Reinforcement learning-based maintenance scheduling for a stochastic deteriorating fuel cell considering stack-to-stack heterogeneity

Jian Zuo, Nadia Yousfi Steiner, Zhongliang Li, Catherine Cadet, Christophe Bérenguer, Daniel Hissel

Highlights

Reinforcement learning-based maintenance scheduling for a stochastic deteriorating fuel cell considering stack-to-stack heterogeneity

Jian Zuo, Nadia Yousfi Steiner, Zhongliang Li, Catherine Cadet, Christophe Bérenguer, Daniel Hissel

- A condition-based maintenance scheduling problem for fuel cells is established.
- Both stationary and time-varying maintenance policies are built.
- Both periodic and non-periodic inspection policies are investigated.
- Influence of the stack-to-stack deterioration heterogeneity on maintenance is investigated.
- Solution of the maintenance problem using reinforcement learning techniques.

Reinforcement learning-based maintenance scheduling for a stochastic deteriorating fuel cell considering stack-to-stack heterogeneity

Jian Zuo[a], Nadia Yousfi Steiner[a], Zhongliang Li[a,*], Catherine Cadet[c], Christophe Bérenguer[c], Daniel Hissel[a,b]

[a]*Université de Franche-Comté, FEMTO-ST, UTBM, CNRS, , Belfort, 90000, France*
[b]*Institut Universitaire de France (IUF), France*
[c]*Univ. Grenoble Alpes, CNRS, Grenoble INP, GIPSA-lab, , Grenoble, 38000, France*

Abstract

Maintenance scheduling of stack replacement is indispensable for achieving the durable and reliable operation of fuel cell systems. However, the related maintenance actions for minimizing long-term risk and cost in fuel cells constitute a complex optimization problem. Moreover, the complicated stochastic degradation behavior of fuel cell stacks hid the development of maintenance scheduling policies. Thus, the action phase of maintenance scheduling in fuel cell prognostics and health management (PHM) studies and stochastic degradation modeling are identified as two major research gaps. This paper investigates the maintenance scheduling of a stochastically deteriorating fuel cell stack exposed to deterioration heterogeneity. The maintenance problem consists in finding preventive control limits and inspection intervals that minimize total maintenance costs. A proximal policy optimization (PPO) based method is used to develop the maintenance strategy by describing the maintenance scheduling problem as a Markov decision process (MDP). Investigation results show that a steeper preventive control limit and more conservative inspection intervals are needed when heterogeneity exists. The proposed PPO-based method can solve the MDP and obtain more general maintenance policies considering the stochastic deterioration of fuel cell systems. The formulation of a condition-based maintenance scheduling problem, the reinforcement learning-based solution, and the investigation of stack-to-stack degradation heterogeneity's influence constitute this work's main novelties. This study provides useful insights for the development of fuel cell PHM studies, particularly within the action phase, helping to enhance the durability and reliability of fuel cell systems by better operating them. So the paper could be useful for different fuel cell applications. Meanwhile, the framework discussed in the paper, i.e., developing reinforcement learning for maintenance planning, is generic enough to be referred to in other related applications.

Keywords: Proton exchange membrane fuel cell, maintenance scheduling, proximal policy optimization, Markov decision process, stochastic deterioration

1. Introduction

1.1. Background

Prognostics and health management (PHM) is a dynamic approach that enables the health assessment, diagnostics, prognostics, and health management of fuel cell systems [1]. It is recognized as one of the most effective methods for addressing durability and reliability challenges in fuel cells [2, 3]. PHM can be divided into three distinct phases: observation, analysis, and action. The observation phase involves the acquisition and processing of durability data [4]. Subsequently, the processed data are utilized for the diagnostics and prognostics of the studied fuel cell system. This allows for the determination of both the current and future health state of the fuel cell. Finally, the action phase is undertaken to make operational decisions, such as determining fuel cell output power and evaluating the need for stack replacement. These decisions are based on the outcomes of the analysis phase, with the ultimate goal of optimizing the operation of fuel cell systems to achieve extended durability and enhanced reliability [5, 6]. Unless specified otherwise, the term "fuel cell" used in this work refers to a proton exchange membrane fuel cell.

*Corresponding author.
Email address: zhongliang.li@univ-fcomte.fr (Zhongliang Li)

1.2. Literature review

The observation and analysis phases have received significant attention in the field of fuel cells. Numerous experiments have been conducted to investigate fuel cell degradation. Ren et al. [7] provided a comprehensive summary of studies examining the impact of three common driving conditions (i.e., open-circuit/idling, dynamic load, and startup-shutdown) on fuel cell durability. Several datasets have been constructed based on fuel cell durability tests to support the development of the analysis phase [8, 9]. Extensive studies have been done for diagnostic and prognostic analyses in fuel cells based on the available data. Gong et al. [10] proposed a data-driven diagnosis framework that coupled fuel cell physics, sensor pre-selection, and deep leaning diagnosis to achieve high diagnosis accuracy. Ao et al. [11] conducted a diagnostic study based on electrochemical impedance spectroscopy data. Various data-driven approaches, such as recurrent neural network-based [12, 13, 14], echo state network [15], nonlinear autoregressive neural network and recurrent neural network [16, 17] have been developed for predicting fuel cell performance deterioration.

The action phase has received less attention compared to the previous phases. Yue et al. [18] reviewed the current status of fuel cell prognostics and decision-making, emphasizing that the limited availability of decision-oriented prognostic models and experimental datasets hinders the conduct of large-scale post-prognostic decision studies. Several works have focused on optimizing fuel cell power load decisions through energy management strategies (EMSs) [19]. For instance, Zhang et al. [20] proposed an equivalent consumption minimization strategy for minimizing fuel consumption in a hybrid fuel cell system. Yan et al. [21] optimized power allocation between multiple stacks to minimize overall hydrogen consumption. However, these methods do not consider fuel cell deterioration in their objective function, limiting their ability to address durability concerns. Developing deterioration-aware EMS requires an appropriate deterioration model for control purposes [22]. While deterministic empirical deterioration models proposed by Pei et al. [23], Chen et al. [24] have been widely adopted, they fail to account for the stochastic nature of fuel cells, including stack-to-stack deterioration variability [25]. Generally, the deterioration heterogeneity of fuel cells comes from two sources: i) the manufacturing procedure; ii) the design of the fuel cell system. The deterioration heterogeneity has been observed in experiments [26, 27]. Furthermore, load allocation-based management strategies are often developed independently of diagnostics and prognostics approaches, which may limit their practical applicability.

Previous studies have primarily focused on power load allocation decisions in fuel cell management, overlooking maintenance decisions that are crucial for enhancing the system's availability and reliability. Maintenance encompasses a range of technical and management actions throughout the lifetime of an item, aimed at preserving or restoring its performance to meet required standards [28, 29]. Maintenance strategies are classified into corrective or reactive, proactive, and opportunistic maintenance according to when maintenance is conducted [30]. Corrective maintenance is a failure-based maintenance strategy, namely maintenance is conducted only when a failure has occurred. This strategy can avoid unnecessary inspection. However, it can be impractical for systems with high failure rates thus lowering system reliability. Unexpected failures may also cause more downtime costs. Proactive maintenance is designed to schedule inspection and replacement before failure occurs. Preventive and condition-based maintenance are typical proactive maintenance strategies. Opportunistic maintenance is composed of a group of maintenance activities that are carried out on a component when an opportunity exists during the maintenance activities on other components in a multi-component system [31]. Among the existing maintenance strategies, condition-based maintenance is widely employed due to its efficiency in enhancing the system's reliability. We proposed a classical condition-based inspection replacement policy, with a non-stationary control limit rule policy for the replacement. The originality comes from the method used to optimize the policy rather than from the structure of the policy itself. The solution method allows for deriving automatically the form of the non-stationary control limit policy. The originality comes also from the type of degrading system to which it is applied, namely, with random effects.

Maintenance policy modeling of a fuel cell system requires three important factors, namely, the deterioration process, system structure, and maintenance process [32]. This interdisciplinary study is vital for developing effective maintenance strategies. The deterioration process aims to model fuel cell performance decay using stochastic process functions, e.g., Gamma process [33], Wiener process [34]. System structure directly influences system reliability and availability, such as redundancy structure [35]. To improve system reliability and availability, maintenance operations such as preventive maintenance, and corrective maintenance are needed to restore the system.

Renewal theory and Markov decision process (MDP) are two frameworks for dealing with maintenance optimization. Omshi and Grall [36] studied a maintenance cost model derived from semi-regenerative property (i.e., the expected cost rate of the system is equal to the ratio of the expected cost incurred in a renewal cycle). Castro et al.

[37] compared the computation of the expected cost rate using semi-regenerative and renewal theory. It is revealed that the semi-regenerative techniques are more efficient in computation time. The maintenance decisions can also be optimized by formulating the problem as a MDP. Zhang et al. [38] investigated the influence of deterioration heterogeneity on the optimal maintenance policy using MDP. The optimal maintenance decisions were calculated by solving a Bellman equation based on the proposed MDP through the value iteration method (tabular-based). However, these conventional tabular-based approaches require that the state and action variables are in discrete space, which restricts their application in real-world scenarios. Deep reinforcement learning (DRL) overcomes these limits by incorporating the generalization power of neural networks. Zhang and Si [39] utilized DRL to solve a maintenance planning optimization problem for multi-component systems, demonstrating the advantages in terms of computation efficiency and ability to handle high-dimensional problems. A policy proximal optimization-based DRL method is developed to handle the operation and maintenance problem in renewable energy systems [40]. This strategy is proven to outperform maintenance strategies such as deep Q-network. A DRL framework is developed to derive the cost-optimal maintenance policy for wind turbine components [41]. The dynamic inspection intervals and repair threshold are optimized to formulate the optimal maintenance policy.

1.3. Research gaps and contributions

The following research gaps in fuel cell PHM studies are identified and addressed:

- Existing PHM studies typically address individual phases independently, lacking a comprehensive and integrated methodology to enhance fuel cell durability and reliability. Works such as [13, 14] focus solely on the prognostics of fuel cells in the analysis phase. [10, 11] focus on the diagnosis of fuel cells. Moreover, these studies fail to directly improve fuel cell durability and reliability since no operation actions are optimized.
- The decision-making problem of maintenance scheduling has been overlooked in current PHM studies. Existing works on the action phase in fuel cell PHM mainly deal with the optimization of power load allocation [19, 21].
- Current deterministic deterioration models such as [23, 24] do not consider the stochastic nature of fuel cell deterioration, rendering them inadequate for modeling maintenance scheduling.

The following contributions are made to this work on maintenance scheduling of fuel cells:

- Concerning the gaps in lack of investigation on maintenance scheduling problems and the use of deterministic degradation models, a condition-based maintenance scheduling optimization problem is proposed for a stochastically deteriorating fuel cell.
- The use of a reinforcement learning-based approach enables the possibility of considering non-stationary maintenance decision-making policy whose structure is unknown. The proposed approach can handle stochastic fuel cell degradation model and optimize the maintenance scheduling decisions which contributes to the investigation of analysis and action phases in fuel cell PHM.
- A policy gradient-based reinforcement learning model optimizes the maintenance policy efficiently.
- The influence of the heterogeneity of the deterioration on the maintenance policy is revealed.

After this introduction, in Section 2, a brief introduction of the Gamma process and reinforcement learning methods for fuel cell degradation modeling and maintenance scheduling optimization is provided. Next, in Section 3, we provide an overview of the stochastic models used and describe the fuel cell maintenance scheduling problem. Subsequently, in Section 4, a reinforcement learning-based approach to address the maintenance optimization problem is proposed. To evaluate the performances of the proposed maintenance strategy, we compare it with a linear function-based empirical policy in Section 5. Lastly, in Section 6, we summarize the key findings and present perspectives for the condition-based maintenance of fuel cells.

2. Material and methods

Gamma process (GP) and reinforcement learning are the two approaches to developing the maintenance scheduling problem in PEM fuel cells. The stochastic degradation modeling of fuel cell stacks is developed based on Gamma process models. The maintenance scheduling decisions are optimized using a reinforcement learning-based approach.

