Energy Conversion and Management 251 (2022) 115030

Contents lists available at ScienceDirect

Energy Conversion and Management

journal homepage: www.elsevier.com/locate/enconman

Deep reinforcement learning based energy management strategy of fuel cell hybrid railway vehicles considering fuel cell aging

Kai Deng [a,*], Yingxu Liu [b], Di Hai [a], Hujun Peng [a], Lars Löwenstein [c], Stefan Pischinger [b], Kay Hameyer [a]

[a] Institute of Electrical Machines (IEM), RWTH Aachen University, Germany
[b] Institute for Combustion Engines (VKA), RWTH Aachen University, Germany
[c] Siemens Mobility GmbH, Vienna, Austria

ARTICLE INFO

Keywords:
Energy management
Fuel cell hybrid railway vehicle
Deep Reinforcement learning (DRL)
TD3
Fuel cell aging

ABSTRACT

In the rail transportation industry, growing energy and environmental awareness requires the use of alternatives to combustion engines. These include hybrid electrically driven railway vehicles powered by fuel cells and batteries. The cost of hydrogen consumption and the lifetime of fuel cells are currently the main challenges that need to be addressed before widespread deployment of fuel cell railway vehicles can be realized. With this in mind, this work focuses on the energy management system with emphasis on optimizing the energy distribution to reduce the overall operational cost. The presented energy management strategy (EMS) aims at minimizing hydrogen consumption and fuel cell aging costs while achieving a favorable balance between battery charging and discharging. In order to take fuel cell aging into account in energy management and mitigate fuel cell aging trough power distribution, an online fuel cell aging estimation model based on four operation modes is introduced and applied. Moreover, the advanced deep reinforcement learning method Twin Delayed Deep Deterministic Policy Gradient (TD3) is used to obtain a promising EMS. To improve the adaptability of the strategy, a stochastic training environment, which is based on real measured speed profiles considering passenger numbers is used for training. Assuming different environmental and passenger transport volumes, the results confirm that the proposed TD3-EMS achieves battery charge-sustaining at low hydrogen consumption while slowing down fuel cell degradation.

1. Introduction

1.1. Background

In order to meet the world's massive energy needs while reducing greenhouse gas (GHG) and pollutant emissions, the search for efficient and clean energy sources and propulsion systems for vehicles is being pursued with increasing intensity worldwide. In this context, electric drives powered by fuel cells are considered a promising alternative to combustion engines using conventional fossil fuels, as they have higher efficiency and do not produce any harmful emissions. Fuel cells also have good market prospects, especially in the rail transport sector, where there is a high demand for environmentally friendly drives. Many governments have introduced specific targets and policies to promote the development of clean energy sources such as hydrogen. According to a corresponding roadmap, China will have more than 50,000 fuel cell vehicles and 300 hydrogen refueling stations in operation by 2025, and more than 1 million fuel cell vehicles, and 1000 hydrogen refueling stations will be in service by 2030 [1]. In the United States, hydrogen demand for various applications is expected to exceed 17 million tons by the end of 2030, with sales of 1.2 million fuel cell vehicles and 4,300 fueling stations in operation nationwide [2].

1.2. Literature review

One of the core issues in the development of fuel cell hybrid vehicles

This work is funded by the German Federal Ministry of Transport and Digital Infrastructure (BMVi) under the National Innovation Program Hydrogen and Fuel Cell Technology (NIP) with the funding numbers of 03B10502B and 03B10502B2. It is also funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) -GRK 1856. The authors gratefully acknowledge the support by Siemens AG, NIP and DFG.

* Corresponding author.
E-mail address: kai.deng@iem.rwth-aachen.de (K. Deng).

https://doi.org/10.1016/j.enconman.2021.115030
Received 28 September 2021; Received in revised form 9 November 2021; Accepted 12 November 2021
Available online 24 November 2021

(FCHEVs) is energy management. The goal of a superior energy management strategy (EMS) is to find the optimal distribution of the load power to achieve favorable overall performance in terms of fuel consumption and other related aspects, such as battery charge-sustaining, extended service life of the fuel cell, etc. For hybrid vehicles (HEVs) powered by multiple energy sources, different energy management strategies have been proposed and studied in the literature. They can be divided into two main categories, namely rule-based and optimization-based strategies [3].

The rule-based strategies are based on predetermined heuristic or empirical rules, which are widely used in practice due to their robustness and low computational cost. The early studies for the rule-based strategies follow the state machine method to determine the vehicle operation modes, and then rely on the expert knowledge to define the control logic aimed at improving the fuel economy [4]. Although the rule-based strategy can achieve good fuel economy, it does not guarantee achievement of the global optimum. To address the above problems, some parameter settings in rule-based strategies can be improved by optimization techniques [5]. However, the obtained strategies suffer from a lack of adaptability and the results may vary significantly in different driving environments.

Compared to the rule-based strategies, optimization-based strategies deal with the energy management as an optimization problem. They are typically divided into two types: offline- and online optimization-based strategies. In offline optimization-based strategies, the problem is considered for the whole driving cycle. Among them, Dynamic Programming (DP) can ensure the global optimum and is often used as a basis to evaluate the results of other strategies. Despite its ability to achieve optimal results, DP relies on a fixed driving cycle and its computational complexity limits its application. To realize the online implementation, in [6,7] DP is applied to determine the optimal results and the control characteristics are summarized to the control rules available online. Another offline optimization-based strategy is the Pontryagin minimum principle (PMP). It is applied to the dynamic system by converting it into a local optimization problem and searching for the instantaneous minimum of the Hamiltonian function. Studies in [8,9] have proved that the PMP strategy can obtain a near optimal solution with the prior knowledge of the driving cycle. Despite the low computational cost, the PMP method relies on the global driving information and the optimum is usually found by the offline shooting method. With the goal of providing real-time control, adaptive PMPs have been proposed in [10,11]. In these works, the adaptive estimation of the costate is achieved by the prediction method, and therefore they can be considered as the online optimization based method. Another typical online optimization based method is the model predictive control (MPC), which provides suboptimal control without global information as a prerequisite. For accurate prediction, an MPC-based EMS for HEVs with adaptive Markov chain prediction is presented in [12]. It shows robustness to variation in driving conditions and excellent fuel economy performance. In [13], an online adaptation mechanism of the PMP-derived cost function is introduced into the MPC structure, which shows promising fuel economy and battery charge-sustaining. However, the prediction accuracy and the computational cost are the limitations for the MPC-based strategies.

More recently, reinforcement learning (RL) has been a trendy research area for solving real-time optimization problems. The goal of RL is to maximize the preset reward function through interaction and exploration with a training environment. Its model-free property makes it flexible and adaptable to different systems and driving conditions. Q-learning is a type of RL that determines actions based on Q-values which represent the expected rewards of the actions. EMS derived from Q-learning has shown not only improved fuel economy, but also rapid convergence in [14,15]. These works also demonstrate the effectiveness and adaptability without any prior knowledge. To obtain superior results, the study in [16] discusses the effect of parameter discretization and training settings on Q-learning. Despite the advantages of RL, the problem of "curse of dimensionality" may arise due to the discretization of states and control variables, which limits its use. With the introduction of neural networks, these limitations can be addressed through Deep Reinforcement Learning (DRL). DRL has been implemented in [17,18] and shows excellent convergence and low fuel consumption, which also confirms its adaptability in different driving cycles. In [19], an EMS for series–parallel plug-in hybrid electric buses is implemented based on a novel DRL strategy called deep deterministic policy gradient (DDPG). Thereby, traffic information and passenger flow are considered, which makes the results more robust and generalizable. To improve the training efficiency of DDPG, expert knowledge is incorporated in [20], which results in a fast learning process with improved fuel economy that confirms its generalization. Similarly, a rule-based expert system is introduced to avoid the cold start of the training process in [21]. The work considers battery thermal safety and the results show doubled training efficiency and promising overall driving costs. In [22], soft actor-critic RL is exploited to determine the power distribution of a hybrid bus, which takes into account the thermal safety of the batteries. The results show a significant reduction in training time and total driving costs including fuel consumption and battery degradation.

In addition to the above energy management optimization aspects for general hybrid vehicles, the aging of fuel cells is a key point that must be considered for fuel cell hybrid vehicles. The fuel cell mainly consists of three components: electrode, electrolyte membrane and bipolar plate. Degradation of any of these components can affect the overall lifetime of the fuel cell [23]. In [24], the degradation modes of the different components and several methods for estimating the degradation of a fuel cell are summarized. Accordingly, the fuel cell service life estimation methods are roughly categorized into data-driven methods and model-based methods. Although the fuel cell state and its degradation can not be directly and explicitly determined by the EMS, an analysis of the degradation process for HEVs based on operating conditions proposed in [25] has shown that the operating conditions can affect the fuel cell degradation and its service life. In the model-based approach, different grades of models have varying degrees of accuracy and computational load. By using computational fluid dynamics (CFD) methods, the complex performance of fuel cells can be well studied from geometry [26,27] to degradation [28], especially in material and chemical perspectives, but with a large computational load. Since EMS requires an estimation with low computational cost, an operation-based estimation is often used in the energy management problems. In [29], a simplified electrochemical model that considers the decay of the electrochemical surface area is employed in the EMS. The results show that the EMS can extend the fuel cell service life and achieve low average operational cost. In [30], a stochastic dynamic programming based EMS is introduced. The design of the EMS takes into account the effects on the voltage degradation of the fuel cell, which results in a significant increase in lifetime with only a small increase in hydrogen consumption, and therefore reduces the overall cost.

