, 2018). It learns the optimal solution by evaluating the value function $Q(s, a)$ of the state action pairs as shown in Fig.4. Q-learning is based on the TD technique; it updates $Q(s, a)$ after each step using the reward and does not wait until the episode is terminated. Also, as an off-policy technique, Q-learning updates are based on the optimal action-value function of the next state regardless of the policy followed (Sutton & Barto, 2018). The episodes are the data observations that the agent learns from. They consist of tuples of (s, a, r). At the end of the learning process, the action that has the maximum Q value at each state corresponds to the optimal action for that state. Therefore, the solution that the agent obtains depends on the reward, and different reward function designs lead to different solutions. Therefore, more explicit information about the minimization horizon should be provided to the agent through the reward function to obtain the solution that minimizes the average cost over the defined horizon. The proposed reward function aims to define a solution that minimizes the maintenance cost over the age by simultaneously maximizing the RUL and minimizing the maintenance cost. The following section proposes a case study with real data that applies and validates the proposed CBM model.

7.3 Case Study

This case study deals with the CBM problem of identical metal sheet culverts. The deterioration data of the culverts is collected through reglementary inspections, and the levels of different failure modes are assessed by inspectors. The deterioration level of the culvert is described as a function of the failure mode levels and the age. The failure modes are assessed using either a five-level or three-level scale. In the collected deterioration data, the main failure modes that affect the culverts' deterioration are defects in materials with a five-level scale (level 5 is the best and 1 is the worst) and cracking and assembly defects, with a three-level scale (level 5 is the best and 1 is the worst). The system's deterioration level D takes continuous values between 0 and 100, where a value less than 40 is considered a failure. Failure of the culvert means that it functions inadequately with high risk in terms of safety. The two failure mods FML_1 and FML_2 can be improved through multi-level preventive repair actions; as their levels improve, the system's deterioration level also improves. Different actions comprise different effects and costs. Also, the culvert can be replaced and returned to as-good-as-new condition.

7.3.1 Culver's MDP Model

Table 7.1 shows the MDP model summary for the case study. In Table 7.1, the state is defined by the FML_1 (five levels), FML_2 (three levels), system's deterioration level D (0-100), D < 40 indicates the culver failure, and the age of the culvert. Mostly, the max culver's age is 60 years and in rare cases (less than 15%) it reaches 80 years. Thus, we considered the age span to be between 0 and 80 years. A 10 year time step is used for the age, which is suitable for this kind of application. At the age of 0, the culverts are in as-good-as new condition, so it is clear that no maintenance is needed; therefore, this age is eliminated from the state space. The state-space contains 120 states that consider all combinations of FML_1 , FML_2 , age, and the corresponding D.

For each state, the possible actions are: do nothing a_0, preventive replacement (Pr), corrective replacement (Cr), and set of possible multi-level preventive repair actions a_{xy}, where x is the amount of improvement applied to FML_1 and y is the amount of improvement applied to FML_2 i.e. $x = 1$ means improvement for $FML_1$by the amount of one step, $y = 2$ means improvement for FML_2 by the amount of two steps. Therefore a_{21} means improvement applied to $FML_1$by the amount of two steps and improvement applied to FML_2 by the amount of one step, and a_{10} improvement applied to $FML_1$by the amount of one step and do nothing for to FML_2. The corrective replacement Cr action is performed only in case of failure. The obtained state, actions pairs space size is 840 (state, action) pairs.

The reward $r(s_i, a_{xy})$ incorporates both RUL_{s_i} and $C(s_i, a_{xy})$ defined for unit length (u.L). The estimation method for RUL_{s_i} is proposed in section 2.2. As Table 1 depicts The multi-level preventive repair cost $C(s_i, a_{xy})$ depends on the failure modes levels, the amount of improvement in each failure mode, and the age; where c_i = 100 ($/u.L) is the inspection cost that is incurred, even if no maintenance is performed, c_b = 100 ($/u.L) is the basic maintenance cost, and c = 20 ($/u.L) is a constant. For the replacement actions $C(s_i, \text{Pr})$ = 5000 ($/u.L), and $C(s_i, \text{Cr})$= 15000 ($/u.L). Remarkably, the Cr cost is higher than the Pr cost as a result of the catastrophic effect that failures have.

Table 7.1 MDP model summary.

$\forall s_i \in S = (FML_1, FML_2, \ldots, FML_k,\ D,\ Age) = \forall s_i \in S = (FML_1, FML_2,\ D = f(FML_1, FML_2),\ Age)$
$A(s_i) = \{a_0, Pr, Cr, a_{xy}, \ldots\}\ \forall s_i \in S$
$r(s_i, a_{xy}) = {}^{\mathrm{RUL}_{s_i}}\!/_{C(s_i, a_{xy})}$ $C(s_i, a_{xy}) = c_i + c_b + (c * {FML_1}^2 * x^2 + c * {FML_2}^2 * y^2) * {}^{Age}\!/_{\max Age}$

7.3.2 Culver's Deterioration Prediction and RUL Estimation

Three prediction models are built to learn the deterioration process of FML_1, FML_2, and D. The RF algorithm is used to train the three deterioration prediction models. Table 7.2 shows the inputs, output, optimal hyperparameters, and testing accuracy for the deterioration prediction models. The optimal hyperparameters for the prediction models are obtained using random search, where the criterion is the splitting criteria, max_depth is the maximum depth of any tree, max_features is the number of features to split on, and n_estimator is the total number of trees in the forest. Details about the RF and its implementation are available in (Friedman et al., 2001; James et al., 2013; Pedregosa et al., 2011).

Table 7.2 Failure modes prediction model parameters.

Inputs (independent variables)		$FML_{1_t}, FML_{2_t}, Age, \Delta t$		$FML_{1_{t+\Delta t}}, FML_{2_{t+\Delta t}}, Age$
Output (dependent variable)		$FML_{1_{t+\Delta t}}$	$FML_{2_{t+\Delta t}}$	$D_{t+\Delta t}$
Testing accuracy		81.9%	86.3%	91.3%
Hyperparameters	criterion	'gini'	'entropy'	'mse'
	max_depth	9	7	7
	max_features	'auto'	'sqrt'	'auto'
	n_estimators	750	250	500

After building the deterioration prediction models Cox model can be constructed as it needs inputs from the deterioration models. The covariates of the Cox model are FML_1, FML_2, and their interaction FML_1*FML_2. The failure event is defined based on D since D < 40 is a failure. The Cox model is built using the survival package on R (Therneau & Lumley, 2013). Table 3 summarizes the results for the Cox model.

Table 7.3 Cox survival models summary.

	coefficient	exp(coefficient)	Z	p-value
FML_1	-0.05819	0.94347	-0.979	0.3276
FML_2	-0.14910	0.86148	-2.159	0.1309
FML_1*FML_2	-0.34748	0.70647	-7.834	4.74e-15 *
Schoenfeld residuals for FML_1*FML_2				0.89

From Table 3, we can conclude that the interaction covariate FML_1*FML_2. is the significant covariate. The proportional hazard assumption is tested using Schoenfeld residuals, and Table 3 shows an insignificant p-value for the Schoenfeld residuals of FML_1*FML_2. The residuals are independent of time, which means the proportional hazards assumption holds. This Cox model is used to get the survival probabilities for any possible $FML_1$and FML_2 to be used for estimating the RUL, as proposed in section 7.2.2.

7.3.3 Culver's RL Modeling and Solution

The MDP model, the three deterioration prediction models, and the Cox model are integrated to form the environment for the RL. As described in section 7.2.3, the RL agent is based on a Q-learning algorithm that is trained to obtain the solution by interacting with the environment. The RL agent is trained using 2500 data episodes. The episodes are terminated if failure takes place or if preventive replacement action is performed. The discounting factor γ is 0.9 and the learning rate α is 0.1. In case of explicit termination conditions for the episodes, the discounting factor should be selected near 1 (Pitis, 2019), and with 2500 data episodes, a learning rate of 0.1 is suitable. The exploration rate ε decays over time to allow a high exploration at the beginning of the training that keeps decreasing until the latest episodes. Initially, ε = 1 and it keeps decays by small increments of $(c_1 * \varepsilon)/(c_2 * Number\ of\ episodes\ )$ after each episode. The values for the two constants c_1 and c_2 are equal to 0.25 and 0.5, respectively. Such values allow a moderate decrease for ε. The obtained CBM strategy is presented in Figure 7.5. As discussed earlier, the strategy maps from a state to action.

In Figure 7.5, $FML_1$and FML_2 are shown on the horizontal axis and the age on the vertical axis. It is worth noting that for culverts at the same age and different deterioration conditions, different $FML_1$and FML_2, different actions are selected by the agent. The strategy obtained reaches optimality through the different actions. In the following subsection, the optimality is tested through validation. It is important to note that the strategy is optimal starting from any state and not only a new or as-good-as new state.

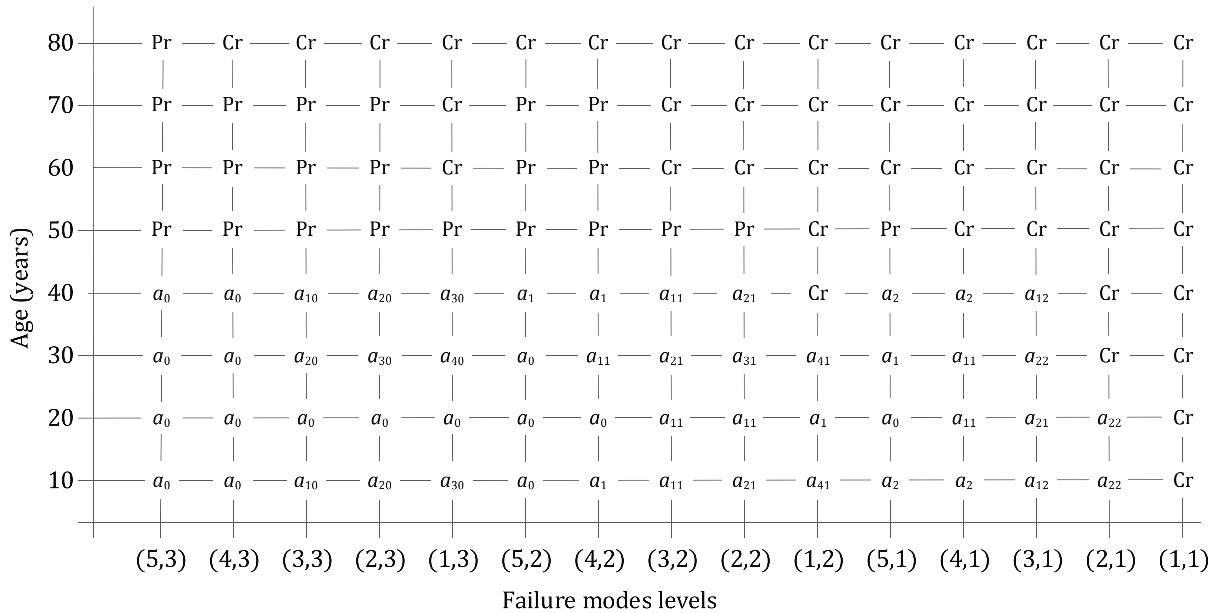


Figure 7.5 Obtained CBM strategy.