Gamma process

A GP is a stochastic process with independent, positive increments that obey a Gamma distribution $Ga(\alpha,\beta)$ characterized by its shape parameter α and scale parameter β. By definition, the increment of a GP $X(t)$ between t_1 and t_2 ($t_2 > t_1 \geq 0$) is given by [33]:

$$\Delta X(t_1, t_2) \sim Ga((v(t_2) - v(t_1)), \beta) \tag{1}$$

where $\Delta X(t_1, t_2) \triangleq X(t_2) - X(t_1)$. v is the coefficient of the shape function. For a stationary GP with constant scale parameter β, $f_{Ga}(x, v(t_2 - t_1), \beta)$ represents the probability density function of the Gamma law with shape function $v(t)$ and scale parameter β:

$$f_{Ga}(x, v(t_2 - t_1), \beta) = \frac{x^{v(t_2-t_1)-1} e^{-x/\beta}}{\beta^{v(t_2-t_1)}\Gamma(v(t_2 - t_1))} \tag{2}$$

where $\Gamma(t) = \int_0^{+\infty} x^{t-1}e^{-x}dx$ is the Gamma function.

The mean and variance of the deterioration increment in the time interval (t_1, t_2) are given by:

$$\text{Mean}(\Delta X(t_1, t_2)) = v(t_2 - t_1) \cdot \beta \tag{3}$$

$$\text{Var}(\Delta X(t_1, t_2)) = v(t_2 - t_1) \cdot \beta^2 \tag{4}$$

Then, the mean deterioration increment in a time interval of length Δt is calculated as $v(\Delta t)\beta$, independent of when the interval begins. The variance of the process increases with the time horizon between t_1 and t_2. Besides, the variance of the process can be tuned independently of the mean.

GP is suitable to model gradual degradation monotonically accumulating over time in a sequence of small increments. For instance, a two-phase Gamma process with fixed change-point is used to model lithium battery voltage decay. This deterioration model can be further used to estimate battery lifetime, state of charge, etc. which contributes to the energy management of lithium batteries [42]. Besides, GP is a well-formulated stochastic process that makes it convenient for mathematical analysis.

Reinforcement learning

Reinforcement learning (RL) focuses on learning how to map situations to actions to maximize a numerical reward. In general, RL handles a sequential decision problem through a trial-and-error approach. This is achieved by enabling an RL agent to iteratively interact with an environment and acquire policy learning. The key elements in RL include environment, agent, policy, reward, and value function.

- The world that the RL agent lives in and interacts with is called the environment. At each interaction step, the state of the environment only changes with the received action from an agent.
- An agent refers to a learning algorithm that senses the state of its environment, takes actions, and perceives reward signals. The goal of the agent is to maximize its cumulative reward.
- A policy is a decision rule for an agent to prescribe actions under perceived states.
- A reward signal defines the objective of an RL problem. At each interaction, the environment sends to the RL agent a single number called the reward.
- Value function refers to the value of a state, or state-action pair. The term "value" means the expected return an agent can expect to accumulate over the long run, starting from that state.

When applying RL methods, the above key elements must be defined to formulate the RL problem and obtain the optimal policy for maximizing the long-term reward.

RL algorithms can be classified into model-free and model-based approaches. Policy optimization and Q-learning are two model-free methods widely adopted in RL. What to learn in RL is a key factor for distinguishing these two approaches. Policy optimization represents a policy explicitly as $\pi_\theta(a|s)$ and optimizes the parameter θ either directly through gradient ascent on the defined performance objective such as advantage actor-critic or indirectly, through maximizing local approximation of the performance objective such as proximal policy optimization. Q-learning learns an approximation for the optimal action-value function for deciding the optimal action.

3. Fuel cell deterioration model and maintenance problem statement

3.1. Fuel cell deterioration model

This section presents stochastic deterioration models of fuel cell stacks using a Gamma process. Here, we present only the main formulas of the proposed models. For further details on the modeling works, please refer to [25, 43, 44].

Gamma process model

Fuel cell stack resistance is chosen as a health indicator, namely a degradation performance indicator, the deterioration of a fuel cell is then characterized by the increment of its resistance. The deterioration due to a fixed load amplitude (power) is modeled as a stationary Gamma process, which means that the increment of the deterioration level due to the load level between time t_1 and t_2 ($t_2 > t_1 \geq 0$) can be modeled as:

$$\Delta R_L(t_1, t_2) = R_L(t_2) - R_L(t_1) \sim Ga(v(t_2 - t_1), \beta) \tag{5}$$

where v is coefficient of the shape function, β is the scale parameter. $Ga(v(t_2 - t_1), \beta)$ represents the Gamma distributed random variable with shape parameter $v(t_2 - t_1)$ and scale parameter β. Its probability density function is:

$$f_{Ga}(x, v(t_2 - t_1), \beta) = \frac{x^{v(t_2-t_1)-1} e^{-x/\beta}}{\beta^{v(t_2-t_1)} \Gamma(v(t_2 - t_1))}. \tag{6}$$

To link fuel cell degradation with operation load, an empirical degradation rate function $D(L)$ is proposed for the GP model [43, 44]. The shape parameter v is modeled as a function of the load L. The mean of resistance increment over a time unit interval, which is also the average deterioration rate, can thus be written as $D(L) = v(L)\beta$.

Random-effect Gamma process model (GP-RE)

The random effect model is studied due to the existence of stack-to-stack deterioration variability in PEM fuel cells. Generally, the deterioration variability comes from two sources: i) the manufacturing procedure, ii) the design of fuel cell systems. The stochastic differences between different cells/elements forming the stack can exit during the manufacturing procedure. For example, the hydrophobicity of bipolar plate surface, membrane electrical properties, and mechanical properties of seals. It is very unlikely to make them perfectly identical for different stacks. A fuel cell stack is much more complicated than a single fuel cell due to the stacked structure. When the reactants pass through the single inlet to the inside cells, it is hard to make them evenly distributed. The same problems exist for the cooling and heating of a stack. These phenomena will affect fuel cell performance which causes deterioration variability. To account for stack-to-stack deterioration variability, a random effect is added to the GP model, by re-parameterizing the scale parameter as a random variable, following a Gamma law.

$$\begin{gathered} \Delta R_L(t_1, t_2) \sim Ga(v \cdot (t_2 - t_1), \beta_s) \\ \beta_s \sim Ga(\delta, \phi) \end{gathered} \tag{7}$$

where δ, ϕ are the shape and scale parameters for defining the β (Eq. (5)). The new shape parameter sampled from $Ga(\delta, \phi)$ is denoted as β_s.

According to [45], the first-hitting-time (T) of GP-RE trajectory to failure threshold FT can be analytically computed as:

$$F_T(t|v, \delta, \phi) = \mathbb{P}(T \leq t) = 1 - F_{2vt,2\delta}\left(\frac{\delta \cdot FT}{\phi v t}\right) \tag{8}$$

where $F_{i,j}(\cdot)$ stands for the cumulative distribution function of F-distribution with degrees of freedom (i, j).

GP-based models have a small number of parameters that can be estimated from fuel cell deterioration data. These models can efficiently generate degradation paths to facilitate the development of decision-making strategies for fuel cells. Fig. 1 shows the fuel cell performance (output power) decay trajectories obtained by GP and GP-RE models. The use of a Gamma process (that is monotonic) is not the only possible choice for stochastic deterioration modeling for fuel cells. The deterioration could have been modeled by another stochastic process, such as Wiener (that allows for non-monotonic paths). We assume that the overall deterioration trend of the tested fuel cell stack is monotonic. The variation in the experimental data is assumed to be caused by variation in operating conditions, e.g., current density levels that do not reflect fuel cell deterioration. Therefore, the GP can be used to model the actual degradation

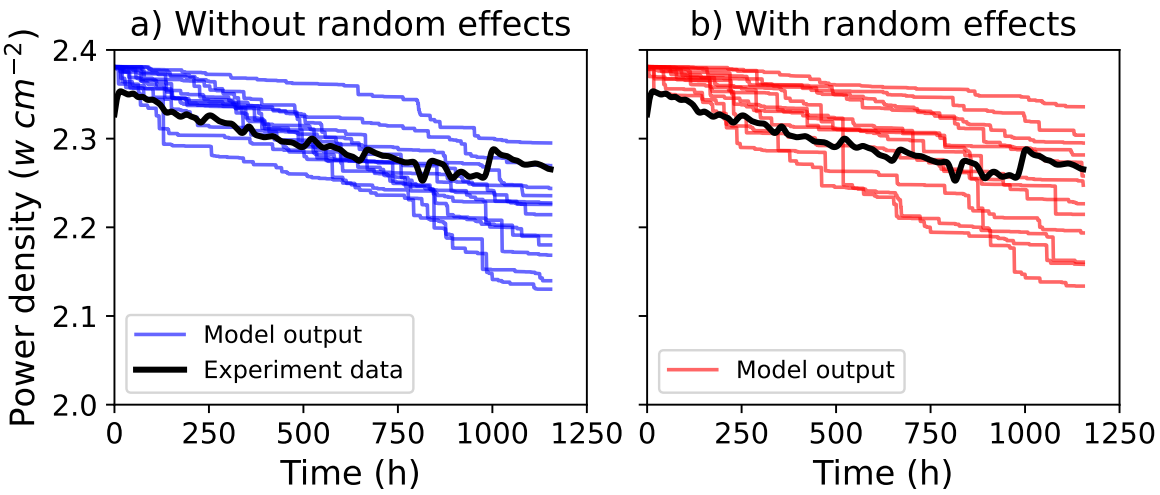

Figure 1: Sampled fuel cell output power deterioration trajectories. When the random effects exist, the unit degradation rates vary from unit-to-unit.

state of fuel cells that is monotonically increasing. The shape and scale parameters of the GP model are estimated using the dataset from [8]. Notably, a noticeable variance in deterioration rates becomes apparent in the presence of random effects.

3.2. Maintenance problem formulation

A schematic of maintenance scheduling for a single fuel cell stack is depicted in Fig. 2. As discussed in Section 3.1, the deterioration is modeled through a stationary Gamma process with or without random effects. The stack is considered to have failed when its deterioration level exceeds the failure threshold FT. The lifetime or failure time of a fuel cell stack is defined as the duration time from initial operation till the output voltage drops by 10% under the rated power conditions [23].

Maintenance assumptions

The maintenance optimization of a single fuel cell stack is based on the following assumptions:

- The stack's degradation level (continuously deteriorating over time) can only be assessed with inspections. The inspection of a fuel cell stack refers to measuring the fuel cell deterioration level which corresponds in our case to the value of the stack resistance. Note that this value cannot be directly measured, but rather it should be estimated from measurements retrieved from experimental polarization curves or electrochemical impedance spectroscopy. In practice, inspecting the fuel cell means performing an impedance spectroscopy or polarization curve, and estimating the value of the internal resistance from these experimental data. The inspection is assumed to be perfect, that is, the deterioration level of a stack can be perfectly known.
- Maintenance operations can only be performed at inspection time.
- Stack failure can only be revealed with an inspection.
- The maintenance of the balance of plant components is not considered, assuming they are durable and reliable enough to support fuel cell system operation.
- The maintenance operation is assumed to be perfect, that is, after maintenance (preventive maintenance or corrective maintenance), the stack's degradation restores to an as-good-as-new (AGAN) state. The PM and CM operations bring back the stack's age to zero. This is a strong assumption but reasonable since we are considering stack replacement as a maintenance action.

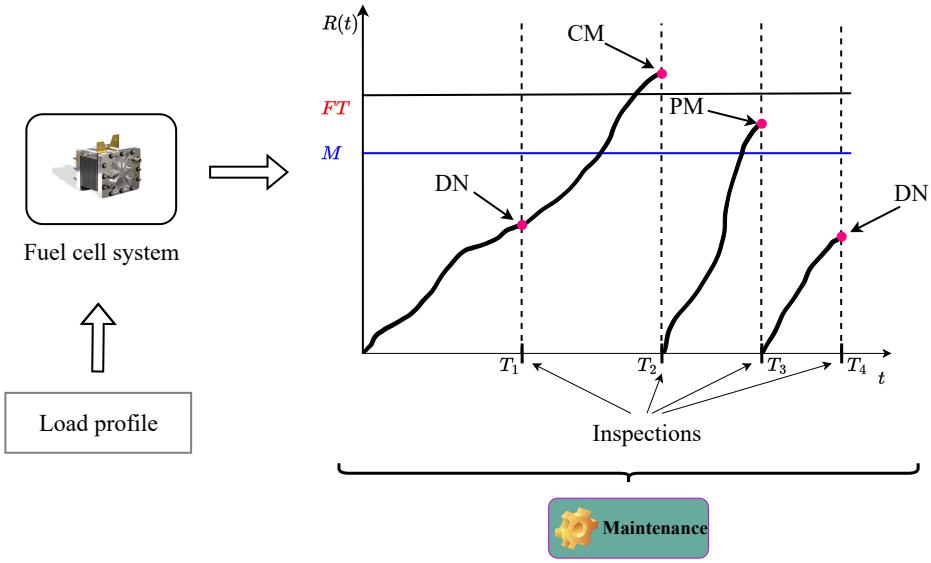

Figure 2: Schematic diagram of maintenance schedule in fuel cells.