1.3. Motivation

In many parts of the world, electrified rail transport with power supply via overhead lines is a widespread solution. However, due to economic advantages, diesel powered railway vehicles operating without overhead lines still dominate on many branch lines in less populated areas. As an environmentally friendly alternative to these diesel railway vehicles, fuel cell hybrid propulsion with zero emissions is of great interest. The electric motors of these vehicles are powered both by fuel cells and batteries, and it is necessary to properly manage the power distribution to achieve battery charge-sustaining and keep the operational costs low. To obtain an adaptive EMS, DRL is a highly prospective approach. To the authors' knowledge, so far there are no RL-based EMS studies for fuel cell railway vehicles in the literature.

In the case of railway vehicles, the number of passengers is very stochastic, which leads to uncertainties in the required power. In

addition, due to the restricted transient response of a fuel cell system, the battery operates as a buffer system to support the dynamic power demand while the fuel cell system covers the average power demand. Due to the large differences between the average and maximum power of a typical railway vehicle, it makes sense to use a large sized battery and a fuel cell with relatively low power. However, this causes the vehicle to have difficulties maintaining the battery's state of charge during long trips, which makes it difficult for the training to converge when using DRL. Moreover, fuel cell degradation is a major obstacle to commercial deployment. Therefore, its loss should be factored into the overall operational costs to optimize the design in advance.

1.4. Contributions

To address the above issues, a DRL-based EMS that accounts for fuel cell aging is proposed in this paper. The main contributions are listed below:

- The Twin Delayed Deep Deterministic Policy Gradient (TD3) approach is used as one of the advanced DRL methods to develop an intelligent energy management strategy for fuel cell and battery railway vehicles.
- An online evaluation model for fuel cell voltage degradation considering four operation modes is introduced to predict fuel cell aging.
- A newly designed reward function involving battery charge-sustaining is developed to stabilize the training process.
- A stochastic training environment considering passenger flows for railway vehicles based on the measured driving data is used to simulate real driving conditions.

1.5. Organization

Section 2 presents the modeling of the vehicle and hybrid power system, including the fuel cell aging. In Section 3, the DRL-based energy management problem is formulated, and the framework of the TD3-EMS is presented. Section 4 presents the training and simulation results, and analysis and discussion are performed. Finally, Section 5 gives a summary of this work.

2. System modeling

2.1. Fuel cell hybrid railway vehicle modeling

The drivetrain of the railway vehicle is shown in Fig. 1. A well-tuned proportional–integral (PI) driver controller is used to control the vehicle speed according to the speed profiles.

As one of the power sources to satisfy the dynamic load demands, a battery system with a rated voltage V_{bat} of 850 V and a capacity Q_{bat} of 207 Ah is provided for the railway vehicle. As shown in Fig. 2, this battery system is modeled with three R-C elements to provide an accurate approximation of the dynamic performance. The main parameters related to the state of charge (SoC) of the battery are shown in Fig. 3.

A proton-exchange membrane (PEM) fuel cell system with an internal DC/DC converter covers the average power demand. It has a maximum power output of 200 kW. The hydrogen mass flow can be calculated by looking up the characteristic curve in Fig. 4.

To adress the energy management problem, the vehicle modeling mainly considers longitudinal vehicle dynamics and battery dynamics. Other powertrain components consisting of electric motors, power electronics, auxiliary systems and fuel cell systems are modeled based on look-up tables. The vehicle model was created in MATLAB/Simulink and validated through a Hardware-in-the-Loop (HiL) test. The associated data and modeling details can be found in [6]. Table 1 presents the main system equations. The related parameters are listed in Table 2 and the variables are summarized in Table 3.

In the fuel cell hybrid vehicle, the power distribution is realized by controlling the fuel cell power P_{fc} through the energy management system. By following a reference speed profile, the power demand P_{dem} is controlled by the driver model and can be obtained by summing up the traction power $P_{traction} = F_t \cdot v_{veh}$, the power losses along the drivetrain P_{loss} and the auxiliary system power P_{aux}. Based on the quasi-steady-state modeling, P_{loss} can be obtained from lookup tables and P_{aux} is assumed to be constant 55 kW. Since the source-side DCDC converter is used to regulate the voltage of the DC link, and the power output of the fuel cell system is controlled by the EMS, the battery system operates as a buffer energy source. Some constraints due to the battery and the fuel cell system are specified in Table 4.

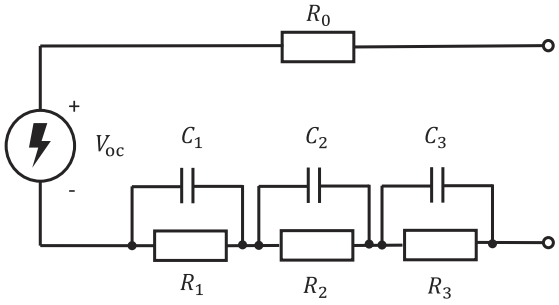

Fig. 2. Battery's equivalent circuit.

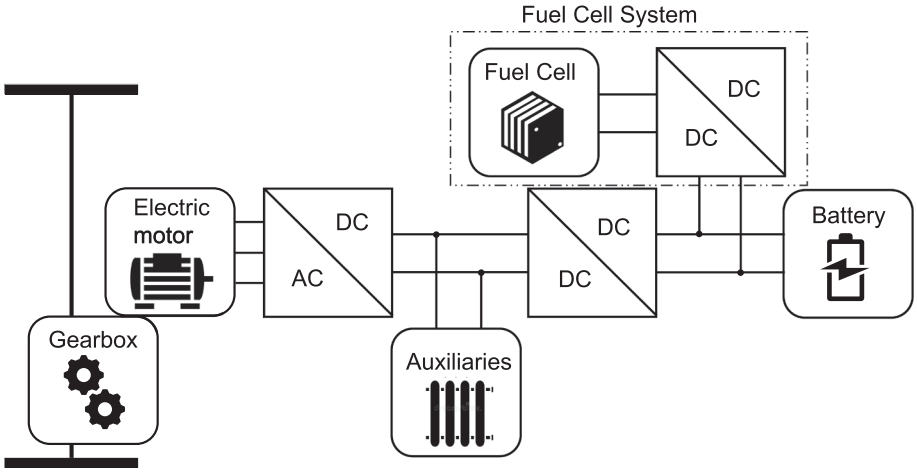

Fig. 1. Structure of the fuel cell and battery railway vehicle model.

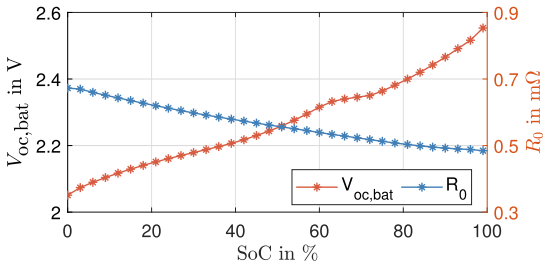

Fig. 3. Parameters of the battery's equivalent circuit at 25 °C: Open circuit voltage V_{oc} in V and internal resistance R_0 in Ω.

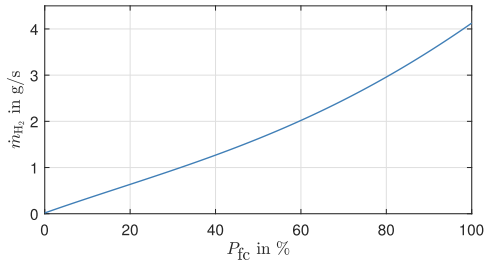

Fig. 4. Hydrogen consumption curve of the fuel cell system.

Table 1
Main system equations for the hybrid railway vehicle.

Description	System equations
Vehicle dynamics	$m\cdot\frac{dv_{veh}}{dt} = F_t - F_m - f_r mg\cos(\alpha) - \frac{1}{2}\rho_{air}C_{air}Av_{veh}^2 - mg\sin(\alpha)$
Total mass	$m = m_{veh} + n_p\cdot m_p$
Hydrogen mass flow	$\dot{m}_{H_2} = f(P_{fc})$
Battery dynamics	$\dot{SoC} = -\frac{I_{bat}}{Q_{bat}} = \frac{V_{oc} - \sum_{i=1}^{3} V_i}{2R_0 Q_{bat}} + \sqrt{\left(\frac{V_{oc} - \sum_{i=1}^{3} V_i}{2R_0 Q_{bat}}\right)^2 - \frac{P_{bat}}{R_0 Q_{bat}^2}}$
Polarization voltage	$\dot{V}_i = \frac{I_{bat}}{C_i} - \frac{V_i}{R_i\cdot C_i}, \quad i = 1,2,3$
Power demand	$P_{dem} = F_t\cdot v_{veh} + P_{loss} + P_{aux}$
Power distribution	$P_{bat} = P_{dem} - P_{fc}$

Table 2
Main parameters for the hybrid railway vehicle.

