

Xiujie ZHAO, Piao CHEN, Loon Ching TANG

Condition-based maintenance via Markov decision processes: A review

© Higher Education Press 2025

Abstract The optimization of condition-based maintenance (CBM) poses challenges due to the rapid advancement of monitoring technologies. Traditional CBM research has mainly relied on theory-driven approaches, which lead to the development of several effective maintenance models characterized by their wide applicability and attractiveness. However, when the system reliability model becomes complex, such methods may run into intractable cost models. The Markov decision process (MDP), a classic framework for sequential decision making, has drawn increasing attention for optimization of CBM optimization due to its appealing tractability and pragmatic applicability across different problems. This paper presents a review of research that optimizes CBM policies using MDP, with a focus on mathematical modeling and optimization methods to enable action. We have organized the review around several key components that are subject to similar mathematical modeling constraints, including system complexity, the availability of system conditions, and diverse criteria of decision-makers. An increase in interest has led to the optimization of CBM for systems possessing increasing numbers of components and sensors. Then, the review focuses on joint optimization problems with CBM. Finally,

as an important extension to traditional MDPs, reinforcement learning (RL) based methods are also reviewed as ways to optimize CBM policies. This paper provides significant background research for researchers and practitioners working in reliability and maintenance management, and gives discussions on possible future research directions.

Keywords reliability, degradation modeling, dynamic programming, reinforcement learning, sequential decision problems.

Received Jul. 17, 2024; revised Sep. 4, 2024; accepted Sep. 7, 2024

Xiujie ZHAO
College of Management and Economics, Tianjin University, Tianjin 300072, China

Piao CHEN (✉)
ZJU-UIUC Institute, Zhejiang University, Haining 314400, China
E-mail: piaochen@intlzju.edu.cn

Loon Ching TANG
Department of Industrial Systems Engineering and Management,
National University of Singapore, Singapore 117576, Singapore

This research was supported by the National Natural Science Foundation of China (Grant Nos. 72401253, 72371182, 72002149, and 72271154), and the National Social Science Fund of China (23CGL018), the State Key Laboratory of Biobased Transportation Fuel Technology, China (Grant No. 512302-X02301) and a start-up grant from the ZJU-UIUC Institute at Zhejiang University (Grant No. 130200-171207711).

1 Introduction

Maintenance is vital for preventing unexpected failures to guarantee system availability. It is the development of modern sensor technology together with the Internet of Things (IoTs) that allows gathering much more critical data about the system health for supporting the process of reliability management. Condition-based maintenance (CBM), due to its flexibility and economic effectiveness, has turned into an overly attractive alternative to traditional time-based maintenance (TBM). CBM modeling and optimization have been widely studied and applied to various systems (Alaswad and Xiang 2017; de Jonge and Scarf, 2020).

Early studies on CBM modeling and optimization usually use renewal process-based methods to model system reliability and maintenance interventions, and the objective is typically the long-term cost rate (Grall et al., 2002). However, these methods have some drawbacks. First, solving the renewal equation for complex failure mechanisms is problematic, therefore, the analytical assessment of the expected number of failures and determination of optimal cost are troublesome. For these two reasons combined, it is usually impossible to find the optimal policy from such equations analytically. Secondly, the complexity of such equations often requires numerical methods; these demand heavy computation,

especially if one has a large number of model parameters. Lastly, renewal process-based methods tend to support only stationary policies, and CBM can be extended to dynamic, state-dependent and time-dependent decisions. These difficulties have motivated researchers to extend models for CBM in a more flexible way.

A Markov decision process (MDP) is a discrete-time stochastic control process where both chance and the actions of the decisionmaker play a part in making decisions. MDPs are able to model dynamic decisions and facilitate the policy optimization (Howard 1960). For maintenance planning, a decision maker usually makes choices sequentially over time in regard to maintenance actions. For example, a widely used MDP model for CBM modeling can be described by a five-tuple (X, A, c, P^A, λ) , where X is the set of system health states, A is the set of maintenance actions, c is the cost functions, P^A is the probability set that describes how system state evolves from one state to another under certain actions, and λ is the discount factor. The most frequently used criterion for optimization is to minimize the long-term maintenance costs. Specifically, given any policy π , the objective function under π is defined in terms of the expected total discounted cost:

$$V^\pi(x) \triangleq \mathbf{E} \left[(1 - \lambda) \cdot \sum_{i=0}^{\infty} \lambda^i c(X_i, \pi_i(X_i)) \mid X_0 = x \right], \quad (1)$$

where X_0 is the initial system state, and $c(X_t, \pi_t(X_t))$ is the value or cost where the decision maker adopts the action $\pi_t(X_t)$ from policy π at time t . The optimal policy that minimizes $V^\pi(x_0)$, i.e., $\pi^* = \operatorname{argmin}_{\pi \in \Pi} V^\pi(x_0)$, where x_0 is the given initial state and Π is the policy space. The objective function is often reformulated in an iterative manner, commonly known as the Bellman equation, to

facilitate programming. The MDP has two popular variants that can be adapted to different scenarios:

- Partially observable Markov decision process (POMDP): A POMDP is a widely used extension of an MDP. In a POMDP, the system dynamics are determined by an MDP, but the decision maker does not have direct visibility of the underlying system state.
- Semi-Markov decision process (SMDP): An SMDP is applied to dynamic systems where the duration of time spent in each state is a continuous random variable, rather than discrete “flat” epochs.

Earlier literature reviews based on CBM by Alaswad and Xiang (2017), Olde Keizer et al. (2017a), and Li et al. (2020) focus on a vast area of CBM optimization techniques instead of focusing on particular methods. Currently, there has been growing research on dynamic maintenance optimization techniques especially those based on MDP and some of its modifications. Further, to investigate this, we searched the keywords “Condition-based maintenance” and “Markov decision process” combined on Web of Science and summarized the trend in papers published in each year. An upward trend of publication is evident from Fig. 1 for the last two decades. This growth has led us to reconsider both past and current literature discussing CBM with the help of MDP approaches and lay out future research.

While several of the reviews have focused on CBM research, none have been devoted to the optimization of CBM using the MDP-based techniques. This review aims to capture the literature from the year 2001 to 2024, and therefore the analysis provides a coverage of the papers. Particular stress is given to investigations that deploy degradation models for defining maintenance issues. Notably, Table 1 lists the names of journals that contain more than two papers included in this review.

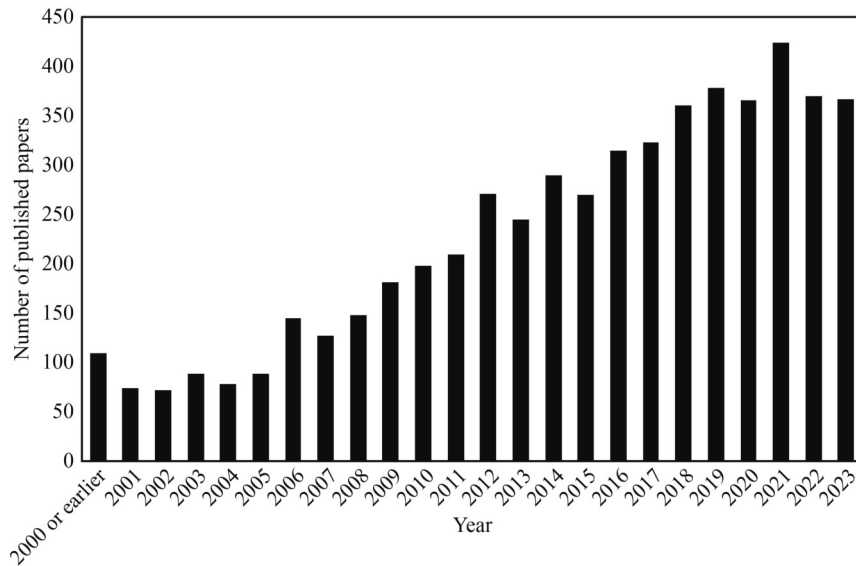


Fig. 1 Number of papers indexed by Web of Science with keywords CBM and MDP until 2023.

