creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학박사학위논문

시계열 데이터 분석 통합 심층 결정적 정책 경사법을 활용한 연료전지 차량의 플러딩과 드라잉 진단 기반 에너지 효율 주행 시스템 개발

Energy-efficient Driving based on Diagnosis of Flooding and Drying in Fuel Cell Electric Vehicle using Sequential Data Analysis-integrated Deep Deterministic Policy Gradient

2023 년 8 월

서울대학교 대학원

기계항공공학부

김 경 현

시계열 데이터 분석 통합 심층 결정적 정책 경사법을 활용한 연료전지 차량의 플러딩과 드라잉 진단 기반 에너지 효율 주행 시스템 개발

Energy-efficient Driving based on Diagnosis of Flooding and Drying in Fuel Cell Electric Vehicle using Sequential Data Analysis-integrated Deep Deterministic Policy Gradient

지도교수 차 석 원

이 논문을 박사 학위논문으로 제출함
2023 년 4 월

서울대학교 대학원
기계항공공학부
김 경 현

김경현의 공학박사 학위논문을 인준함
2023 년 6 월

위 원 장 _____ 안 성 훈 _____ (인)

부위원장 _____ 차 석 원 _____ (인)

위 원 _____ 송 한 호 _____ (인)

위 원 _____ 박 태 현 _____ (인)

위 원 _____ 이 희 윤 _____ (인)

Abstract

Energy-efficient Driving based on Diagnosis of Flooding and Drying in Fuel Cell Electric Vehicle using Sequential Data Analysis-integrated Deep Deterministic Policy Gradient

The issue of global warming caused by rapid climate change has gained worldwide attention. The automotive industry is exploring various solutions to reduce greenhouse gas emissions. One such solution is the development of autonomous driving technology to achieve energy-efficient driving with minimal fuel consumption based on vehicle-to-infrastructure (V2I). Another approach is the hybridization of vehicles by replacing the power source with environmentally-friendly fuel cells, specifically, Polymer Electrolyte Membrane Fuel Cells (PEMFC). However, these solutions present challenges that must be addressed.

In order to achieve energy-efficient driving in fuel cell hybrid electric vehicles (FCHEVs), it is essential to develop optimal speed control and power distribution strategies among the PEMFC and battery based on the vehicle's internal and external information. These approaches require optimization techniques for achieving the best performance; however, the long calculation times of these techniques make them difficult to apply to actual vehicles, requiring appropriate control strategies. Furthermore, the PEMFC, which is the energy source of FCHEVs, has critical flaws such as flooding and drying, occurring for a long-time operation under extreme conditions. Therefore, it is necessary to operate the fuel cell under stable conditions by implementing driving control and batter power assist. In this paper, a driving

system for FCHEVs is proposed that considers energy-efficient speed control, power distribution strategies, and moisture-related errors of PEMFC.

Initially, modeling of the FCHEV is performed using backward-looking simulation. Additionally, a semi-empirical PEMFC model is developed based on a single fuel cell experimental result. Using the developed vehicle model, the driving system is trained with deep deterministic policy gradient (DDPG), a type of reinforcement learning. This system produces the target speed of the vehicle and reference values for power distribution to the controller of the fuel cell and battery. By updating the action space of the DDPG so as not to exceed the limiting conditions of the powertrain at every step, the vehicle model increases the possibility of performing the actions suggested by the system. Through parameter optimization, the performance of the model is improved by applying parameters suitable for the DDPG. Furthermore, the model evaluates fuel consumption and operational point by considering the road gradient applied to learning phase. The proposed system exhibits 97.11 % optimality compared to DP, a global optimization method and outperformed control based on cruise control and rule-based strategy by 36.52 %.

A model for diagnosing the defects is developed to determine whether flooding or drying occurs in FCHEVs of an energy-efficient driving system. An experimental procedure is conducted to deliberately trigger flooding and drying, and the electrochemical data acquired during the experiment are subsequently analyzed. Using the sequential data obtained, a diagnostic model is created utilizing long-short term memory (LSTM) technique and bootstrap aggregation (bagging) ensemble method. The diagnosis rate for flooding and drying achieved 88.11%. The output value of the diagnostic model is incorporated into the reward function of the DDPG method to develop an energy-efficient driving system, considering flooding and drying of the PEMFC. The integration of fuel cell condition diagnostics into the

driving system was verified to decrease the likelihood of flooding and drying in the flow channel. The average time taken for the system to recover from these occurrences was 0.5956 seconds. Additionally, due to the avoidance of errors, the fuel consumption rate improved by approximately 1.25% when compared to the driving system without diagnosis system.

To evaluate the generality of the suggested driving system, the car model undergoes testing in diverse road conditions. As reinforcement learning relies on the Bellman equation, which updates future Q values, a change in the environment may cause a decline in optimality. Therefore, online-learning is performed to prevent performance degradation. In addition, the effect of online learning according to the convergence of offline learning is verified. The model that undergoes online learning exhibits a fuel consumption reduction of 5.59% compared to the offline model.

This study developed a novel system using a single DDPG algorithm, which simultaneously presents the target value of the optimal speed control and power distribution strategy. The reinforcement learning model was effective in reducing the occurrence of fatal defects in PEMFC, such as flooding and drying, and controlled them to quickly return to normal. The system demonstrated excellent generalization and was improved through online learning. Thus, the proposed energy-efficient driving system for a fuel cell hybrid vehicle, which considers the stability of a polymer electrolyte fuel cell, presented in this study has contributed towards the environmentally friendly development of autonomous driving, a critical focus area of the automotive industry. Additionally, the study has presented a methodology for developing a high-performance power distribution strategy.

Keyword: Energy-efficient driving, Power distribution, Fuel cell hybrid electric vehicle (FCHEV), Deep deterministic policy gradient (DDPG), Polymer electrolyte membrane fuel cells (PEMFC), Fault diagnosis

Student Number: 2017-28650

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1. Background

Global warming is a critical environmental issue that confronts humanity. The increase in the emissions of greenhouse gases, such as carbon dioxide, has led to an elevation in global temperatures, which has adversely affected the global ecosystem and human health. The temperature increase trend is evident in Figure 1, which displays the global average temperature from 1850 to 2022, indicating a steep rise in recent times. The escalation of temperatures because of climate change has caused a loss of biodiversity, alterations in water quality, coastal erosion, drought, and changes in the abundance of pests and diseases [1]. In response to these challenges, the international community has taken concrete steps to address climate change. The Paris Agreement, signed during the United Nations Climate Change Conference in 2015, aims to restrict global warming to below 2°C above pre-industrial levels and make endeavors to limit the temperature rise to 1.5°C. To this end, countries have submitted nationally determined contributions (NDCs) outlining their respective plans to mitigate greenhouse gas emissions. As of 2020, 196 countries have developed NDCs [2]. The Korean government, in the NDC presented in accordance with Article 9 of the Special Act on Climate Change Response in 2015, aims to reduce greenhouse gas emissions by 40% in comparison to 2018 levels by 2030. Germany's NDC targets a reduction of 55% by 2030, 70% by 2040, and 85% by 2050 compared to 1990 levels [3].