7.3.4 Validation

A validation process is performed to examine the optimality of the proposed CBM model and to test four main points considered in the proposed CBM model. The first two points to test are related to the reward function design. The third point is the dependency between the failure modes. Finally, the fourth point is related to the possible maintenance actions. The strategy obtained is compared against four strategies based on four scenarios that tackle the mentioned points. The comparison is based on the objective of the proposed CBM model, which is the average maintenance cost. Figure 7.6 and Figure 7.7 show the algorithm that was followed to obtain the average maintenance cost starting from any state.

Inputs	The Maintenance strategy (mapping from state to action) The cost of maintenance actions. The deterioration prediction models. Cox model.
Initiate	cost=0, i = 1 while i < state space size $s = s_i$ age = Age of s_i
Starting from s_i follow the maintenance strategy to determine the action and follow the process in Fig. 7, where ${FML_1}_t{}'$, ${FML_2}_t{}'$ are the failure modes levels after performing the maintenance action and $D_t{}'$ is the corresponding system's deterioration level. cost= cost+ performed maintenance action cost age=age+Δt continue until preventive replacement or corrective replacement takes place. Average cost = cost/age ($/u.L/year)	

Figure 7.6 Algorithm to obtain the average maintenance cost starting from any state.

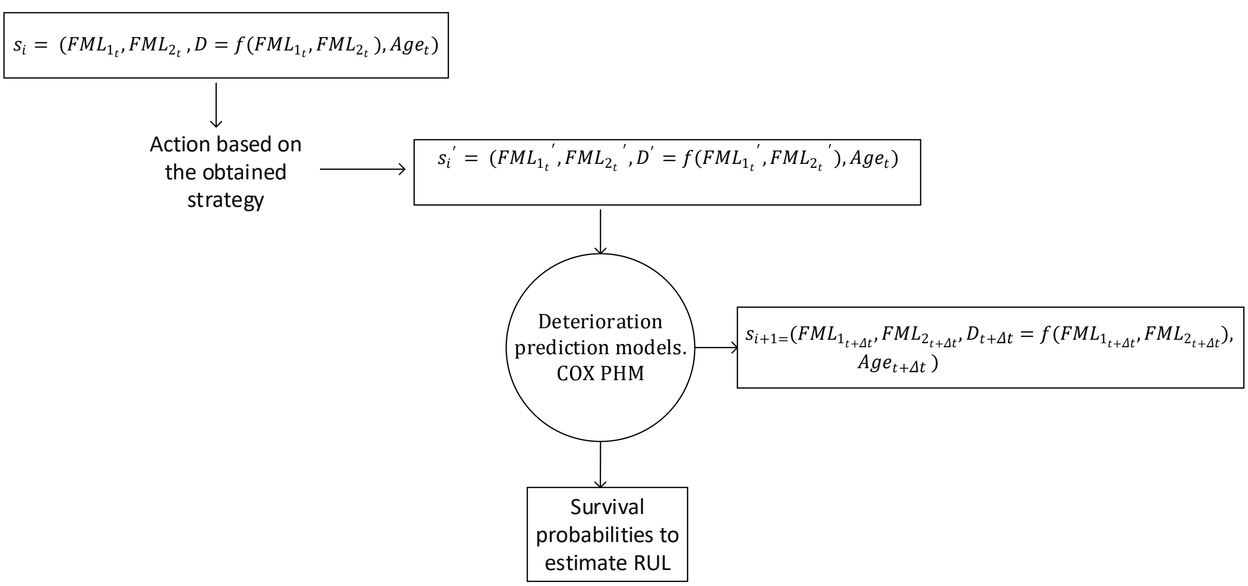


Figure 7.7 The obtained maintenance strategy application.

Comparison 1. The first comparison examines the effect of using a different reward function where RUL is estimated based on the current state only. The strategies obtained are based on the two scenarios. Scenario (a) is based on the proposed model, and the proposed RUL estimation section 7.2.2. Scenario (b) estimates the RUL based on the current state only. Figure 7.8 presents the actions and the average cost for each state based on the two scenarios.

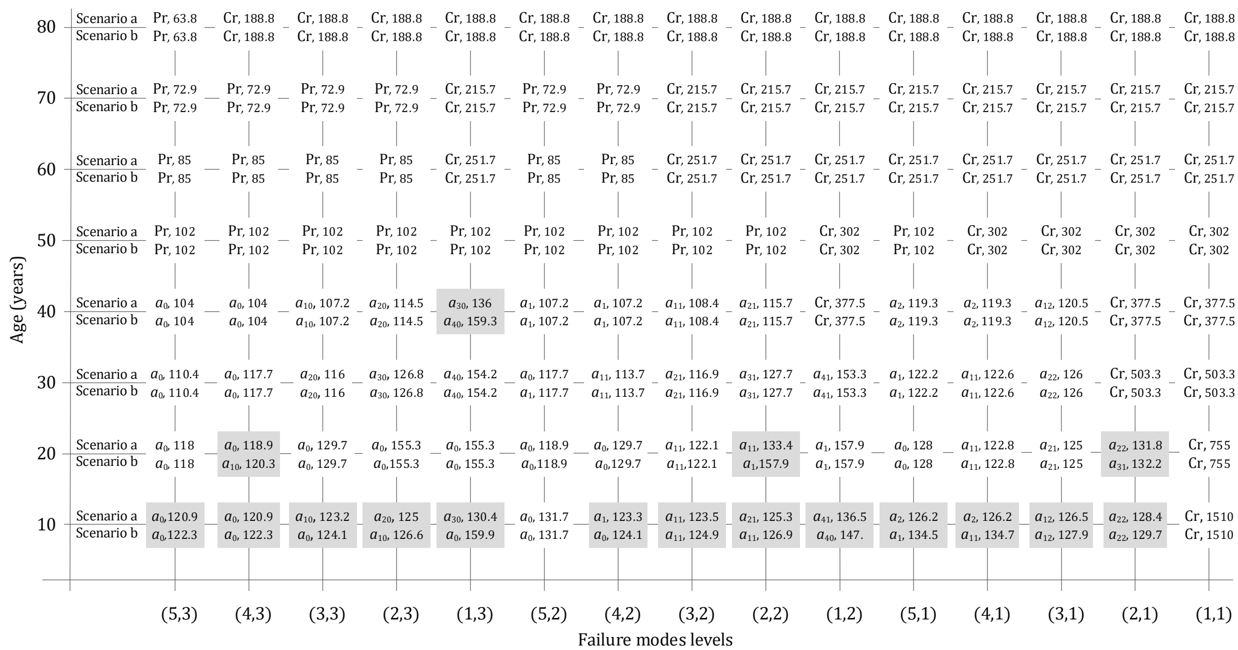


Figure 7.8 Action and average maintenance cost ($/u.L/year) comparison between scenarios (a) and (b).

In Figure 7.8, the states at which the average maintenance cost is different in the two scenarios are highlighted. Table 7.4 summarizes the differences in the average cost.

Table 7.4 Summary of the differences between the scenarios (a) and (b).

	scenario (a)	scenario (b)
Number of states with a lower average cost	17	0
Saving average per state ($/u.L/year/state)	6.9	0

Figure 7.8 and Table 7.4 depict that estimating the RUL using the proposed process yield a more cost-effective strategy that minimizes the average cost in 118 states by the average amount of 6.6 ($/u.L/year/state). The interpretation behind this finding is that estimating the RUL based on the current state only assumes that the system will stay on the current deterioration conditions until failure. Following this assumption leads to an optimistically high value for the RUL estimation. This assumption is not accurate, as the system will keep deteriorating over time. The proposed method for estimating the RUL does not follow this assumption, as it considers the system's

deterioration over time, which leads to a more realistic RUL estimation. Optimistic RUL estimation in the reward function leads to a strategy with action selected based on an inaccurate optimistic reward function; such a strategy does not yield the minimum average cost.

Comparison 2. In this comparison, a new scenario (c) that uses negative maintenance cost as a reward function is compared against scenario (a). Figure 7.9 presents the obtained results based on scenario (c) compared to scenario (a).

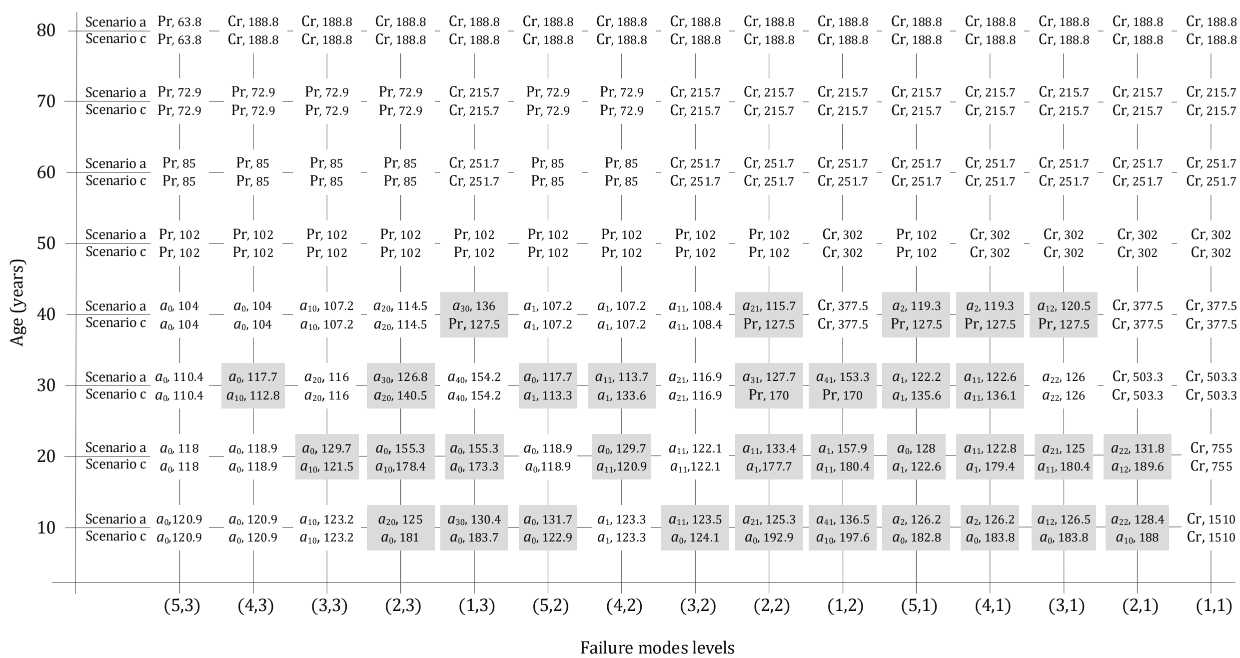


Figure 7.9 Action and average maintenance cost ($/u.L/year) comparison between scenario a and scenario c.

Table 7.5 Summary of the differences between scenarios (a) and (c).