Maintenance problem statement

As demonstrated in Fig. 2, three potential maintenance actions, denoted as $\mathbb{A} = \{\text{PM, CM, DN}\}$, are available based on the observed deterioration level of the stack. These actions include preventive maintenance (PM), corrective maintenance (CM), and doing nothing (DN). The decision variable for prescribing different maintenance actions is the preventive control limit $M(t_a)$ ($0 < M(t_a) < FT$). DN is assigned when the stack's deterioration level lies below $M(t_a)$ (inspections T_1, T_4). PM is assigned at inspection T_3, and CM is assigned at inspection T_2.

The overall maintenance costs are accumulated based on the consequences of the prescribed actions. Let t_a represent the age of the studied fuel cell stack from the beginning of its lifetime, and let the stack's deterioration level be denoted as $R(t_a)$. Assuming the system degradation $R(t_a)$ is Markovian [36], $(t_a, R(t_a))$ together form a discrete time continuous state Markov chain. At each inspection, a maintenance decision must be made from the action space $\mathbb{A}$.

- If $R(t_a) \geq FT$, a CM action has to be performed at a cost of C_c. After the CM, The deterioration level $R(t_a)$ is restored to the initial state R^0 (AGAN state). Besides, the unavailability cost is incurred due to stack failure.
- If $M(t_a) \leq R(t_a) < FT$, a PM action is taken, the deterioration level $R(t_a)$ restores to R^0 at a cost of C_p.
- If $0 < R(t_a) < M(t_a)$, a DN action is selected, and no maintenance cost is incurred. However, as the fuel cell continues to deteriorate, possible unavailable costs will be incurred at the next inspection time $t_a + d$, where d is the inspection interval.

The challenge lies in striking a balance between two competing objectives. On one hand, we aim to maximize the utilization of a fuel cell stack to avoid shorting its remaining useful lifetime (RUL). On the other hand, it is crucial to replace a deteriorated stack before it fails, as the cost associated with corrective replacements can be high. This situation creates a tradeoff between continuing to use a fuel cell stack, which may be at risk of system failure, and replacing the stack, which results in the loss of its RUL. To address this tradeoff, maintenance scheduling optimization seeks to identify the optimal inspection time (controlled by t_{ins}) and maintenance actions (controlled by $M(t_a)$) that minimize overall maintenance costs $C(t)$.

To assign a sequence of maintenance decisions during the period of time $[0, H]$, an overall cumulative maintenance cost is defined:

$$C(H) = C_i N_i(H) + C_p N_p(H) + C_c N_c(H) + C_u d(H) \tag{9}$$

Table 1: Summary of major variables used in Section 3.

Variables	Description
C_i	Unit inspection cost
C_p	Unit preventive maintenance cost
C_c	Unit corrective replacement cost
C_u	System unavailability cost per time unit
$N_p(H)$	Number of preventive replacement in [0,H]
$N_c(H)$	Number of corrective replacement in [0,H]
$d(H)$	System unavailable time during [0,H]
$C(H)$	Overall maintenance cost during [0,H]
v	Coefficient of the Gamam shape function
β	Gamma scale parameter
FT	Fuel cell stack failure threshold
M	Preventive control limit
T_{ins}	Inspection interval

where H is the considered operating time horizon, $N_i(H)$ is the total number of inspections in $[0, H]$, C_i is the unit inspection cost. $N_p(H)$ is the number of preventive replacements in $[0, H]$, C_p is the unit preventive replacement cost. $N_c(H)$ is the number of corrective replacement in $[0, H]$, C_c is the unit corrective replacement cost. The unit preventive maintenance is assumed to be smaller than the unit corrective maintenance, namely, $C_p \ll C_c$. C_u is system unavailability (downtime) cost per time unit and $d(H)$ stands for system unavailable time during $[0, H]$.

Then, the maintenance optimization problem consists in finding the optimal inspection T_{ins} and $M(t_a)$ that minimize overall maintenance cost $C(H)$, which writes,

$$\underset{T_{ins}, M(t_a)}{\operatorname{argmin}} \quad C(H) \tag{10}$$

where t_{ins} is inspection interval decision, $M(t_a)$ is used to prescribe the maintenance actions from $\mathbb{A}$.

Note that both $M(t_a)$ and T_{ins} can be either constant or time-varying during the operating period $[0, H]$. In Fig. 2, periodic maintenance refers to the scenarios when inspection intervals are equal, namely, $T_2 - T_1 = T_3 - T_2 = T_4 - T_3$. Non-periodic inspection releases this restriction. As a result, two types of maintenance scheduling problems are formulated:

- Periodic inspection-based maintenance scheduling: Periodic inspections are scheduled during system operation, and the problem is to optimize the preventive replacement threshold, that is, the control limit $M(t_a)$ for deciding maintenance operations. T_{ins} is considered as a given constant value. Periodic inspection plans are easy to implement. However, they are limited in reaching the optimal inspection policy for minimizing the maintenance cost.

- Non-periodic inspection-based maintenance scheduling: At each inspection, the next inspection time is treated as a decision variable and has to be optimized regarding the overall maintenance costs. The control limit $M(t_a)$ is assumed to be a constant and its value is considered to be given. Non-periodic inspection plans are more flexible than periodic inspections which allows to avoid wasting useful lifetime and decreasing inspection labor costs. However, those strategies are difficult to implement.

The primary goal of studying non-periodic inspection is to reduce inspection expenses [46, 47]. Regular inspections can result in accumulated inspection costs, but they can also reduce system failure and unavailability costs. Conversely, less frequent inspections decrease the cost of inspection but increase the costs associated with system failures and unavailability.

A summary of major variables used in this section is presented in Table 1. The next step is to design an optimization algorithm for solving the proposed maintenance optimization problems.

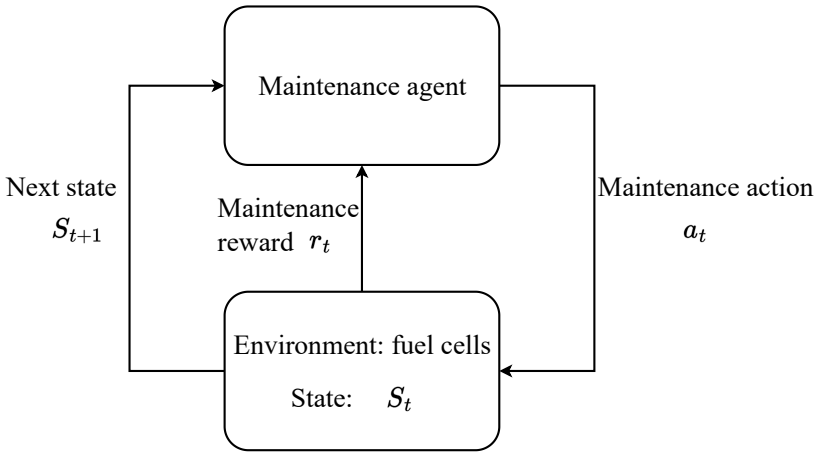

Figure 3: Schematic of reinforcement learning loop for maintenance.

4. Reinforcement learning-based maintenance policy optimization

This section presents the RL-based maintenance policy for solving the proposed maintenance scheduling problem. The RL-based approach gives us the possibility to consider non-stationary maintenance decision-making policy whose structure is unknown or chosen a prior. Let RL-M(t) denote the RL-based maintenance policy for the periodic inspection-based maintenance problem, and RL-d(t) denote the RL-based maintenance policy for the non-periodic inspection-based maintenance problem.

Fig. 3 illustrates the classical agent-environment loop in RL. The maintenance agent, represented by an RL algorithm, interacts with the fuel cell environment by performing maintenance actions. These actions influence the state of the fuel cell, transitioning it from S_t to S_{t+1}, while also receiving an immediate reward r_t. Subsequently, the maintenance agent generates the next action by considering updated observations and rewards. This entire process is repeated iteratively to maximize the accumulated future return G_t, which is expressed as:

$$G_t = \sum_{k=0}^{H} \gamma^k r_{t+k} \tag{11}$$

where γ is the discount factor, H is the considered horizon. In the context of fuel cell studies, H is a finite variable. Note that the accumulated return G_t can be viewed as a negative form of the accumulated maintenance cost $C(H)$ as introduced in Eq. (9). Thereby, maximizing G_t leads to minimizing $C(H)$.

4.1. Reinforcement learning environment

The proposed fuel cell deterioration model constitutes the RL environment. This environment determines the transition of the fuel cell deterioration state in the RL loop. The deterioration state transition follows a stochastic pattern based on the Gamma process model (Eqs. (5, 7)). Note that the deterioration state remains within the range of R^0 to FT. Before stack failure, the deterioration state continues to increase when DN action is received. However, the deterioration state of a stack restores to AGAN state (R^0) following a PM or CM action.

For each state-action pair (S, a) in the maintenance optimization problem, a reward is assigned to indicate the immediate gain or loss resulting from the selected action a in state S. To maximize the expected trajectory return (Eq. 11), an RL agent is required to interact with the fuel cell environment. The trajectory refers to the collection of transition histories during $[0, t]$. The RL agent learns an optimized maintenance policy $\pi(a|S) : S \rightarrow a$, which determines the optimal maintenance action to be taken under specific states.

4.2. Reinforcement learning agent

As described in Fig. 3, an agent's role is to make optimal decisions regarding operational actions with the aim of maximizing the long-term expected reward. To achieve this objective, a proximal policy optimization agent is developed and utilized to optimize the maintenance decisions.

Proximal policy optimization

Proximal policy optimization (PPO) is a policy gradient-based method to maximize the expected return given some proximal constraints [48]. In PPO, a key focus is how to effectively utilize the collected interaction data to facilitate policy improvement without accidentally leading to performance collapse. Studies in RL have indicated that the clipped version of PPO (PPO-Clip) demonstrates superior performance and is relatively easier to implement compared to other policy gradient algorithms such as trust region policy optimization [49, 50]. Thus PPO-Clip is chosen to optimize the maintenance policy. We use PPO to refer to the PPO-Clip.

PPO follows an advantage actor-critic approach, which combines both value-based and policy-based methods to enhance the stability of policy learning. The architecture includes:

- An actor that controls how the agent behaves (policy-based);
- A critic that measures the quality of the action taken (value-based).

The design of a clipped global loss ensures that policy updates that would produce significant shifts from the previous configuration are discouraged whether the advantage is positive or negative. Specifically, the clipped objective is:

$$Loss_t^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}, \text{clip}(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}, 1-\epsilon, 1+\epsilon)\right)\hat{A}_t(s,a)\right] \tag{12}$$

where $\text{clip}(x, l, u) := \max(\min(x, u), l)$. ϵ is the hyperparameter for controlling the clip range. θ are the policy parameters to be optimized at the current time, and θ_k are the parameters of the policy at iteration k. $\pi(a|s)$ is denoted as the policy function which indicates the probability of selecting action a under state s. $\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}$ is the probability ratio. $\hat{A}_t(s,a)$ stands for the estimated advantage function which characterizes the advantage of taking a specific action. The clipped objective can be further simplified to

$$Loss_t^{CLIP}(\theta) = \mathbb{E}_t\left[\min\left(\frac{\pi_\theta(a|s)}{\pi_{\theta_k}(a|s)}\hat{A}_t(s,a), g(\epsilon, \hat{A}_t(s,a))\right)\right] \tag{13}$$

where

$$g(\epsilon, \hat{A}) = \begin{cases} (1+\epsilon)\hat{A} & \hat{A} \geq 0 \\ (1-\epsilon)\hat{A} & \hat{A} < 0. \end{cases}$$

PPO adopts the generalized advantage estimator to compute $\hat{A}_t$:

$$\hat{A}_t = \delta_t + (\gamma\lambda)\delta_{t+1} + \cdots + (\gamma\lambda)^{T-t+1}\delta_{T-1} \tag{14}$$

where δ_t is the Bellman residual term, also known as temporal difference error, which is computed as $\delta_t = r_t + \gamma V_{\hat{\theta}}(S_{t+1}) - V_{\hat{\theta}}(S_t)$. $r_t + \gamma V_{\hat{\theta}}(S_{t+1})$ is the target for the temporal difference update. γ is the discounting factor. t specifies the time index in [0, T], within a given length T trajectory segment. $V_{\hat{\theta}}$ is an approximate value function composed of a deep neural network model with parameter $\hat{\theta}$. $\lambda = 0.95$. The value target is calculated as: $V_{\hat{\theta}}^{tar} = V_{\hat{\theta}}(S_t) + \hat{A}_t$, and the value loss is defined as:

$$Loss_t^{VF} = \frac{1}{2}(V_\theta(S_t) - V_{\hat{\theta}}^{tar})^2 \tag{15}$$

where $V_\theta(S_t)$ is the estimated value function at state S_t and θ denotes the deep neural network model's parameters. The total PPO loss is expressed as:

$$Loss_t^{tot} = \mathbb{E}_t\{Loss_t^{CLIP}(\theta) + Loss_t^{VF}(\theta)\}. \tag{16}$$

Algorithm 1 summarizes the main calculation steps in PPO.