The transportation industry is a significant contributor to the overall greenhouse gas emissions, which is a major driver of global warming. As shown in Figure 2, greenhouse gas emissions by sector in 2020 were distributed among various sectors,

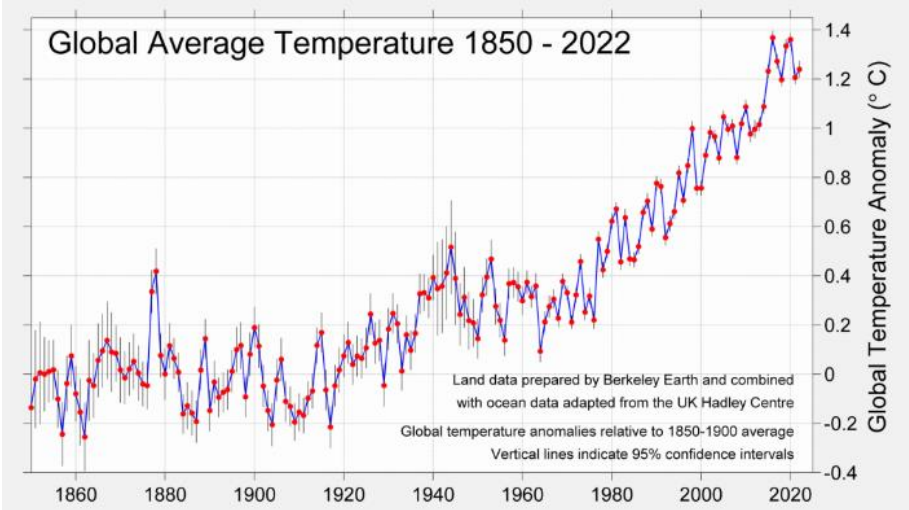


Figure 1. Global average temperature 1850-2022 [4]

with the transportation industry accounting for approximately 16.2% of total emissions. This highlights the need for the transportation sector to take a more proactive role in reducing greenhouse gas emissions. Moreover, within the transportation, the road transport sector is responsible for the majority of greenhouse gas emissions. The increasing concern over climate change and the need to reduce greenhouse gas emissions has prompted the automobile industry to develop energy-efficient driving system and eco-friendly vehicles.

1.1.1. Energy-efficient Driving System for Connected and Autonomous Vehicles

One solution to reducing greenhouse gas emissions is the development of autonomous vehicles that can operate energy efficiently. Connected and autonomous vehicles (CAVs) use advanced technologies such as sensors, GPS, and artificial

intelligence (AI) to navigate roads and make driving decisions. Figure 3 presents statistical data forecasting the size of the autonomous vehicle market by 2030, demonstrating the significant potential of the autonomous driving market.

The feasibility of realizing energy-efficient driving for autonomous vehicles through the implementation of Vehicle-to-Information (V2I) communication technology is evident [7]. Vehicle-to-Infrastructure (V2I) is a technology that enables communication between vehicles and infrastructure systems. It is a critical component of the Intelligent Transportation System (ITS) and aims to improve the safety and efficiency of transportation systems. V2I technology uses wireless

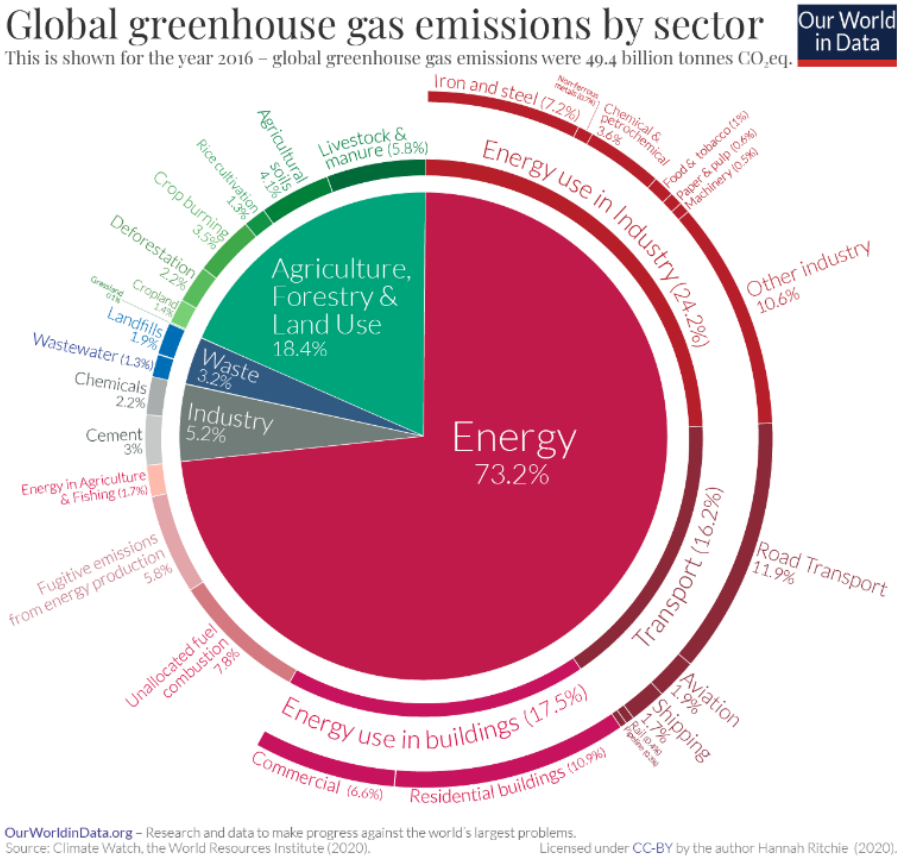


Figure 2. Global greenhouse gas emissions by sector [5]

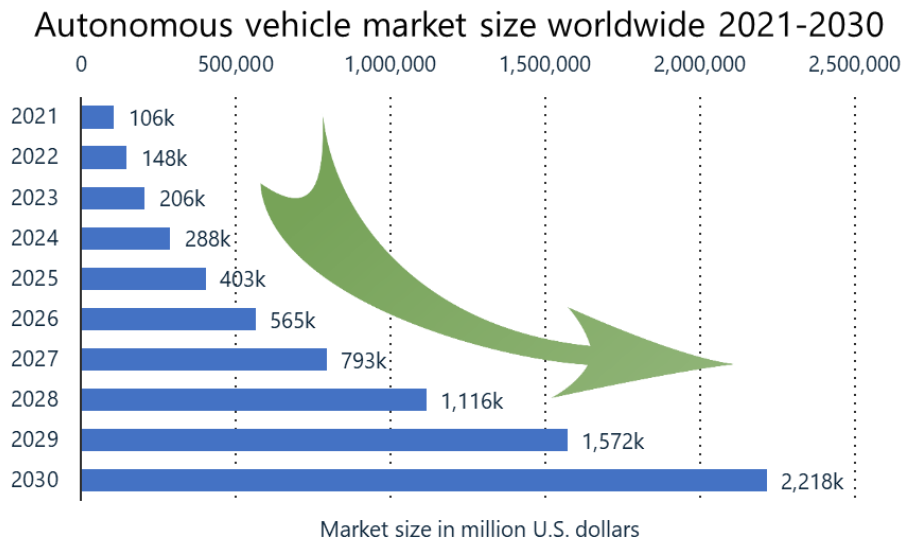


Figure 3. Global autonomous vehicle market size trends in 2021 and 2022, with a forecast through 2030 [6]

communication systems, such as Dedicated Short-Range Communications (DSRC) and Cellular Vehicle-to-Everything (C-V2X), to exchange information between vehicles and infrastructure [8]. Figure 4 shows the V2I and V2V communication. V2I technology can be used to improve the efficiency of transportation systems. By providing real-time information about the altitude and curvature of the road ahead, V2I can help drivers optimize their driving behavior and reduce fuel consumption. It can also provide information about the location of steep inclines and declines, allowing drivers to adjust their speed and acceleration to conserve fuel. This results in more efficient use of energy and reduced emissions, as the vehicles can avoid idling and take optimal routes, among other things.