	Scenario (a)	Scenario (c)
Number of states with a lower average cost	26	7
Saving average per state ($/u.L/year/state)	34.7	7.024

From Figure 7.9 and Table 7.5, we can conclude that the proposed design for the reward function yields better results in 26 states with an average savings of 34.7 ($/u.L/year/state), and using a reward function that depends on cost only yields better results in 7 states, with an average savings of 7.024 ($/u.L/year/state). In summary, the proposed design for the reward function returns a more cost-effective strategy. As an interpretation, the proposed design for the reward function provides explicit information about the effect of the action on the system's RUL besides the action cost. This kind of information related to the expected RUL helps the RL agent return a strategy that minimizes the average maintenance cost. The reason for the better results in 7 states when only the cost is used in the reward function is the accuracy of the RUL estimation. More accurate estimation for the RUL is expected to lead to a lower number of states in which a cost only reward function performs better.

Comparison 3. The proposed maintenance model stresses the importance of considering the dependency between the different failure modes in the maintenance modeling. The scenario for this comparison is designed to test and emphasize the importance of this point. The proposed model treats the dependency between the failure modes in both the deterioration prediction model and the Cox model. Table 7.2 shows that both FML_1 and FML_2 are used as inputs to the three deterioration prediction models for FML_1, FML_2, and D. Also, Table 7.3 shows that an interaction covariate FML_1*FML_2 is considered in the Cox model, which is found to have a significant covariate. scenario (d) omits the dependency between the failure modes. New deterioration prediction models and the Cox model are constructed without considering dependency. Table 7.6 shows the inputs for the deterioration prediction models and the covariates considered for scenario (d).

Table 7.6 Failure mode prediction model parameters for scenario (d).

Inputs (independent variables)		$FML_{1_t}, Age, \Delta t = 10$	$FML_{2_t}, Age, \Delta t = 10$	$FML_{1_{t+\Delta t}}, Age$	$FML_{2_{t+\Delta t}}, Age$
Output (dependent variable)		$FML_{1_{t+\Delta t}}$	$FML_{2_{t+\Delta t}}$	$D_{t+\Delta t}$	$D_{t+\Delta t}$
Testing accuracy		79.2%	83.2%	76.39 %.	77.39 %.
Hyperparameters	criterion	'gini'	'entropy'	'mse'	'mse'
	max_depth	9	5	7	7
	max_features	'auto'	'None'	'log2'	"sqrt"
	n_estimators	750	1000	100	750

Table 7.7 Cox model summary for scenario (d).

	coefficient	exp(coefficient)	Z	p-value
FML_1	-0.24789	0.78044	-11.16	<2e-16 *
FML_2	-0.50605	0.60287	-27.77	<2e-16 *
	Schoenfeld residuals for FML_1			0.229
	Schoenfeld residuals for FML_2			0.387

The deterioration prediction model for each failure mode level depends on the failure mode level itself and the age. Two deterioration prediction models for D are constructed, one for each failure mode. The definition of the failure of the system does not change ($D < 40$ is failure). The main difference is that the FML_1 and FML_2 effects are separated. Each of the two failure modes is predicted and maintained individually without considering the other failure mode. The Cox model does not include the interaction between failure modes. Notably, the prediction models that do not respect the dependency between the FML_1 and FML_2 have less testing accuracy than those that respect the dependency. In the developed Cox model for scenario (d), both FML_1 and FML_2 are statistically significant in the absence of the interaction covariate, as Table 7.7 shows. Also, the proportional hazards assumption is justified using the Schoenfeld residuals as shown in Table 7.7. The states, reward function, and RL agent have the same configuration. The obtained strategy and its average cost starting from each state are shown and compared to the strategy obtained based on the proposed maintenance model in Figure 7.10.

Figure 7.10 shows the two strategies as a mapping from state to action and the average cost in (\$/u.L/year) starting from each state. The highlighted states are those with different average costs. Both Table 7.8 and Figure 7.10 show that considering the dependency between the failure modes has a significant effect on the obtained strategy and therefore on the average cost. Treating the failure modes separately provokes early corrective replacement, as Figure 7.10 shows. These corrective replacements are avoided by considering the failure modes' dependency and defining the failure using both FML_1 and $FML_2$and not each one separately. Independent failure mode modeling results in the improvement of the action selected in one state only, whereas significant cost savings are observed in 35 states when the dependency between the failure modes is considered.

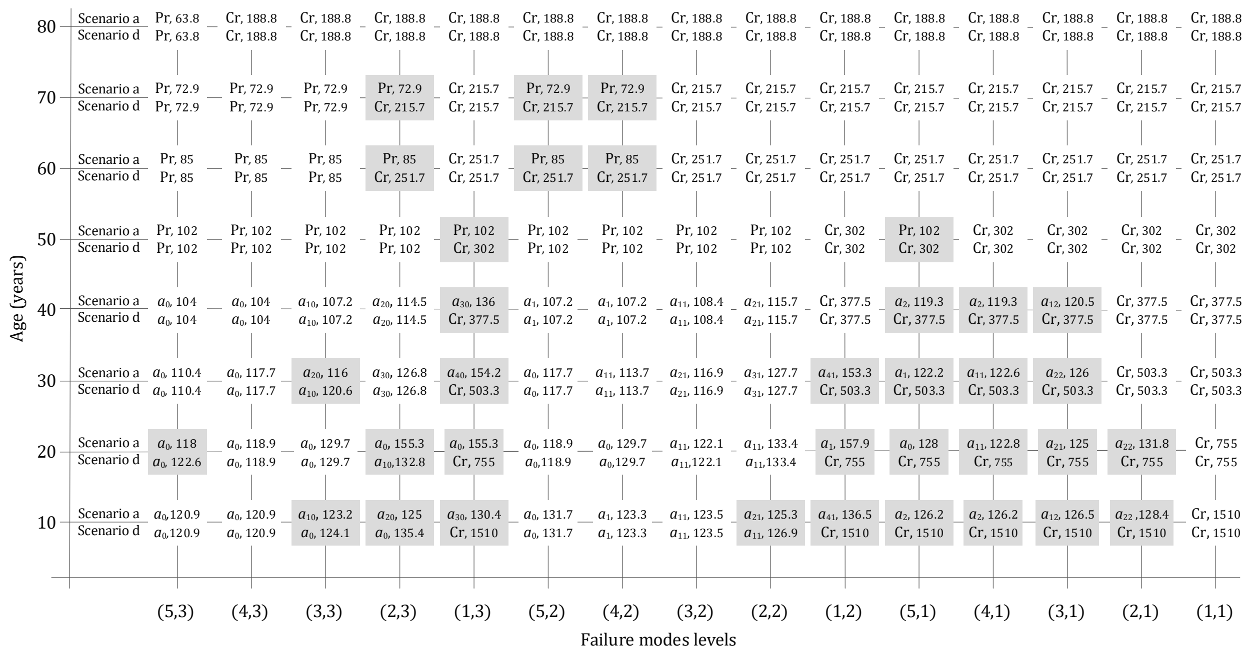


Figure 7.10 Action and average maintenance cost ($/u.L/year) comparison between scenario a and scenario d.

Table 7.8 Summary of the differences between scenario a and scenario (d).

	Scenario a	Scenario d
Number of states with a lower average cost	35	1
Savings average per state ($/u.L/year/state)	462.9	22.5

Comparison 4. The effect of considering the multi-level preventive repair action is studied in this comparison. In scenario e, a strategy that only includes the preventive repair, corrective repair, and do-nothing action is obtained and compared to the one obtained through scenario a. To obtain this strategy, the same maintenance model is used by removing the multi-level preventive repair action from it. Figure 7.11 and Table 7.9 show the obtained results and the comparison between scenario a and scenario (e).

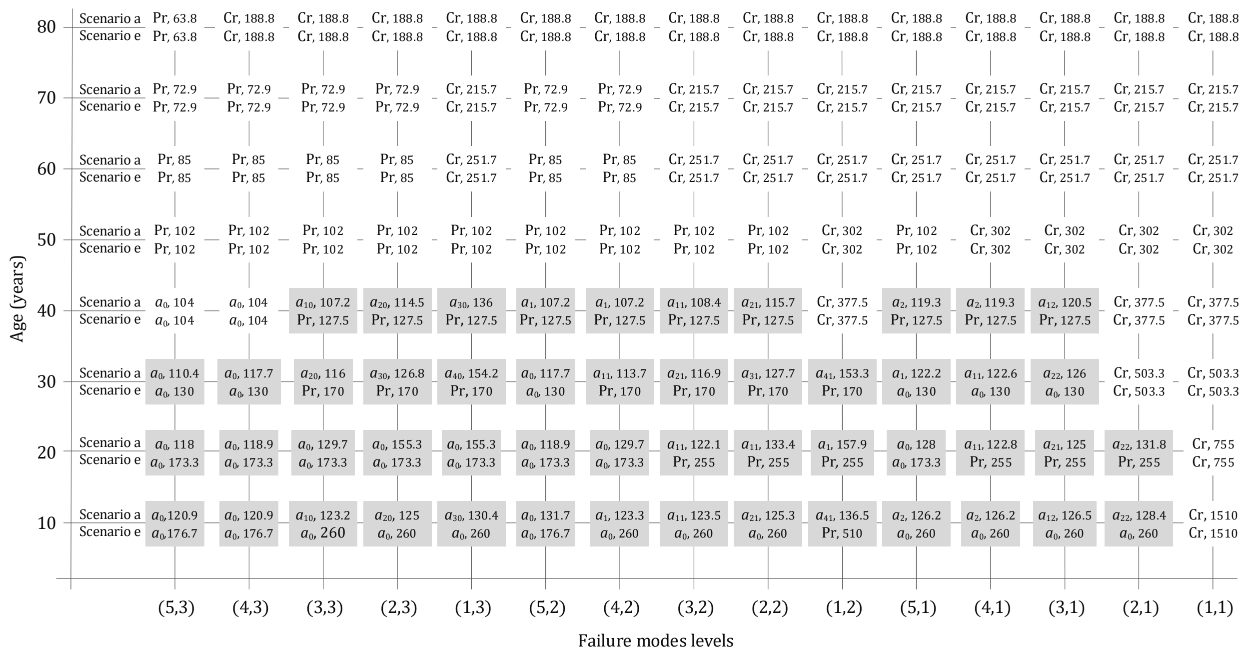


Figure 7.11 Action and average maintenance cost ($/u.L/year) comparison between scenario a and scenario e.

Table 7.9 Summary of the differences between scenario a and scenario (e).

	Scenario a	Scenario e
Number of the states with a lower average cost	50	1
Savings average per state ($/u.L/year/state)	68.3	8.5

From Figure 7.11 and Table 7.9, it is remarkable that considering the possible multi-level preventive repair action leads to a more cost-effective maintenance strategy than considering only the preventive repair action.

7.4 Conclusion

Regardless of the growing interest in CBM strategies, most of the proposed models overlooked the existence of multi deterioration failure modes for the same system. Most of the studies consider a single deterioration failure mode; others considered a sudden failure mode in addition to the deterioration failure mode. Even when flexible models like MDP are used with solution methods like RL, the same limitations still exist.