In periodic inspection-based maintenance, the PPO agent is tasked with determining whether to take DN or PM actions. For non-periodic inspection-based maintenance, the PPO agent is responsible for determining the appropriate inspection intervals. The inspection interval spans from 20 to 1000 h, with a step size of 20 h, resulting in a total of 50 possible inspection actions. Note that this is a design choice of this work and the focus here is to develop the PPO agent for obtaining the optimal maintenance policy under this design choice.

Algorithm 1: On-policy PPO

Output: Value network V_η, policy π_θ

Initialize network parameters: θ for policy network, η for value network; maximum training epochs N_{epo} and training batch B.
for *training iterations* = 1 **to** N_{epo} **do**
    Clear training batch B.
    for *each RL collect step* **do**
        Observe the environment state S_t, select action a_t according to the current state policy $\pi_\theta(a_t|S_t)$. Execute a_t, obtain the reward r_t, and the environment transitions to the next state S_{t+1}.
    end
    Compute advantage estimation $\hat{A}$ through Eq. (14) and add the experiences (S_t, a_t, r_t, S_{t+1}) to training batch B.
    for *each RL training step* **do**
        Recompute advantage estimate $\hat{A}$ through Eq. (14).
        Split the training batch B into K mini-batches according to the batch size.
        for *mini-batch* $k = 1$ **to** K **do**
            Compute the total PPO loss. Update critic and actor networks with respect to the $Loss_t^{tot}$ (The gradient calculation of both networks is detailed in [48]).
        end
    end
end

Reward function design

The expression of the reward function is derived for the proposed periodic maintenance and non-periodic maintenance.

- Basic reward function for periodic inspection-based maintenance

The reward function $r(t)$ is taken as the negative form of the maintenance cost $C(t)$ as introduced in Section 3. The reward function $r(t)$ is derived based on different maintenance actions that are:

1) DN action. The corresponding cost is the summation of the inspection cost and the potential unavailability (failure before the next inspection) cost.

The calculation of the above unavailability cost requires the conditional failure probability of the investigated deterioration process. In GP, the probability of stack failure at the next inspection time $a + d$, given an observed deterioration $R(a)$ can be calculated as follows [33]:

$$\begin{aligned} P_d(t = a + d) &= \mathbb{P}\left(T \leq t | R(a)\right) \\ &= \frac{\Gamma\left(\alpha(L_i) \cdot (a + d), (FT - R(a))\beta\right)}{\Gamma\left(\alpha(L_i) \cdot (a + d)\right)} \end{aligned} \tag{17}$$

where t is the time variable. For GP-RE model, P_d can be computed based on Eq. (8), which writes:

$$P_d(t = a + d) = \mathbb{P}(T \leq t | R(a)) = 1 - F_{2vt,2\delta}\left(\frac{\delta \cdot (FT - R(a))}{\phi v(a + d)}\right) \tag{18}$$

Then, the reward for taking DN is computed as:

$$r_{DN} = -C_u \cdot \frac{P_d(a + d) \cdot d}{k_1} - C_i \tag{19}$$

where k_1 is the hyperparameter for turning the value of r_{DN}. $P_d(a + d) \cdot d$ is used to denote the estimated potential unavailable time. $C_u \cdot \frac{P_d(a+d) \cdot d}{k_1}$ is the proposed formula for potential unavailable time cost.

2) PM action. The fixed maintenance cost term C_p is directly used to compute the reward.

$$r_{PM} = -C_p - C_u \cdot \frac{P_d(a+d) \cdot d}{k_1} \tag{20}$$

where $R(a) = R^0$ in term $P_d(a+d)$. The potential unavailable time cost term is the same as computed in Eq. (19).

3) CM action. The maintenance cost is composed of a fixed cost term C_c and unavailability cost.

Assuming the inspected deterioration level is $R(t_1)$ ($R(t_1) \geq FT$). The stack's deterioration level at the last inspection t_0 is $R(t_0)$. Then the reward of taking CM is

$$r_{CM} = -C_c - C_u \frac{R(t_1) - FT}{R(t_1) - R(t_0)}(t_1 - t_0) \tag{21}$$

where $\frac{R(t_1)-FT}{R(t_1)-R(t_0)}(t_1 - t_0)$ is the unavailable time calculated using the inspected degradation level $R(t_1)$.

The final reward function during $[0, t]$ is expressed as:

$$r(t) = \sum_{i=0}^{t} r_{DN}(i) + r_{PM}(i) + r_{CM}(i) \tag{22}$$

- Additional reward term for non-periodic inspection-based maintenance

Non-periodic inspection introduces a tradeoff between inspection and the cost related to system failures and unavailability, known as the inspection-failure tradeoff. We can schedule a longer inspection interval to save inspection costs for stacks with lower levels of deterioration, while shorter inspection intervals can be scheduled for stacks with higher levels of deterioration. The reward function in Eq. (22) does not contain any reward feedback for taking larger inspection intervals, thus failing to optimize the inspection-failure tradeoff. An additional reward term is necessary to quantify the reward associated with performing appropriate subsequent inspections.

One possibility is to design an additional reward term for encouraging "effective" inspections. "Effective" means when a control limit M is decided, the inspected deterioration R lies between M and failure threshold FT. Moreover, when this R is approaching but not exceeding FT, the more efficient the inspection is. Then, the additional reward term is expressed as:

$$r_e = \begin{cases} -k_2(M - (R(t) + \alpha\beta d)) & \text{if } R(t) + \alpha\beta t_d \leq M \\ k_3 \frac{R(t)+\alpha\beta d}{FT} & \text{if } FT > R(t) + \alpha\beta d > M \\ -k_4(R(t) + \alpha\beta d - FT) & \text{if } R(t) + \alpha\beta d > FT \end{cases} \tag{23}$$

where k_2, k_3, k_4 are positive coefficients for balancing maintenance cost terms. d is the inspection interval.

These three terms leverage the estimated future deterioration as well as the coefficients to encourage an effective inspection. k_2 corresponds to the term in which the estimated future deterioration is below M. In this case, a relatively longer inspection interval d will be assigned to ensure that the estimated deterioration $R(t) + \alpha\beta t_d$ approaches the control limit M, thus avoiding a significant negative reward. k_3 is the coefficient for assigning positive rewards for taking longer inspections when the conditions are met. According to Eq. (23), a positive reward is assigned when the estimated deterioration level is between M and FT, namely, "effective" decisions. The assigned reward is proportional to the value of the estimated deterioration. k_4 corresponds to the penalty term when the estimated deterioration exceeds FT.

To sum up, the reward function for a non-periodic inspection is computed as:

$$r(t) = \sum_{i=0}^{t} r_{DN}(i) + r_{PM}(i) + r_{CM}(i) + r_e(i) \tag{24}$$

5. Results and discussion

This section focuses on evaluating the performance of the proposed maintenance policy. An empirical strategy based on a linear function is introduced at first. Then, the maintenance optimization results are presented and compared for all strategies. The key findings and discoveries derived from the obtained results will be presented in terms of periodic inspection-based and non-periodic inspection-based maintenance.

5.1. Comparison strategy - a linear function-based maintenance policy

A linear function-based maintenance policy is proposed as follows:

$$M(t) = \frac{A}{1000} \cdot t + b \ \ (\Omega.cm^2), \quad t \geq 0 \tag{25}$$

$$d(R) = a \cdot R + c \ \ (h), \quad R \geq 0 \tag{26}$$

where t is the age of a fuel cell stack during the operation period. R is the stack's degradation level. $M(t)$ is the control limit function ($FT > M(t) \geq 10^{-5}\ \Omega.cm^2$), and $d(R)$ is the inspection time interval function ($d(R) \geq 10$ h). A, a are the slope parameters. b, c are the intercept parameters.

According to the dynamic properties of M and d, the above strategy can be categorized into:

- Both M and d are constant (periodic maintenance), that is, M-d;
- M is a linear function of stack age (time-varying), d is a constant (periodic maintenance), that is, M(t);
- d is a linear function of stack degradation level (time-varying), M is a constant (non-periodic maintenance), that is, d(t).

5.2. Simulation settings

The shape and scale parameters of the Gamma process are estimated from a real fuel cell experimental dataset using the method of moments method, $v = 0.01252, \beta = 4.38\times10^{-3}$ [25, 43]. The estimation accuracy can be accessed by comparing the estimated resistances and the measurements. The mean square error between the estimated average resistances and the measured resistances is 1.4×10^{-5}. For the random effects, the initial shape and scale parameters are set as $\delta = 43.8$, $\phi = 10^{-4}$ to simulate fuel cell stacks with relatively small heterogeneity. Note that the random effects parameters are set after the parameters of the GP model are decided. Due to the lack of available experimental data for estimating the random effects, the random effects parameters are set empirically by comparing with the experimental degradation trajectory as shown in Fig. 1. The value of FT is computed as 0.0972 $\Omega.cm^2$ when the initial degradation level $R^0 = 0.0\ \Omega.cm^2$. In our study, the fuel cell stack operates at a nominal current density, i.e., 0.7 $A.cm^{-2}$ [8].

Currently, there is very limited data on the specific maintenance costs for fuel cell systems. In general, the maintenance cost of a fuel cell stack reflects the cost of labor and parts of fuel cell stack components [51]. Then the maintenance cost coefficient settings are decided based on the assigned maintenance operations.

- C_i in inspection. We assume the inspection costs are mainly caused by the labor which measures/diagnoses fuel cell stack health state. The labor cost rate is set as \$60/h [52]. We assume that the inspection work takes around half an hour on average, thus C_i = \$30.
- C_c in corrective maintenance. CM involves labor costs and the cost of corrective replacement of the fuel cell stack. In practice, stack cost varies with system size. The cost of a new stack is around \$5000/kW [53]. For the investigated fuel cell stack (500 W), the cost rate per corrective maintenance C_c is set as \$3000 (stack cost plus labor cost).
- C_p in preventive maintenance. PM involves labor as well as component repair costs which are cheaper than CM. C_p = \$2100 is set as the average PM cost.
- C_u for unavailability costs. The unavailability costs vary depending on different applications. Here we target stationary applications and C_u = \$150/h is used to represent a medium unavailability cost.

For calculation convenience, the above costs are converted into ratio costs, that is, $C_i = 1, C_p = 70, C_c = 100, C_u = 5$. The original cost is obtained by multiplying by \$30.

5.3. Results for periodic inspection-based maintenance

Under the periodic inspection, the linear function-based maintenance policy studied in this section is restricted to a time-varying control limit, namely $M(t)$.

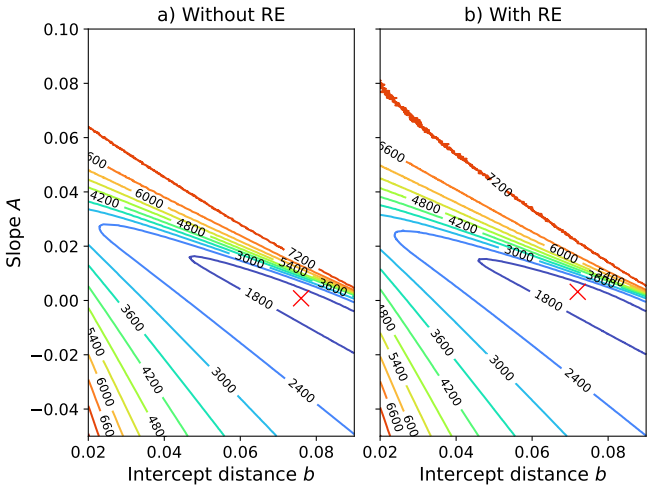

Figure 4: Contour plot of the overall maintenance costs under periodic inspection-time-varying control limit $M(t) = A \cdot t + b$: the optimized maintenance policy is $(A = 0.0008, b = 0.076)$ for GP model and $(A = 0.0032, d = 0.072)$ for GP-RE model. The red cross markers denote the data points with minimum maintenance costs.