Symbol	Parameters	Values
m_v	Vehicle mass	51000 kg
m_p	Passenger mass	75 kg
g	Gravitational acceleration	9.81 m/s^2
f_r	Rolling resistance coefficient	0.0015
ρ_{air}	Air density	1.2 kg/m^3
C_{air}	Aerodynamic coefficient	0.15
A	Vehicle front area	10 m^2

2.2. Fuel cell aging simulation

The application of the DRL based approach to obtain a promising EMS requires a timely reward function for the evaluation of the fuel cell aging. From an energy management perspective, the output of the EMS, i.e., the desired fuel cell power, has an impact on the fuel cell aging [25]. Therefore, an operation mode based fuel cell aging evaluation is appropriate. In this work, a fuel cell degradation model considering four common operation modes is introduced. These four modes of operation include: open circuit operation (OCV), start-stop cycles (SSC), voltage cycles (VC) and calendric aging (CA) [31,32]. The aging of fuel cells is regarded as the sum of the degradations caused by the four operation modes:

$$\Delta U_{total} = \Delta U_{OCV} + \Delta U_{SSC} + \Delta U_{VC} + \Delta U_{CA}, \tag{1}$$

In the following introductions for each degradation, the inputs of the degradation model, which are the cell voltage degradation U_{cell} and the increment of the cell voltage ΔU_{cell}, are in V, while the intermediate values $\Delta U_{OCV}, \Delta U_{SSC}, \Delta U_{VC}, \Delta U_{CA}$ and the output ΔU_{total} are in μV.

When the fuel cell is operated at open circuit voltage, the degradation of the fuel cell is accelerated [32]. In the degradation model, when the voltage is higher than 0.8 V, the fuel cell is considered to work at open circuit voltage [32]. The voltage loss caused by open circuit operation ΔU_{OCV} can be calculated by:

$$\Delta U_{OCV} = C_1\cdot t_{OCV}, \tag{2}$$

where t_{OCV} is the time in hours under the OCV working condition.

The fuel cell system is automatically shut down when the system power is below 12 kW. Each time the fuel cell system shuts down, the cell voltage degradation ΔU_{SSC} is:

$$\Delta U_{SSC} = C_2\cdot n_{SSC}, \tag{3}$$

where n_{SSC} denotes the number of starts and stops of the fuel cell system.

When the fuel cell system is working at a constant voltage, the performance of the fuel cell continues to decrease [33]. ΔU_{CA} caused by the calendric aging is related to the cell voltage U_{cell}:

$$\Delta U_{CA} = (C_3 + C_4\cdot e^{C_5\cdot U_{cell}})\cdot t_{CA}, \tag{4}$$

where t_{CA} is the time in hours under the CA working condition.

During the operation of the fuel cell system, the voltage cycling can lead to voltage degradation [32]. It is assumed that the voltage degradation caused by the voltage cycling is only related to ΔU_{cell}. It is counted once, when the voltage remains constant for 5 s, or either changes from rising to falling or from falling to rising. Using n_{VC} as the number of counts, the formula is given:

Table 3
Main variables for the hybrid railway vehicle.

Variables	Description	Variables	Description
α	Railroad slope	t	Time
v_{veh}	Vehicle speed	m	Total mass
F_t	Electric traction or brake force	F_m	Mechanical brake force
n_p	Passenger numbers	$\dot{m}_{H_2}$	Hydrogen mass flow
SoC	State of charge	P_{bat}	Battery power
R_0	Internal ohmic resistance	I_{bat}	Battery current
V_{oc}	Open circuit voltage	V_i	Polarization voltage
C_i	Polarization capacitance	R_i	Polarization resistance
P_{dem}	Demand power	P_{aux}	Auxiliary systems' power
P_{loss}	Power losses along the drivetrain	P_{fc}	Fuel cell power

Table 4
System constraints.

Min	Variable	Max
0 kW	P_{fc}	200 kW
−5 kW/s	$\Delta P_{fc}/\Delta t$	5 kW/s
0.15	SoC	0.95
−900 A	I_{bat}	900 A

$$\Delta U_{VC} = C_6 \cdot e^{C_7 \Delta U_{cell}} \cdot n_{VC}, \quad (5)$$

Due to the difference in structure and materials of different fuel cells, the voltage degradation caused by the four operation modes varies greatly [34]. In this work, the Ballard Mark 513 fuel cell is utilized. The parameters of the fuel cell degradation model are adapted accordingly, as shown in Table 5.

The model prediction results for the Ballard Mark 513 fuel cell aging are compared with available experimental aging data [35–37] in Fig. 5. The experimental data in cases 1–3 are the average voltage degradation obtained by operating for a long time at different voltages. As for cases 4–8, the operating voltage of the fuel cells varies constantly between high and low voltages for 21000 cycles. Each cycle lasts one minute. The operating voltage and testing duration for the aging tests are shown in Table 6.

In Fig. 6, the polarization curve of the fuel cell is given. The studied fuel cell system has 1320 cells with each 300 cm^2 cell areas. As a result, the hydrogen consumption curve is extended considering the cell voltage degradation as shown in Fig. 7.

3. Deep reinforcement learning strategy for energy management system

Relying on the EMS, this module determines the power distribution between the fuel cell system and the battery system. Considering the constraints of the system, the EMS in this work aims to achieve best fuel economy, maintain battery SoC and slow down the fuel cell aging. To obtain an adaptive online available strategy, the TD3-based deep reinforcement learning is applied.

Generally, the energy management problem in the form of RL is represented as a Markov decision process (MDP). In the following subsections, the TD3 agent is presented and the settings for training are explained. After the training process, the trained policy obtained from the TD3 agent is utilized as the EMS. In Fig. 8, an overview of the TD3-based energy management for the fuel cell hybrid railway vehicle is displayed.

3.1. TD3 algorithm

In the area of RL, an MPD is applied to represent the interaction between an agent and its environment. According to the current state s of the environment, the agent performs an action a that follows a policy for the environment. Meanwhile, the agent receives a reward r for performing the action and a new state s' from the environment. Based on this feedback, the agent updates the policy. Its target is to find the policy π which maximizes the action-value function. Hereby, the action-value function, also known as the Q function is specified as the expected discounted sum of rewards:

$$Q(s,a) = \mathbb{E}[\sum_{k=0}^{T} \gamma^k r_k(s,a)], \quad (6)$$

where γ is the discount factor.

Table 5
Parameters of fuel cell degradation model based on Ballard Mark 513.

Parameters	Values
C_1	11.4
C_2	14
C_3	0.5057
C_4	0.07866
C_5	2.965
C_6	0.0045
C_7	16.46

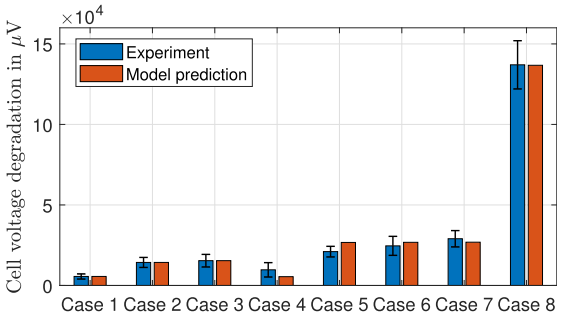

Fig. 5. Comparison of the experimental results and the prediction results from the fuel cell degradation model.

Table 6
Aging test conditions for case 1–8 based on Ballard Mark 513.

Case number	Operating voltage	Time
1	0.62 V	5600 h
2	0.78 V	11000 h
3	0.82 V	11000 h
4	0.6 –0.8 V	21000 cycles, each has 25 s for 0.6 V, 30 s for 0.8 V and 5 s for changing
5	0.6 –0.9 V	21000 cycles, each has 45 s for 0.6 V, 10 s for 0.9 V and 5 s for changing
6	0.6 –0.9 V	21000 cycles, each has 25 s for 0.6 V, 30 s for 0.9 V and 5 s for changing
7	0.6 –0.9 V	21000 cycles, each has 5 s for 0.6 V, 50 s for 0.9 V and 5 s for changing
8	0.6 –1.0 V	21000 cycles, each has 25 s for 0.6 V, 30 s for 1.0 V and 5 s for changing

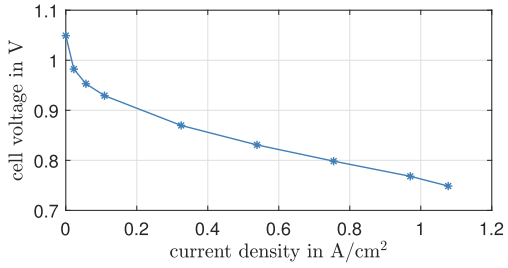

Fig. 6. Fuel cell polarization curve.

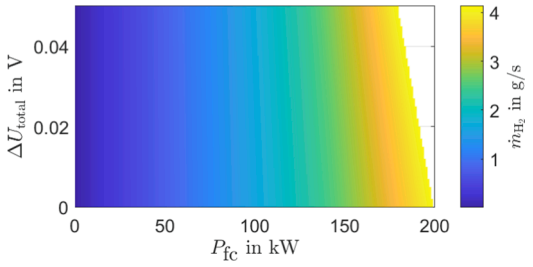

Fig. 7. Aging considered hydrogen consumption map of the fuel cell system with the inputs of the fuel cell system power and the total cell voltage degradation.