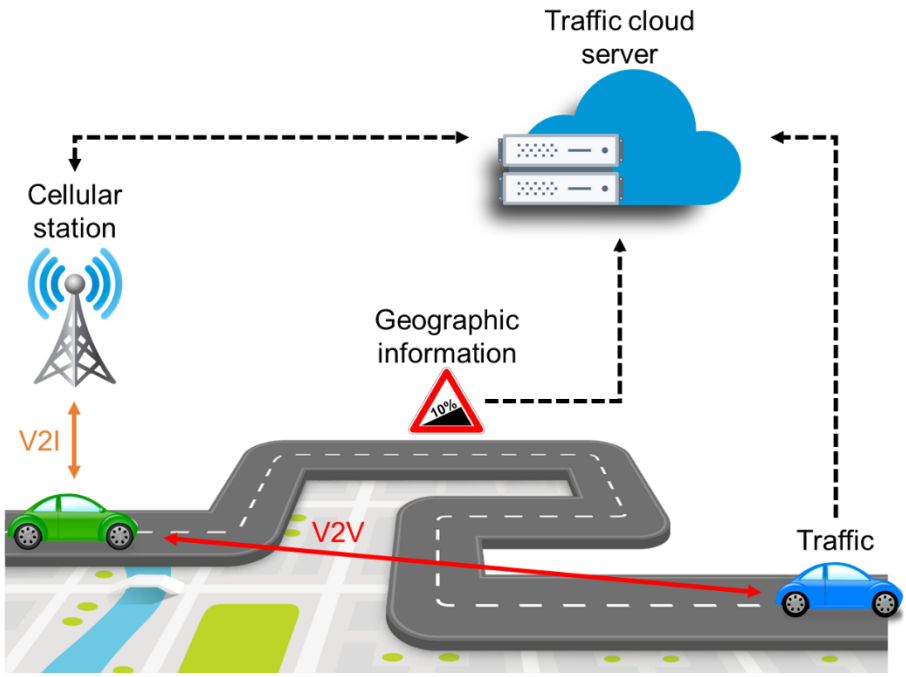


Figure 4. Vehicle-to-Infrastructure (V2I) and Vehicle-to-Vehicle (V2V) communication

1.1.2. Fuel Cell Hybrid Electric Vehicle

Fuel Cell Hybrid Electric Vehicles (FCHEVs) have garnered substantial attention in recent years as a promising technology that offers a potential solution to the challenges associated with traditional internal combustion engine (ICE) vehicles, such as air pollution and greenhouse gas emissions. FCHEVs operate by converting the energy from hydrogen fuel into electricity, which is used to power an electric motor. One of the primary advantages of FCHEVs is that they emit only water vapor as a byproduct.

The powertrain of FCHEVs consists of a fuel cell stack, an energy storage system, an electric motor, and an energy management system (EMS). The fuel cell

stack generates electricity by combining hydrogen fuel and oxygen from the air, which is facilitated by the electrochemical process in the stack. The energy storage system in FCHEVs is typically a high-capacity battery pack that can store energy generated by the fuel cell stack or recapture energy during braking or coasting. The role of the electric motor in FCHEVs is to transform the electrical energy sourced from both the fuel cell stack and battery pack into mechanical energy, which is then utilized to drive the vehicle forward. The EMS in FCHEVs manages the flow of power between the fuel cell stack, battery, and electric motor [9].

The primary objective of the EMS in an FCHEVs is to ensure that the power demands of the vehicle are met, while at the same time minimizing energy losses and maximizing the overall efficiency of the system. To achieve this goal, the EMS continually monitors the vehicle's power consumption and assesses the driving conditions, such as the road gradient, speed, and acceleration, to determine the optimal power distribution strategy. This involves ensuring that the power generated by the fuel cell stack and battery is sufficient to meet the vehicle's power demands, while also considering the energy efficiency of each component.

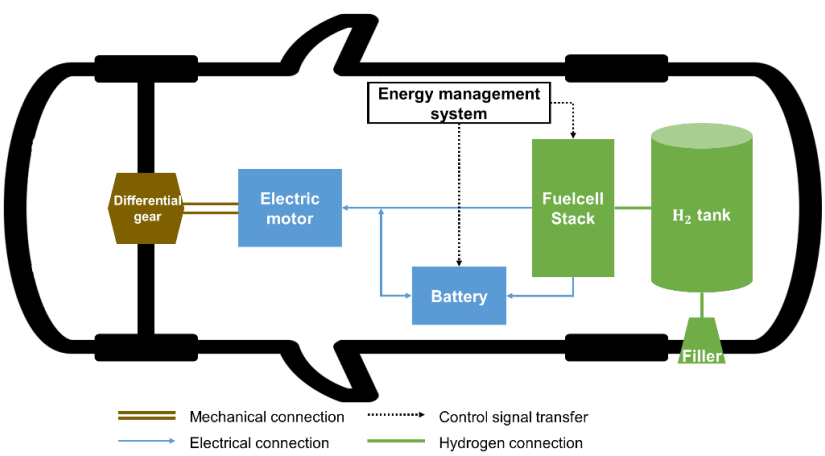


Figure 5. Structure of fuel cell hybrid electric vehicles (FCHEVs)

Given that FCHEVs operate under conditions of comparatively lower temperature, they predominantly depend on proton exchange membrane fuel cells (PEMFCs) as their primary energy source. This is owing to the ability of PEMFCs to operate effectively at lower temperatures, making them a favorable choice for powering FCHEVs. Typical PEMFCs are composed of an anode, cathode, and electrolyte membrane as illustrated in Figure 6. The chemical formula for the reaction at each electrode is as follows:

$$Anode: H_2 \rightarrow 2H^+ + 2e^- \tag{1}$$

$$\text{Cathode:} \frac{1}{2}O_2 + 2H^+ + 2e^- \rightarrow H_2O \tag{2}$$

The anode is the negative electrode and is responsible for oxidizing the hydrogen fuel to release electrons. The cathode is the positive electrode and facilitates the reduction of oxygen from the air, creating water vapor as a byproduct. The anode and cathode are divided by an electrolyte membrane, which selectively permits the passage of positively charged ions. This selective transfer generates an electric current that can be harnessed to propel the vehicle. The hydrogen fuel is then supplied to the anode side of the fuel cell stack, where it reacts with the anode catalyst to produce electrons and protons. The electrons travel along an external circuit to reach the cathode, generating an electric current that drives the vehicle's electric motor. The protons move through the electrolyte membrane to the cathode section of the fuel cell. There, they combine with oxygen from the atmosphere to create water vapor as the sole emission produced by FCHEVs.