This paper has proposed a CBM strategy that considers multi deterioration failure modes with multi-level preventive repair actions. The proposed modeling and solution methods are both generic, data-driven combining a machine learning deterioration prediction model, a RL solution method, and a Cox model. The dependency between the failure modes is preserved in both the deterioration prediction models and the Cox model. Unlike the widely used RL reward function that only considers the cost, a reward function that comprises both the RUL and the cost is proposed. This is to ensure that the maintenance strategy obtained by the RL results in a minimum average maintenance cost. An estimation process for the RUL based on a reliability approach is proposed in the CBM model to be used by the proposed reward function. In contrast to widely considered replacement actions that return the system to as-good as-new condition, multi-level preventive repair actions are considered. The multi-level preventive repair actions can improve the deterioration failure modes and return the system to a condition that is between as-good-as-new and as-bad-as-old condition.

The proposed CBM modeling and solution method are applied to a case study on the maintenance of culvers. The solution obtained by the proposed model in the case study is a CBM strategy, which is a mapping from state to action. This strategy is compared with four other strategies to prove its optimality. Four comparisons are intended to test the four main points with respect to the CBM model. The first point is related to the proposed RUL estimation process. The second is related to the reward function design. The third is concerned with the dependency between the failure modes. The final one is related to the multi-level preventive repair actions. The first comparison tested two different processes of RUL estimation. The second test is on the proposed reward function against the ordinary reward function, which uses only the cost. In the third comparison, the obtained strategy is compared against a strategy that does not consider the dependency between the multi deterioration failure modes. It is obtained by modeling the two failure modes separately in both the deterioration prediction models and the Cox model. Finally, in the fourth comparison, a traditional strategy with only a preventive repair action is obtained and compared with the optimal CBM strategy. The four comparisons showed that the proposed CBM strategy is the most cost-efficient in terms of minimizing the average maintenance cost.

The proposed CBM modeling and solution method are both generic and data-driven. Therefore, they can be widely applied to different cases whenever deterioration data is available. Interesting points for future research include studying the effect of different RUL estimation methods and the effect of imperfect maintenance.

7.5 Appendix

Demonstration that $\text{RUL(t, X)} = \frac{1}{R_o(t)^{\exp(\psi X)}} \int_t^{\infty} R_o(\tau)^{\exp(\psi X)}\, d\tau$

$f(\tau) = \frac{-dR(\tau)}{d\tau}$, then $f(\tau)d\tau = -dR(\tau)$	Eq. 7
$\text{RUL(t)} = \frac{1}{R(t)} \int_t^{\infty} \tau f(\tau) d\tau - t$	Eq. 8
$\text{RUL(t)} = \frac{1}{R(t)} \int_t^{\infty} -\tau dR(\tau) d\tau - t$	Eq. 9
$\text{RUL(t)} = \frac{1}{R(t)} (\int_t^{\infty} R(\tau)d\tau + [-\tau R(\tau)]_t^{\infty})$	Eq. 10
$\text{RUL(t)} = \frac{1}{R(t)} \int_t^{\infty} R(\tau)d\tau + \frac{1}{R(t)}(tR(t)) - t = \frac{1}{R(t)} \int_t^{\infty} R(\tau)d\tau$	Eq. 11
Considering the Cox PHM model $R(t) = \exp(-\int_0^t h(\tau)d\tau)$	Eq. 12
$R(t, X) = \exp(-\int_0^t h_o(\tau)\exp(\psi X)\, d\tau)$	Eq. 13
Based on Cox's proportional hazards assumption $R(t, X) = \exp(-\exp(\psi X) \int_0^t h_o(\tau)d\tau)$	Eq. 14
$R(t, X) = \exp(-\int_0^t h_o(\tau)d\tau)^{\exp(\psi X)}$	Eq. 15
$R(t, X) = R_o(t)^{\exp(\psi X)}$	Eq. 16
Substitute Eq. 16 in Eq. 11 $\text{RUL(t, X)} = \frac{1}{R_o(t)^{\exp(\psi X)}} \int_t^{\infty} R_o(\tau)^{\exp(\psi X)}\, d\tau$	

CHAPTER 8 GENERAL DISCUSSION

8.1 Comments on the Methodology

The methodology proposed in this research is intended to be a generic data-driven methodology to obtain optimal CBM strategies. The methodology has been progressively developed in the three main contributions presented in Chapters 5 through 7. These contributions are related, yet they can be applied separately to solve maintenance problems with different levels of complexity.

The methodology treated the modeling and the solution components of the CBM problem based on data-driven approaches and it discussed a critical point of RL application in maintenance and how critical the design of the reward function is.

This thesis proposed a case based on real data to test the three contributions of the proposed methodology. The case study is assumed to be representative and general enough that if the methodology is applied to other cases, it should work properly.

8.2 Research Limitations

The proposed research in this thesis has certain limitations. Those are related to ignoring imperfect maintenance actions, assuming fixed inspection frequency, the selection of the deterioration prediction model, and testing the methodology on a single case study.

All of the maintenance actions considered in this research are assumed to be perfect. This means that the actions have a guaranteed effect on the system. In reality, the effect of maintenance actions - especially multi-level repair actions - is not granted because of the different variables of human labor. In many cases, these actions may have less effect on the system deterioration condition; in other words, the effect of the action has a stochastic nature to a certain level and in this research, the effects of the action are assumed to be deterministic.

Also, the inspection intervals are assumed to have a fixed distribution through the age of the system. The inspection interval is selected based on the application of the case study and it has to be adjusted according to the application. The reason to fix the interval inspections is to separately test the effect of the new reward function design that encompasses the RUL. Flexible inspection intervals were proven to have positive effects on cost minimization.

The deterioration prediction model was based only on the random forest approach. Other prediction approaches as deep neural networks may result in different performance; this research does not study the effect of considering different prediction approaches. In this research, we mainly focused on the proof of concept. Using different prediction approaches may have certain improvements on the methodology performance, but this does not help much with the proof of concept. Also, the size of the data in the case study used in this research is not sufficient for training models as deep neural networks, and only a single case study is used to test the methodology.

Overall, and regardless of the discussed limitations, we believe that we have proposed a practical working methodology for CBM that can make use of widely available deterioration and maintenance data. This methodology should work and yield optimal results with other applications, beyond just the one proposed in the case study.

CHAPTER 9 CONCLUSION AND RECOMMENDATIONS

Condition-based maintenance has been receiving increasing attention from both the scientific and industrial sectors. This attention is driven by its superiority and applicability from both scientific and industrial points of view. Despite this increasing attention, whether the condition-based maintenance models are realistically applicable has been overlooked.

This research presents a generic data-driven methodology for optimizing multi-level CBM strategies that applies to real-world problems. The proposed methodology integrates several data-driven approaches and methods as prediction models, reinforcement learning, nonparametric survival approaches, and reliability-based estimation technique for the remaining useful life. The contribution of the proposed methodology has scientific and practical aspects. Scientifically, the proposed methodology solves an important problem in maintenance doming by adopting data-driven approaches; namely, machine learning-based approaches. Adopting a machine learning approach with industrial problems has a notable impact on efficiently enabling the use of widely available data from industry. Moreover, these approaches are not limited to certain assumptions that are practically hard to fulfill. Practically, the proposed methodology applies to real problems. This methodology considers aspects such as the different possible maintenance actions, treating the modeling and solution through machine learning methods that use available data. In addition, the proposed methodology is generic and can be applied to a wide range of real-world problems.

In Chapter 5, the first part of this thesis contribution, the proposed methodology integration between prediction models and reinforcement learning for optimizing condition-based maintenance strategies is proposed. Multi-level preventive repair actions are considered in this condition-based maintenance strategy to have contact with more realistic problems. This proposal does not suffer from the usual limitations of parameters and TPM estimating or assuming, and it is not related to a certain application. The proposed methodology is applied to real data on sheet metal culverts, and the optimality of the obtained strategy is tested.

In the second part of this thesis contribution, Chapter 6, the proposed methodology is enhanced by proposing integration between a nonparametric survival technique and reliability-based approach for remaining useful life estimation and adding it to the original methodology. The remaining useful life estimation is utilized in a novel way via a new design for the reinforcement learning reward function. Comparing the condition-based maintenance strategies that are based on this proposal with other strategies that use reward function based on cost only has shown that considering the remaining useful life as a part of the optimization criterion leads to more cost-effective maintenance strategies.

In the third part of this thesis contribution, Chapter 7, the previously proposed methodology is reformed to allow multi deteriorating failure modes for the same system to be considered. In this proposal, the overlooked problem of existing multi deteriorating failure modes and their dependency is addressed. The dependency is addressed from the deterioration perspective and the maintenance perspective. The reward function also considered the remaining useful life through a nonparametric survival technique that takes into account multi failure modes and a reliability-based approach. After applying the methodology to the case study, comprehensive comparisons are carried out to test all the anchor points of the proposed methodology. All of the comparisons performed showed that the obtained strategy based on the proposed methodology is the most cost-efficient in terms of minimizing the average maintenance cost.

The proposed research in this thesis opens the door for more aspects to be considered in the future: i. the effect of using different deterioration prediction models on the obtained CBM strategies. ii. The imperfect maintenance action effects on the RL learning process. iii. The improvement that could be achieved in the cost reduction if the inspection intervals are considered as a decision variable by the RL. iv. It would be of interest to have different real-world case studies with applications from different domains.

REFERENCES

AbdelHaleem, B., & Yacout, S. (1998). Simulation of components replacement policies for a fleet of military trucks. *Quality Engineering, 11*(2).

Aboura, K., Agbinya, J. I., & Eskandarian, A. (2014). Weibull Decision Support Systems in Maintenance. *Organizacija*, *47*(2), 81-89. https://doi.org/10.2478/orga-14-0008

Adsule, A., Kulkarni, M., & Tewari, A. (2020). Reinforcement learning for optimal policy learning in condition-based maintenance. *IET Collaborative Intelligent Manufacturing, 2*(4), 182-188. Research Article.

Ahmad, R., Kamaruddin, S. J. C., & Engineering, I. (2012). An overview of time-based and condition-based maintenance in industrial application. *63*(1), 135-149.

Ahsan, S., Lemma, T. A., & Gebremariam, M. (2020). Reliability analysis of gas turbine engine by means of bathtub-shaped failure rate distribution. *Process Safety Progress, 39*, e12115.

Ait-Kadi, D., & Chelbi, A. (2010). Inspection strategy for availability improvement. *Journal of Intelligent Manufacturing*, *21*(2), 231-235. https://doi.org/10.1007/s10845-008-0192-3

Alaswad, S., & Yisha, X. (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, *157*, 54-63. http://dx.doi.org/10.1016/j.ress.2016.08.009

Alves, F., Badikyan, H., Moreira, H. J. A., Azevedo, J., Moreira, P. M., Romero, L., & Leitão, P. (2020, 17-19 June 2020). *Deployment of a Smart and Predictive Maintenance System in an Industrial Case Study.* 2020 IEEE 29th International Symposium on Industrial Electronics (ISIE) (pp. 493-498). https://doi.org/10.1109/ISIE45063.2020.9152441

Apinantanapong, N., & Nivesrangsan, P. (2021). *Application of Artificial Neural Networks to Fault Detection of Rolling Ball Bearing.* 2021 6th International Conference on Business and Industrial Research (ICBIR), 20-21 May 2021, Piscataway, NJ, USA (pp. 147-152). https://doi.org/10.1109/ICBIR52339.2021.9465860

Bai, B., Li, Z., Wu, Q., Zhou, C., & Zhang, J. (2020). Fault data screening and failure rate prediction framework-based bathtub curve on industrial robots. *Industrial Robot: the international journal of robotics research application*.