The twin delayed deep deterministic policy gradient algorithm (TD3) is among the state-of-the-art RL algorithms with an actor-critic configuration [38]. The policy is approximated by an actor neural network π_ϕ. This network maps states into continuous actions and has the

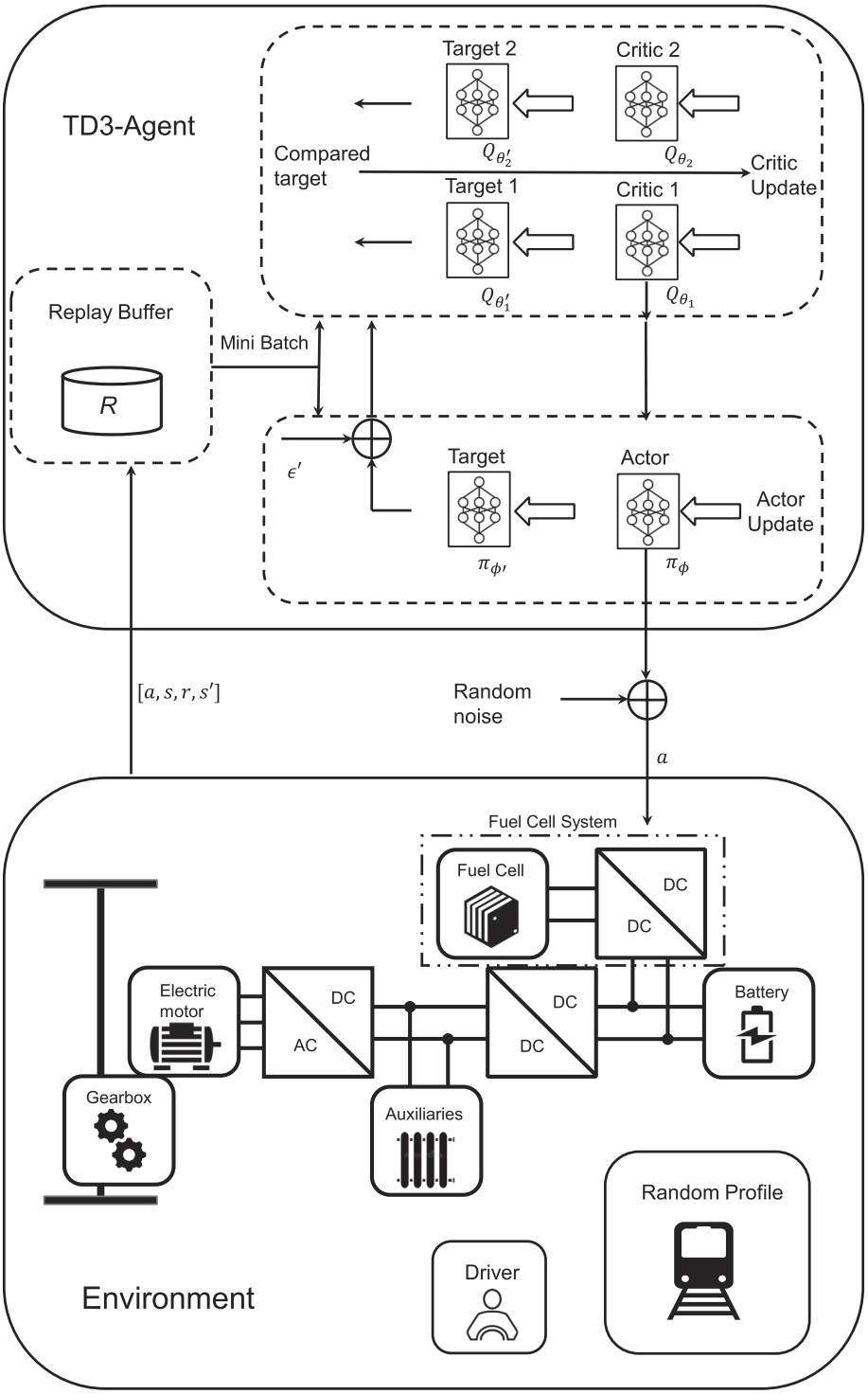

Fig. 8. TD3-based energy management for the fuel cell hybrid railway vehicle.

parameters denoted by ϕ. The action-value function Q is represented by two critic neural networks Q_{θ_1} and Q_{θ_2} with the network parameters θ_1 and θ_2. Its role is to evaluate actions and lead the actor to a favorable performance. These two critic networks constitute a clipped double-Q learning which can reduce the overestimation bias. Meanwhile, the target network approach is employed. The actor neural network π_ϕ and the two critic neural networks $Q_{\theta_{1,2}}$ have their corresponding target networks $\pi_{\phi'}$ and $Q_{\theta'_{1,2}}$. The target network parameters ϕ', θ'_1 and θ'_2 initially have the same values as the corresponding original networks and are slowly updated, which provides consistent goals and stabilizes the training process.

The training of the two critic neural networks aims at minimizing the following loss function:

$$L = [y - Q_{\theta_i}(s,a)]^2, \quad i = 1,2, \tag{7}$$

whereby a single target value y is computed from the target neural networks:

$$y = r + \gamma \min_{i=1,2} Q_{\theta_i'}(s', \pi_{\phi'}(s') + \epsilon'), \tag{8}$$

where, $\epsilon' \sim \mathrm{clip}(\mathcal{N}(0, 0.2), -0.5, 0.5)$ is a clipped noise.

The policy in the actor neural network is set to be updated less frequently than the critic neural networks. The objective is to minimize the loss function before updating the policy. Using the deterministic policy gradient algorithm [39], the policy can be updated by taking the gradient of the expected return:

$$\nabla_\phi J(\phi) = \mathbb{E}[\nabla_a Q_{\theta_1}(s,a)|_{a=\pi_\phi(s)} \nabla_\phi \pi_\phi(s)]. \tag{9}$$

3.2. TD3-based energy management strategy

As in the DRL configuration shown in Fig. 8, the training environment including the entire vehicle model interacts with the TD3-based agent. The settings of agent actions, environment states, and rewards are critical to the interaction and learning of the agent. In this subsection, the setup of the TD3-based energy management is explained in detail.

3.2.1. TD3-based agent actions

The agent controls the power output of the fuel cell system. Thus, the desired fuel cell power is a direct control input to the environment, where the fuel cell system is automatically controlled by its own DCDC converter to achieve the desired power output. Therefore, the action is defined as

$$a = [P_{\mathrm{fc}}]. \tag{10}$$

3.2.2. Environment states

The agent needs proper states information to reasonably manage the power. The power demand of the vehicle, the state of the fuel cell system, and the state of the battery system are used to form the environment states. In this case, the power demand P_{dem}, the fuel cell system power P_{fc}, and the state of charge SoC can represent the power flow of the power system, the state of the fuel cell system and the dynamics of the battery, respectively. The vehicle velocity v_{veh}, the acceleration $\dot{v}_{\mathrm{veh}}$ and the passenger numbers n_{p} characterize the vehicle dynamics considering the passenger flow. Thus, the states of the environment are set to be:

$$s = [P_{\mathrm{dem}}, P_{\mathrm{fc}}, SoC, v_{\mathrm{veh}}, \dot{v}_{\mathrm{veh}}, n_{\mathrm{p}}]. \tag{11}$$

3.2.3. Rewards

The reward setting is crucial because the reward not only provides feedback to the agent on the effectiveness of the action, but also strongly influences the success of the training convergence. In this work, the reward combines the cost of hydrogen consumption $L_{\mathrm{H_2}}$, the cost of fuel cell aging $L_{\mathrm{fc_{aging}}}$, and the penalty for battery charge-sustaining G_{bat}:

$$r = -L_{\mathrm{H_2}} - L_{\mathrm{fc_{aging}}} + G_{\mathrm{bat}}. \tag{12}$$

Since the cost parts $L_{\mathrm{H_2}}$ and $L_{\mathrm{fc_{aging}}}$ are kept as small as possible and the goal of the TD3-agent is to maximize the reward, $L_{\mathrm{H_2}}$ and $L_{\mathrm{fc_{aging}}}$ are assigned minus signs in the reward function.

Accordingly, at a time step $\Delta t = 1\,\mathrm{s}$ with the hydrogen price $p_{\mathrm{H_2}}$, the hydrogen cost $L_{\mathrm{H_2}}$ is defined as follows:

$$L_{\mathrm{H_2}} = p_{\mathrm{H_2}} \cdot \dot{m}_{\mathrm{H_2}} \cdot \Delta t. \tag{13}$$

To analyze the effects of fuel cell aging in a meaningful way, $L_{\mathrm{H_2}}$ denotes the hydrogen cost under non-degradation conditions. The additional hydrogen cost caused by fuel cell degradation $\dot{m}_{\mathrm{H_2,aging}}$ can be calculated by looking up Fig. 7 and subtracting the hydrogen cost under non-degradation conditions. Then the fuel cell aging cost is:

$$L_{\mathrm{fc_{aging}}} = p_{\mathrm{fc}} \cdot \frac{\Delta U_{\mathrm{deg}}}{\Delta U_{\mathrm{tolerance}}} + p_{\mathrm{H_2}} \cdot \dot{m}_{\mathrm{H_2,aging}} \cdot \Delta t, \tag{14}$$

where p_{fc} is the fuel cell system price and $\Delta U_{\mathrm{tolerance}}$ is the tolerable voltage degradation. According to [40], $\Delta U_{\mathrm{tolerance}}$ is defined as the 10% voltage degradation at $1.0 - 1.5\,\mathrm{A/cm^2}$, which in this work corresponds to 74 mV.