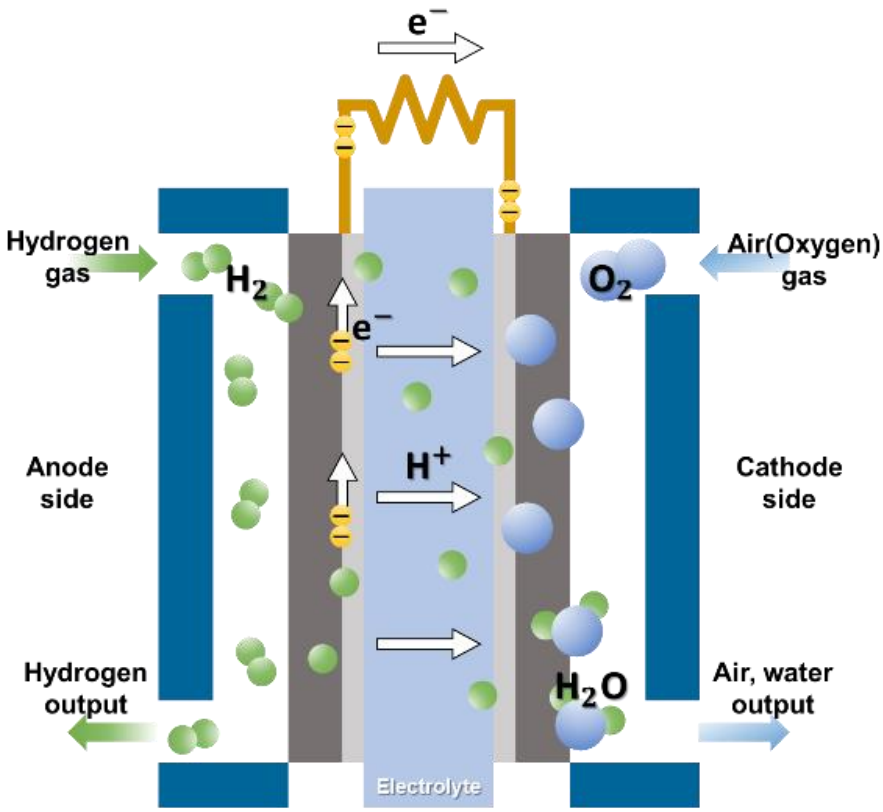


Figure 6. Schematic of typical proton exchange membrane fuel cells

1.2. Motivation

The utilization of V2I technology can enable energy-efficient driving to leverage road information for optimizing driving behavior in autonomous vehicles, leading to a reduction in fuel consumption. The environmental advantages of FCHEVs stem from their ability to generate only water during electricity production without emitting harmful gases, while an appropriate power distribution control strategy through the EMS can minimize hydrogen consumption. The combined benefits of energy-efficient driving in FCHEVs, including reduced fuel consumption

and minimized emissions, yield significant environmental benefits. However, the optimization of vehicle speed to enhance energy-efficient driving, as well as the optimization of power distribution between batteries and fuel cells, represents significant challenges that require attention. The execution of both optimizations simultaneously presents a formidable challenge due to the large computational requirements involved. Moreover, the implementation of optimal control strategies in real-world driving scenarios is complicated by the need for advanced knowledge of future conditions, as well as the long computation times required.

To overcome these obstacles, current research in this area have focused on the development of driving speed control and power distribution strategies utilizing a variety of methods such as rule-based approaches, optimization techniques, and neural networks. Wang et al. developed a system that uses road gradient information to optimize the energy consumption of connected and automated vehicles (CAVs) by generating speed profiles through model predictive control (MPC) [10]. Lee et al. presents an eco-driving control strategy based on a model-based RL method for electric vehicles. The model-free RL is utilized for evaluating driving conditions, while incorporating domain knowledge of vehicle dynamics to improve learning efficiency. [11]. The development of a power distribution control strategy for hybrid systems, such as fuel cell vehicles, has been investigated using a methodology similar to that employed in the field of energy-efficient speed control. Ahmadi et al. proposed a new power sharing method to develop an intelligent control technology for energy management of fuel cell vehicles and used a genetic algorithm to adjust the parameters of Fuzzy Logic Control [12]. Lin et al. conducted a study on the power distribution optimization of fuel cell vehicles when driving uphill, utilizing dynamic programming based on the equivalent consumption minimization strategy (ECMS) [13]. Zheng et al. developed an power distribution strategy for a fuel cell electric bus by utilizing reinforcement learning with prioritized experience replay

(PER) [14]. Zhou et al. developed a state-of-charge (SOC) regulator for a hybrid system that consists of multiple fuel cell stacks using deep deterministic policy gradient [15]. The study of methods for jointly addressing speed control for energy-efficient driving and power distribution strategies for hybrid systems primarily have employed optimization techniques. Kim et al. introduced co-optimization and sequential optimization techniques for optimal speed profiling and power management, employing Pontryagin's minimum principle (PMP) optimization method to ascertain the appropriate control actions for the vehicle, including regeneration, braking, propulsion, coasting, and cruising [16]. Nie et al. employed model predictive control to tackle the reference speed and fuel management challenges of fuel cell vehicles, incorporating battery degradation into their approach [17].

The studies mentioned above demonstrate that EMS and energy-efficient speed control methods, which prioritize fuel efficiency, make a significant contribution to reducing fuel consumption while maintaining the driving performance of the vehicle. However, simultaneously addressing the issues of an energy-efficient speed control system and the power distribution strategy of a hybrid system with an optimization algorithm presents several challenges. Firstly, although the optimal solution can demonstrate the potential of the system, its real-time applicability is limited due to the extended time required for calculations. Secondly, using a discrete action can lead to errors in presenting a solution, as it can result in unrealistic simulations of vehicle speed control and battery power output. In addition, the instability of the PEMFC, which is the primary power source for FCHEVs, can significantly affect the performance of the vehicle. Failure to consider PEMFC instability can lead to large decreases in performance. Common PEMFC issues include flooding and drying. Flooding is a common issue in PEMFCs that occurs at high current densities, and it reduces the efficiency of fuel usage by blocking the gas diffusion path and

catalyst layer. The flooding phenomenon leads to an accumulation of excess water in the membrane electrode assembly (MEA), which restricts the diffusion of reactants and inhibits their access to active sites. As a result, the performance of the PEMFC deteriorates, and the fuel consumption increases. The excess water in the MEA also creates a back pressure that restricts the removal of product water from the cell. This back pressure further exacerbates the flooding problem, as it causes more water to accumulate in the cell. Moreover, the accumulation of excess water in the cell can also lead to a drop-in cell voltage, which further reduces the performance of the PEMFC [18-21]. Conversely, when the PEMFC experiences dehydration, the membrane's water content diminishes, leading to a considerable decrease in the PEMFC's electrical conductivity. This increase in ohmic resistance can limit the power output and reduce the cell efficiency. Drying also affects the mechanical properties of the membrane, leading to cracks, delamination, and a reduction in its durability. Additionally, drying can cause the membrane to shrink, leading to a loss of contact between the catalyst layer and the membrane, which can result in a further reduction in the performance of the PEMFC [22, 23]. Therefore, it is crucial to develop effective strategies to mitigate flooding and drying in PEMFC to ensure their stable and efficient operation. In this regard, power control strategies for the electronic control unit (ECU) of PEMFCs have been identified as an effective approach to avoid flooding and drying. By adjusting the required power to the PEMFC, the water content of the cell can be optimized to prevent both flooding and drying [24, 25]. By lowering the required power for PEMFC through vehicle speed control and battery assist, the water content of the cell can be optimized to prevent both flooding and drying.

As previously noted, there are challenges to be addressed in deploying an energy-efficient driving assistance system for FCHEVs. This thesis presents a model that provides concurrent references for vehicle driving speed and power distribution,

which is capable of real-time control and yields continuous action. Additionally, the model is designed to consider errors associated with moisture in PEMFCs, with the aim of minimizing performance degradation during driving.