Results for the linear function-based maintenance policy (M-d and M(t)-d)

For periodic inspection-based maintenance, the inspection interval d is set as a constant value. The baseline results are obtained by the linear function-based maintenance policy using Monte Carlo simulation. The number of Monte Carlo histories is set as 10^4 to ensure the convergence of the estimated maintenance costs. Heuristic search is used to search for the best combination of parameters that gives the lowest maintenance cost. Both the constant and linear function-based time-varying control limits are investigated.

When M is constant, namely $A = 0$ in Eq. (25), the searching range of M is from 0.01 to 0.09, with a value increment step of 0.002 $\Omega.cm^2$. d ranges from 100 to 1025, with a value increment step of 25 h (M-d). In the GP model, the combination $(M = 0.078, d = 150)$ gives the lowest maintenance cost, 1640.74. For the GP-RE model, the optimized maintenance policy is given by $(M = 0.076, d = 150)$ with the lowest maintenance cost of 1628.08.

A general form of the control limit (Eq. (25)), that is, time-varying, is further studied (M(t)). For A different from 0, the searching range of A is from -0.05 to 0.1, with an increment step value of 10^{-4}, excluding the value $A = 0$ from the study. b is from 0.02 to 0.09 $\Omega.cm^2$ with a value increment step of 0.002. It is noted that the slope satisfies $A \neq 0$ since the constant type control limit is optimized independently. Based on the results with M constant, the constant inspection interval is fixed as $d = 150$ h. Fig. 4 presents the contour plot of the obtained maintenance costs under GP and GP-RE models. The minimum maintenance cost for the GP model is 1644.31, with 218.19 inspections, 18.61 PMs, and 0.58 CMs (on average). The corresponding maintenance policy is given by $(A = 0.0008, b = 0.076)$. For the GP-RE model with deterioration variability, the optimized maintenance cost is 1618.77, with 217.84 inspections, 18.28 PMs, and 0.56 CMs. Compared to the results with M constant, the time-varying control limit policy decreases maintenance costs when deterioration variability exists.

Results for the reinforcement learning-based maintenance policy (RL-M(t))

As a deep RL approach, the hyperparameters defined in the PPO model need to be tuned. The relatively larger search space and the high computation cost of the training and evaluation processes highlight the need for an effective hyperparameter optimization (HPO) method. We applied the Tree-structured Parzen Estimator algorithm for efficiently sampling hyperparameters and the Median pruning algorithm for pruning unpromising trails [54, 55, 56].

Table 2 collects the hyperparameters to be optimized for periodic maintenance. Note that '[]' in the table represents the hidden layers and the number of neurons in each layer. For instance, '[140, 64]' is a neural network with two hidden layers, 140 neurons for the first layer and 64 neurons for the second layer. The HPO results are summarized in Fig. 5. From Fig. 5 a), c), we see that the objective values, which are taken as negative of episode return, converge quickly during the optimization process. The importance weights for the objective value results show that

Table 2: Hyperparameters considered for PPO agent under periodic maintenance

Hyperparameter	Meaning	Searching range
k_1	Reward tuning coefficients (Eq. (19))	(1.0, 4.0), with increment step of 0.1
ϵ	The clip ratio used in the PPO objective	0.1, 0.15, 0.2
$N_{hid,act}$	Number of hidden layers in the actor neural network	[32], [64], [100], [128], [256], [128, 32], [128, 64], [256, 32], [256, 64]
$N_{hid,crit}$	Number of hidden layers in the critic neural network	[32], [64], [100], [128], [256], [128, 32], [128, 64], [256, 32], [256, 64]
lr_{act}	Learning rate for the actor neural network	$(10^{-6}, 10^{-3})$
lr_{crit}	Learning rate for the critic neural network	$(10^{-6}, 10^{-3})$
N_{epo}	Number of training epochs for one episode data	10, 15, 20, 25
γ	Discount factor for future rewards	0.9, 0.95, 0.99

Table 3: Optimized hyperparameter values for periodic maintenance

	k_1	lr_{act}	$N_{hid,act}$	γ	lr_{crit}	$N_{hid,crit}$	N_{epo}	ϵ
Values of GP	2.6	7.47×10^{-5}	[128]	0.95	3.23×10^{-4}	[256, 64]	25	0.2
Values of GP-RE	1.7	2.17×10^{-4}	[100]	0.9	3.92×10^{-4}	[128, 64]	15	0.2

lr_{act}, k_1, and $N_{hid,act}$ are the three most important hyperparameters under the GP model. The lr_{act} receives the highest importance weight of 0.84. In the GP-RE model, k_1 is the most important hyperparameter, followed by lr_{act} and γ. The final hyperparameters for the GP and GP-RE models are summarized in Table 3. The high-importance weight in the reward coefficient k_1 justifies the importance of reward design (Section 4.2). Fig. 6 shows the obtained training curves using the optimal hyperparameters. The PPO agents converge quickly for both models.

The evaluation results show that the average maintenance cost for the GP model is 1641.42, with 218.16 inspections, 18.58 PMs, and 0.58 CMs. This result is closer to the average maintenance cost in the constant M, slightly better than the linear function-based empirical strategy. For the GP-RE model with random effects, the proposed maintenance by PPO further reduces the average maintenance cost to 1615.83 (with 217.71 inspections, 18.11 PMs, and 0.54 CMs) compared to the linear function-based maintenance policy.

The obtained maintenance policies are depicted in Fig. 7, including the linear function-based policies. The optimal maintenance policies under the GP-RE model show a growing trend with stack age. While the optimal control limits are nearly constant under the GP model. The existence of heterogeneity tends to vary the deterioration rate of different stacks. As a result, a time-varying control limit is needed to replace fast-deteriorating stacks early to prevent high failure costs and replace the slow-deteriorating stacks late to avoid watering RUL. This explains why the maintenance cost is decreased by developing a time-varying control limit. However, the average deterioration rate remains constant for the GP model, thus there is no need to assign a time-varying control limit. The constant control limit is the optimal one for the GP model [38]. That's why the results of RL and linear function-based policies are both slightly worse than that of the constant M.

Sensitivity analysis of random effect parameter

This part investigates how optimal maintenance policy varies in random effect parameters. The results are obtained from the linear function-based strategy M(t). Recall that $\delta = \delta_0 = 43.8, \phi = \phi_0 = 10^{-4}$ in the original GP-RE model (Eq. (7)). The expected mean increment and variance of a Gamma process $Ga(\delta, \phi)$ are respectively, $\delta\phi$ and $\delta\phi^2$. Thus, the expected deterioration rates remain identical in the two comparison groups. $Ga(\delta_0/5, \phi \cdot 5)$ exposes a relatively high deterioration heterogeneity, and $Ga(\delta_0 \cdot 5, \phi/5)$ exposes a relatively smaller deterioration heterogeneity compared with the original GP-RE model.

Fig. 8 shows the obtained optimal control limit curves. The control limit tends to be steeper when the deterioration heterogeneity is higher. The existence of deterioration variability caused varying deterioration rates among stacks. Thus, a relatively lower control limit is assigned to screen out stacks with high deterioration rates at an early age.

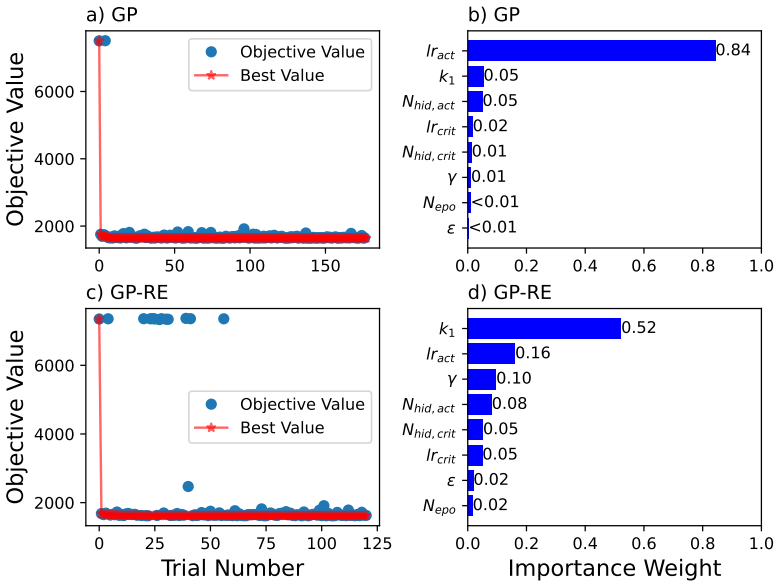

Figure 5: Hperparameter optimization results under periodic maintenance. a) Optimization history for GP model, b) Importance weight for objective value in GP model, c) Optimization history for GP-RE model, d) Importance weight for objective value in GP-RE model.

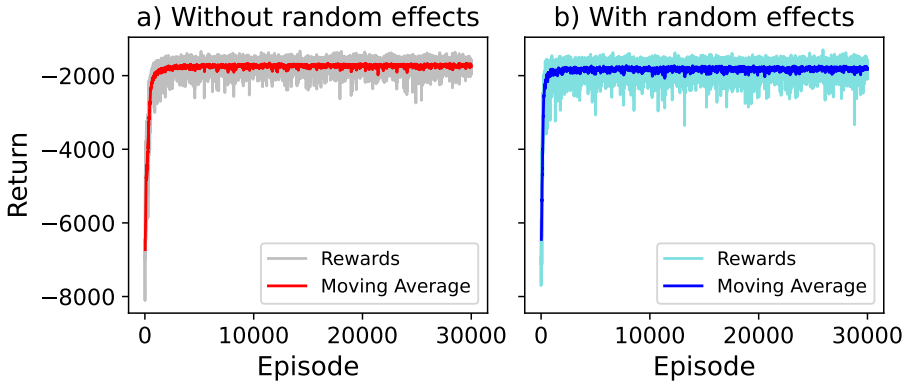

Figure 6: Return of PPO training under periodic maintenance.

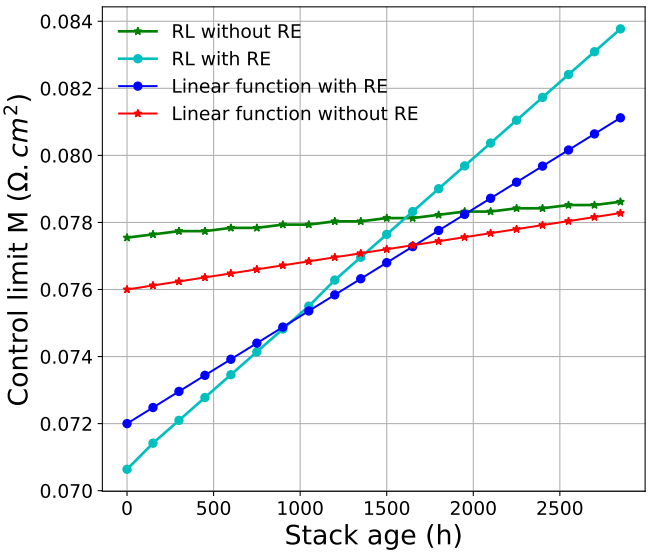

Figure 7: Optimized time-varying maintenance control limit shape.

On the other hand, a relatively larger control limit is decided for those with lower deterioration rates at a later age to avoid wasting stacks' remaining useful lifetime. We are more certain about the true deterioration rates when the deterioration heterogeneity is low. Thus, a less steep control limit is assigned.