Furthermore, G_{bat} represents the penalty term for the battery. In contrast to the general method of using the squared error of the SoC to realize the sustained charging, a penalty term is defined that provides more efficient immediate feedback to the agent:

$$G_{\mathrm{bat}} = w_{\mathrm{bat}} \cdot (SoC - SoC_{\mathrm{ref}}) \cdot I_{\mathrm{bat}}, \tag{15}$$

where w_{bat} is the weighting factor of the battery's charge-sustaining. It has the unit of \$/kA, which keeps G_{bat} dimensionally compatible with the hydrogen cost. SoC_{ref} represents the reference value of SoC.

3.3. Environment settings for training

In the RL configuration, the environment should simulate the complete vehicle behavior to provide sufficient training information to the agent. To obtain a properly adaptive TD3-EMS and avoid overfitting in the training, a comprehensive set of speed profiles is required. In this work, recorded railroad data is used. The raw data were collected on the railroad line between Aachen and Cologne, Germany. In total, four speed profiles were recorded for different time periods during the day. In order to fully utilize the data to represent different driving conditions and to maintain a certain degree of realism, the speed profiles between every two adjacent stations are randomly selected and finally combined into one speed profile. Fig. 9 shows the generation of the random speed profiles based on the recorded ones. The number of passengers is randomly selected from 0 to 150, taking into account the variation of the total vehicle mass due to the passenger flow.

4. Training and simulation results

4.1. Training settings

For the proposed TD3-based energy management strategy (TD3-EMS), a training procedure should be performed before testing. The settings for training the TD3-agent are summarized in Table 7. To fully explore the policy, the normally distributed noise added to the action is large at the beginning and then its variance decreases with a 98% discount rate per training episode i.

As in the reward settings, the hydrogen consumption and aging costs are unified as prices. To calculate this part of the reward, the hydrogen price $p_{\mathrm{H_2}}$ is given as 0.002 \$/g [41]. According to [42], the total stack cost for a 100 kW system is \$70,000, thus p_{fc} is \$140,000 for the fuel cell system under study. As for the battery system, SoC for each training episode always starts at 0.6 and the reference value of SoC is also 0.6. Hereby, the weighting factor w_{bat} determines the impact of the battery's SoC. To achieve proper charge-sustaining, this factor is 0.03 after repeated tuning.

At the beginning of the training process, the replay buffer acquires the memory data in the first ten training episodes. The speed profiles and the passenger flow are randomly generated for each episode. Then the training of the TD3 agent begins from the eleventh episode. During the training, the speed profile and the number of passengers are varied every five episodes, allowing the agent to experience different environmental situations while maintaining a consistent training environment.

4.2. Impact of the reward settings on convergence

The reward setting is critical for training convergence. In this work, the SoC of the battery can vary greatly due to the large power dynamics of the rail vehicle. When the SoC reaches the system limits shown in Table 4, the training episode will be automatically aborted. To achieve adaptive maintenance of the SoC and fast convergence of the training, a

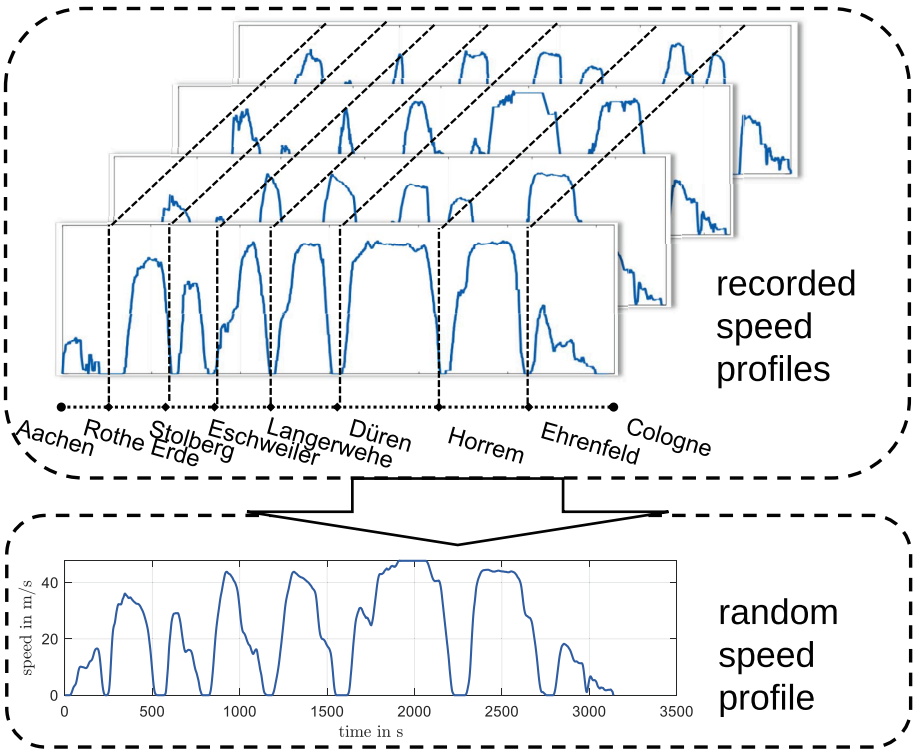

Fig. 9. Stochastic environment generated by randomly combining the speed profiles of the corresponding railroad sections from Aachen to Cologne in Germany.

Table 7
Settings for training the TD3 agent.

Networks and parameters	Settings and values
Actor networks	Fully-connected layers 400/200/100 units
Actor learning rate	0.0001
Critic networks	Fully-connected layers 400/200/100 units
Critic learning rate	0.001
Optimizer type	Adam
Batch size	64
Delayed policy update iterations	10
Target update rate	0.005
Discount factor	0.99
Exploration policy	$\mathcal{N}(0, var)$, $var = 3{\cdot}0.98^i$

penalty term (15) is introduced. To show the improvement, another TD3-agent with a commonly used penalty term G^*_{bat} is trained for comparison:

$$G^*_{\text{bat}} = w^*_{\text{bat}} \cdot (SoC - SoC_{\text{ref}})^2, \tag{16}$$

where w^*_{bat} is a penalty factor with a value of 20 to penalize the action appropriately when the SoC deviates from the reference value. In Fig. 10, the proposed TD3-EMS converges at about 150 episodes and then remains stationary. The normal TD3-EMS using (16) suffers from intense fluctuations and is difficult to converge. This shows that the commonly used term (16) does not give valid feedback. For the large sized battery in this work, the squared error of the SoC represents the deviation from the reference value. As the SoC increment at each time step is small, it has a minor effect on the squared error of the SoC no matter what action the agent takes. Therefore, this reward term is almost constant and hardly reflects the quality of the action. The new reward function term (15) is directly related to the battery current and can give timely feedback on the effect of the action. The results show that it can accelerate and stabilize the training process. In the following subsections, only the proposed TD3-EMS with (15) is considered.

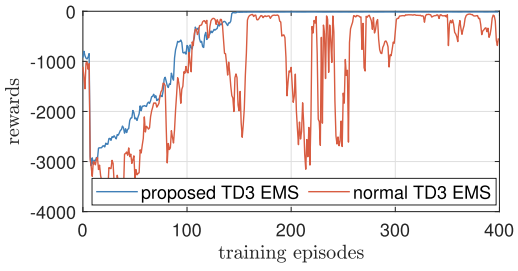

Fig. 10. Comparison of total training rewards by using the proposed reward term (15) and the normal reward term (16) for battery ch.arge-sustaining.

4.3. Simulation results and analysis

In order to evaluate the proposed TD3-EMS, a test speed profile of the railroad route from Aachen to Cologne (test cycle 1) shown in Fig. 11a is utilized, which is different from the random speed profiles introduced in Section 3.3 for training. A typical fault in the reinforcement learning is the overfitting, which results in promising performance on the training cycles while not on the test cycles. Despite the different speed profiles between the test cycle 1 and the training cycles, they share the same railroad route. To investigate the transferability and adaptability of the proposed TD3-EMS, another test profile of the railroad route in Baden-Württemberg (test cycle 2) shown in Fig. 11b is used. In the following subsections, results on passenger flow disturbance, battery's charge-sustaining, fuel cell aging and optimality of the operational cost are presented.