1.3. Thesis Outlines

This thesis proposes the use of deep deterministic policy gradient (DDPG), a variant of deep Q-network (DQN), to address the aforementioned issues related to energy-efficient driving for FCHEVs. Specifically, the DDPG model is utilized to simultaneously optimize energy-efficient speed control, as well as power distribution between the battery and fuel cell. The reinforcement learning model of the DDPG framework is trained to present the driving speed and the distribution of battery power, based on the state of the vehicle and road geographical information obtained from V2I communication. Additionally, a diagnostic model is developed to detect flooding and drying in PEMFC, which can then be incorporated into an energy-efficient driving assistance system to prevent performance degradation. Chapter 2 of this thesis introduces a model for FCHEVs that is based on actual PEMFC experiments. The developed vehicle model is then validated. In Chapter 3, a driving control and power distribution model for a fuel cell hybrid vehicle is trained using DDPG. Chapter 4 analyzes the influence of PEMFC flooding and drying and develops a model that considers performance degradation. Chapter 5 evaluates the generality of the proposed driving assistance system and assesses its performance with and without online learning. Finally, Chapter 6 summarizes the conclusions of this thesis and outlines future plans for research in this area.

- Chapter 2 presents the modeling of FCHEVs with a parallel hybrid system that utilizes a PEMFC as the main power source. The power of

each component is calculated based on the vehicle's required acceleration and the efficiency of each component. The PEMFC and battery are designed to adequately meet the required power based on the specific hybrid system. To ensure the accuracy of the modeling, experiments are conducted under various conditions. The experimental data collected includes the temperature of the cell and the relative humidity of each electrode. Based on this data, models of the PEMFC are developed using various coefficients and map data. To validate the accuracy of the developed PEMFC modeling, the outcomes are compared to the experimental results.

- Chapter 3 describes the development of controllers for energy-efficient driving and power distribution in FCHEVs. The controllers are learned using a deep reinforcement learning algorithm called DDPG, which is based on the vehicle state and geographical information of the road. The FCHEVs model developed in Chapter 2 is driven on test roads with varying grades to investigate the optimal driving speed for minimizing fuel consumption, and to determine the optimal power distribution between the fuel cell and energy storage system. A reward function is designed to include the fuel consumption, driving time, and SOC reference, and the control model is trained through continued driving to obtain the optimal reward. The trained model's performance is evaluated against the optimal results and typical control methods to assess its effectiveness. This chapter provides a detailed account of the DDPG algorithm and the methodology used to develop the controllers for energy-efficient driving and power distribution in FCHEVs.

- In Chapter 4, a driving control and power distribution system is developed for FCHEVs that considers the issues of flooding and drying, which are known to be critical defects in PEMFCs. Experiments on PEMFC cells under extreme conditions conducted, deliberately causing flooding and drying, in order to investigate their effects on the cells' performance. Through the analysis of the voltage pattern during flooding and drying, an approximate model for performance degradation is developed. Additionally, a system for diagnosing flooding and drying is established using the Long Short-Term Memory (LSTM) and bootstrap aggregation (bagging) method, based on time-series experimental data. By integrating the flooding and drying diagnosis model with the DDPG learning process, an optimal driving technique and energy management system strategy that considers flooding and drying are established.

- Chapter 5 focuses on evaluating the generalizability of the system developed in the preceding chapters, as well as investigating the performance of a DDPG-based online learning system for driving control and power distribution. The objective of this chapter is to evaluate the models in real-world road environment and determine their capacity to adjust to different driving conditions. The existing model is compared and analyzed with the model incorporating online learning.

- In Chapter 6, the summary of the results obtained from the development of energy-efficient driving assistance systems for FCHEVs is presented. The main contributions of each of the previous chapters are reviewed and their significance in advancing the practical

implementation of FCHEVs is discussed. The limitations of the research are acknowledged and areas for future exploration are suggested, particularly in light of the expanding availability of information and technology. The chapter concludes with a discussion of the potential implications of the work for the broader fields of transportation and sustainable energy.

Chapter 2. Modeling of Fuel Cell Hybrid Vehicles based on PEMFC Experiments

2.1. Introduction

In this chapter, modeling of FCHEVs, which are target vehicles of energy-efficient driving assistance system, is performed based on experiments conducted on a PEMFC, which is a popular choice for FCHEVs power sources due to its high efficiency, low operating temperature, and fast response. The FCHEVs structure considered in this study is composed of a parallel hybrid system, as depicted in Figure 7, where the fuel cell stack and battery are responsible for outputting power. This hybrid system is a common configuration for FCHEVs and helps improve their overall efficiency by utilizing both the fuel cell stack and battery as power sources. The efficiency of each component, including the fuel cell stack, battery, and motor,

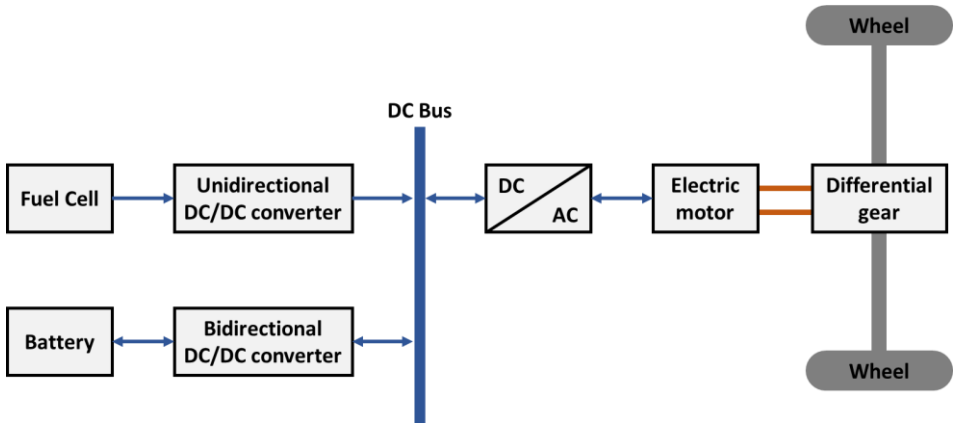


Figure 7. Powertrain structure of an FCHEVs

is considered in the calculation of the vehicle's traction power based on vehicle dynamics. The specifications of the vehicle, such as its weight, dimensions, and maximum power output, are presented in Table 1. To develop the vehicle model, the data used was referenced from a commercial program called Autonomie, which was developed at the Argonne National Laboratory and is widely used for FCHEVs modeling.

The experimental setup used for the PEMFC is thoroughly described. This includes a detailed account of the fuel cell, as well as the hydrogen and air supply systems. In order to evaluate the performance of the PEMFC, various test protocols were utilized, including polarization curves and power density measurements. These tests were conducted to measure the output power of the fuel cell under various conditions. Based on the experimental data, a mathematical model was created to simulate the operation of a PEMFC. The model considers the complex

Table 1. Specifications of the FCHEVs model

Parameter	Value
Fuel Cell stack power	50 kW
Battery capacity	6.5 Ah
Electric motor power	60 kW
Vehicle mass	1500 kg
Final differential gear	Ratio: 10 / Efficiency: 98 %
Converter efficiency	95 %
Tire radius	0.305 m
Front area	1.8 m^2
Air drag coefficient	0.29
Rolling resistance coefficient	0.007

electrochemical reactions, as well as the transfer of mass, momentum, and heat. It is validated against experimental data. Overall, the chapter provides a comprehensive overview of the experimental and modeling approaches used to investigate the performance of a PEMFC-based FCHVs.