5.4. Results for non-periodic inspection-based maintenance

A non-periodic inspection-based maintenance policy is expected to decrease inspection costs further by assigning time-varying inspections. The linear function-based maintenance policy discussed in this section refers to a time-varying inspection, namely $d(t)$. The control limit M is a predefined constant, i.e., $M = 0.08\ \Omega.cm^2$ which allows investigating the gain by optimizing the inspection decisions. According to the results obtained in Section 5.3, the optimal control limits (constant type) for the GP and GP-RE models are 0.078 and 0.076 $\Omega.cm^2$, respectively. Thereby, we select a constant control limit M that is not optimal for both GP and GP-RE models to justify better the objective of this section, namely to investigate the gain in overall maintenance costs by further optimizing the inspection policies.

Results for the linear function-based maintenance policy (d(t))

Linear function-based inspection policy requires two parameters a, c as defined in Eq. (26). The number of the Monte Carlo simulation histories is set as 10^4. The searching space of a ranges from -10^4 to -10^3 with an increment step of 10, and c is from 300 to 1000 with an increment step of 50. The optimized maintenance policies are summarized in Fig. 9. The optimal maintenance costs are 1505.14 for the GP model and 1485.32 for the GP-RE model.

Results for the reinforcement learning-based maintenance policy (RL-d(t))

Table 4 summarizes the hyperparameters of PPO under non-periodic maintenance. The obtained HPO results are demonstrated in Fig. 10. The optimization histories are shown in Figs. 10 a), c). The hyperparameters are successfully optimized for GP and GP-RE models. Based on the results presented in Figs. 10 b), d), it can be concluded that both the GP and GP-RE models identify K_3 and lr_{act} as the two most significant hyperparameters. Nevertheless, the GP-RE model assigns a relatively higher weight to the remaining parameters such as $N_{hid,crit}$ and lr_{crit} compared to the GP model. This suggests that the HPO process is more demanding under the GP-RE model. According to Eq. (23), k_3 characterizes the extra reward term for taking 'effective' decisions, that is, encouraging a relatively large inspection step for a less deteriorated fuel cell stack. The HPO results verify this design which is crucial for deciding effective inspections.

The hyperparameter decisions for the GP and GP-RE models are shown in Table 5. It is noticed that the value of k_3 in the GP-RE model is smaller than that in the GP model. This indicates that when random effects exist, a more conservative inspection is preferred for minimizing the overall maintenance costs.

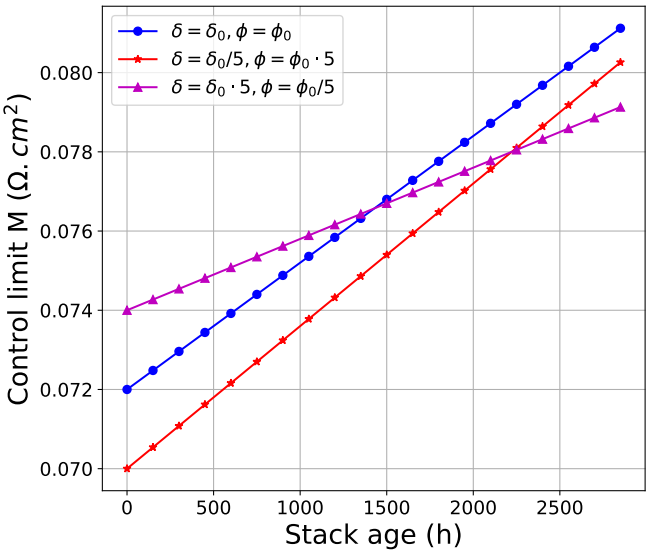

Figure 8: Sensitivity of control limits to different random effects parameters.

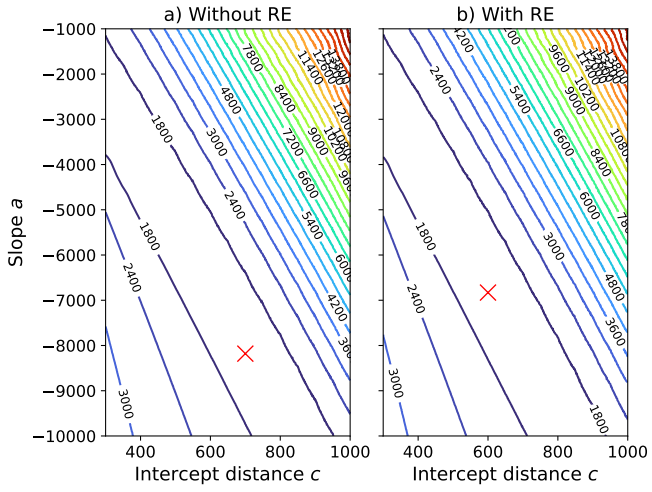

Figure 9: Contour plot of the overall maintenance costs under non-periodic inspection-time-varying control limit $d(R) = a \cdot R + c$: the optimized maintenance policy is $(a = -8180,\ c = 700)$ for GP model and $(a = -6830, c = 600)$ for GP-RE model. The red cross markers denote the data points with minimum maintenance costs.

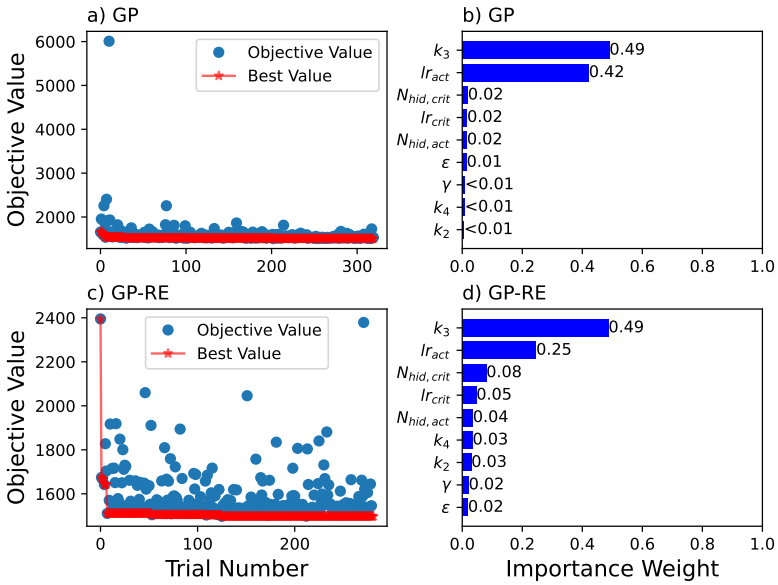

Figure 10: Hperparameter optimization results under non-periodic maintenance. a) Optimization history for GP model, b) Importance weight for objective value in GP model, c) Optimization history for GP-RE model, d) Importance weight for objective value in GP-RE model.

Fig. 11 illustrates the inspection policies obtained by the PPO agent and the linear function-based strategy. Both the comparison strategy and the RL-based strategy demonstrate that the inspection policy of the GP-RE model is lower than that of the GP model below a certain deterioration level. However, as the deterioration level approaches FT, the inspection decisions become similar under both the GP and the GP-RE models. These results validate the impact of random effects on inspection decisions, highlighting the necessity of adopting more conservative inspections in the presence of such effects. The average maintenance cost obtained for the GP model is 1506.26, and 1497.63 for the GP-RE model. It is noted that the obtained maintenance cost is slightly larger than that of linear function-based inspections. This may be caused by the fact that the true optimal inspections under the studied problem are nearly linear, thus the empirical liner function-based policy can better approach the optimal policy. Compared to the baseline results in periodic inspection which are 1640.74 and 1628.08 for the GP and GP-RE models, the proposed RL policy further reduces the maintenance by 8.20% and 8.0% for the GP and the GP-RE model, respectively. Fig. 12 summarizes the average maintenance cost results for all investigated policies.

Finally, example deterioration paths under the baseline strategy with periodic maintenance and the non-periodic maintenance by PPO agent are shown in Fig. 13. The number of inspections was reduced in Fig. 13b) by using the PPO agent without causing additional stack failure.

Sensitivity analysis of random effect parameter

The optimal inspections are presented in Fig. 14. When fuel cell stacks exhibit a larger heterogeneity, the maintenance policy schedules smaller inspections to prevent the populations with high deterioration rates from causing extra failure costs. By contrast, for stacks with smaller heterogeneity, their deterioration rates exhibit smaller variances. Therefore, relatively larger inspection intervals are assigned to reduce inspection costs. Nevertheless, in all scenarios, the inspection intervals tend to converge to a similar value as the stacks become more deteriorated. On average, the surviving populations of more deteriorated stacks show lower deterioration rates. Consequently, there is no need for distinct inspection assignments among these stacks.

5.5. Results with increased heterogeneity

This section investigates the performance of the proposed PPO policy under increased deterioration heterogeneity ($\delta = \delta_0/10, \phi = \phi \cdot 10$).

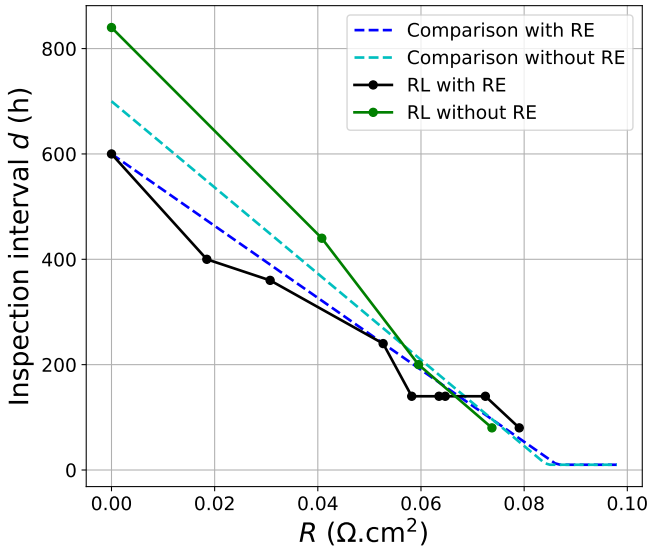

Figure 11: Non-periodic inspection policies for the comparison strategy and the RL-based strategy (sampled from the trained PPO agent).

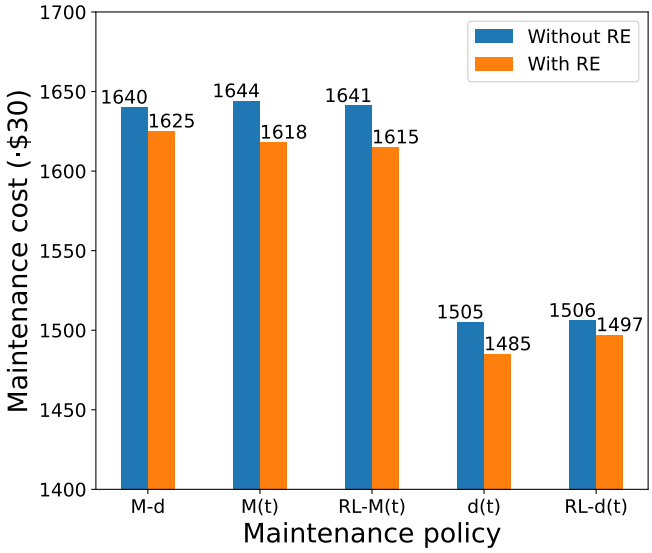

Figure 12: Average maintenance costs optimized by the investigated maintenance policies.

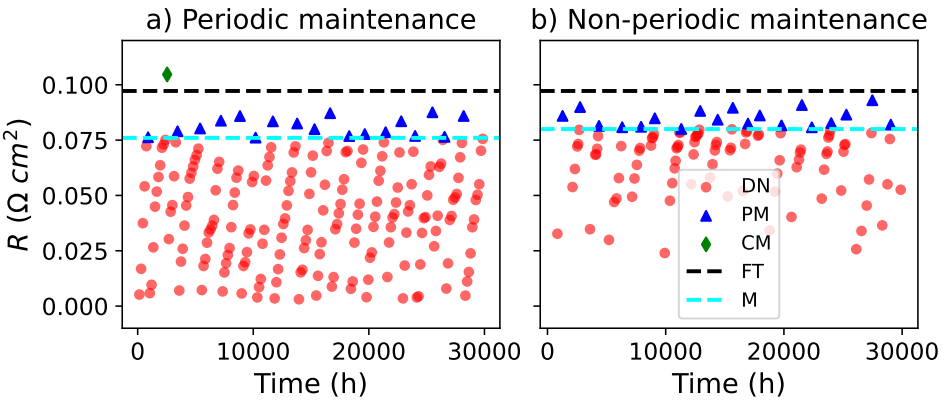

Figure 13: Deterioration trajectory comparison for GP-RE model. a) For baseline strategy; b) for PPO agent.