Baldin, I. A. E. (1986). Technical diagnostic- and condition-based maintenance for better plant availability. *Measurement*, *4*(1), 7-22. https://doi.org/https://doi.org/10.1016/0263-2241(86)90024-2

Barde, S., Shin, H., & Yacout, S. (2016). *Opportunistic preventive maintenance strategy of a multi-component system with hierarchical structure by simulation and evaluation.* 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 1-8).

Barde, S., Yacout, S., & Hayong, S. (2019). Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. *Journal of Intelligent Manufacturing, 30*(1), 147-161. https://doi.org/10.1007/s10845-016-1237-7

Barlow, R., & Hunter, L. (1960). Optimum preventive maintenance policies. *Operations research, 8*(1), 90-100.

Blischke, W. R., & Murthy, D. P. (2003). *Case studies in reliability and maintenance* (Vol. 480). John Wiley & Sons.

Boros, E., Hammer, P. L., Ibaraki, T., Kogan, A., Mayoraz, E., & Muchnik, I. (2000). An implementation of logical analysis of data. *IEEE Transactions on knowledge and Data Engineering, 12*(2), 292-306.

Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2018). Review, analysis and synthesis of prognostic-based decision support methods for condition based maintenance. *Journal of Intelligent Manufacturing, 29*(6), 1303-1316. https://doi.org/10.1007/s10845-015-1179-5

Bracke, S. (2020). *Preventive maintenance planning based on Weibull distribution models: The impact of the random scatter behaviour of the threshold parameter.* 4th IFAC Workshop on Advanced Maintenance Engineering, Services and Technologies, AMEST 2020, September 10, 2020 - September 11, 2020, Cambridge, United kingdom (Vol. 53, pp. 161-166). https://doi.org/10.1016/j.ifacol.2020.11.026

Braga, J. A. P., & Andrade, A. R. (2019). Optimizing maintenance decisions in railway wheelsets: a Markov decision process approach. *Proceedings of the Institution of Mechanical Engineers, Part O (Journal of Risk and Reliability), 233*(2), 285-300.

Breslow, N. (1972). Discussion of the paper by DR Cox cited below. *Journal of the Royal Statistical Society, Series B, 34,* 187-220.

Breslow, N. E., & Wellner, J. A. (2007). Weighted likelihood for semiparametric models and two-phase stratified samples, with application to Cox regression. *Scandinavian Journal of Statistics, 34*(1), 86-102.

Bulla, C., & Birje, M. N. J. J. o. R. I. E. (2021). Improved data-driven root cause analysis in fog computing environment. 1-19.

Buoniu, L., de Bruin, T., Toli, D., Kober, J., & Palunko, I. (2018). Reinforcement learning for control: Performance, stability, and deep approximators. *Annual Reviews in Control, 46,* 8-28. https://doi.org/10.1016/j.arcontrol.2018.09.005

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*(4), 1823-1834. https://doi.org/10.1109/TPWRS.2010.2043269

Castanier, B., Grall, A., & Berenguer, C. (2005). A condition-based maintenance policy with non-periodic inspections for a two-unit series system. *Reliability Engineering & System Safety, 87*(1), 109-203. https://doi.org/10.1016/j.ress.2004.04.013

Chen, C., Liu, Y., Wang, S., Sun, X., Di Cairano-Gilfedder, C., Titmus, S., & Syntetos, A. A. (2020). Predictive maintenance using cox proportional hazard deep learning. *Advanced Engineering Informatics, 44*. https://doi.org/10.1016/j.aei.2020.101054

Chen, N., Ye, Z.-S., Xiang, Y., & Zhang, L. (2015). Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research*, *243*(1), 190-199. https://www.sciencedirect.com/science/article/pii/S0377221714009527

Cholette, M. E., Yu, H., Borghesani, P., Ma, L., & Kent, G. (2019). Degradation modeling and condition-based maintenance of boiler heat exchangers using gamma processes. *Reliability Engineering and System Safety, 183*, 184-196. https://doi.org/10.1016/j.ress.2018.11.023

Compare, M., Bellani, L., Cobelli, E., & Zio, E. (2018). Reinforcement learning-based flow management of gas turbine parts under stochastic failures. *International Journal of Advanced Manufacturing Technology, 99*, 2981-2992. http://dx.doi.org/10.1007/s00170-018-2690-6

Compare, M., Bellani, L., Cobelli, E., Zio, E., Annunziata, F., Carlevaro, F., & Sepe, M. (2020). A reinforcement learning approach to optimal part flow management for gas turbine maintenance. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 234*(1), 52-62. https://doi.org/10.1177/1748006X19869750

Cox, D. R. (1972). Regression Models and Life-Tables. *Journal of the Royal Statistical Society. Series B (Methodological), 34*(2), 187-220. http://www.jstor.org/stable/2985181

de Jonge, B., & Scarf, P. A. (2020). A review on maintenance optimization. *European Journal of Operational Research, 285*(3), 805-824. https://doi.org/10.1016/j.ejor.2019.09.047

de Jonge, B., Teunter, R., & Tinga, T. (2017). The influence of practical factors on the benefits of condition-based maintenance over time-based maintenance. *Reliability Engineering & System Safety, 158*, 21-30. https://doi.org/https://doi.org/10.1016/j.ress.2016.10.002

Deloux, E., Fouladirad, M., & Bérenguer, C. (2015). Health-and-usage-based maintenance policies for a partially observable deteriorating system. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability, 230*(1), 120-129. https://doi.org/10.1177/1748006X15609863

Dettenborn, T., Hartikainen, A., & Korkiala-Tanttu, L. (2020). Pavement Maintenance Threshold Detection and Network-Level Rutting Prediction Model Based on Finnish Road Data. *Journal of Infrastructure Systems, 26*(2). https://doi.org/10.1061/(ASCE)IS.1943-555X.0000539

Deutsch, J., & He, D. (2018). Using Deep Learning-Based Approach to Predict Remaining Useful Life of Rotating Components. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 48*(1), 11-20. https://doi.org/10.1109/TSMC.2017.2697842

Deutsch, J., He, M., & He, D. (2017). Remaining Useful Life Prediction of Hybrid Ceramic Bearings Using an Integrated Deep Learning and Particle Filter Approach. *Applied Sciences, 7*(7). https://doi.org/10.3390/app7070649

Diyin, T., Jinsong, Y., Xiongzi, C., & Makis, V. (2015). An optimal condition-based maintenance policy for a degrading system subject to the competing risks of soft and hard failure. *Computers & Industrial Engineering, 83*, 100-110. https://doi.org/10.1016/j.cie.2015.02.003

Diyin, T., Makis, V., Jafari, L., & Jinsong, Y. (2015). Optimal maintenance policy and residual life estimation for a slowly degrading system subject to condition monitoring. *Reliability Engineering & System Safety, 134*, 198-207. https://doi.org/10.1016/j.ress.2014.10.015

El Naqa, I., & Murphy, M. J. (2015). What is machine learning? In *machine learning in radiation oncology* (pp. 3-11). Springer.

Elsayed, E. A. (2003). Mean residual life and optimal operating conditions for industrial furnace tubes. In Wallace R. Blischke & D. N. P. Murthy (Eds.), *Case Studies in Reliability and Maintenance* (pp. 497-515).

Elsayed, E. A., Haitao, L., & Ling-Yau, C. (2006). Maintenance of continuously monitored degrading systems. *European Journal of Operational Research, 175*(2), 821-835. https://doi.org/10.1016/j.ejor.2005.05.017

Elsheikh, A., Yacout, S., & Ouali, M.-S. J. N. (2019). Bidirectional handshaking LSTM for remaining useful life prediction. *323*, 148-156.

Elwany, A. H., Gebraeel, N. Z., & Maillart, L. M. (2011). Structured replacement policies for components with complex degradation processes and dedicated sensors. *Operations Research*, *59*(3), 684-695. https://doi.org/10.1287/opre.1110.0912

Falamarzi, A., Moridpour, S., & Nazem, M. (2019). A review of rail track degradation prediction models. *Australian Journal of Civil Engineering, 17*(2), 152-166. https://doi.org/10.1080/14488353.2019.1667710

Fan, R., Feng, D., Yang, C., & Lin, S. (2019). *Optimization of Condition-based Maintenance for Traction Power Supply Equipment based on Partially Observable Semi-Markov Decision Process.* 10 th Prognostics and System Health Management Conference, PHM-Qingdao 2019, October 25-27, 2019 Qingdao, China. https://doi.org/10.1109/PHM-Qingdao46334.2019.8942840

Farnoosh, N. Z. G., & Makis, V. (2015). Optimal condition-based maintenance policy for a partially observable system with two sampling intervals. *International Journal of Advanced Manufacturing Technology*, *78*(5-8), 795-805. https://doi.org/10.1007/s00170-014-6651-4

Fouladirad, M., & Grall, A. (2014). On-line change detection and condition-based maintenance for systems with unknown deterioration parameters. *IMA Journal of Management Mathematics, 25*(2), 139-158. https://doi.org/10.1093/imaman/dps032

Fouladirad, M., & Grall, A. (2015). Monitoring and condition-based maintenance with abrupt change in a system's deterioration rate. *International Journal of Systems Science, 46*(12), 2183-2194. https://doi.org/10.1080/00207721.2013.859327

Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning. In (Vol. 1, pp. 303-332). Springer series in statistics New York.

Fu, Y., Zhu, X., & Ma, X. (2020). Optimum component reallocation and system replacement maintenance for a used system with increasing minimal repair cost. *Reliability Engineering and System Safety*, *204*. https://doi.org/10.1016/j.ress.2020.107137

Gao, K., Peng, R., Qu, L., & Wu, S. (2020). Jointly optimizing lot sizing and maintenance policy for a production system with two failure modes. *Reliability Engineering and System Safety, 202*. https://doi.org/10.1016/j.ress.2020.106996

Ghasemi, A., & Esmaeili, S. (2015). An Optimal Equipment Replacement Model Using Logical Analysis of Data. *International Journal of Strategic Decision Sciences, 6*(2), 83-96. https://doi.org/10.4018/ijsds.2015040105

Golmakani, H. R. (2020). A novel approach for optimal replacement threshold and inspection scheme for condition based maintenance. https://doi.org/10.1108/JQME-05-2020-0036

Guo, C., Wang, W., Guo, B., & Si, X. (2013). A maintenance optimization model for mission-oriented systems based on Wiener degradation. *Reliability Engineering & System Safety, 111*, 183-194.

Guo, J., Wang, C., Cabrera, J., & Elsayed, E. A. (2018). Improved inverse Gaussian process and bootstrap: Degradation and reliability metrics. *Reliability Engineering & System Safety, 178*, 269-277. https://doi.org/https://doi.org/10.1016/j.ress.2018.06.013

Hafez, M., Ksaibati, K., & Atadero, R. A. (2019). Optimizing Expert-Based Decision-Making of Pavement Maintenance using Artificial Neural Networks with Pattern-Recognition Algorithms. https://doi.org/10.1177/0361198119851085

Hamaide, V., & Glineur, F. (2019). *Predictive Maintenance of a Rotating Condenser Inside a Synchrocyclotron.* BNAIC/BENELEARN.