2.2. Design and experiments of PEMFC

PEMFC, the main power source of FCHEVs, is composed of components such as stack, channel of each electrode, and humidifier. In order to develop and validate a mathematical model for each component, a single cell is fabricated, and experiments are conducted to collect data. The single cell is illustrated in Figure 8 and consists of an endplate, a current collector with channel, gaskets, gas diffusion layers (GDLs), an MEA, and a bipolar plate. The specifications of the PEMFC single

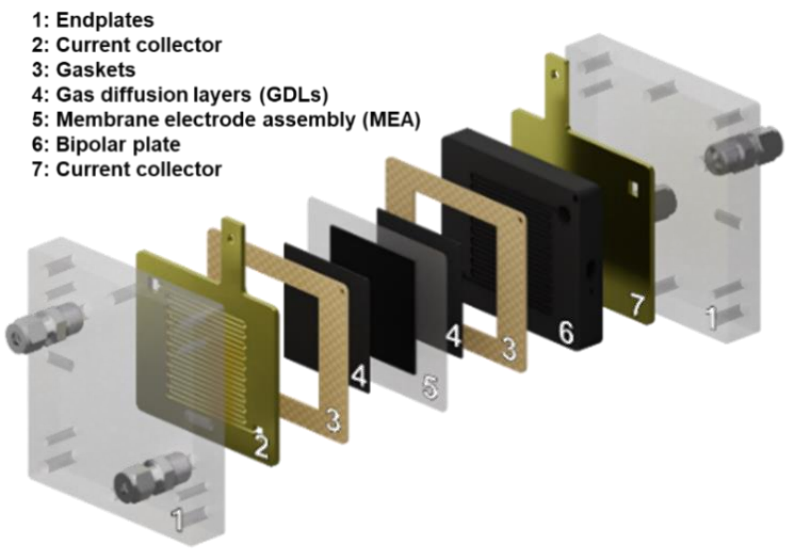


Figure 8. Schematic of a single PEMFC

cell are outlined in Table 2. The experimental setup, depicted in Figure 9, includes fuel and air processing systems, temperature sensors in the fuel cells and inlet gas pipes, and a potentiostat for measuring voltage and impedance modulus. During the experiments, the PEMFC is operated using this setup, and data are recorded and analyzed. Experimental investigations are carried out, involving the manipulation of three independent variables, namely, the relative humidity within the range of 50% to 120%, the temperature of the cell spanning from 40 to 80 degrees, and the current density ranging from 0 to 0.8. The data collected include cell temperature, humidity of the anode and cathode, cell voltage, impedance modulus, and current density, and were gathered over a 100-hour period. The data obtained from these experiments are analyzed and utilized to develop mathematical models for each component of PEMFCs, which are subsequently validated through further model driving test.

Table 2. Specifications of PEMFC for experiments

Specifications of a single PEMFC	
Endplates	Hard-anodized aluminum
Current collector	Gold-plated stainless steel
Flow-field plate	Graphite 51-channel parallel channel (channel width: 0.3 mm)
Gas diffusion layer	SGL® 36BB (PTFE treated)
Membrane	Nafion® 211
Catalyst loading amount (mg/cm^2)	Anode 0.12 / Cathode 0.12
Active area (cm^2)	9.0

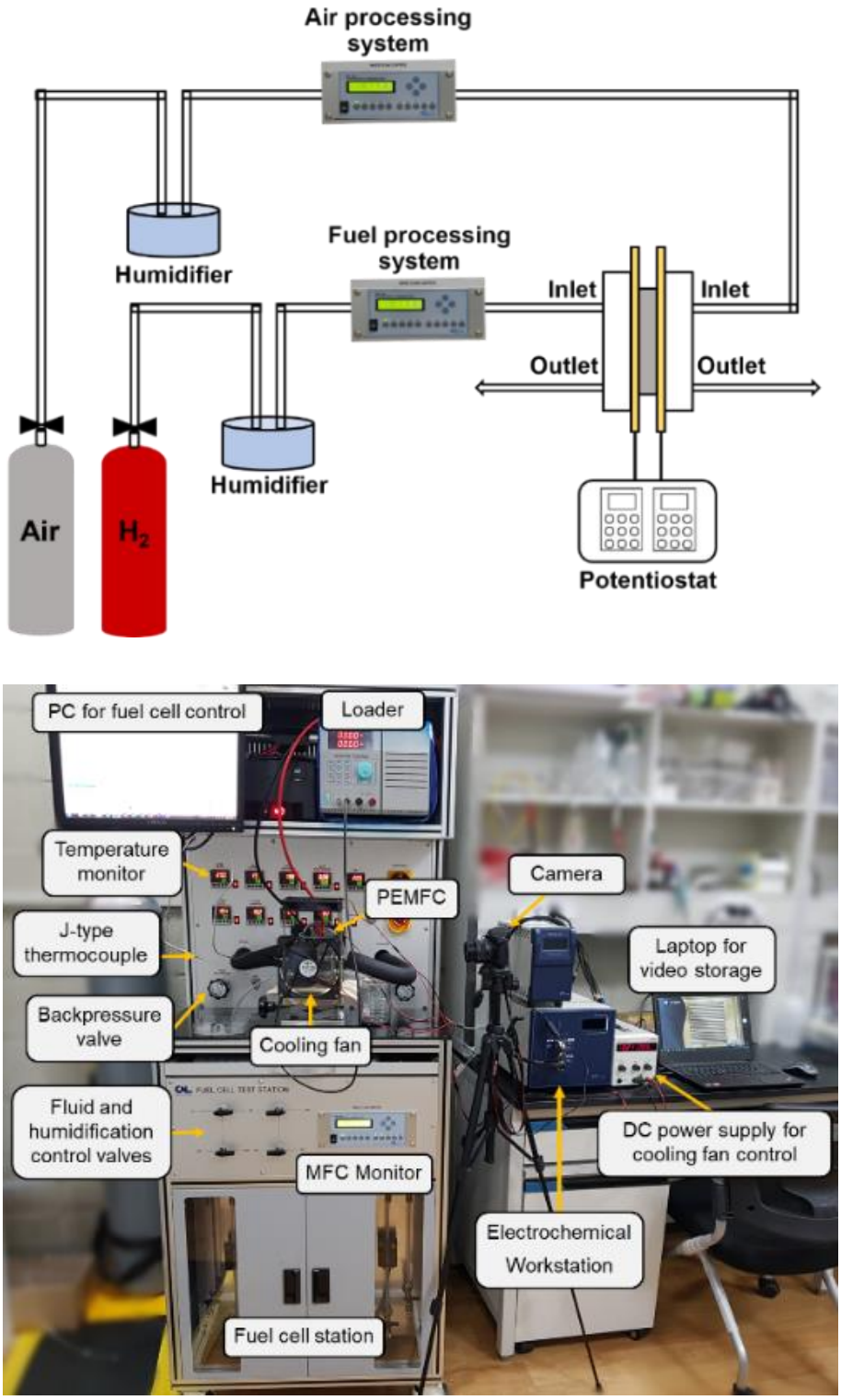


Figure 9. Configuration of experimental apparatus for PEMFC operations

2.3. FCHEVs Modeling

This thesis describes the development of a backward-looking vehicle simulator. In a forward-looking simulator, the driver model issues commands for the accelerator and brake pedals to follow a predefined target speed profile, and the vehicle's components respond accordingly. In contrast, the backward-looking simulator analyzes the operation of each vehicle component to determine how they function together. Once the required speed of the vehicle is determined, it is calculated backward from the wheels to the power source without using a driver model. The backward-looking simulator is commonly used for developing control algorithms or evaluating the fuel economy performance of the vehicle, as it is more time-efficient compared to the forward-looking simulator [26]. The adoption of quasi-steady models in the backward-looking simulator results in the omission of transient dynamics in the vehicle powertrain. To validate the proposed control strategy, the driving assistance system and power distribution strategy are evaluated in the backward-looking vehicle simulator.