The related works classification is presented in Fig. 2. From a system structure perspective, previous research has focused on single-component and multi-component dependence, and different types of dependence of systems have been studied as well. It is worth noting that, depending on the needs of decision makers, various forms of MDP models have been applied to tackle given problems. More research has been focused on solving MDP problems with more emphasis on integrated decision making and reinforcement learning (RL) based methods more recently.

This paper starts with an outline of the work focused on single-compartment models, described in Section 2, where MDP modeling is introduced jointly with concrete criteria. Section 3 reviews the publications about multi-component systems with the relevant focus on the maintenance strategies for systems with dependent components and the efficiency of algorithms for solving MDPs. Section 4 investigates the studies that simultaneously solve CBM policies with other decisions. Section 5 is a

general discussion of innovation of the current related works in the field of CBM using RL. Last, in Section 6 researcher summarizes the paper and gives recommendations for future research.

2 Single-component systems

2.1 MDP modeling of maintenance problems

Early literature on maintenance optimization did not explicitly use the term “condition-based maintenance”, but rather focused on addressing system deterioration to support maintenance decisions. During the period from the 1950s to the 1980s, rapid advancements in decision sciences and computer sciences led to the emergence of pioneering works in maintenance optimization through the use of MDP. System deterioration is typically described by a Markov chain with transition probabilities. Tijms and van der Duyn Schouten (1985) demonstrated that, in many applications, the optimal policy is a control limit policy, with the policy iteration algorithm demonstrating its advantages. Albright (1979) examined the monotonicity results for a general class of POMDPs, optimizing a system with two states (good or bad) for maintenance. The early papers published in journals on operations research and management science significantly contributed to the application of MDP in maintenance problems. It is important to note, however, that the limitations of these early studies were often related to the following: a) the absence of a convincing mechanism for establishing the link between failure modes and the Markov model, b) a rather limited size of either the state set or the action set, and c) deficient algorithms for solving

Table 1 Journals that contain more than two papers included in the review

Journal Name	Number of papers
Reliability Engineering & System Safety	27
European Journal of Operational Research	17
Computers & Industrial Engineering	11
IEEE Transactions on Reliability	7
Journal of Manufacturing Systems	5
Expert Systems with Applications	4
IEEE Transactions on Automation Science and Engineering	3
IISE Transactions	3

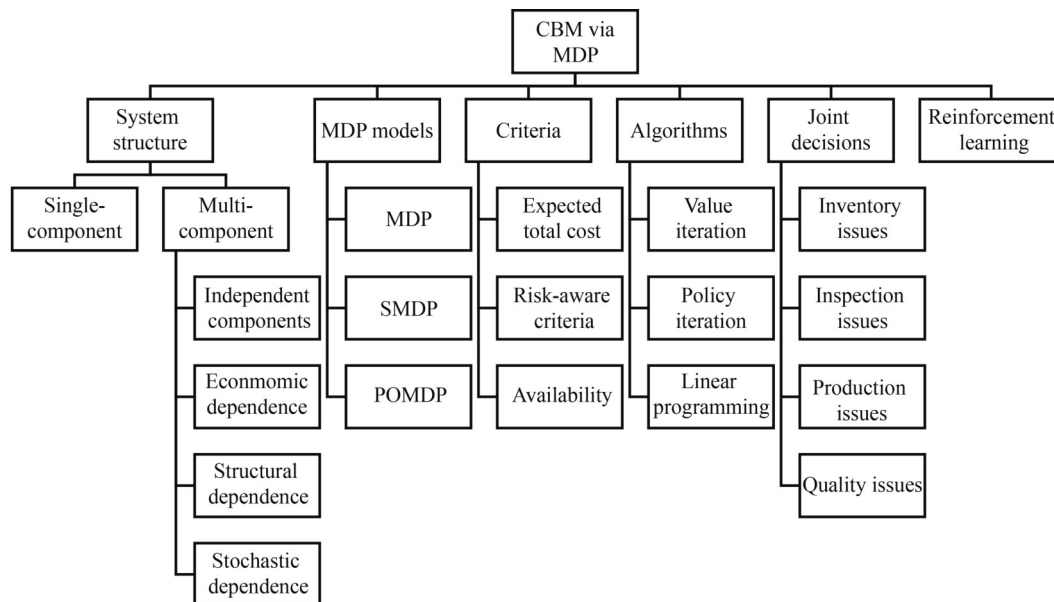


Fig. 2 A broad classification of related works.

the problem and related simulation techniques. In recent decades, extensive efforts have been made to extend these models.

The modeling paradigm of CBM via MDP based on stochastic degradation processes has been widely adopted in various studies. A representative modeling framework is presented by Chen et al. (2015). At each decision epoch k , the value function $V_{k,\delta}(u, v)$ is defined for all time epochs $u = 0, \delta, 2\delta, \dots$, and system state $v \in \mathcal{R}^+$, where δ is the inspection interval. Note that $V_{k,\delta}(u, v)$ is the minimum total discounted cost starting from epoch k until the end of the planning horizon. If an infinite horizon is considered, k can be dropped from the model. Defining c_p and c_f as the cost of preventive maintenance and corrective maintenance, respectively, the value function can be written in the manner of Bellman equation:

$$V_{\delta}(u, v) = \begin{cases} \min\{e^{-r\delta}U_{\delta}(u, v) + \\ W_{\delta}(u, v), c_p + V_{\delta}(0, 0)\}, & v \leq D, \\ c_f + V_{\delta}(0, 0), & v > D, \end{cases} \quad (2)$$

where $U_{\delta}(u, v)$ is the expected value function with one period transition from the state (u, v) , $W_{\delta}(u, v)$ is the expected downtime cost until the next epoch, and r is the discount factor. The monotonicity properties of the value function with respect to time and system state can then be elegantly gleaned by analyzing the stochastic properties of $W_{\delta}(u, v)$ and $U_{\delta}(u, v)$, which is usually achieved via induction and stochastic dominance (Shaked and Shanthikumar 2007). Consequently, the structural properties of the optimal policy can be determined. In most cases, the optimal policy is proven to be a specific control limit policy.

Glazebrook et al. (2005) conducted a study on maintenance performed by a group of repairmen and modeled the problem as an MDP within the restless bandit framework, which accounts for stochastic changes in machine states. Kuhn and Madanat (2005) focused on managing infrastructure systems under uncertainty and formulated the problem as an MDP to optimize the selection and scheduling of maintenance, repair, and rehabilitation activities in a cost-effective manner. Abeygunawardane et al. (2013) developed an adaptive maintenance optimization model for aging equipment, considering delays in inspection and maintenance activities. Chan and Asgarpoor (2006) utilized an MDP to determine the optimal maintenance policy for a component, taking into account random and deterioration-induced failures. Batun and Maillart (2012) reevaluated the trade-offs between maintenance and production planning by incorporating random yields and product mix constraints in their constrained MDP model. Zhou and Li (2023) investigated a novel maintenance problem involving two competing dependent risks: minor failures and major failures. They modeled the problem as an MDP and included the

cumulative number of minor failures as a covariate in the hazard function of major failures to accurately evaluate system degradation.

Motivated by industrial settings where systems operate under varying environmental conditions, Gan et al. (2023) developed a maintenance optimization strategy for multi-state systems in dynamic settings. The study emphasizes that conventional strategies often overlook the interdependence between the system and its environment, leading to ineffective maintenance. Similarly, Luo et al. (2024) explored a CBM policy for systems experiencing stochastic degradation in a dynamic environment. They model this degradation using a Markov process that affects the degradation drift parameter.