Hamidi, M., Szidarovszky, F., & Szidarovszky, M. (2016). New one cycle criteria for optimizing preventive replacement policies. *Reliability Engineering & System Safety, 154*, 42-48. https://doi.org/10.1016/j.ress.2016.04.010

Han, X., Xia, Y., Ye, F., & Wang, Y. (2020). Ageing models and maintenance strategy for road tunnels. *Structure and Infrastructure Engineering, 16*(5), 831-846. https://doi.org/10.1080/15732479.2019.1670680

He, Y., Gu, C., Chen, Z., & Han, X. (2017). Integrated predictive maintenance strategy for manufacturing systems by combining quality control and mission reliability analysis. *International Journal of Production Research, 55*(19), 5841-5862. https://doi.org/10.1080/00207543.2017.1346843

Hong, H. P., Zhou, W., Zhang, S., & Ye, W. (2014). Optimal condition-based maintenance decisions for systems with dependent stochastic degradation of components. *Reliability Engineering & System Safety, 121*, 276-288. https://doi.org/10.1016/j.ress.2013.09.004

Hong, P., Xiaosheng, S., Changhua, H., Zhaoqiang, W., Dangbo, D., & Zhenan, P. (2018). A multi-stage Wiener process-based prognostic model for equipment considering the influence of imperfect maintenance activities. *Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology, 34*(6), 3695-3705. https://doi.org/10.3233/JIFS-169544

Hongzhou, W. (2002). A survey of maintenance policies of deteriorating systems. *European Journal of Operational Research, 139*(3), 469-489. https://doi.org/10.1016/S0377-2217(01)00197-7

Hu, J., & Zhang, L. (2014). Risk based opportunistic maintenance model for complex mechanical systems. *Expert Systems with Applications, 41*(6), 3105-3115. https://doi.org/https://doi.org/10.1016/j.eswa.2013.10.041

Hu, Y., Miao, X., Zhang, J., Liu, J., & Pan, E. (2021). Reinforcement learning-driven maintenance strategy: A novel solution for long-term aircraft maintenance decision optimization. *Computers and Industrial Engineering, 153*. https://doi.org/10.1016/j.cie.2020.107056

Huang, J., Chang, Q., & Arinez, J. (2020). Deep reinforcement learning based preventive maintenance policy for serial production lines. *Expert Systems with Applications, 160*. https://doi.org/10.1016/j.eswa.2020.113701

Huy, T.-B., Borghesani, P., Cholette, M. E., & Lin, M. (2020). Optimization of condition-based maintenance considering partial opportunities. *Quality and Reliability Engineering International, 36*(2), 529-546. https://doi.org/10.1002/qre.2585

Huynh, K. T. (2021). An adaptive predictive maintenance model for repairable deteriorating systems using inverse Gaussian degradation process. *Reliability Engineering and System Safety, 213*. https://doi.org/10.1016/j.ress.2021.107695

Huynh, K. T., Barros, A., Bérenguer, C., & Castro, I. T. (2011). A periodic inspection and replacement policy for systems subject to competing failure modes due to degradation and traumatic events. *Reliability Engineering & System Safety, 96*(4), 497-508. https://doi.org/https://doi.org/10.1016/j.ress.2010.12.018

Huynh, K. T., Castro, I. T., Barros, A., & Bérenguer, C. (2014). On the Use of Mean Residual Life as a Condition Index for Condition-Based Maintenance Decision-Making. *IEEE Transactions on Systems, Man, and Cybernetics: Systems, 44*(7), 877-893. https://doi.org/10.1109/TSMC.2013.2290772

Huynh, K. T., Grall, A., & Bérenguer, C. (2019). A Parametric Predictive Maintenance Decision-Making Framework Considering Improved System Health Prognosis Precision. *IEEE Transactions on Reliability, 68*(1), 375-396. https://doi.org/10.1109/TR.2018.2829771

Ingemarsdotter, E., Kambanou, M. L., Jamsin, E., Sakao, T., & Balkenende, R. (2021). Challenges and solutions in condition-based maintenance implementation - A multiple case study. *Journal of Cleaner Production, 296*. https://doi.org/10.1016/j.jclepro.2021.126420

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). Tree based methods. In *An introduction to statistical learning* (Vol. 112, pp. 303-332). Springer.

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing, 20*(7), 1483-1510. https://doi.org/https://doi.org/10.1016/j.ymssp.2005.09.012

Jian, Z., Xiaoyan, H., Youtong, F., Jing, Z., He, Z., & Jing, L. (2016). Optimal inspection-based preventive maintenance policy for three-state mechanical components under competing failure modes. *Reliability Engineering & System Safety, 152*, 95-103. https://doi.org/10.1016/j.ress.2016.02.007

Jingjing, W., Makis, V., & Xian, Z. (2019). Optimal condition-based and age-based opportunistic maintenance policy for a two-unit series system. *Computers and Industrial Engineering, 134*, 1-10. https://doi.org/10.1016/j.cie.2019.05.020

Jocelyn, S., Chinniah, Y., Ouali, M.-S., & Yacout, S. (2020). *Application of logical analysis of data to prevent machinery-related accidents.* 2016 Industrial and Systems Engineering Research Conference, ISERC 2016, May 21, 2016 - May 24, 2016, Anaheim, CA, United states (pp. 813-818).

Kaparthi, S., & Bumblauskas, D. (2020). Designing predictive maintenance systems using decision tree-based machine learning techniques. *International Journal of Quality & Reliability Management, 37*(4), 659-686. https://doi.org/10.1108/IJQRM-04-2019-0131

Khaleghei, A., & Makis, V. (2015). Reliability estimation of a production system subject to condition monitoring with two modes of failures. *IFAC-PapersOnLine*, *48*(3), 472-477. https://doi.org/https://doi.org/10.1016/j.ifacol.2015.06.126

Khan, K., Sohaib, M., Rashid, A., Ali, S., Akbar, H., Basit, A., . . . Engineering. (2021). Recent trends and challenges in predictive maintenance of aircraft's engine and hydraulic system. *J Journal of the Brazilian Society of Mechanical Sciences, 43*(8), 1-17.

Khorsheed, R. M., & Beyca, O. F. (2020). An integrated machine learning: Utility theory framework for real-time predictive maintenance in pumping systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture, 235*(5), 887-901. https://doi.org/10.1177/0954405420970517

Kim, J., Ahn, Y., & Yeo, H. (2016). A comparative study of time-based maintenance and condition-based maintenance for optimal choice of maintenance policy. *Structure and Infrastructure Engineering*, *12*(12), 1525-1536. https://doi.org/10.1080/15732479.2016.1149871

Klein, J. P., & Moeschberger, M. L. (2006). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Klutke, G. A., & Yoonjung, Y. (2002). The availability of inspected systems subject to shocks and graceful degradation. *IEEE Transactions on Reliability, 51*(3), 371-374. https://doi.org/10.1109/TR.2002.802891

Kuhnle, A., Jakubik, J., & Lanza, G. (2019). Reinforcement learning for opportunistic maintenance optimization. *Production Engineering - Research and Development, 13*(1), 33-41. https://doi.org/10.1007/s11740-018-0855-7

Kumar, G., Jain, V., & Gandhi, O. P. (2018). Availability analysis of mechanical systems with condition-based maintenance using semi-Markov and evaluation of optimal condition monitoring interval. *Journal of Industrial Engineering International, 14*(1), 119-131. https://doi.org/10.1007/s40092-017-0212-z

Kurt, M., & Kharoufeh, J. P. (2010). Monotone optimal replacement policies for a Markovian deteriorating system in a controllable environment. *Operations Research Letters, 38*(4), 273-279. https://doi.org/10.1016/j.orl.2010.03.001

Leukel, J., González, J., & Riekert, M. (2021). Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. *Journal of Manufacturing Systems, 61*, 87-96. https://doi.org/https://doi.org/10.1016/j.jmsy.2021.08.012

Li, X., Makis, V., Zhao, Z., Zuo, H., Duan, C., & Zhang, Y. (2019). Optimal Bayesian maintenance policy for a gearbox subject to two dependent failure modes. *Quality and Reliability Engineering International*, *35*(2), 659-676. https://doi.org/10.1002/qre.2428

Li, X., Makis, V., Zuo, H., & Cai, J. (2018). Optimal Bayesian control policy for gear shaft fault detection using hidden semi-Markov model. *Computers and Industrial Engineering, 119*, 21-35. https://doi.org/10.1016/j.cie.2018.03.026

Li, Y., Xiaobing, M., & Yu, Z. (2017). A condition-based maintenance model for a three-state system subject to degradation and environmental shocks. *Computers & Industrial Engineering, 105,* 210-222. https://doi.org/10.1016/j.cie.2017.01.012

Liang, Z., Liu, B., Xie, M., & Parlikad, A. K. (2019). Condition-based maintenance for long-life assets with exposure to operational and environmental risks. *International Journal of Production Economics*. https://doi.org/10.1016/j.ijpe.2019.09.003

Lin, D. (2007). On the Breslow estimator. *Lifetime data analysis, 13*(4), 471-480.

Lin, D., Banjevic, D., & Jardine, A. K. (2006). Using principal components in a proportional hazards model with applications in condition-based maintenance. *Journal of the Operational Research Society, 57*(8), 910-919.

Lin, L., Luo, B., & Zhong, S. (2018). Multi-objective decision-making model based on CBM for an aircraft fleet with reliability constraint. *International Journal of Production Research, 56*(14), 4831-4848.

Lin, L., Luo, B., & Zhong, S. (2019). Maintenance decision-making based on condition-based maintenance for fleet. *Computer Integrated Manufacturing Systems, 25*(3), 661-672. https://doi.org/10.13196/j.cims.2019.03.013

Lin, S., Fan, R., Feng, D., Yang, C., Wang, Q., & Gao, S. (2020). Condition-Based Maintenance for Traction Power Supply Equipment Based on Partially Observable Markov Decision Process. *IEEE Transactions on Intelligent Transportation Systems*, 1-15. https://doi.org/10.1109/TITS.2020.3009175

Ling, Z., Wang, X., & Qu, F. (2018). *Reinforcement Learning-Based Maintenance Scheduling for Resource Constrained Flow Line System.* 4th IEEE International Conference on Control Science and Systems Engineering, ICCSSE 2018, August 21, 2018 - August 23, 2018, Wuhan, China (pp. 364-369). https://doi.org/10.1109/CCSSE.2018.8724807

Lior, R. (2014). *Data mining with decision trees: theory and applications* (Vol. 81). World scientific.