4.3.1. Performance with different passenger flows

Fig. 12 shows the trajectories of fuel cell power and battery's SoC under the proposed TD3-EMS. As a common disturbance in the rail transportation, the passenger flow is considered. From Fig. 12a and Fig. 12b, it can be seen that the proposed TD3-EMS controls the fuel cell power above 100 kW when the railway vehicle is in motion and around 60 kW when the railway vehicle is stopping at the station. Especially in

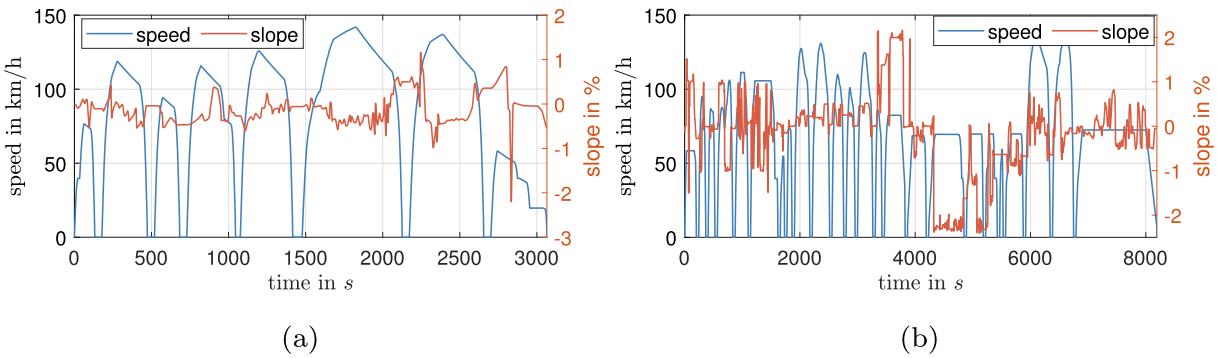

Fig. 11. Speed and railroad slope profiles of: (a) test cycle 1, (b) test cycle 2.

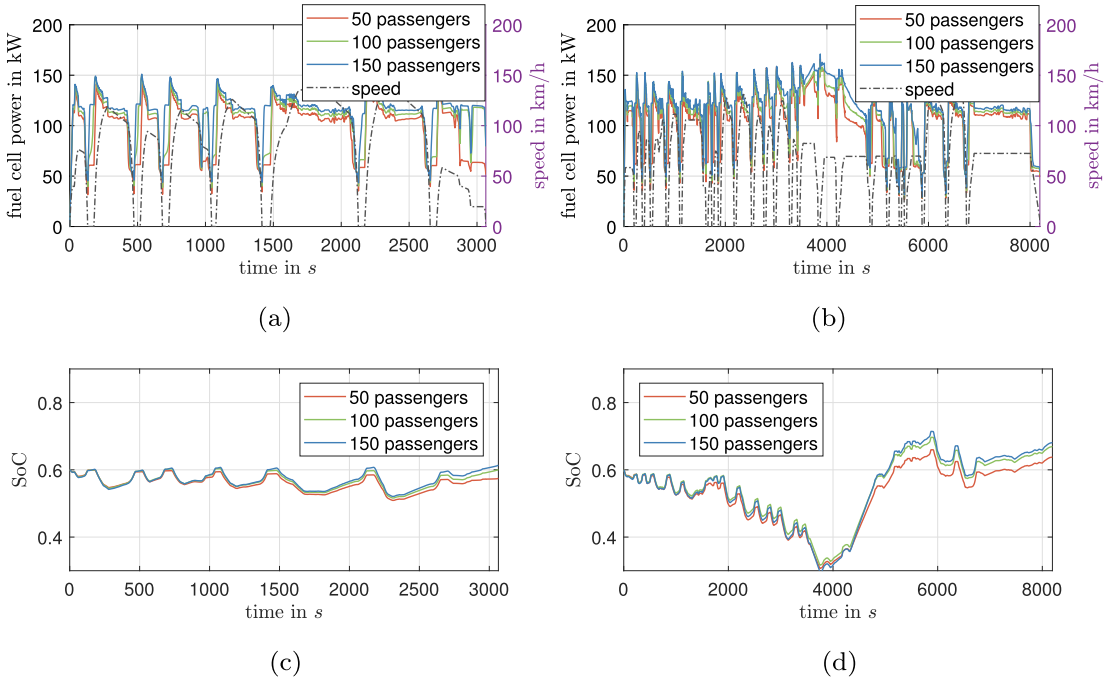

Fig. 12. Simulation results under the proposed TD3-EMS for various passenger numbers of: (a) fuel cell power in test cycle 1, (b) fuel cell power in test cycle 2, (c) battery's SoC in test cycle 1, (d) battery's SoC in test cycle 2.

the acceleration phase of the vehicle, the fuel cell power can be boosted up to 150 kW. This makes sense because the driving situation causes different power demands. The fuel cell power due to different driving situations exactly reflects the adaptability of the proposed TD3-EMS. Furthermore, the fuel cell power increases as the number of passengers increases, indicating the ability of the proposed TD3-EMS to adapt to passenger flow. Despite the different passenger flows, the ending SoC is kept around 0.6 in Fig. 12c and Fig. 12d. This demonstrates that the proposed TD3-EMS has achieved charge-sustaining.

4.3.2. Battery's charge-sustaining

To further investigate the performance of the TD3-EMS in terms of battery's charge-sustaining, the simulations are performed with different initial SoC values while the reference SoC remains at 0.6. The results of the fuel cell power and the SoC trajectories are displayed in Fig. 13. As expected, from Fig. 13a and Fig. 13b, it can clearly be seen that the fuel cell power is generally lower when the initial SoC is 0.8 compared to the case where the initial SoC is 0.4. The resulting SoC trajectories shown in Fig. 13c and d indicate a trend of SoC trajectories approaching 0.6. In detail, under test cycle 1, the SoC changes at initial values of 0.4 and 0.8 are $+0.06$ and -0.15, respectively. While under test cycle 2, the SoC changes at initial values of 0.4 and 0.8 are $+0.21$ and -0.14, respectively. Hence, a promising adaptability of the TD3-EMS in terms of charge-sustaining is demonstrated.

4.3.3. Fuel cell aging

To evaluate the aging behavior of fuel cells under the proposed TD3-EMS, a reference TD3-EMS without the fuel cell aging term in the reward function is used for comparison. The simulation results with 100 passengers are presented in Fig. 14. From the fuel cell power trajectories in Fig. 14a and b, it can be seen that the proposed TD3-EMS controls the power output of fuel cell system less dynamically than the reference TD3-EMS. Fig. 14c and d display the resulting fuel cell voltage degradation in four different types. Since the fuel cell does not operate at low power under the proposed TD3-EMS, the voltage degradation caused by open circuit operation (OCV) is significantly reduced. The less dynamic fuel cell power contributes to the reduction of the voltage degradation caused by the voltage cycles (VC). In total, the voltage degradation is reduced by 40.2% from 5.07 μV to 3.03 μV in test cycle 1. While in test cycle 2, it is reduced by 49.2% from 13.7 μV to 6.9 μV. In Table 8, the detailed voltage degradation with varying passenger numbers is presented. In summary, the proposed TD3-EMS can stabilize the fuel cell

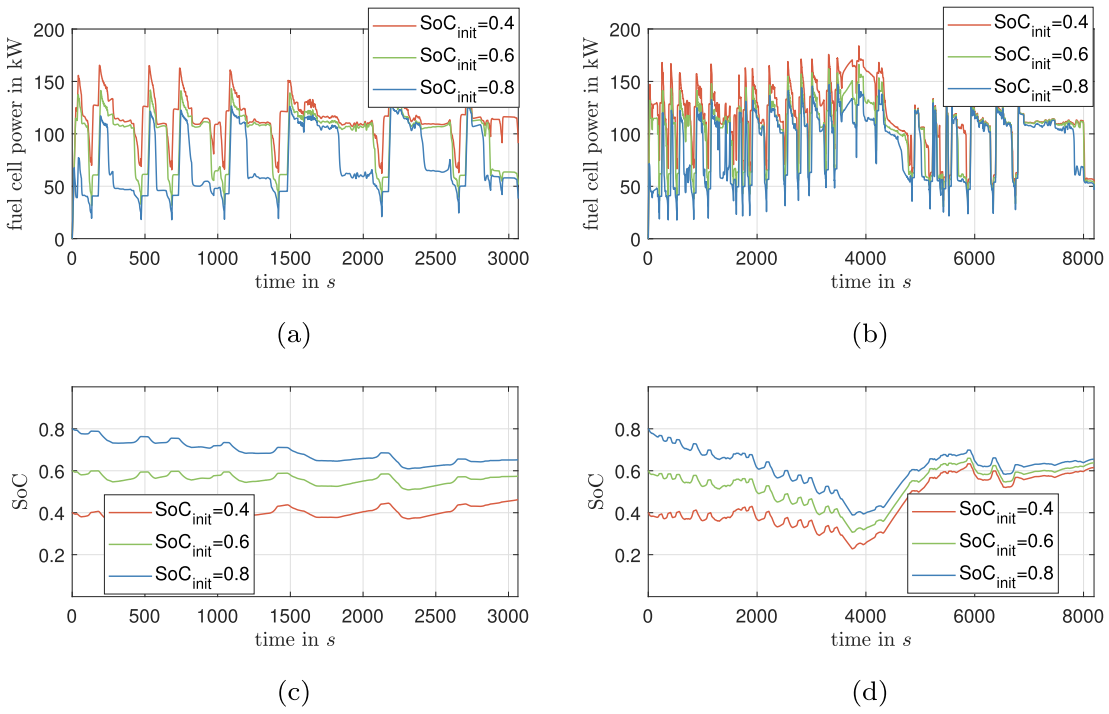

Fig. 13. Simulation results under the proposed TD3-EMS for various initial SoC values of: (a) fuel cell power in test cycle 1, (b) fuel cell power in test cycle 2, (c) battery's SoC in test cycle 1, (d) battery's SoC in test cycle 2.