2.2 SMDP and POMDP

Although the standard MDP is capable of handling various types of maintenance optimization problems, certain inherent assumptions restrict its applicability. First, standard MDPs usually only support deterministic periodic transitions and decision epochs, whereas in practice, aperiodic cases can be quite common. Secondly, the uncertainties in observability cannot be adequately reflected in a standard MDP model. As a result, two commonly used variants, namely SMDP and POMDP, have been developed to address these limitations.

SMDP provided the flexibility and efficiency required to deal with aperiodic transitions, bringing maintenance problems closer in alignment with real-world events. For instance, Guo and Liang (2022a) incorporated an integrated approach that combines physical and stochastic models into the SMDP framework for reinforced concrete structures. Similarly, Duan et al. (2023) presented an adaptive reliability-based maintenance policy specifically designed for mechanical systems operating under varying operational environments. Hu et al. (2022) developed an SMDP model to determine the optimal maintenance policy for production systems that experience random production waits. The decision maker in this model can choose to replace the system either during a production wait or when a time-based threshold is reached.

In practice, decision making is often affected by uncertainties due to the inability to observe the exact state of a system. To address this issue of incomplete information, the framework of POMDPs is commonly used. Ivy and Pollock (2005) employed POMDPs to model a deteriorating system with multiple states, incorporating probabilistic monitoring and silent failures. They derive the optimal policy structure by utilizing the concept of ‘‘marginal monotonicity.’’

In the context of a finite-horizon, discrete-time Markovian deteriorating system, Kuo (2006) and Liu et al. (2021) modeled the optimization of maintenance and quality control as a POMDP, considering the unobservable states. Celen and Djurdjanovic (2020) discussed a

decision-making method for maintenance scheduling and production operations in flexible manufacturing systems (FMSs). In FMSs, the states cannot be directly observed and can only be inferred probabilistically from sensor readings. POMDPs are proposed by the authors to develop an integrated policy for maintenance and production sequencing. Van Staden and Boute (2021) investigated the influence of multi-sensor data quality on CBM strategies, utilizing POMDPs. Lin et al. (2022) concentrated on traction power supply equipment (TPSE) in high-speed railways and propose a maintenance model that employs a POMDP to address the inaccuracies in assessing equipment state caused by inherent uncertainties. Morato et al. (2022) proposed a combined framework that utilizes dynamic Bayesian networks (DBNs) and POMDPs for efficient inspection and maintenance planning of civil and maritime engineering systems. Deep et al. (2023) tackled the challenge of partial observability by modeling the degradation process with time-dependent observations. They developed an optimal maintenance strategy using a POMDP, which takes into account the time-varying nature of condition monitoring signals. Arcieri et al. (2024) designed a framework to develop optimal maintenance strategies for railway infrastructure. They infer POMDP parameters using Markov Chain Monte Carlo (MCMC) sampling and address model uncertainty by incorporating domain randomization during RL training. Guo and Liang (2022b) proposed a predictive Markov decision process (PMDP), which extends the traditional POMDP. The PMDP utilizes the Forward algorithm for predicting inspection timings and the Baum-Welch algorithm for model parameter estimation.

2.3 Criteria and algorithms

Conventional problems in MDPs typically focus on the expected value function. When it comes to maintenance optimization, the optimal policy criterion involves minimizing the expected maintenance cost. However, there are other criteria that decision makers may find more relevant. For instance, maintaining critical engineering systems like power systems requires not only minimizing costs, but also ensuring availability and safety to meet service level requirements. In such cases, risk-aware criteria can be integrated into the MDP modeling of CBM. Several studies have explored CBM policies optimized by MDPs with criteria that go beyond minimizing expected maintenance costs. Xu et al. (2022) proposed a risk-aware maintenance model that incorporates risk functions to assess safety levels and formulates safety constraints within the MDP framework. MDP frameworks have proven effective in addressing maintenance management challenges in mission-oriented systems. For example, Zhao et al. (2022) proposed a robust MDP framework for dynamic decision-making in mission-critical environments. Additionally, Zheng et al. (2024a) developed a

multilevel preventive replacement policy within the MDP framework, focusing on internal deterioration, external shocks, and the completion of various mission types in mission-oriented systems. Gosavi (2006) incorporated risk sensitivity into the total productive maintenance (TPM) and utilizes both linear and dynamic programming approximations to solve the MDP model, addressing maintenance and production planning in manufacturing systems.

Value iteration and policy iteration are the two most commonly used algorithms for solving MDPs. In policy iteration, the algorithm typically begins with an arbitrary policy π and iteratively evaluates and improves it until convergence is achieved. On the other hand, value iteration computes the optimal value by iteratively updating the value function, employing the Bellman equation. Both policy iteration and value iteration are dynamic programming algorithms. Generally, policy iteration outperforms value iteration in terms of efficiency, requiring fewer iterations to reach optimality. However, the value iteration algorithm is generally easier to construct and implement.

Chen and Trivedi (2005) investigated the inclusion of both degradation and Poisson failures in the context of SMDP. They utilize a value iteration algorithm to determine a dynamic threshold-based policy that maximizes system availability. Durango-Cohen and Sarutipand (2009) proposed a quadratic programming framework to address the optimization of maintenance policies for multifacility transportation systems. Their objective is to minimize average maintenance cost. Zhou and Li (2023) employed an iterative convergence algorithm to achieve this objective. Jin et al. (2023) presented a novel maintenance strategy for complex industrial systems using approximate dynamic programming (ADP). They estimate the system state in the next decision period through simulation techniques and use an approximate dynamic programming approach to solve the maintenance problem. They also employ a post-decision state method to reduce the number of system states, resulting in enhanced computational efficiency and optimized maintenance strategies under uncertainty. Zheng et al. (2023) optimized the maintenance policy by combining stochastic dynamic programming with the Nelder-Mead search algorithm. Li et al. (2023) proposed a dual-value iteration algorithm to efficiently solve the maintenance strategy model, addressing the limitations of traditional value iteration algorithms.

It is widely recognized that dynamic programming algorithms, such as value iteration and policy iteration, often face the challenge of high dimensionality, known as the curse of dimensionality. To overcome this challenge, some researchers have attempted to transform MDP-based maintenance problems into other types of programming. For example, Borrero and Akhavan-Tabatabaei (2013) converted MDP models into linear programming models to obtain optimal policies. Similarly, Xu et al.

(2022) transformed risk-aware MDP problems into linear programming problems.

3 Multi-component systems

Nowadays, real engineering systems are usually complex and composed of multiple key components that behave nonhomogeneous degradation characteristics and require different maintenance plans and actions. Although past reviews have discussed maintenance policies for multi-component systems, such as those by Olde Keizer et al. (2017a) and de Jonge and Scarf (2020), there are currently no recent reviews specifically focusing on the optimization of maintenance through MDP for multi-component systems.

Xia et al. (2008) have addressed the issue of maintenance for safety-critical components in joint replacements using a MDP model. They have highlighted the effectiveness of the “shortest-remaining-lifetime-first” rule in optimizing maintenance policies. Zhang et al. (2023b) have considered the impact of a dynamic environment on a multi-component system and have presented a degradation model that takes into account both intrinsic characteristics and the common operating environment. Sun et al. (2018) have optimized inspection and replacement policies for multi-component degrading systems, with a specific example of a 1-out-of-2:G system. Hao et al. (2024) have focused on an optimal maintenance policy that involves component reallocations and preventive replacements in a system with multi-state components.

The literature also extensively discusses SMDP and POMDP models. Wang et al. (2023a) have explored maintenance optimization for a balanced system with interchangeable components using an SMDP approach. They have emphasized the benefits of rearrangement actions and imperfect preventive maintenance. Safieh Mahmoodi et al. (2020) have studied maintenance policies for parallel unit systems in semi-Markov environments with dynamic conditions, with a particular focus on converting the semi-Markov process to a Markov process. For POMDP models, Karabağ et al. (2024) have presented an efficient procedure for optimal maintenance intervention and spare part quantity decisions in partially observable systems with identical components.