Additionally, the PEMFC semi-empirical modeling is performed based on experimental data. The experimental data is utilized to validate the fuel cell model, and determine the coefficient of the model. Semi-empirical models typically include simplified versions of the physical and chemical processes occurring within the fuel cell, resulting in reduced computational complexity compared to mechanistic models [27].

2.3.1. Powertrain Modeling

Based on the mathematical model of longitudinal vehicle dynamics, it is calculated from the vehicle's required speed to the main power source [28]. The wheel torque T_{wheel}, calculated by vehicle acceleration $\dot{v}$, is given by the following equation.

$$T_{wheel} = \frac{m_{veh}\dot{v} + F_{loss}}{R_{tire}} \tag{3}$$

$$F_{loss} = \frac{1}{2}C_d A_f \rho_{air} v^2 + \mu_{roll} m_{veh} g cos\theta + m_{veh} g sin\theta \tag{4}$$

where m_{veh} is the total weight of the vehicle, and F_{loss} is the resistance loss occurring while the vehicle is driving. The first term of F_{loss} represents the aerodynamic resistance, which is dependent on the aerodynamic drag coefficient C_d, the frontal area of the vehicle A_f, and the density of the surrounding air ρ_{air}. The second term represents the rolling resistance, which is determined by the rolling resistance coefficient μ_{roll}, the gravity g, and the road gradient θ. This term accounts for the resistance encountered by the vehicle due to the deformation of the tires and the road surface as well as the gravitational force acting on the vehicle. The third term represents the gradient resistance encountered when driving a vehicle uphill or on an incline.

The required power equation for PEMFC and batteries is determined based on the motor's torque and angular velocity, which is calculated as follows:

$$P_{req} = \frac{\eta_{mot}^{-sgn(T_{mot})} T_{mot} \omega_{mot}}{\eta_{inv}} \tag{5}$$

$$P_{req} = \left(\eta_{dc}^{sgn(P_{bat})} P_{bat} + \eta_{dc} P_{fc}\right) \tag{6}$$

where $T_{mot}(= \eta_f T_{wheel}/r_f)$ is the motor torque, which is calculated by dividing the wheel torque by the final differential fear ratio r_f, $\omega_{\mathrm{mot}}(= r_f v/\mathrm{R_{tire}})$ is the angular speed of the motor, η_{inv} is the efficiency of the AC/DC inverter, and η_{mot} is the efficiency of the motor, which is defined as a function of motor speed and torque, as shown in Figure 10. The power requirements are satisfied by distributing power in a suitable manner between the battery and the PEMFC stack.

Equivalent circuit models are frequently utilized in battery modeling to represent the charging and discharging dynamics since they are capable of expressing internal variations using specific electrical equations and possess clear physical interpretations. This study employs a first-order resistance model, which comprises an internal resistance and an open circuit voltage source, as illustrated in Figure 11 [29, 30]. The dynamics for SOC, which is derived from battery power, can be expressed as follows:

$$S\dot{O}C = -\frac{1}{Q_{bat}} \frac{\left(V_{oc} - \sqrt{V_{oc} - 4P_{bat}R_{int}}\right)}{2R_{int}} \tag{7}$$

where Q_{bat} is the battery capacitance, V_{oc} and R_{int} are open-circuit voltage and internal resistance respectively, which vary according to SOC based on the map depicted in Figure 12.