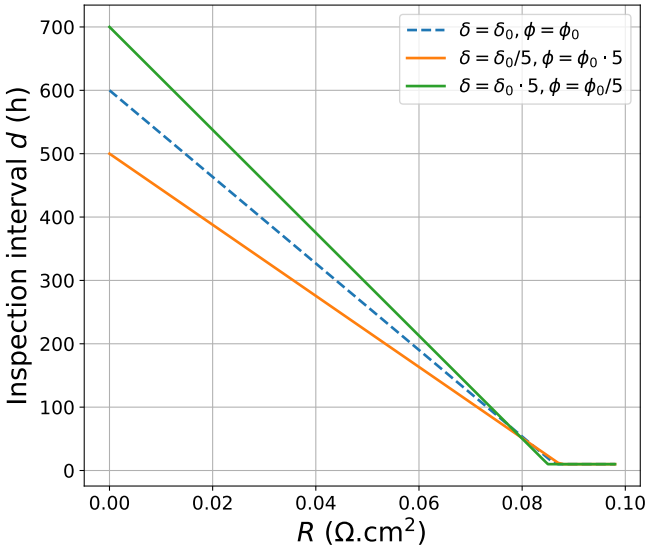

Figure 14: Sensitivity of inspection intervals to different random effects parameters.

Table 4: Hyperparameters considered for PPO agent under non-periodic maintenance

Hyperparameter	Meaning	Searching range
k_2		(100, 1000), with increment step of 1
k_3	Reward tuning coefficients (Eq. (23))	(50, 700), with increment step of 0.25
k_4		(100, 1000), with increment step of 1
ϵ	The clip ratio used in the PPO objective	0.1, 0.15, 0.2
$N_{hid,act}$	Number of hidden layers in the actor neural network	[64], [128], [140], [248], [140, 64], [248, 64], [248, 128], [248, 128, 64]
$N_{hid,crit}$	Number of hidden layers in the critic neural network	[64], [128], [140], [248], [140, 64], [248, 64], [248, 128]
lr_{act}	Learning rate for the actor neural network	$(10^{-6}, 10^{-3})$
lr_{crit}	Learning rate for the critic neural network	$(10^{-6}, 10^{-3})$
γ	Discount factor for future rewards	0.9, 0.95, 0.99

Table 5: Optimized hyperparameter values for non-periodic maintenance

	k_3	lr_{act}	$N_{hid,crit}$	lr_{crit}	$N_{hid,act}$	k_4	k_2	γ	ϵ
Values of GP	170	5.87×10^{-4}	[32]	2.99×10^{-5}	[248, 64]	965	886	0.95	0.1
Values of GP-RE	125.25	2.48×10^{-4}	[128]	3.24×10^{-4}	[248, 64]	891	952	0.95	0.15

Results for periodic inspection

In the periodic inspection, the average maintenance cost of the M-d policy is 1401.85. The optimal policy is given by $M = 0.076\ \Omega.cm^2$, $d = 150$ h. The minimal average maintenance is further reduced to 1365.60 by applying the linear function-based control limits (Eq. (25)). The optimal maintenance policy is found at $A = 0.0022, b = 0.072$.

Table 6 lists the optimal hyperparameter values of the PPO agent under periodic inspection. The average maintenance cost evaluated with the PPO agent is 1315.51. The proposed PPO algorithm reduces the overall maintenance costs by 6.16% compared to the constant policy, and by 3.67% compared to the linear function-based policy.

Fig. 15 presents the obtained control limits under the three maintenance policies. Different from the policy discussed in Section 5.3, an obvious nonlinear trend is observed in the control limit obtained by the PPO agent. The investigated fuel cell stack populations exist higher deterioration variability which requires screening out faster-deteriorating stacks at an early age. This is further demonstrated in the deterioration trajectories (Fig. 16). We select two representative paths, namely path 1 for a rapidly deteriorating stack and path 2 for a slowly deteriorating stack, to compare the distinct behaviors of the obtained policies. The RL-based policy tends to replace path 1 at an early age while replacing the slowly deteriorating stack (path 2) late. Moreover, the average CMs for the RL-based policy is 0.20, much lower than that of the linear function policy, that is, 0.55. This verifies the efficiency of the proposed RL-based strategy.

Results for non-periodic inspection

The optimized maintenance cost for the linear function-based policy (d(t)) is 1287.42. The HPO results for the PPO agent are listed in Table 7. The optimal maintenance cost is 1260.0, which is lower than that of the linear function-based policy. Fig. 17 shows the optimized inspection policies. It is seen that the policy obtained by PPO tends to be more conservative compared to the linear function-based policy.

Table 6: Optimized hyperparameter values for periodic maintenance (increased heterogeneity)

	k_1	lr_{act}	$N_{hid,act}$	γ	lr_{crit}	$N_{hid,crit}$	N_{epo}	ϵ
Values	2.1	1.7×10^{-4}	[128, 32]	0.9	5.33×10^{-5}	[256, 64]	15	0.15

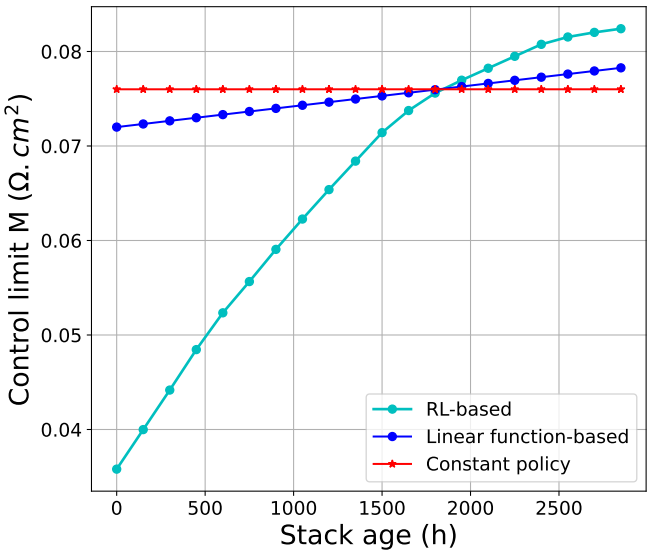

Figure 15: Control limits under increased heterogeneity.

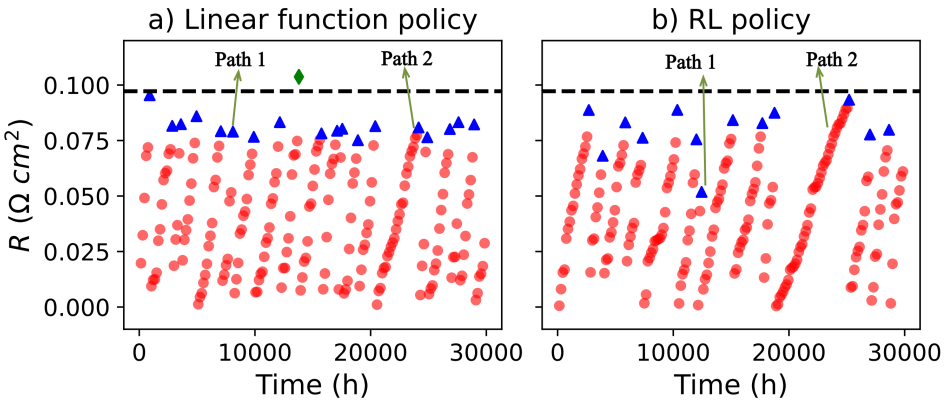

Figure 16: Deterioration trajectory comparison for GP-RE model with increased heterogeneity: red round dots are inspections, blue triangle dots are PMs, and green diamond dots are CMs; one path corresponds to the recorded deterioration of a stack from its initial state until the maintenance replacement was performed.

Table 7: Optimized hyperparameter values for non-periodic maintenance (increased heterogeneity)

	k_3	lr_{act}	$N_{hid,crit}$	lr_{crit}	$N_{hid,act}$	k_4	k_2	γ	ϵ
Values	403.5	7.85×10^{-4}	[140, 64]	7.79×10^{-6}	[140, 64]	157	870	0.9	0.1

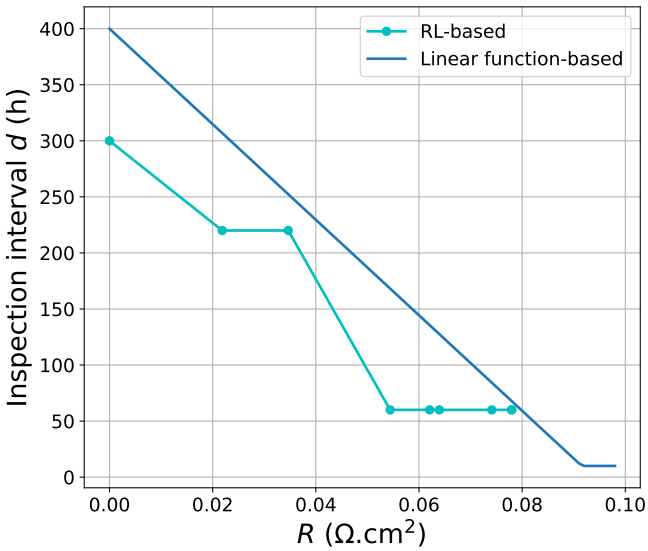

Figure 17: Inspections under increased heterogeneity.

5.6. Remarks on the obtained results

- Advantages of the proposed condition-based maintenance scheduling framework.

Firstly, a condition-based maintenance scheduling problem is formulated, for the first time, in the field of fuel cells. This problem is helpful for further investigating maintenance scheduling for fuel cell systems in practice to decrease operational costs. Secondly, the developed RL-based maintenance scheduling policy proves its flexibility and optimality through comparative case studies. Under the GP and GP-RE degradation models, the RL-based methods achieve better or comparable performance than the linear function-based strategy, which verifies its optimality. Moreover, the RL-based method is more flexible than existing approaches such as renewal theory which can be extended to new problem settings such as multi-stack fuel cell systems.

- Advantages and disadvantages of periodic and non-periodic inspection planning.

Through the comparison of the obtained maintenance cost results under periodic and non-periodic inspection, the advantage of non-periodic inspection in maintenance cost reduction is verified. As shown in Fig. 12, both the linear function-based and RL-based strategies achieve lower maintenance costs under non-periodic inspection. However, periodic inspection-based maintenance is easy to implement in the sense that fewer optimization parameters are required.

6. Conclusion

This paper investigated a novel condition-based maintenance scheduling problem for a single fuel cell stack. The Gamma process models the stochastic deterioration of the studied stack that operates under a constant power load. The stack-to-stack deterioration variability is modeled by a random effect-based GP model. Both periodic and non-periodic maintenance policies are investigated. An intelligent RL-based method is proposed to solve the maintenance problems. The main discoveries include:

1) A time-varying control limit is required when random effects exist.
2) More conservative inspections are preferred when random effects exist.
3) Reward design and hyperparameter tuning play an important role in training/deploying deep reinforcement learning algorithms such as PPO.

4) The RL-based method can handle more complicated maintenance optimization such as fuel cell stacks with high heterogeneity, providing improved maintenance policy.

The proposed GP-based fuel cell degradation model considers realistic factors such as stochasticity. Investigating both periodic and non-periodic can be beneficial for deciding the optimal inspection according to practical conditions. More specifically, the proposed approach for the solution of the maintenance problem allows to select automatically the optimal tuning for the non-stationary control limit rule for maintenance decision-making.

The decision-making-related studies in fuel cell PHM is a challenging task that contains several aspects of decisions such as maintenance scheduling decisions and load allocation decisions. The decision-making under multi-stack systems and the related load allocation decisions have not been explored. Therefore, we conclude that a perspective is to investigate a joint maintenance and load allocation decision-making problem for a multi-stack fuel cell system. In [57], we have formulated and provided a basic solution for the joint maintenance scheduling and load allocation problem in multi-stack fuel cells. This progress verifies the value of our current work and supports the development of our proposed future work.

Acknowledgments

This work has been supported by Plan France 2030/PEPR Hydrogène décarboné/Project DuraSyS-PAC managed by the French Research Agency (Reference ANR-22-PEHY-0002), the EIPHI Graduate School (contract ANR-17-EURE-0002) and the Region Bourgogne Franche-Comté.