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. https://doi.org/10.1016/j.ejor.2021.03.010

Liu, B., Xie, M., Xu, Z., & Kuo, W. (2016). An imperfect maintenance policy for mission-oriented systems subject to degradation and external shocks. *Computers & Industrial Engineering, 102,* 21-32. https://doi.org/https://doi.org/10.1016/j.cie.2016.10.008

Liu, Q., Dong, M., Lv, W., & Ye, C. (2019). Manufacturing system maintenance based on dynamic programming model with prognostics information. *Journal of Intelligent Manufacturing, 30*(3), 1155-1173. http://dx.doi.org/10.1007/s10845-017-1314-6

Liu, Y., Chen, Y., & Jiang, T. (2019). Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. *European Journal of Operational Research*

Lo, H., Ghasemi, A., Diallo, C., & Newhook, J. (2019). Comparative study on logical analysis of data (LAD), artificial neural networks (ANN), and proportional hazards model (PHM) for maintenance prognostics. *Journal of Quality in Maintenance Engineering*, *25*(1), 2-24. https://doi.org/10.1108/JQME-07-2017-0051

Lopes Gerum, P. C., Altay, A., & Baykal-Gürsoy, M. (2019). Data-driven predictive maintenance scheduling policies for railways. *Transportation Research Part C: Emerging Technologies, 107,* 137-154. https://doi.org/https://doi.org/10.1016/j.trc.2019.07.020

Lu, C., Xu, T.-X., & Cong, L.-H. (2019). Condition-based maintenance decision based on inverse Gaussian deterioration process under the condition of regular detection and maintenance. *Journal of Intelligent and Fuzzy Systems, 37*(4), 5767-5775. https://doi.org/10.3233/JIFS-181580

Malek, A., Abbasi-Yadkori, Y., & Bartlett, P. (2014). *Linear programming for large-scale Markov decision problems.* International Conference on Machine Learning (pp. 496-504).

Mandiartha, P., Duffield, C. F., Razelan, I. S. B. M., & Ismail, A. B. H. (2017). *Pavement maintenance optimization model using Markov decision processes.* 1st International Conference on Applied & Industrial Mathematics and Statistics 2017 (ICoAIMS), 8-10 Aug. 2017, UK (Vol. 890, pp. 012104 (012106 pp.)). https://doi.org/10.1088/1742-6596/890/1/012104

Mandiartha, P., Duffield, C. F., Thompson, R. G., & Wigan, M. R. (2017). Measuring pavement maintenance effectiveness using Markov Chains analysis. *Structure and Infrastructure Engineering*, *13*(7), 844-854. https://doi.org/10.1080/15732479.2016.1212901

Mattila, V., & Virtanen, K. (2011). *Scheduling fighter aircraft maintenance with reinforcement learning.* Winter Simulation Conference (WSC 2011), 11-14 Dec. 2011, Piscataway, NJ, USA (pp. 2535-2546). http://dx.doi.org/10.1109/WSC.2011.6147962

Mei, W., & Jie, W. (2012). CHMM for tool condition monitoring and remaining useful life prediction. *International Journal of Advanced Manufacturing Technology, 59*(5-8), 463-471. https://doi.org/10.1007/s00170-011-3536-7

Mimi, Z., Zhisheng, Y., & Min, X. (2014). A condition-based maintenance strategy for heterogeneous populations. *Computers & Industrial Engineering*, *77*, 103-114. https://doi.org/10.1016/j.cie.2014.09.001

Mingcheng, E., Bing, L., Zengqiang, J., & Qi, L. (2018). An optimal reprofiling policy for high-speed train wheels subject to wear and external shocks using a semi-Markov decision process. *IEEE Transactions on Reliability, 67*(4), 1468-1481. https://doi.org/10.1109/TR.2018.2870276

Minghui, W., Xuemin, W., & Xianzhong, L. (2020). *On Condition Maintenance Model for Complex Electromechanical Equipments Based on Remaining Useful Life and Wiener Process.* 2020 International Conference on Mechanical Engineering and Intelligent Manufacturing, 18-20 Sept. 2020, UK (Vol. 1678, pp. 012014 (012013 pp.)). https://doi.org/10.1088/1742-6596/1678/1/012014

Mortada, M.-A., Yacout, S., & Lakis, A. (2014). Fault diagnosis in power transformers using multi-class logical analysis of data. *Journal of Intelligent Manufacturing*, *25*(6), 1429-1439.

Muhamad, N. A., Visa Musa, I., Abdul Malek, Z., & Salah Mahdi, A. (2020). Classification of Partial Discharge Fault Sources on SF6 Insulated Switchgear Based on Twelve By-Product Gases Random Forest Pattern Recognition. *IEEE Access*, *8*, 212659-212674. https://doi.org/10.1109/ACCESS.2020.3040421

Nan, C., Zhi-Sheng, Y., Yisha, X., & Linmiao, Z. (2015). Condition-based maintenance using the inverse Gaussian degradation model. *European Journal of Operational Research, 243*(1), 190-199. https://doi.org/10.1016/j.ejor.2014.11.029

Nowaczyk, S., Prytz, R., Rögnvaldsson, T., & Byttner, S. (2013). *Towards a machine learning algorithm for predicting truck compressor failures using logged vehicle data.* 12th Scandinavian Conference on Artificial Intelligence, Aalborg, Denmark, November 20–22, 2013 (pp. 205-214).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Dubourg, V. J. t. J. o. m. L. r. (2011). Scikit-learn: Machine learning in Python. *12*, 2825-2830.

Peng, Y., Dong, M., & Zuo, M. J. (2010). Current status of machine prognostics in condition-based maintenance: A review. *International Journal of Advanced Manufacturing Technology, 50*(1-4), 297-313. https://doi.org/10.1007/s00170-009-2482-0

Phuc, D., Iung, B., & Cavalcante, C. (2019). *Reliability and Maintenance Cost Forecasting for Systems with Multistate Components Using Artificial Neural Networks.* 2019 4th International Conference on System Reliability and Safety (ICSRS), 20-22 Nov. 2019, Piscataway, NJ, USA (pp. 181-185). https://doi.org/10.1109/ICSRS48664.2019.8987700

Phuc, D., Voisin, A., Levrat, E., & Iung, B. (2015). A proactive condition-based maintenance strategy with both perfect and imperfect maintenance actions. *Reliability Engineering & System Safety, 133*, 22-32. https://doi.org/10.1016/j.ress.2014.08.011

Pitis, S. (2019). *Rethinking the discount factor in reinforcement learning: A decision theoretic approach.* Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 33, pp. 7949-7956).

Puterman, M. L. (2014). *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons.

Qingan, Q., Lirong, C., & Hongda, G. (2017). Availability and maintenance modelling for systems subject to multiple failure modes. *Computers & Industrial Engineering, 108*, 192-198. https://doi.org/10.1016/j.cie.2017.04.028

Qiuyu, Z., & Meiju, L. (2021). *An Optimal Maintenance Cycle Decision of Relay Protection Device Based on Weibull Distribution Model.* Second International Symposium on Simulation and Process Modelling (ISSPM 2020), 29-30 Aug. 2020, Singapore, Singapore (pp. 409-418). https://doi.org/10.1007/978-981-33-4575-1_39

Quatrini, E., Costantino, F., Di Gravio, G., & Patriarca, R. (2020). Condition-Based Maintenance—An Extensive Literature Review. *8*(2), 31. https://www.mdpi.com/2075-1702/8/2/31

Rabbanian, S. S., Nemati, M., & Knapp, G. M. (2021). *A deep reinforcement learning approach for maintenance planning.* IISE Annual Conference and Expo 2021, May 22, 2021 - May 25, 2021, Virtual, Online (pp. 932-937).

Ragab, A., Yacout, S., Ouali, M.-S., & Osman, H. (2019). Prognostics of multiple failure modes in rotating machinery using a pattern-based classifier and cumulative incidence functions. *Journal of Intelligent Manufacturing*, *30*(1), 255-274.

Rahimikelarijani, B., Hamidi, M., Mohassel, A., & Craig, B. (2020). Imperfect Condition-Based Maintenance Strategy for a Deteriorating Rail Track System with Multiple Competitive Failure Modes. *Journal of Transportation Engineering, Part A: Systems*, *146*(9), 04020088 (04020012 pp.). https://doi.org/10.1061/JTEPBS.0000412

Ramírez-Hernández, J. A., & Fernandez, E. (2007). *An approximate dynamic programming approach for job releasing and sequencing in a reentrant manufacturing line.* 2007 IEEE International Symposium on Approximate Dynamic Programming and Reinforcement Learning (pp. 201-208).

Renqing, L., Xiaoxi, L., Yan, S., Zigang, C., & Jin, L. (2017). *An optimal condition based maintenance strategy using inverse-Gaussian degradation process.* 2017 Second International Conference on Reliability Systems Engineering (ICRSE), 10-12 July 2017, Piscataway, NJ, USA (pp. R075 (075 pp.)). https://doi.org/10.1109/ICRSE.2017.8030762

Rocchetta, R., Bellani, L., Compare, M., Zio, E., & Patelli, E. (2019). A reinforcement learning framework for optimal operation and maintenance of power grids. *Applied Energy*, *241*, 291-301. https://doi.org/10.1016/j.apenergy.2019.03.027

Rui, Z., & Makis, V. (2020). Optimal condition-based maintenance with general repair and two dependent failure modes. *Computers & Industrial Engineering*, *141*, 282-292. https://doi.org/10.1016/j.cie.2020.106322

Sakib, N., & Wuest, T. (2018). *Challenges and Opportunities of Condition-based Predictive Maintenance: A Review.* 6th CIRP Global Web Conference, CIRPe 2018, October 23, 2018 - October 25, 2018 (Vol. 78, pp. 267-272). https://doi.org/10.1016/j.procir.2018.08.318

Salih, A. (2020). *Optimizing the Maintenance Strategy Through Predicting Remaining Useful Life (RUL) for Critical Medical Equipment Based on Degradation Analysis.* Advances in Asset Management and Condition Monitoring. COMADEM 2019, 3-5 Sept. 2019, Cham, Switzerland (pp. 1483-1494). https://doi.org/10.1007/978-3-030-57745-2_120

Sancho, L. C. B., Braga, J. A. P., & Andrade, A. R. (2021). Optimizing Maintenance Decision in Rails: A Markov Decision Process Approach. *ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering*, *7*(1). https://doi.org/10.1061/AJRUA6.0001101

Sanner, S., & Boutilier, C. (2012). Approximate linear programming for first-order MDPs. *arXiv preprint arXiv:*.

Santolamazza, A., Cesarotti, V., & Introna, V. (2018). *Anomaly detection in energy consumption for Condition-Based maintenance of Compressed Air Generation systems: an approach based on artificial neural networks.* 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018: Bergamo, Italy, 11-13 June 2018 (Vol. 51, pp. 1131-1136). https://doi.org/10.1016/j.ifacol.2018.08.439

Sarazin, A., Bascans, J., Sciau, J.-B., Song, J., Supiot, B., Montarnal, A., . . . Truptil, S. (2021). Expert system dedicated to condition-based maintenance based on a knowledge graph approach: Application to an aeronautic system. *Expert Systems with Applications, 186*, 115767. https://doi.org/https://doi.org/10.1016/j.eswa.2021.115767

Sewak, M. (2019). Temporal Difference Learning, SARSA, and Q-Learning. In *Deep Reinforcement Learning* (pp. 51-63). Springer.