3.1 Component Dependencies

Different types of dependencies can exist among components. In recent years, some efforts recorded in the literature have tried to model the maintenance of component dependencies in the paradigm of MDPs. Economic dependency is an important type of dependency in maintaining multi-component systems. Li and Wu (2024) have presented a maintenance optimization model for a two-component system using an SMDP. Their model aims to

find optimal maintenance thresholds, with a focus on economic dependency and failure dependency. Hu et al. (2024) have explored a preventive replacement policy for a two-component series system, considering masked failure causes and economic dependence. They have utilized an SMDP framework to minimize long-term maintenance costs per time unit. Liang and Parlikad (2020) have investigated a five-stage approach to optimize predictive group maintenance in a hierarchical multi-system multi-component network. They have considered the influence of the planning horizon and have emphasized the effect of dual economic dependencies.

Some other works focus on the stochastic dependency among components that characterizes the nature of dependent degradation. Kivanç et al. (2022) conducted an evaluation of the application of POMDPs in optimizing maintenance strategies for multi-component systems with stochastic dependence. Andersen et al. (2022) presented a unified framework for optimizing replacement strategies in multi-component systems, considering both TBM and CBM. Their focus is on minimizing long-term maintenance costs and improving computational efficiency. Xu et al. (2021) investigated the generalized CBM optimization problem for multi-component systems by taking into account stochastic dependency among components and imperfect maintenance. Xu et al. (2018) proposed an extended proportional hazard model with stochastic dependence for multi-component systems. They develop a novel state discretization algorithm and integrate cost optimization for minimal repair. Wang et al. (2023b) obtained the optimal maintenance strategy by considering load-sharing constraints, limited maintenance capacity, maintenance setup costs, and structural characteristics. Uit Het Broek et al. (2021) explored joint CBM and load-sharing optimization for two-unit systems with economic dependency. They discuss maintenance optimization and load distribution in production facilities with interchangeable equipment units.

In addition to the common dependencies in reliability models, recent research has also examined new types of dependencies, such as balanced systems (Cui et al., 2018; Guo and Elsayed 2019; Zhao et al., 2020). Zhao and Wang (2022) considered a two-unit balanced system and utilized a bivariate Wiener process to model the degradation paths. The Bellman equation for the model is general and can be expressed as follows:

$$V_{\delta}(x_1, x_2) = \begin{cases} c_c + c_p + V_{k\delta}(0, 0), & \text{if } x_1 > L_1 \text{ or } x_2 > L_2, \\ \min\{C_{\text{replace}}(x_1, x_2), C_{\text{repair}}(x_1, x_2), C_{\text{dn}}(x_1, x_2)\}, & \\ \text{otherwise,} & \end{cases} \quad (3)$$

where x_1 and x_2 represent the degradation level of components 1 and 2, respectively, in the balanced system, and L_1 and L_2 are the corresponding failure thresholds.

Further, c_p and c_c are the costs of preventive maintenance and the additional cost incurred by corrective maintenance. Then the value function in (3) can be further elaborated by

$$\begin{aligned} C_{\text{replace}}(x_1, x_2) &= c_p + V_{k\delta}(0, 0), \\ C_{\text{repair}}(x_1, x_2) &= C_r(x_1, x_2) \\ &\quad + W_\delta(\max\{x_1, x_2\}, \max\{x_1, x_2\}) \\ &\quad + U_\delta(\max\{x_1, x_2\}, \max\{x_1, x_2\}), \\ C_{\text{dn}}(x_1, x_2) &= W_\delta(x_1, x_2) + U_\delta(x_1, x_2), \end{aligned} \quad (4)$$

and $C_{\text{replace}}(\cdot)$, $C_{\text{repair}}(\cdot)$, and $C_{\text{dn}}(\cdot)$ denote the costs of replacement, repair, and doing nothing, respectively, given the inspected degradation levels. W_δ and U_δ are the expected downtime cost and the expected value function at the next inspection epoch, correspondingly. Specifically, W_δ can be given by

$$W_\delta(x_1, x_2) = \mathbb{E} \left[\int_{k\delta}^{(k+1)\delta} \beta(|X_1(t) - X_2(t)| - D) dt \mid X(k\delta) = (x_1, x_2) \right] \quad (5)$$

where $X_1(t)$ and $X_2(t)$ are the time-dependent degradation levels at time t of the two components, respectively, and the term $\beta(|X_1(t) - X_2(t)| - D)$ describes the cost incurred by system imbalances, which is evaluated by the absolute degradation differences, and D is the maximum difference in degradation levels that brings no additional cost. One interesting modification in this model, compared to previous studies, is that the cost functions take into account the cost brought by imbalances, in addition to system failures. The structural properties of the optimal policy are analyzed and it is proven that a two-dimensional control limit policy is optimal.

Other works also address similar problems. Chai et al. (2024) proposed an integrated condition-based reallocation for a 1-out-of-2 pairs balanced system and compare the benefits of combined policies against individual reallocation or maintenance strategies. Wang et al. (2020) proposed a model for a performance-balanced system operating in a shock environment, incorporating a new maintenance policy that considers preventive, corrective, and opportunistic maintenance strategies to optimize system performance. Wang et al. (2023c) proposed a joint policy of component reassignment and preventive maintenance in a balanced multi-state system to improve system reliability.

3.2 Algorithms

For multi-component systems, solving the maintenance problem becomes more challenging due to the increased number of system states and action selections. Dynamic programming can be computationally demanding in such cases. Okogbaa et al. (2008) used binary integer programming (BIP) and linear programming (LP) to obtain the maintenance plan for the entire system.

Karabağ et al. (2024) proposed a novel linear programming approach to optimize a reduced-state MDP. Drent et al. (2024) demonstrated substantial cost savings by applying a modified policy iteration algorithm that decomposes complex MDPs into two-dimensional MDPs. Zhang et al. (2013) developed an efficient modified iterative aggregation procedure to optimize the maintenance policy. uit het Broek et al. (2021) proposed a modified policy iteration algorithm that combines value iteration with policy iteration to find stationary ϵ -optimal policies for the value functions. Vora et al. (2023) presented an algorithm for finding an optimal policy for a multi-component budget-constrained POMDP in the context of infrastructure maintenance and inspection. Hoffman et al. (2022), based on the policy iteration algorithm, used a genetic algorithm (GA) to improve the policy online via Monte Carlo tree search (MCTS), overcoming the curse of dimensionality.

4 Joint optimization with other decisions

It is highly beneficial to optimize maintenance policies in conjunction with other decisions. In various engineering applications, maintenance is closely interconnected with other factors. For instance, when planning maintenance for production systems, it is important to consider production schedules, lot sizing, and other variables that could influence production yield, quality, and inventory. Numerous existing studies have examined production and inventory-related issues alongside maintenance using MDP-based methods (Farahani and Tohidi 2021; Van Horenbeek et al. 2013). One critical aspect of joint optimization is inventory planning. Olde Keizer et al. (2017b) argue that the conventional maintenance and ordering strategy structure utilized in single-component systems is not optimal for multi-component systems, where components share a pool of spare parts. They propose incorporating spare parts planning into maintenance optimization for multi-component systems. Wang and Zhu (2021) propose a novel modeling approach for the joint optimization of CBM and inventory control in k-out-of-n: F systems. Their approach leverages the number of parts in discrete degradation states for modeling, leading to a reduction in the solution space. In the case of a multi-component system procuring spare parts from two suppliers, Zheng et al. (2024a) propose an optimal maintenance and ordering strategy. They discover that when two components exhibit similar degradation levels, replacement is delayed to determine which component should be prioritized. Zhang et al. (2022) develop a joint CBM and spare parts inventory optimization strategy for general series-parallel systems, accounting for both hard and soft failure modes. Hao et al. (2023) extended the joint optimization problem of CBM and inventory to larger state and action spaces. They develop an improved deep RL algorithm based on stochastic policy and a behavioral

criticism framework to address the curse of dimensionality. Considering a degradation system with hidden but partially observable states, Tang et al. (2024) explored the optimal maintenance and spare parts inventory ordering policy within the framework of POMDP. Finally, Drent et al. (2024) investigate optimal data pooling for shared learning in CBM and spare parts management.