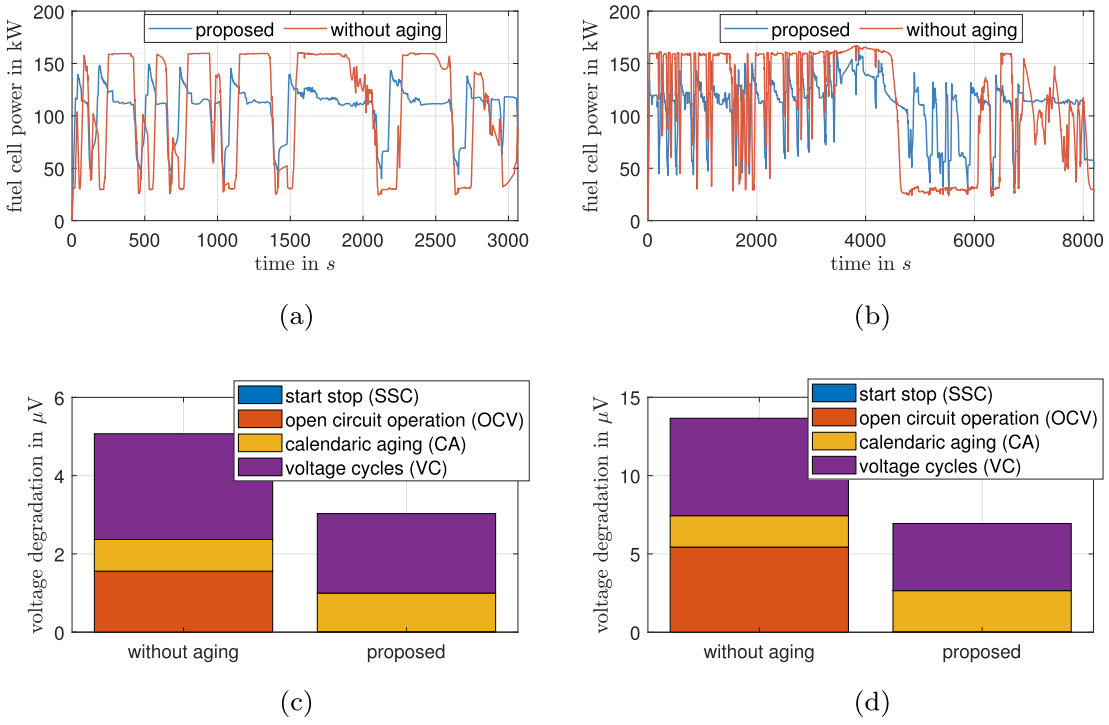

Fig. 14. Simulation results and the comparison between the proposed TD3-EMS and the one without considering aging: (a) fuel cell power in test cycle 1, (b) fuel cell power in test cycle 2, (c) fuel cell voltage degradation in test cycle 1, (d) fuel cell voltage degradation in test cycle 2.

power compared to the reference one, reducing the cell voltage degradation by up to 50.6%. It shows that the proposed TD3-EMS has the ability to slow down the fuel cell aging.

4.3.4. *Optimality of operational cost*

The operational cost of the railway vehicle consists of the hydrogen cost and the fuel cell aging cost. To verify the optimality in the operational cost of the proposed TD3-EMS, an offline global optimization strategy is used as the benchmark. The offline strategy is based on PMP and is verified by dynamic programming in [8]. To show the improvement of the proposed TD3-EMS as an online strategy, the TD3-EMS without considering aging is used for comparison. The simulation

Table 8
Simulation results for the comparison of the voltage degradation caused by start-stop cycle cycles (SSC), open circuit operation (OCV), calendaric aging (CA) and voltage cycles (VC) under test cycle 1 with 1). the reference TD3-EMS without considering fuel cell aging, 2). the proposed TD3-EMS.

Cycle	Strategies	n_p	Voltage degradation [μV]					Difference [%]
			SSC	OCV	CA	VC	Total	
1	Reference	50	0	1.99	0.78	2.72	5.49	–
1	Proposed	50	0	0.03	1.00	1.81	2.84	−48.3%
1	Reference	100	0	1.56	0.81	2.70	5.07	–
1	Proposed	100	0	0.01	0.99	2.03	3.03	−40.2%
1	Reference	150	0	1.12	0.85	2.77	4.74	–
1	Proposed	150	0	0.01	0.98	2.09	3.08	−35.0%
2	Reference	50	0	5.35	2.05	6.25	13.65	–
2	Proposed	50	0	0.04	2.66	4.05	6.75	−50.6%
2	Reference	100	0	5.44	2.00	6.22	13.66	–
2	Proposed	100	0	0.03	2.62	4.29	6.94	−49.2%
2	Reference	150	0	5.05	2.02	5.69	12.76	–
2	Proposed	150	0	0.02	2.59	4.13	6.74	−47.1%

results are presented in Table 9. It can be seen that the hydrogen cost with the proposed TD3-EMS is generally close to the benchmark. However, there are still some gaps in the total operational costs. It is because of the lack of global information, the proposed TD3-EMS reacts to the timely power demands and manages higher power dynamics than the benchmark, resulting in larger voltage degradation, especially in voltage cycles (VC). Compared to the strategy without considering aging, the proposed TD3-EMS indicates up to 28% improvement in the total operational cost. On the one hand, due to the consideration of the fuel cell aging, the power output of the fuel cell system trends to be more stable than the strategy without considering aging. On the other hand, due to the convexity of the hydrogen consumption characteristic curve shown in Fig. 4, the stabilized fuel cell power can benefit the operational cost under the condition of charge-sustaining mode. In summary, the proposed TD3-EMS shows superiority as an online strategy in terms of the operational cost.

5. Conclusions

In this work, a deep reinforcement learning-based energy management strategy considering fuel cell aging is proposed for fuel cell and battery hybrid rail vehicles. Firstly, a fuel cell and battery railway vehicle model is introduced. To improve the forecast for the aging of the fuel cell system, an operation mode-oriented estimation model is introduced. Based on the vehicle model, an energy management problem is formulated under the reinforcement learning structure and one of the state-of-the-art DRL methods TD3 is used to solve the problem. To achieve more realistic simulations and to overcome overfitting, a stochastic training environment based on measured vehicle speed profiles and random passenger numbers is created. The training is done through the interaction between the TD3 agent and the environment. Thereby, a new reward term for the battery's charge-sustaining is introduced. The training results indicate a more stable convergence than the one with the commonly used error squared SoC. After training, the proposed TD3-EMS is initially tested with two real world speed profiles and different numbers of passengers, which shows promising capabilities for SoC maintenance. The proposed TD3-EMS achieves up to 50.6% reduction in fuel cell voltage degradation compared to the reference TD3-EMS without considering aging. Moreover, the overall operational cost is investigated for the two strategies and a benchmark strategy. The proposed TD3-EMS show results closer to the benchmark with an improvement of up to 28% compared to another online strategy. In future work, more online available information that facilitates approaching the benchmark optimum will be considered to further reduce overall operational costs.

Table 9
Simulation results in terms of hydrogen cost, fuel cell degradation cost and operational cost for 1). offline benchmark strategy, 2). proposed TD3-EMS and 3). TD3-EMS without aging.

Cycle	Strategies	n_p	H_2[$]	Aging [$]	Cost [$]	Difference [%]
1	Benchmark	50	9.71	3.61	13.21	–
1	Proposed	50	9.81	5.39	15.20	+14.1
1	Without aging	50	10.60	10.42	21.02	+57.8
1	Benchmark	100	10.84	3.29	14.13	–
1	Proposed	100	10.93	5.75	16.68	+18.1
1	Without aging	100	11.78	9.63	21.41	+51.6
1	Benchmark	150	11.78	3.84	15.62	–
1	Proposed	150	11.82	5.84	17.66	+13.1
1	Without aging	150	12.67	9.00	21.67	+38.7
2	Benchmark	50	27.78	8.33	36.11	–
2	Proposed	50	28.16	12.82	40.98	+13.5
2	Without aging	50	29.99	25.94	55.93	+54.9
2	Benchmark	100	30.17	8.50	38.67	–
2	Proposed	100	30.52	13.19	43.71	+13.0
2	Without aging	100	32.34	25.95	58.29	+50.7
2	Benchmark	150	32.31	9.16	41.47	–
2	Proposed	150	32.69	12.81	45.50	+9.7
2	Without aging	150	34.56	24.24	58.80	+41.8

CRediT authorship contribution statement

Kai Deng: Conceptualization, Methodology, Software, Formal-analysis, Data-curation, Validation, Writing-original-draft. **Yingxu Liu:** Formal-analysis, Data-curation, Validation, Writing-review-editing. **Di Hai:** Writing-review-editing. **Hujun Peng:** Writing-review-editing. **Lars Löwenstein:** Writing-review-editing. **Stefan Pischinger:** Writing-review-editing. **Kay Hameyer:** Supervision, Writing-review-editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Factsheet: hydrogen and fuel cell technology in china, https://www.now-gmbh.de/wp-content/uploads/2020/09/Factsheet-China-FC-EN.pdf; 2020 [Accessed June 14, 2021].