Sgarbossa, F., Zennaro, I., Florian, E., & Persona, A. (2018). *Impacts of weibull parameters estimation on preventive maintenance cost.* 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018: Bergamo, Italy, 11-13 June 2018 (Vol. 51, pp. 508-513). https://doi.org/10.1016/j.ifacol.2018.08.369

Shafiee, M., & Finkelstein, M. (2015). An optimal age-based group maintenance policy for multi-unit degrading systems. *Reliability Engineering & System Safety, 134*, 230-238. https://doi.org/https://doi.org/10.1016/j.ress.2014.09.016

Sharifi, M., & Taghipour, S. (2020). *Joint optimizing the Production Sequence and Maintenance Plan for a Single-Machine Multi-Failure System.* 2020 Annual Reliability and Maintainability Symposium (RAMS), 27-30 Jan. 2020, Piscataway, NJ, USA (pp. 6 pp.). https://doi.org/10.1109/RAMS48030.2020.9153629

Sharifi, M., & Taghipour, S. (2021). Optimal production and maintenance scheduling for a degrading multi-failure modes single-machine production environment. *Applied Soft Computing, 106*. https://doi.org/10.1016/j.asoc.2021.107312

Shey-Huei, S., Griffith, W. S., & Nakagawa, T. (1995). Extended optimal replacement model with random minimal repair costs. *European Journal of Operational Research, 85*(3), 636-649. http://dx.doi.org/10.1016/0377-2217(93)E0364-4

Shin, J.-H., & Jun, H.-B. (2015). On condition based maintenance policy. *Journal of Computational Design and Engineering, 2*(2), 119-127. https://doi.org/10.1016/j.jcde.2014.12.006

Silva, W., & Capretz, M. (2019, 8-11 July 2019). *Assets Predictive Maintenance Using Convolutional Neural Networks.* 2019 20th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD) (pp. 59-66). https://doi.org/10.1109/SNPD.2019.8935752

Soonsung, H., Jongpil, J., & Youngbin, K. (2018). *SVM-RBM based Predictive Maintenance Scheme for IoT-enabled Smart Factory.* 2018 Thirteenth International Conference on Digital Information Management (ICDIM), 24-26 Sept. 2018, Piscataway, NJ, USA (pp. 162-167). https://doi.org/10.1109/ICDIM.2018.8847132

Sutton, R. S., & Barto, A. G. (2011). Tabular Solution Methods Chapter 4-6. In *Reinforcement learning: An introduction*

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Tan, Y., Niu, C., Tian, H., Hou, L., & Zhang, J. (2019). A one-class SVM based approach for condition-based maintenance of a naval propulsion plant with limited labeled data. *Ocean Engineering, 193*. https://doi.org/10.1016/j.oceaneng.2019.106592

Tangbin, X., Lifeng, X., Ershun, P., & Jun, N. (2017). Reconfiguration-oriented opportunistic maintenance policy for reconfigurable manufacturing systems. *Reliability Engineering & System Safety*, *166*, 87-98. https://doi.org/10.1016/j.ress.2016.09.001

Therneau, T., & Lumley, T. (2013). *R survival package*. https://cran.r-project.org/web/packages/survival/survival.pdf

Tran, V. T., & Yang, B.-S. (2012). An intelligent condition-based maintenance platform for rotating machinery. *Expert Systems with Applications, 39*(3), 2977-2988. https://doi.org/https://doi.org/10.1016/j.eswa.2011.08.159

Van Otterlo, M., & Wiering, M. (2012). *Reinforcement learning and markov decision processes*. Springer.

Vu, H. C., Do, P., & Barros, A. (2016). A Stationary Grouping Maintenance Strategy Using Mean Residual Life and the Birnbaum Importance Measure for Complex Structures. *IEEE Transactions on Reliability*, *65*(1), 217-234. https://doi.org/10.1109/TR.2015.2455498

Waghen, K., & Ouali, M.-S. J. E. S. w. A. (2019). Interpretable logic tree analysis: A data-driven fault tree methodology for causality analysis. *136*, 376-391.

Wang, H. (2002). A survey of maintenance policies of deteriorating systems. *J European journal of operational research, 139*(3), 469-489.

Wang, H., Liao, H., Ma, X., & Bao, R. (2021). Remaining Useful Life Prediction and Optimal Maintenance Time Determination for a Single Unit Using Isotonic Regression and Gamma Process Model. *Reliability Engineering and System Safety, 210*. https://doi.org/10.1016/j.ress.2021.107504

Wang, J., Makis, V., & Zhao, X. (2019). Optimal condition-based and age-based opportunistic maintenance policy for a two-unit series system. *Computers and Industrial Engineering, 134*, 1-10. https://doi.org/10.1016/j.cie.2019.05.020

Wang, X., Wang, H., & Qi, C. (2016). Multi-agent reinforcement learning based maintenance policy for a resource constrained flow line system. *Journal of Intelligent Manufacturing, 27*(2), 325-333. https://doi.org/10.1007/s10845-013-0864-5

Wei, S., Bao, Y., & Li, H. (2020). Optimal policy for structure maintenance: A deep reinforcement learning framework. *Structural Safety, 83*. https://doi.org/10.1016/j.strusafe.2019.101906

Wiering, M., & Van Otterlo, M. (2012). *Reinforcement learning* (Vol. 12).

Wu, F., Niknam, S. A., & Kobza, J. E. (2015). A cost effective degradation-based maintenance strategy under imperfect repair. *Reliability Engineering & System Safety, 144*, 234-243. https://doi.org/https://doi.org/10.1016/j.ress.2015.08.002

Wu, Z., Guo, B., Axita, Tian, X., & Zhang, L. (2020). A Dynamic Condition-Based Maintenance Model Using Inverse Gaussian Process. *IEEE Access, 8*, 104-117. https://doi.org/10.1109/ACCESS.2019.2958137

Xanthopoulos, A. S., Kiatipis, A., Koulouriotis, D. E., & Stieger, S. (2018). Reinforcement learning-based and parametric production-maintenance control policies for a deteriorating manufacturing system. *IEEE Access, 6*, 576-588. https://doi.org/10.1109/ACCESS.2017.2771827

Xia, F., Ning, J., Huang, X., & biostatistics. (2018). Empirical comparison of the Breslow Estimator and the Kalbfleisch Prentice Estimator for survival functions. *Journal of biometrics, 9*(2).

Xiang, S., Huang, D., & Li, X. (2018, 28-31 Oct. 2018). *A Generalized Predictive Framework for Data Driven Prognostics and Diagnostics using Machine Logs.* TENCON 2018 - 2018 IEEE Region 10 Conference (pp. 0695-0700). https://doi.org/10.1109/TENCON.2018.8650152

Xiao, W., Hongwei, W., & Chao, Q. (2016). Multi-agent reinforcement learning based maintenance policy for a resource constrained flow line system. *Journal of Intelligent Manufacturing, 27*(2), 325-333. https://doi.org/10.1007/s10845-013-0864-5

Xuejing, Z., Fouladirad, M., Berenguer, C., & Bordes, L. (2010). Condition-based inspection/replacement policies for non-monotone deteriorating systems with environmental covariates. *Reliability Engineering & System Safety, 95*(8), 921-934. https://doi.org/10.1016/j.ress.2010.04.005

Yacout, S. (2010). *Fault detection and diagnosis for condition based maintenance using the logical analysis of data.* The 40th International Conference on Computers & Indutrial Engineering (pp. 1-6).

Yang, L., Ma, X., & Zhao, Y. (2017). A condition-based maintenance model for a three-state system subject to degradation and environmental shocks. *Computers and Industrial Engineering, 105*, 210-222. https://doi.org/10.1016/j.cie.2017.01.012

Yousefi, N., Tsianikas, S., & Coit, D. W. (2020). Reinforcement learning for dynamic condition-based maintenance of a system with individually repairable components. *Quality Engineering, 32*(3), 388-408. https://doi.org/10.1080/08982112.2020.1766692

Zeyang, T., Chengke, Z., Wei, J., Wenjun, Z., Xiaoping, J., Jianhui, Y., . . . Bojie, S. (2014). Analysis of Significant Factors on Cable Failure Using the Cox Proportional Hazard Model. *IEEE Transactions on Power Delivery, 29*(2), 951-957. https://doi.org/10.1109/TPWRD.2013.2287025

Zhang, C., & Tee, K. F. J. E. P. (2019). Application of gamma process and maintenance cost for fatigue damage of wind turbine blade. *158*, 3729-3734.

Zhang, N., & Si, W. (2020). Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks. *Reliability Engineering & System Safety, 203*, 107094. https://doi.org/https://doi.org/10.1016/j.ress.2020.107094

Zhang, N., & Yang, Q. (2015). Optimal maintenance planning for repairable multi-component systems subject to dependent competing risks. *IIE Transactions, 47*(5), 521-532. https://doi.org/10.1080/0740817X.2014.974115

Zhang, S.-x., Zan, X., Li, H., & Han, C.-s. (2015). Weibull proportional hazards model based determination of the test intervals for condition based maintenance. *Fire Control and Command Control, 40*(7), 151-154.

Zhang, X., & Gao, H. (2012). Road maintenance optimization through a discrete-time semi-Markov decision process. *Reliability Engineering and System Safety*, *103*, 110-119. https://doi.org/10.1016/j.ress.2012.03.011

Zhang, X., Li, Y., & Wang, X. (2017). Maintenance strategy of corroded oil-gas pipeline based on inverse Gaussian process. *Shiyou Xuebao/Acta Petrolei Sinica*, *38*(3), 356-362. https://doi.org/10.7623/syxb201703013

Zhao, X., Yang, J., & Qin, Y. (2021). Optimal Condition-Based Maintenance Strategy via an Availability-Cost Hybrid Factor for a Single-Unit System During a Two-Stage Failure Process. *IEEE Access, 9*, 45968-45977.

Zhen, C., Yaping, L., Tangbin, X., & Ershun, P. (2019). Hidden Markov model with auto-correlated observations for remaining useful life prediction and optimal maintenance policy. *Reliability Engineering & System Safety*, *184*, 123-136. https://doi.org/10.1016/j.ress.2017.09.002

Zheng, R., & Makis, V. (2020). Optimal condition-based maintenance with general repair and two dependent failure modes. *Computers and Industrial Engineering*, *141*. https://doi.org/10.1016/j.cie.2020.106322

Zhenyu, L., Xiaobing, M., Jun, Y., & Yu, Z. (2014). Reliability Modeling for Systems with Multiple Degradation Processes Using Inverse Gaussian Process and Copulas. *Mathematical Problems in Engineering*, 829597 (829510 pp.). https://doi.org/10.1155/2014/829597

Zhou, P., & Yin, P. T. (2019). An opportunistic condition-based maintenance strategy for offshore wind farm based on predictive analytics. *Renewable and Sustainable Energy Reviews, 109*, 1-9. https://doi.org/10.1016/j.rser.2019.03.049

Zhou, Z.-J., Hu, C.-H., Wang, W.-B., Zhang, B.-C., Xu, D.-L., & Zheng, J.-F. (2012). Condition-based maintenance of dynamic systems using online failure prognosis and belief rule base. *Expert Systems with Applications, 39*(6), 6140-6149. https://doi.org/10.1016/j.eswa.2011.11.068