In addition to inventory planning, production scheduling and planning are key factors that impact the profitability of manufacturers. These factors are often optimized in conjunction with maintenance policies (Sun et al., 2023; uit het Broek et al., 2020). Aramon Bajestani et al. (2014) proposed an integer planning model to solve the integrated maintenance and production scheduling problem in multi-machine production systems. Their objective is to minimize long-term maintenance costs and production loss costs. Ao et al. (2019) presented a two-step strategy for integrating production and maintenance decision-making in semiconductor production. The aim is to optimize production performance and reduce costs effectively. Kang and Subramaniam (2018) proposed an integrated SMDP model for controlling production rates and preventive maintenance schedules in a single-machine manufacturing system. To address the need for balancing production efficiency with maintenance requirements, Zheng et al. (2021) integrated SMDP with production lot sizing decisions and maintenance schedules. Their proposed model is based on the proportional hazards model with a continuous-state covariate process. Celen and Djurdjanovic (2020) developed a comprehensive strategy for integrating maintenance scheduling and production sequencing for highly complex manufacturing equipment. Their approach, implemented under a POMDP framework, demonstrates superior performance compared to separate maintenance and production decision-making. Koopmans and de Jonge (2023) investigate the joint optimization problem of maintenance and production for a production system. They believe that the production rate affects the system's degradation and aim to maximize the total production output.

Research on optimizing maintenance strategies, along with other elements such as quality and monitoring, has also been conducted. Nguyen et al. (2019) focused on tuning the quality of condition monitoring to optimize CBM and simultaneously optimize maintenance planning and monitoring quality using a POMDP approach. Zheng et al. (2024b) considered the quality of spare parts provided by suppliers and study the joint optimization of CBM and spare parts supply strategies. Zhao et al. (2022) aimed to improve both preventive maintenance and performance control concurrently. Their focus is on mission-critical systems where deterioration is controlled by adjusting the performance level. The objective is to balance mission success and system reliability. Zhao et al. (2023) optimized maintenance and warranty policies for different types of products using the MDP paradigm.

5 Maintenance optimization via reinforcement learning

An inherent limitation of classic MDP models and their variants is that decision makers are required to specify all model parameters, thereby determining the system and decision dynamics. However, in practice, estimating these parameters can be challenging due to limited historical information. To address this issue, the RL framework allows the learning model to estimate and update parameters within the MDP model, enabling unsupervised decision making. RL has gained traction in various fields, including operations research, finance, computing intelligence, and healthcare engineering. The RL-driven CBM policy has become increasingly popular due to its adaptiveness to uncertain dynamic systems. Ogunfowora and Najjaran (2023) review the research on RL and deep reinforcement learning (DRL) in maintenance planning and optimization, providing classifications and summaries of existing studies using taxonomies.

For single-component systems, RL-based methods can be employed to dynamically determine the optimal maintenance policy based on uncertain parameters. Mikhail et al. (2024) combined machine learning (ML), RL, and a reliability-based approach to develop a data-driven method for optimizing CBM strategies. This approach simultaneously minimizes average maintenance costs and maximizes the remaining useful life of the system. Peng and Feng (2021) modeled the CBM problem as a continuous-state MDP in discrete time without discretizing the system's deterioration condition. They use Gaussian process regression for function approximation in RL to model the state transition and state value functions and propose an RL algorithm to minimize the long-run average cost. Regarding the aviation industry, Ruan et al. (2021) addressed the operational aircraft maintenance routing problem (OAMRP), aiming to generate optimal routes for each aircraft that satisfy maintenance requirements and comply with regulations. They propose an integer linear programming (ILP) framework based on network flow and a novel RL algorithm while considering multiple critical maintenance constraints. Additionally, the continuous development and extension of DRL and its variants have further advanced the study of maintenance strategy optimization (Lee et al. 2024).

The advantages of RL-based methods compared to traditional methods are particularly notable for multi-component systems. In a study by Liu et al. (2020), a unique selective maintenance optimization method was proposed that integrates DRL techniques based on AC algorithms. The main objective of this method is to dynamically select and execute a set of feasible maintenance actions for multi-state systems that are capable of executing multiple consecutive missions within a defined time horizon. The aim is to maximize the expected number of future mission successes under limited maintenance

resources.

Cheng et al. (2023) designed a DRL framework to explore cost-optimal CBM strategies for offshore wind turbines, using MDP models. The method employs DQN and proximal policy optimization to derive dynamic inspection intervals and adaptive repair thresholds. In another study by Mohammadi and He (2022), an optimization method for railway renewal and maintenance planning was proposed, utilizing DRL techniques. The method effectively addresses environmental uncertainty by considering both cost effectiveness and risk reduction. Historical data are used to simulate the railway environment, and the double deep Q-network (DDQN) and prioritized replay memory mechanism are applied. For complex equipment maintenance problems, such as aero-engines, Wei et al. (2024) proposed a maintenance framework that integrates a decomposition strategy, a neighborhood-based parameter-transfer policy, and DRL for multi-objective optimization. Su et al. (2022) proposed a new multi-agent RL approach for optimizing preventive maintenance (PM) strategies in serial production lines with multiple levels of PM actions. The method utilizes an adaptive learning framework based on the value-decomposition multi-agent AC algorithm. In the context of opportunistic maintenance optimization of multi-component systems, Zhang et al. (2023a) developed a modified proximal policy optimization approach using a DRL algorithm. They design a parameterized action space structure and a multi-task RL framework based on an infinite horizon MDP. Considering the stochastic and economic dependencies between components, Do et al. (2024) designed a multi-agent DRL-based maintenance method for manufacturing systems. Additionally, Goby et al. (2023) developed a prescriptive analytic approach based on DRL for solving sequential decision problems with large, noisy state spaces and combinatorial action spaces. Joint optimization with other decisions can also be easily extended by RL for CBM optimization. Paraschos et al. (2020) investigated a stochastic production/inventory system that considers several deterioration failures and product quality. They proposed an RL-based approach to optimize the total expected profit of the system, which includes various aspects such as production, maintenance, and product quality control. Zheng et al. (2024a) designed a hybrid DRL algorithm for the joint optimization of maintenance and spare part ordering from multiple suppliers for multi-component systems. Their aim is to minimize the expected cost while addressing the limitations of the value iteration algorithm in large-scale problems. Ye et al. (2023) developed a deep deterministic policy gradient (DDPG) algorithm based on an MDP optimization model to achieve optimal reliability-quality joint control, using a machine-level dynamic reliability and quality model to describe the complex interactions of the manufacturing network. Paraschos et al. (2020) proposed an RL approach to determine the optimal joint

production, maintenance, and quality control strategy.

Since this paper mainly reviews MDP-based CBM optimization methods, we do not list too many RL-driven works. More related works can be referred to in the literature (Barde et al., 2019; Ferreira Neto et al., 2024; Geurtsen et al., 2023; Hamida and Goulet, 2023; Hu et al., 2023; Lv et al., 2023; Tseremoglou and Santos 2024; Zhang and Si 2020). However, RL-based methods for CBM problems face several challenges. First, the maintenance decision models can be quite complex, making it difficult to derive analytical results in many circumstances, thereby increasing the difficulty in solving RL problems. Secondly, obtaining data that reflect system health and maintenance actions in a structural manner can be onerous. Problems such as incomplete or censored data are common, and decision makers must deal with coarse data to support their decisions. Lastly, convergence issues, particularly in terms of analytical efforts, can be very challenging in classic RL settings.