[2] Road map to a us hydrogen economy, https://cafcp.org/sites/default/files/ Road+Map+to+a+US+Hydrogen+Economy+Full+Report.pdf, accessed September 17, 2021 (2020).
[3] Xu N, Kong Y, Chu L, Ju H, Yang Z, Xu Z, Xu Z. Towards a smarter energy management system for hybrid vehicles: a comprehensive review of control strategies. Appl Sci 2019;9(10):2026.
[4] Shaohua L, Changqing D, Fuwu Y, Jun W, Zheng L, Yuan L. A rule-based energy management strategy for a new bsg hybrid electric vehicle. In: 2012 Third global congress on intelligent systems. IEEE; 2012. P. 209–12.
[5] Pam A, Bouscayrol A, Fiani P, Noth F. Rule-based energy management strategy for a parallel hybrid electric vehicle deduced from dynamic programming. In: 2017 IEEE vehicle power and propulsion conference (VPPC), IEEE; 2017. P. 1–6.
[6] Peng H, Li J, Thul A, Deng K, Ünlübayir C, Löwenstein L, Hameyer K. A scalable, causal, adaptive rule-based energy management for fuel cell hybrid railway vehicles learned from results of dynamic programming. eTransportation 4:2020; 100057.
[7] Liu C, Wang Y, Wang L, Chen Z. Load-adaptive real-time energy management strategy for battery/ultracapacitor hybrid energy storage system using dynamic programming optimization. J Power Sour 2019;438:227024 .
[8] Peng H, Chen Z, Li J, Deng K, Dirkes S, Gottschalk J, Ünlübayir C, Thul A, Löwenstein L, Pischinger S, et al. Offline optimal energy management strategies considering high dynamics in batteries and constraints on fuel cell system power rate: from analytical derivation to validation on test bench. Appl Energy 2021;282: 116152 .
[9] Zhang N, Ma X, Jin L. Energy management for parallel hev based on pmp algorithm. In: 2017 2nd International conference on robotics and automation engineering (ICRAE), IEEE; 2017. P. 177–82.
[10] Kim N, Jeong J, Zheng C. Adaptive energy management strategy for plug-in hybrid electric vehicles with pontryagin's minimum principle based on daily driving patterns. Int J Precis Eng Manuf-Green Technol 2019;6(3):539–48.
[11] Li X, Wang Y, Yang D, Chen Z. Adaptive energy management strategy for fuel cell/battery hybrid vehicles using pontryagin's minimal principle. J Power Sour 2019; 440:227105 .
[12] Xiang C, Ding F, Wang W, He W, Qi Y. Mpc-based energy management with adaptive markov-chain prediction for a dual-mode hybrid electric vehicle. Sci China Technol Sci 2017;60(5):737–48.
[13] Deng K, Peng H, Dirkes S, Gottschalk J, Ünlübayir C, Thul A, Löwenstein L, Pischinger S, Hameyer K. An adaptive pmp-based model predictive energy management strategy for fuel cell hybrid railway vehicles. eTransportation 7:2021; 100094.
[14] Liu C, Murphey YL. Optimal power management based on q-learning and neuro-dynamic programming for plug-in hybrid electric vehicles. IEEE Trans Neural Networks Learn Syst 2019;31(6):1942–54.
[15] Du G, Zou Y, Zhang X, Kong Z, Wu J, He D. Intelligent energy management for hybrid electric tracked vehicles using online reinforcement learning. Appl Energy 2019;251:113388 .
[16] Xu B, Rathod D, Zhang D, Yebi A, Zhang X, Li X, Filipi Z. Parametric study on reinforcement learning optimized energy management strategy for a hybrid electric vehicle. Appl Energy 2020;259:114200 .
[17] He D, Zou Y, Wu J, Zhang X, Zhang Z, Wang R. Deep q-learning based energy management strategy for a series hybrid electric tracked vehicle and its adaptability validation. In: 2019 IEEE transportation electrification conference and expo (ITEC), IEEE; 2019. P. 1–6.
[18] Hu Y, Li W, Xu K, Zahid T, Qin F, Li C. Energy management strategy for a hybrid electric vehicle based on deep reinforcement learning. Appl Sci 2018;8(2):187.
[19] Wu Y, Tan H, Peng J, Zhang H, He H. Deep reinforcement learning of energy management with continuous control strategy and traffic information for a series-parallel plug-in hybrid electric bus. Appl Energy 2019;247:454–66.
[20] Lian R, Peng J, Wu Y, Tan H, Zhang H. Rule-interposing deep reinforcement learning based energy management strategy for power-split hybrid electric vehicle. Energy 2020;197:117297 .
[21] Wu J, Wei Z, Liu K, Quan Z, Li Y. Battery-involved energy management for hybrid electric bus based on expert-assistance deep deterministic policy gradient algorithm. IEEE Trans Vehicul Technol 2020;69(11):12786–96.
[22] Wu J, Wei Z, Li W, Wang Y, Li Y, Sauer D. Battery thermal-and health-constrained energy management for hybrid electric bus based on soft actor-critic drl algorithm. IEEE Trans Ind Inf.
[23] Wu J, Yuan XZ, Martin JJ, Wang H, Zhang J, Shen J, Wu S, Merida W. A review of pem fuel cell durability: degradation mechanisms and mitigation strategies. J Power Sour 2008;184(1):104–19.
[24] Yue M, Jemei S, Gouriveau R, Zerhouni N. Review on health-conscious energy management strategies for fuel cell hybrid electric vehicles: degradation models and strategies. Int J Hydrogen Energy 2019;44(13):6844–61.
[25] Li J, Hu Z, Xu L, Ouyang M, Fang C, Hu J, Cheng S, Po H, Zhang W, Jiang H. Fuel cell system degradation analysis of a chinese plug-in hybrid fuel cell city bus. Int J Hydrogen Energy 2016;41(34):15295–310.
[26] Carcadea E, Ismail MS, Ingham DB, Patularu L, Schitea D, Marinoiu A, Ion-Ebrasu D, Mocanu D, Varlam M. Effects of geometrical dimensions of flow channels of a large-active-area pem fuel cell: a cfd study. Int J Hydrogen Energy 2021;46 (25):13572–82.
[27] Chellehbari YM, Adavi K, Amin JS, Zendehboudi S. A numerical simulation to effectively assess impacts of flow channels characteristics on solid oxide fuel cell performance. Energy Convers Manage 2021;244:114280 .
[28] Fink C, Gößling S, Karpenko-Jereb L, Urthaler P. Cfd simulation of an industrial pem fuel cell with local degradation effects. Fuel Cells 2020;20(4):431–52.
[29] Wang Y, Moura SJ, Advani SG, Prasad AK. Power management system for a fuel cell/battery hybrid vehicle incorporating fuel cell and battery degradation. Int J Hydrogen Energy 2019;44(16):8479–92.
[30] Fletcher T, Thring R, Watkinson M. An energy management strategy to concurrently optimise fuel consumption & pem fuel cell lifetime in a hybrid vehicle. Int J Hydrogen Energy 2016;41(46):21503–15.
[31] Pei P, Chang Q, Tang T. A quick evaluating method for automotive fuel cell lifetime. Int J Hydrogen Energy 2008;33(14):3829–36.
[32] Bonitz S. Zur Lebensdauerabschätzung von Brennstoffzellen mit den Methoden der Betriebsfestigkeit. Technische Universität Clausthal 2016.
[33] De Bruijn F, Dam V, Janssen G. Durability and degradation issues of pem fuel cell components. Fuel Cells 2008;8(1):3–22.
[34] Placca L, Kouta R. Fault tree analysis for pem fuel cell degradation process modelling. Int J Hydrogen Energy 2011;36(19):12393–405.
[35] St-Pierre J, Jia N. Successful demonstration of ballard pemfcs for space shuttle applications. J New Mater Electrochem Syst 2002;5(4):263–72.
[36] St-Pierre J, Wilkinsor D, Knights S, Bos M. Relationships between water management, contamination and lifetime degradation in pefc. J New Mater Electrochem Syst 2000;3(2):99–106.
[37] Zihrul P, Hartung I, Kirsch S, Huebner G, Hasché F, Gasteiger HA. Voltage cycling induced losses in electrochemically active surface area and in h2/air-performance of pem fuel cells. J Electrochem Soc 2016;163(6):F492.
[38] Fujimoto S, Hoof H, Meger D. Addressing function approximation error in actor-critic methods. In: International conference on machine learning, PMLR; 2018. P. 1587–96.
[39] Silver D, Lever G, Heess N, T. Degris, D. Wierstra, M. Riedmiller, Deterministic policy gradient algorithms. In: International conference on machine learning, PMLR; 2014. P. 387–95.
[40] Doe technical targets for polymer electrolyte membrane fuel cell components, https://www.energy.gov/eere/fuelcells/doe-technical-targets-polymer-electrolyte-membrane-fuel-cell-components, accessed June 14, 2021.
[41] Doe technical targets for hydrogen production from electrolysis, https://www.energy.gov/eere/fuelcells/doe-technical-targets-hydrogen-production-electrolysis [Accessed June 14, 2021].
[42] Manufacturing cost analysis of 100 and 250 kw fuel cell systems for primary power and combined heat and power applications, https://www.energy.gov/eere/fuelcells/downloads/manufacturing-cost-analysis-100-and-250-kw-fuel-cell-systems-primary-power; 2016 [Accessed June 14, 2021].