References

[1] S. Dirkes, J. Leidig, P. Fisch, S. Pischinger, Prescriptive lifetime management for PEM fuel cell systems in transportation applications, Part I: State of the art and conceptual design, Energy Conversion and Management 277 (2023) 116598.
[2] Y. Wang, D. F. R. Diaz, K. S. Chen, Z. Wang, X. C. Adroher, Materials, technological status, and fundamentals of PEM fuel cells–a review, Materials today 32 (2020) 178–203.
[3] X. Lin, X. Xu, H. Lin, Predictive-ECMS based degradation protective control strategy for a fuel cell hybrid electric vehicle considering uphill condition, eTransportation 12 (2022) 100168.
[4] Z. Hua, Z. Zheng, E. Pahon, M.-C. Péra, F. Gao, A review on lifetime prediction of proton exchange membrane fuel cells system, Journal of Power Sources 529 (2022) 231256.
[5] M. Bahrami, J.-P. Martin, G. Maranzana, S. Pierfederici, M. Weber, S. Didierjean, Fuel cell management system: An approach to increase its durability, Applied Energy 306 (2022) 118070.
[6] X. Zhao, L. Wang, Y. Zhou, B. Pan, R. Wang, L. Wang, X. Yan, Energy management strategies for fuel cell hybrid electric vehicles: Classification, comparison, and outlook, Energy Conversion and Management 270 (2022) 116179.
[7] P. Ren, P. Pei, Y. Li, Z. Wu, D. Chen, S. Huang, Degradation mechanisms of proton exchange membrane fuel cell under typical automotive operating conditions, Progress in Energy and Combustion Science 80 (2020) 100859.
[8] R. Gouriveau, M. Hilairet, D. Hissel, S. Jemei, M. Jouin, E. Lechartier, S. Morando, E. Pahon, M. Pera, N. Zerhouni, IEEE PHM 2014 data challenge: Outline, experiments, scoring of results, winners, IEEE 2014 PHM Challenge, Tech. Rep (2014).
[9] J. Zuo, H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, C. Zhang, Long-term dynamic durability test datasets for single proton exchange membrane fuel cell, Data in Brief 35 (2021) 106775.
[10] Z. Gong, B. Wang, Y. Xing, Y. Xu, Z. Qin, Y. Chen, F. Zhang, F. Gao, B. Li, Y. Yin, et al., High-precision and efficiency diagnosis for polymer electrolyte membrane fuel cell based on physical mechanism and deep learning, eTransportation (2023) 100275.
[11] Y. Ao, S. Laghrouche, D. Depernet, Diagnosis of proton exchange membrane fuel cell system based on adaptive neural fuzzy inference system and electrochemical impedance spectroscopy, Energy Conversion and Management 256 (2022) 115391.
[12] R. Ma, T. Yang, E. Breaz, Z. Li, P. Briois, F. Gao, Data-driven proton exchange membrane fuel cell degradation predication through deep learning method, Applied energy 231 (2018) 102–115.
[13] J. Zuo, H. Lv, D. Zhou, Q. Xue, L. Jin, W. Zhou, D. Yang, C. Zhang, Deep learning based prognostic framework towards proton exchange membrane fuel cell for automotive application, Applied Energy 281 (2021) 115937.
[14] C. Wang, M. Dou, Z. Li, R. Outbib, D. Zhao, J. Zuo, Y. Wang, B. Liang, P. Wang, Data-driven prognostics based on time-frequency analysis and symbolic recurrent neural network for fuel cells under dynamic load, Reliability Engineering & System Safety 233 (2023) 109123.
[15] Z. Hua, Z. Zheng, E. Pahon, M.-C. Péra, F. Gao, Remaining useful life prediction of PEMFC systems under dynamic operating conditions, Energy Conversion and Management 231 (2021) 113825.
[16] C. Wang, M. Dou, Z. Li, R. Outbib, D. Zhao, B. Liang, A fusion prognostics strategy for fuel cells operating under dynamic conditions, eTransportation 12 (2022) 100166.
[17] W. Zhu, B. Guo, Y. Li, Y. Yang, C. Xie, J. Jin, H. B. Gooi, Uncertainty quantification of proton-exchange-membrane fuel cells degradation prediction based on Bayesian-Gated Recurrent Unit, eTransportation 16 (2023) 100230.
[18] M. Yue, S. Jemei, N. Zerhouni, R. Gouriveau, Proton exchange membrane fuel cell system prognostics and decision-making: Current status and perspectives, Renewable Energy 179 (2021) 2277–2294.

[19] H. Peng, Y. Chen, Z. Chen, J. Li, K. Deng, A. Thul, L. Löwenstein, K. Hameyer, Co-optimization of total running time, timetables, driving strategies and energy management strategies for fuel cell hybrid trains, eTransportation 9 (2021) 100130.
[20] W. Zhang, J. Li, L. Xu, M. Ouyang, Optimization for a fuel cell/battery/capacity tram with equivalent consumption minimization strategy, Energy Conversion and Management 134 (2017) 59–69.
[21] Y. Yan, Q. Li, W. Chen, W. Huang, J. Liu, J. Liu, Online control and power coordination method for multistack fuel cells system based on optimal power allocation, IEEE Transactions on Industrial Electronics 68 (2020) 8158–8168.
[22] J. Zhao, X. Li, C. Shum, J. McPhee, Control-oriented computational fuel cell dynamics modeling–model order reduction vs. computational speed, Energy 266 (2023) 126488.
[23] P. Pei, Q. Chang, T. Tang, A quick evaluating method for automotive fuel cell lifetime, International Journal of Hydrogen Energy 33 (2008) 3829–3836.
[24] H. Chen, P. Pei, M. Song, Lifetime prediction and the economic lifetime of proton exchange membrane fuel cells, Applied Energy 142 (2015) 154–163.
[25] J. Zuo, C. Cadet, Z. Li, C. Bérenguer, R. Outbib, Fuel cell stochastic deterioration modeling for energy management in a multi-stack system, in: 2022 13th International Conference on Reliability, Maintainability, and Safety (ICRMS), IEEE, 2022, pp. 104–108.
[26] Y. Chatillon, C. Bonnet, F. Lapicque, Heterogeneous aging within PEMFC stacks, Fuel Cells 14 (2014) 581–589.
[27] A. Macias, M. Kandidayeni, L. Boulon, H. Chaoui, A novel online energy management strategy for multi fuel cell systems, in: 2018 IEEE International Conference on Industrial Technology (ICIT), IEEE, 2018, pp. 2043–2048.
[28] M. Rausand, A. Hoyland, System reliability theory: models, statistical methods, and applications, volume 396, John Wiley & Sons, 2003.
[29] L. Pinciroli, P. Baraldi, E. Zio, Maintenance optimization in Industry 4.0, Reliability Engineering & System Safety (2023) 109204.
[30] Z. Ren, A. S. Verma, Y. Li, J. J. Teuwen, Z. Jiang, Offshore wind turbine operations and maintenance: A state-of-the-art review, Renewable and Sustainable Energy Reviews 144 (2021) 110886.
[31] C. Zhang, W. Gao, T. Yang, S. Guo, Opportunistic maintenance strategy for wind turbines considering weather conditions and spare parts inventory management, Renewable Energy 133 (2019) 703–711.
[32] N. Zhang, F. Qi, C. Zhang, H. Zhou, Joint optimization of condition-based maintenance policy and buffer capacity for a two-unit series system, Reliability Engineering & System Safety 219 (2022) 108232.
[33] J. M. van Noortwijk, A survey of the application of Gamma processes in maintenance, Reliability Engineering & System Safety 94 (2009) 2–21.
[34] Z. Chen, T. Xia, Y. Li, E. Pan, A hybrid prognostic method based on gated recurrent unit network and an adaptive Wiener process model considering measurement errors, Mechanical Systems and Signal Processing 158 (2021) 107785.
[35] W. Wang, M. Lin, Y. Fu, X. Luo, H. Chen, Multi-objective optimization of reliability-redundancy allocation problem for multi-type production systems considering redundancy strategies, Reliability Engineering & System Safety 193 (2020) 106681.
[36] E. M. Omshi, A. Grall, Replacement and imperfect repair of deteriorating system: Study of a CBM policy and impact of repair efficiency, Reliability Engineering & System Safety 215 (2021) 107905.
[37] I. T. Castro, R. J. Basten, G.-J. Van Houtum, Maintenance cost evaluation for heterogeneous complex systems under continuous monitoring, Reliability Engineering & System Safety 200 (2020) 106745.
[38] L. Zhang, Y. Lei, H. Shen, How heterogeneity influences condition-based maintenance for Gamma degradation process, International Journal of Production Research 54 (2016) 5829–5841.
[39] N. Zhang, W. Si, Deep reinforcement learning for condition-based maintenance planning of multi-component systems under dependent competing risks, Reliability Engineering & System Safety 203 (2020) 107094.
[40] L. Pinciroli, P. Baraldi, G. Ballabio, M. Compare, E. Zio, Optimization of the operation and maintenance of renewable energy systems by deep reinforcement learning, Renewable Energy 183 (2022) 752–763.
[41] J. Cheng, Y. Liu, W. Li, T. Li, Deep reinforcement learning for cost-optimal condition-based maintenance policy of offshore wind turbine components, Ocean Engineering 283 (2023) 115062.
[42] C. P. Lin, M. H. Ling, J. Cabrera, F. Yang, D. Y. W. Yu, K. L. Tsui, Prognostics for lithium-ion batteries using a two-phase Gamma degradation process model, Reliability Engineering & System Safety 214 (2021) 107797.
[43] J. Zuo, C. Cadet, Z. Li, C. Berenguer, R. Outbib, Post-prognostics decision-making strategy for load allocation on a stochastically deteriorating multi-stack fuel cell system, Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability 237 (2023) 40–57.
[44] J. Zuo, C. Cadet, Z. Li, C. Bérenguer, R. Outbib, A deterioration-aware energy management strategy for the lifetime improvement of a multi-stack fuel cell system subject to a random dynamic load, Reliability Engineering & System Safety 241 (2024) 109660.
[45] J. Lawless, M. Crowder, Covariates and random effects in a Gamma process model with application to degradation and failure, Lifetime data analysis 10 (2004) 213–227.
[46] C. Guo, Z. Liang, A predictive markov decision process for optimizing inspection and maintenance strategies of partially observable multi-state systems, Reliability Engineering & System Safety 226 (2022) 108683.
[47] Y. Shen, M. Fouladirad, A. Grall, Impact of dust and temperature on photovoltaic panel performance: A model-based approach to determine optimal cleaning frequency, Heliyon 10 (2024).
[48] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal policy optimization algorithms, arXiv preprint arXiv:1707.06347 (2017).
[49] L. Engstrom, A. Ilyas, S. Santurkar, D. Tsipras, F. Janoos, L. Rudolph, A. Madry, Implementation matters in deep policy gradients: A case study on ppo and trpo, arXiv preprint arXiv:2005.12729 (2020).
[50] A. Al-Hilo, M. Samir, C. Assi, S. Sharafeddine, D. Ebrahimi, Uav-assisted content delivery in intelligent transportation systems-joint trajectory planning and cache management, IEEE Transactions on Intelligent Transportation Systems 22 (2020) 5155–5167.
[51] G. Wang, M. Miller, L. Fulton, Estimating maintenance and repair costs for battery electric and fuel cell heavy duty trucks (2022).
[52] M. Wei, T. Lipman, A. Mayyas, J. Chien, S. H. Chan, D. Gosselin, H. Breunig, M. Stadler, T. McKone, P. Beattie, et al., A total cost of ownership model for low temperature PEM fuel cells in combined heat and power and backup power applications, Technical Report,

Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 2014.
[53] https://www.fuelcellstore.com/fuel-cell-stacks/high-power-fuel-cell-stacks, 2023.
[54] T. Akiba, S. Sano, T. Yanase, T. Ohta, M. Koyama, Optuna: A next-generation hyperparameter optimization framework, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, 2019, pp. 2623–2631.
[55] S. Watanabe, Tree-structured parzen estimator: Understanding its algorithm components and their roles for better empirical performance, arXiv preprint arXiv:2304.11127 (2023).
[56] Y. He, P. Liu, Z. Wang, Z. Hu, Y. Yang, Filter pruning via geometric median for deep convolutional neural networks acceleration, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019, pp. 4340–4349.
[57] J. Zuo, N. Y. Steiner, Z. Li, C. Cadet, C. Bérenguer, D. Hissel, Optimal post-prognostics decision making for multi-stack fuel cells in transportation: toward joint load allocation and maintenance scheduling, IEEE Transactions on Transportation Electrification (2024).