6 Conclusions and future research directions

In this paper, we review the literature on CBM modeling and optimization using MDP methods. Based on a deep search and review of related papers published in top-tier journals, we can see that there is a sharp growth of related studies in the past decades. More and more interest is given to handle problems with larger states and decision sets in the literature. In addition, more researchers are devoted to joint optimization problems and RL-based methods together with CBM modeling and optimization.

From this intensive literature review, it follows that many attractive research topics remain open and challenging in this area. Some possible directions are given as follows:

- 1) From a maintenance modeling perspective, the system dynamics along with the maintenance actions and their interactions can be extended in various directions in order to adapt the models to more application scenarios. Of interest could be the inclusion in the degradation models of time-varying dynamic covariates (Hong et al., 2015) characterizing a wide class of systems and enabling a decision maker to plan the maintenance actions not only based on system state but also on external covariates.

- 2) Joint policies that consider maintenance and other decisions may have more general benefits to decision makers. Literature has shown the good flexibility an MDP provides in various decision-making problems. Previous studies have carried out some in-depth investigation with regard to joint policies that take into consideration production planning, lot sizing, and inventory. We point out quality-related issues in those topics (Ben-Daya and Duffuaa 1995). First, for production systems, dynamic quality metrics integrated at the time of

manufacturing may provide significant support for maintenance decisions. Second, for sold systems, the optimization of maintenance and warranty policies can be made jointly by using MDP as shown by Zhao et al. 2023. More specifically, adaptive and customer-specific maintenance-warranty policies hold great potential to be explored. Third, more realistic optimization criteria can be considered for the CBM optimization via MDP.

3) If the problem scale is large, further gains in these algorithms for solving MDPs can be obtained by exploiting the system reliability characteristics. For example, modern electricity distribution systems are complex networks composed of dependent subsystems. The maintenance problem of such systems is of tremendous scale. Nevertheless, the dependencies of the subsystems can be quite sparse in some aspects, which make the factored MDP modeling possible to reduce the problem scale and enable the computation of optimal policies. Moreover, we find that most of the existing RL methods applied in the CBM optimization mainly aim to find the optimal policy in numerical manners. Scant research outputs have considered the convergence of regrets, which have been popular in other theoretical disciplines (Cheung et al., 2021). It could be a good interest for the researchers to investigate the structural properties of the CBM policies using the RL framework.

Competing Interests The authors declare that they have no competing interests.

References

- Abeygunawardane S K, Jirutitijaroen P, Xu H (2013). Adaptive maintenance policies for aging devices using a Markov decision process. *IEEE Transactions on Power Systems*, 28(3): 3194–3203
- Alaswad S, Xiang Y (2017). A review on condition-based maintenance optimization models for stochastically deteriorating system. *Reliability Engineering & System Safety*, 157: 54–63
- Albright S C (1979). Structural results for partially observable Markov decision processes. *Operations Research*, 27(5): 1041–1053
- Andersen J F, Andersen A R, Kulahci M, Nielsen B F (2022). A numerical study of Markov decision process algorithms for multi-component replacement problems. *European Journal of Operational Research*, 299(3): 898–909
- Ao Y, Zhang H, Wang C (2019). Research of an integrated decision model for production scheduling and maintenance planning with economic objective. *Computers & Industrial Engineering*, 137: 106092
- Arcieri G, Hoelzl C, Schwery O, Straub D, Papakonstantinou K G, Chatzi E (2024). POMDP inference and robust solution via deep reinforcement learning: An application to railway optimal maintenance. *Machine Learning*, 1–29
- Aramon Bajestani M, Banjevic D, Beck J C (2014). Integrated maintenance planning and production scheduling with Markovian deteriorating machine conditions. *International Journal of Production Research*, 52(24): 7377–7400
- Barde S R A, Yacout S, Shin H (2019). Optimal preventive maintenance policy based on reinforcement learning of a fleet of military trucks. *Journal of Intelligent Manufacturing*, 30(1): 147–161
- Batun S, Maillart L M (2012). Reassessing tradeoffs inherent to simultaneous maintenance and production planning. *Production and Operations Management*, 21(2): 396–403
- Ben-Daya M, Duffuaa S O (1995). Maintenance and quality: The missing link. *Journal of Quality in Maintenance Engineering*, 1(1): 20–26
- Borrero J S, Akhavan-Tabatabaei R (2013). Time and inventory dependent optimal maintenance policies for single machine workstations: An MDP approach. *European Journal of Operational Research*, 228(3): 545–555
- Celen M, Djurdjanovic D (2020). Integrated maintenance and operations decision making with imperfect degradation state observations. *Journal of Manufacturing Systems*, 55: 302–316
- Chai X, Kilic O A, Veldman J, Teunter R H, Zhao X (2024). Condition-based reallocation and maintenance for a 1-out-of-2 pairs balanced system. *European Journal of Operational Research*, 318(2): 618–628
- Chan G K, Asgarpoor S (2006). Optimum maintenance policy with Markov processes. *Electric Power Systems Research*, 76(6–7): 452–456
- Chen D, Trivedi K S (2005). Optimization for condition-based maintenance with semi-Markov decision process. *Reliability Engineering & System Safety*, 90(1): 25–29
- 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
- Cheng J, Liu Y, Li W, Li T (2023). Deep reinforcement learning for cost-optimal condition-based maintenance policy of offshore wind turbine components. *Ocean Engineering*, 283: 115062
- Cheung W C, Simchi-Levi D, Zhu R (2021). Hedging the drift: Learning to optimize under nonstationarity. *Management Science*, 68(3): 1696–1713
- Cui L, Gao H, Mo Y (2018). Reliability for k -out-of- n : F balanced systems with m sectors. *IIE Transactions*, 50(5): 381–393
- de Jonge B, Scarf P A (2020). A review on maintenance optimization. *European Journal of Operational Research*, 285(3): 805–824
- Deep A, Zhou S, Veeramani D, Chen Y (2023). Partially observable Markov decision process-based optimal maintenance planning with time-dependent observations. *European Journal of Operational Research*, 311(2): 533–544
- Do P, Nguyen V T, Voisin A, Iung B, Neto W A F (2024). Multi-agent deep reinforcement learning-based maintenance optimization for multi-dependent component systems. *Expert Systems with Applications*, 245: 123144
- Drent C, Drent M, van Houtum G J (2024). Optimal data pooling for shared learning in maintenance operations. *Operations Research Letters*, 52: 107056
- Duan C, Deng T, Song L, Wang M, Sheng B (2023). An adaptive reliability-based maintenance policy for mechanical systems under variable environments. *Reliability Engineering & System Safety*, 238: 109396
- Durango-Cohen P L, Sarutipand P (2009). Maintenance optimization for transportation systems with demand responsiveness. *Transportation Research*, 43(1): 1–11

- tion Research Part C, *Emerging Technologies*, 17(4): 337–348
- Farahani A, Tohidi H (2021). Integrated optimization of quality and maintenance: A literature review. *Computers & Industrial Engineering*, 151: 106924
- Ferreira Neto W A, Virgínio Cavalcante C A, Do P (2024). Deep reinforcement learning for maintenance optimization of a scrap-based steel production line. *Reliability Engineering & System Safety*, 249: 110199
- Gan S, Hu H, Coit D W (2023). Maintenance optimization considering the mutual dependence of the environment and system with decreasing effects of imperfect maintenance. *Reliability Engineering & System Safety*, 235: 109202
- Geurtsen M, Didden J B H C, Adan J, Atan Z, Adan I (2023). Production, maintenance and resource scheduling: A review. *European Journal of Operational Research*, 305(2): 501–529
- Glazebrook K D, Mitchell H M, Ansell P S (2005). Index policies for the maintenance of a collection of machines by a set of repairmen. *European Journal of Operational Research*, 165(1): 267–284
- Goby N, Brandt T, Neumann D (2023). Deep reinforcement learning with combinatorial actions spaces: An application to prescriptive maintenance. *Computers & Industrial Engineering*, 179: 109165
- Gosavi A (2006). A risk-sensitive approach to total productive maintenance. *Automatica*, 42(8): 1321–1330
- Grall A, Dieulle L, Berenguer C, Roussignol M (2002). Continuous-time predictive-maintenance scheduling for a deteriorating system. *IEEE Transactions on Reliability*, 51(2): 141–150
- Guo C, Liang Z (2022a). Semi-Markovian maintenance optimization for reinforced concrete enabled by a synthesized deterioration model. *IEEE Transactions on Reliability*, 71(4): 1577–1589
- Guo C, Liang Z (2022b). A predictive Markov decision process for optimizing inspection and maintenance strategies of partially observable multi-state systems. *Reliability Engineering & System Safety*, 226: 108683
- Guo J, Elsayed E A (2019). Reliability of balanced multi-level unmanned aerial vehicles. *Computers & Operations Research*, 106: 1–13
- Hamida Z, Goulet J A (2023). Hierarchical reinforcement learning for transportation infrastructure maintenance planning. *Reliability Engineering & System Safety*, 235: 109214
- Hao S, Zheng J, Yang J, Sun H, Zhang Q, Zhang L, Jiang N, Li Y (2023). Deep reinforce learning for joint optimization of condition-based maintenance and spare ordering. *Information Sciences*, 634: 85–100
- Hao Y, Zhu X, Kuo W (2024). Optimization of condition-based maintenance with multiple times of component reallocation using Markov decision process. *IEEE Transactions on Reliability*, 73(1): 131–141
- Hoffman M, Song E, Brundage M P, Kumara S (2022). Online improvement of condition-based maintenance policy via monte carlo tree search. *IEEE Transactions on Automation Science and Engineering*, 19(3): 2540–2551
- Hong Y, Duan Y, Meeker W Q, Stanley D L, Gu X (2015). Statistical methods for degradation data with dynamic covariates information and an application to outdoor weathering data. *Technometrics*, 57(2): 180–193
- Howard R A (1960). *Dynamic programming and Markov processes*. MIT Press
- Hu J, Huang Y, Shen L (2024). Maintenance optimization of a two-component series system considering masked causes of failure. *Quality and Reliability Engineering International*, 40(1): 388–405
- Hu J, Sun Q, Ye Z S (2022). Replacement and repair optimization for production systems under random production waits. *IEEE Transactions on Reliability*, 71(4): 1488–1500
- Hu J, Wang H, Tang H K, Kanazawa T, Gupta C, Farahat A (2023). Knowledge-enhanced reinforcement learning for multi-machine integrated production and maintenance scheduling. *Computers & Industrial Engineering*, 185: 109631
- Ivy J S, Pollock S M (2005). Marginally monotonic maintenance policies for a multi-state deteriorating machine with probabilistic monitoring, and silent failures. *IEEE Transactions on Reliability*, 54(3): 489–497
- Jin H, Song X, Xia H (2023). Optimal maintenance strategy for large-scale production systems under maintenance time uncertainty. *Reliability Engineering & System Safety*, 240: 109594
- Kang K, Subramaniam V (2018). Integrated control policy of production and preventive maintenance for a deteriorating manufacturing system. *Computers & Industrial Engineering*, 118: 266–277
- Karabağ O, Bulut Ö, Toy A Ö, Fadıloğlu M M (2024). An efficient procedure for optimal maintenance intervention in partially observable multi-component systems. *Reliability Engineering & System Safety*, 244: 109914
- Kıvanç İ, Özgür-Ünlüakın D, Bilgiç T (2022). Maintenance policy analysis of the regenerative air heater system using factored POMDPs. *Reliability Engineering & System Safety*, 219: 108195
- Koopmans M, de Jonge B (2023). Condition-based maintenance and production speed optimization under limited maintenance capacity. *Computers & Industrial Engineering*, 179: 109155
- Kuhn K D, Madanat S M (2005). Model uncertainty and the management of a system of infrastructure facilities. *Transportation Research Part C, Emerging Technologies*, 13(5–6): 391–404
- Kuo Y (2006). Optimal adaptive control policy for joint machine maintenance and product quality control. *European Journal of Operational Research*, 171(2): 586–597
- Lee J S, Yeo I H, Bae Y (2024). A stochastic track maintenance scheduling model based on deep reinforcement learning approaches. *Reliability Engineering & System Safety*, 241: 109709
- Li M, Wu B (2024). Optimal condition-based opportunistic maintenance policy for two-component systems considering common cause failure. *Reliability Engineering & System Safety*, 250: 110269
- Li S, Yang Z, He J, Li G, Yang H, Liu T, Li J (2023). A novel maintenance strategy for manufacturing system considering working schedule and imperfect maintenance. *Computers & Industrial Engineering*, 185: 109656
- Li Y, Peng S, Li Y, Jiang W (2020). A review of condition-based maintenance: Its prognostic and operational aspects. *Frontiers of Engineering Management*, 7(3): 323–334
- Liang Z, Parlikad A K (2020). Predictive group maintenance for multi-system multi-component networks. *Reliability Engineering & System Safety*, 195: 106704
- Lin S, Fan R, Feng D, Yang C, Wang Q, Gao S (2022). Condition-Based maintenance for traction power supply equipment based on partially observable Markov decision process. *IEEE Transactions*

- on Intelligent Transportation Systems, 23(1): 175–189
- Liu X, Sun Q, Ye Z S, Yildirim M (2021). Optimal multi-type inspection policy for systems with imperfect online monitoring. *Reliability Engineering & System Safety*, 207: 107335
- Liu Y, Chen Y, Jiang T (2020). Dynamic selective maintenance optimization for multi-state systems over a finite horizon: A deep reinforcement learning approach. *European Journal of Operational Research*, 283(1): 166–181
- Luo Y, Zhao X, Liu B, He S (2024). Condition-based maintenance policy for systems under dynamic environment. *Reliability Engineering & System Safety*, 246: 110072
- Lv Y, Guo X, Zhou Q, Qian L, Liu J (2023). Predictive maintenance decision-making for variable faults with non-equivalent costs of fault severities. *Advanced Engineering Informatics*, 56: 102011
- Mahmoodi S, Hamed Ranjkesh S, Zhao Y Q (2020). Condition-based maintenance policies for a multi-unit deteriorating system subject to shocks in a semi-Markov operating environment. *Quality Engineering*, 32(3): 286–297
- Mikhail M, Ouali M S, Yacout S (2024). A data-driven methodology with a nonparametric reliability method for optimal condition-based maintenance strategies. *Reliability Engineering & System Safety*, 241: 109668
- Mohammadi R, He Q (2022). A deep reinforcement learning approach for rail renewal and maintenance planning. *Reliability Engineering & System Safety*, 225: 108615
- Morato P G, Papakonstantinou K G, Andriotis C P, Nielsen J S, Rigo P (2022). Optimal inspection and maintenance planning for deteriorating structural components through dynamic Bayesian networks and Markov decision processes. *Structural Safety*, 94: 102140
- Nguyen K T P, Do P, Huynh K T, Bérenguer C, Grall A (2019). Joint optimization of monitoring quality and replacement decisions in condition-based maintenance. *Reliability Engineering & System Safety*, 189: 177–195
- Ogunfowora O, Najjaran H (2023). Reinforcement and deep reinforcement learning-based solutions for machine maintenance planning, scheduling policies, and optimization. *Journal of Manufacturing Systems*, 70: 244–263
- Okogbaa O G, Otieno W, Peng X, Jain S (2008). Transient analysis of maintenance intervention of continuous multi-unit systems. *IIE Transactions*, 40(10): 971–983
- Olde Keizer M C A, Flapper S D P, Teunter R H (2017a). Condition-based maintenance policies for systems with multiple dependent components: A review. *European Journal of Operational Research*, 261(2): 405–420
- Olde Keizer M C A, Teunter R H, Veldman J (2017b). Joint condition-based maintenance and inventory optimization for systems with multiple components. *European Journal of Operational Research*, 257(1): 209–222
- Paraschos P D, Koulinas G K, Koulouriotis D E (2020). Reinforcement learning for combined production-maintenance and quality control of a manufacturing system with deterioration failures. *Journal of Manufacturing Systems*, 56: 470–483
- Peng S, Feng Q (2021). Reinforcement learning with Gaussian processes for condition-based maintenance. *Computers & Industrial Engineering*, 158: 107321
- Ruan J H, Wang Z X, Chan F T S, Patnaik S, Tiwari M K (2021). A reinforcement learning-based algorithm for the aircraft maintenance routing problem. *Expert Systems with Applications*, 169: 114399
- Shaked M, Shanthikumar J G (2007). *Stochastic orders*. New York: Springer
- Su J, Huang J, Adams S, Chang Q, Beling P A (2022). Deep multi-agent reinforcement learning for multi-level preventive maintenance in manufacturing systems. *Expert Systems with Applications*, 192: 116323
- Sun Q, Chen P, Wang X, Ye Z (2023). Robust condition-based production and maintenance planning for degradation management. *Production and Operations Management*, 32(12): 3951–3967
- Sun Q, Ye Z, Chen N (2018). Optimal inspection and replacement policies for multi-unit systems subject to degradation. *IEEE Transactions on Reliability*, 67(1): 401–413
- Tang X, Xiao H, Kou G, Xiang Y (2024). Joint optimization of condition-based maintenance and spare parts ordering for a hidden multi-state deteriorating system. *IEEE Transactions on Reliability*, 1–12
- Tijms H C, van der Duyn Schouten F A (1985). A Markov decision algorithm for optimal inspections and revisions in a maintenance system with partial information. *European Journal of Operational Research*, 21(2): 245–253
- Tseremoglou I, Santos B F (2024). Condition-based maintenance scheduling of an aircraft fleet under partial observability: A deep reinforcement learning approach. *Reliability Engineering & System Safety*, 241: 109582
- uit het Broek M A J, Teunter R H, de Jonge B, Veldman J (2021). Joint condition-based maintenance and load-sharing optimization for two-unit systems with economic dependency. *European Journal of Operational Research*, 295(3): 1119–1131
- uit het Broek M A J, Teunter R H, de Jonge B, Veldman J, Van Foreest N D (2020). Condition-based production planning: Adjusting production rates to balance output and failure risk. *Manufacturing & Service Operations Management*, 22: 792–811
- Van Horenbeek A, Buré J, Cattrysse D, Pintelon L, Vansteenwegen P (2013). Joint maintenance and inventory optimization systems: A review. *International Journal of Production Economics*, 143(2): 499–508
- van Staden H E, Boute R N (2021). The effect of multi-sensor data on condition-based maintenance policies. *European Journal of Operational Research*, 290(2): 585–600.
- Vora M, Thangeda P, Grussing M N, Ornik M (2023). Welfare maximization algorithm for solving budget-constrained multi-component POMDPs. *IEEE Control Systems Letters*, 7: 1736–1741
- Wang J, Liu H, Lin T (2023a). Optimal rearrangement and preventive maintenance policies for heterogeneous balanced systems with three failure modes. *Reliability Engineering & System Safety*, 238: 109429
- Wang J, Wang Y, Fu Y (2023b). Joint optimization of condition-based maintenance and performance control for linear multi-state consecutively connected systems. *Mathematics*, 11(12): 2724
- Wang J, Zhu X (2021). Joint optimization of condition-based maintenance and inventory control for a k -out-of- n :F system of multi-state degrading components. *European Journal of Operational Research*, 290(2): 514–529
- Wang S, Zhao X, Wu C, Wang X (2023c). Joint optimization of multi-stage component reassignment and preventive maintenance for

- balanced systems considering imperfect maintenance. *Reliability Engineering & System Safety*, 237: 109367
- Wang X, Zhao X, Wang S, Sun L (2020). Reliability and maintenance for performance-balanced systems operating in a shock environment. *Reliability Engineering & System Safety*, 195: 106705
- Wei Z, Zhao Z, Zhou Z, Ren J, Tang Y, Yan R (2024). A deep reinforcement learning-driven multi-objective optimization and its applications on aero-engine maintenance strategy. *Journal of Manufacturing Systems*, 74: 316–328
- Xia L, Zhao Q, Jia Q S (2008). A structure property of optimal policies for maintenance problems with safety-critical components. *IEEE Transactions on Automation Science and Engineering*, 5(3): 519–531
- Xu J, Liang Z, Li Y F, Wang K (2021). Generalized condition-based maintenance optimization for multi-component systems considering stochastic dependency and imperfect maintenance. *Reliability Engineering & System Safety*, 211: 107592
- Xu J, Zhao X, Liu B (2022). A risk-aware maintenance model based on a constrained Markov decision process. *IIEE Transactions*, 54(11): 1072–1083
- Xu M, Jin X, Kamarthi S, Noor-E-Alam M (2018). A failure-dependency modeling and state discretization approach for condition-based maintenance optimization of multi-component systems. *Journal of Manufacturing Systems*, 47: 141–152
- Ye Z, Cai Z, Yang H, Si S, Zhou F (2023). Joint optimization of maintenance and quality inspection for manufacturing networks based on deep reinforcement learning. *Reliability Engineering & System Safety*, 236: 109290
- Zhang C, Li Y F, Coit D W (2023a). Deep reinforcement learning for dynamic opportunistic maintenance of multi-component systems with load sharing. *IEEE Transactions on Reliability*, 72(3): 863–877
- Zhang J, Zhao X, Song Y, Qiu Q (2022). Joint optimization of condition-based maintenance and spares inventory for a series-parallel system with two failure modes. *Computers & Industrial Engineering*, 168: 108094
- Zhang N, Deng Y, Liu B, Zhang J (2023b). Condition-based maintenance for a multi-component system in a dynamic operating environment. *Reliability Engineering & System Safety*, 231: 108988
- 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
- Zhang Z, Wu S, Li B, Lee S (2013). Optimal maintenance policy for multi-component systems under Markovian environment changes. *Expert Systems with Applications*, 40(18): 7391–7399
- Zhao X, He Z, Wu Y, Qiu Q (2022). Joint optimization of condition-based performance control and maintenance policies for mission-critical systems. *Reliability Engineering & System Safety*, 226: 108655
- Zhao X, Liu B, Xu J, Wang X L (2023). Imperfect maintenance policies for warranted products under stochastic performance degradation. *European Journal of Operational Research*, 308(1): 150–165
- Zhao X, Wang Z (2022). Maintenance policies for two-unit balanced systems subject to degradation. *IEEE Transactions on Reliability*, 71(2): 1116–1126
- Zhao X, Wu C, Wang X, Sun J (2020). Reliability analysis of k -out-of- n : F balanced systems with multiple functional sectors. *Applied Mathematical Modelling*, 82: 108–124
- Zheng M, Su Z, Wang D, Pan E (2024a). Joint maintenance and spare part ordering from multiple suppliers for multicomponent systems using a deep reinforcement learning algorithm. *Reliability Engineering & System Safety*, 241: 109628
- Zheng M, Ye H, Wang D, Pan E (2024b). Joint decisions of components replacement and spare parts ordering considering different supplied product quality. *IEEE Transactions on Automation Science and Engineering*, 21(2): 1952–1964
- Zheng R, Xing Y, Ren X (2023). Multilevel preventive replacement for a system subject to internal deterioration, external shocks, and dynamic missions. *Reliability Engineering & System Safety*, 239: 109507
- Zheng R, Zhou Y, Gu L, Zhang Z (2021). Joint optimization of lot sizing and condition-based maintenance for a production system using the proportional hazards model. *Computers & Industrial Engineering*, 154: 107157
- Zhou H, Li Y (2023). Optimal replacement in a proportional hazards model with cumulative and dependent risks. *Computers & Industrial Engineering*, 176: 108930