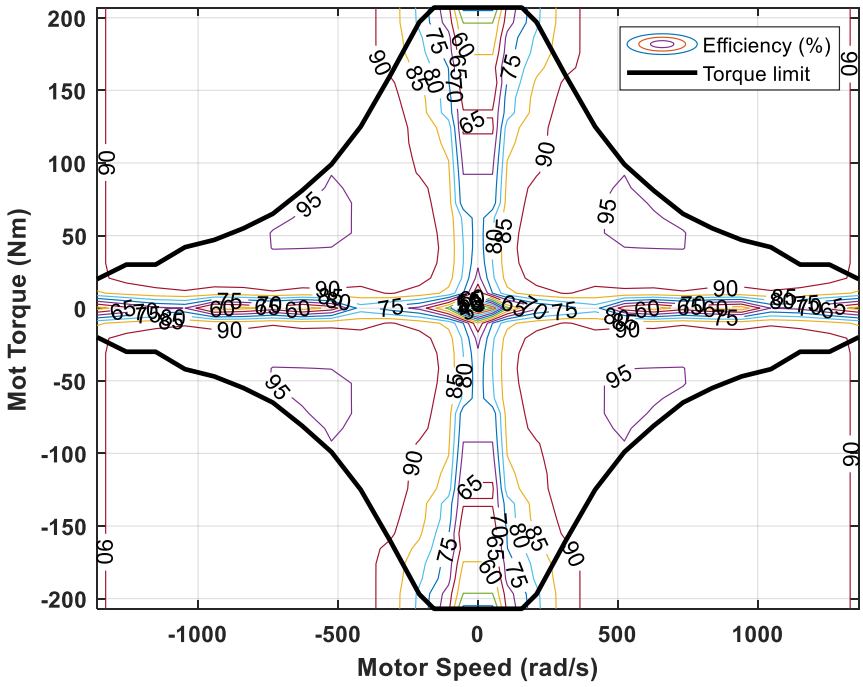


Figure 10. Efficiency and torque limit of the electric motor

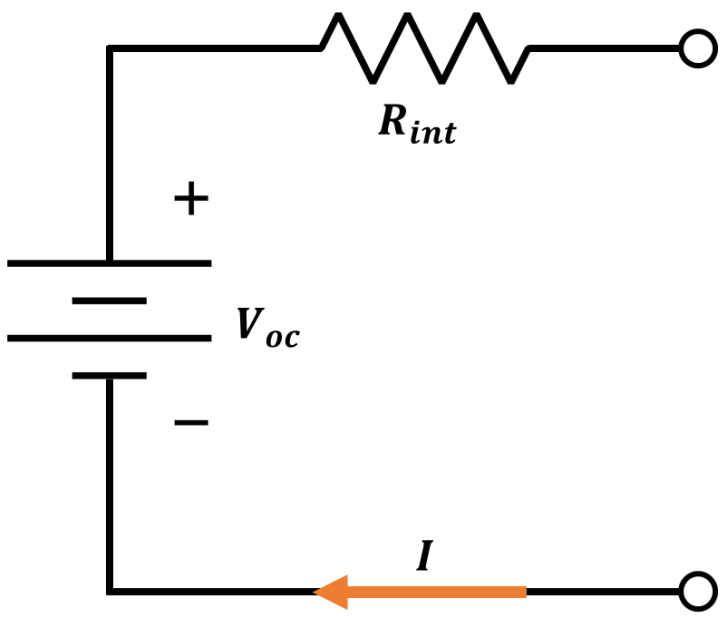


Figure 11. Equivalent circuit diagram of battery

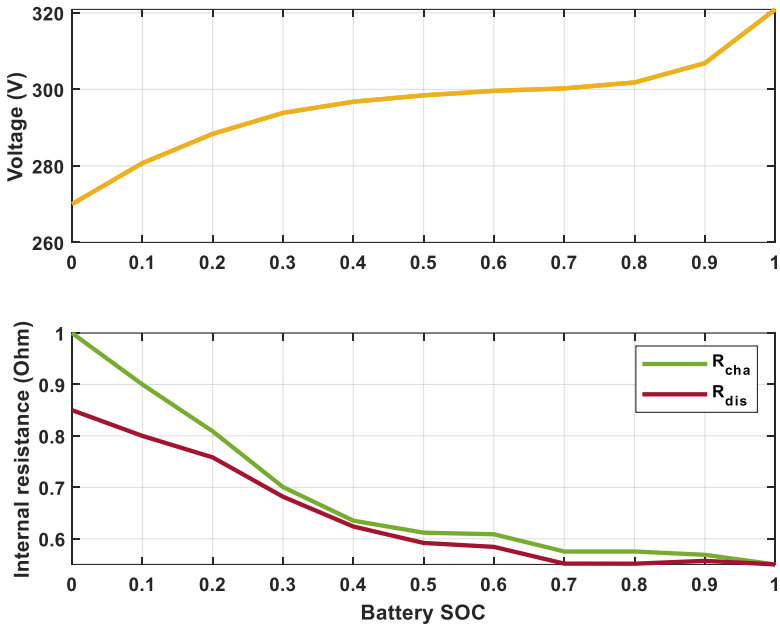


Figure 12. open-circuit voltage and internal resistance according to SOC

2.3.2. Semi-empirical Model of PEMFC

Practical implementation of the PEMFC is limited by several factors, including the lack of accurate models for predicting its performance. Among the available models, the semi-empirical approach is widely used due to its simplicity and accuracy [31-33]. In this model, the PEMFC is treated as a set of interconnected processes, and the overall performance is calculated by combining experimental data and theoretical calculations. Table 3 displays the parameters utilized in the model. The fuel cell performance model relates the output power of the fuel cell to the operating conditions, such as the pressure, flow rate, and temperature of the reactants. This model is based on empirical data obtained from experimental measurements. The semi-empirical model of PEMFC predicts the performance of the fuel cell under

Table 3. Parameters of PEMFC model

Properties	Values
Number of cells	400
Active area (cm^2)	270
Target temperature (K)	352.983
Membrane thickness (μm)	125
Gas channel width (m)	0.01
Number of channels	8
Density of dry membrane (kg/cm^3)	0.002
Equivalent weight of dry membrane (kg/mol)	1.1
Gas diffusion layer thickness (μm)	250
Overall specific heat of MEA (J/kgK)	870

various operating conditions. This is done by fitting the experimental data to the theoretical model and adjusting the parameters to match the experimental results.

The voltage of a PEMFC can be calculated by considering the Nernst equation and the various sources of voltage losses [34]. The Nernst equation relates the cell voltage of an electrochemical reaction to the standard potential, the activities of the reactants and products, and the quantity of electrons involved in the reaction. For the PEMFC, the Nernst equation can be written as:

$$E_{nernst} = E^0 - \frac{RT}{nF} ln \frac{p_{H_2O}}{p_{H_2} p_{O_2}^{1/2}} \tag{8}$$

where E^0 is the standard-state reversible voltage, R is the gas constant, T is the stack temperature, n is the number of electrons transferred in the reaction, F

is the Faraday constant, p_{H_2O} is the water partial pressure, p_{H_2} is the hydrogen partial pressure, and p_{O_2} is the oxygen partial pressure.

The voltage loss can be categorized into three main types: activation loss, ohmic loss, and concentration loss. Activation loss is caused by the kinetic limitations of the electrochemical reactions taking place at the electrodes. Various factors like temperature, catalyst activity, and reactant concentration can affect the activation overpotential needed to initiate the electrochemical reactions. The activation loss can be expressed as:

$$\eta_{act} = \frac{RT}{\alpha nF} ln \frac{j_{cell}}{j_0} \tag{9}$$

where j_{cell} is the current density of the cell, and j_{leak} represents the parasitic losses resulting from various sources, such as current leakage, gas crossover, and undesired side reactions. α and j_0 are the transfer coefficient and exchange current density respectively, which can be determined by Tafel equation. Specifically, the Tafel slope and the intercept of the Tafel equation obtained from experimental data, as shown in Figure 13, can be used to calculate α and j_0.

Ohmic loss, also known as the internal resistance of the fuel cell, is the voltage drop that occurs across the membrane and the electrodes due to the resistance of the materials and interfaces. Several factors contribute to the magnitude of the ohmic loss, such as the membrane's thickness and conductivity, the electrodes' surface area and thickness, and the contact resistance between the electrodes and the membrane. The ohmic loss is dependent on the conductivity of the Nafion membrane, which in turn is influenced by its water content [35, 36]. The ohmic loss can be quantified using the following equation:

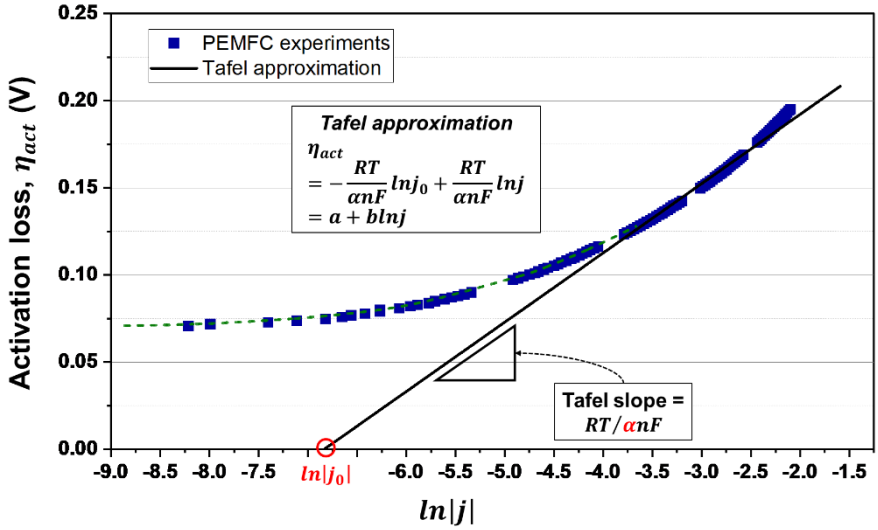


Figure 13. Tafel approximation based on PEMFC experiments

$$\lambda_{an,ca} = \begin{cases} 0.043 + 17.81\varphi_{an,ca} - 39.85\varphi_{an,ca}^2 + 36.0\varphi_{an,ca}^3; & \text{for } 0 < \varphi_{an,ca} \leq 1 \\ 14 + 1.4(\varphi_{an,ca} - 1) & \text{for } 1 < \varphi_{an,ca} \leq 3 \end{cases} \quad (10)$$

$$\sigma_{303K}(\lambda_{an,ca}) = 0.005193\lambda_{an,ca} - 0.00326 \quad (11)$$

$$\sigma(T,\lambda) = \sigma_{303K}(\lambda_{an,ca}) \exp\left[1268\left(\frac{1}{303} - \frac{1}{T}\right)\right] \quad (12)$$

$$ASR_{ohmic} = t_m/\sigma(T,\lambda) \quad (13)$$

$$\eta_{ohmic} = j_{cell} ASR_{ohmic} \quad (14)$$

where $\lambda_{an,ca}$ is the water content, $\varphi_{an,ca}$ refers to the relative humidity, σ indicates the conductivity of the membrane, t_m represents the membrane thickness, and ASR_{ohmic} is the area-specific resistance.

Concentration loss is a common phenomenon that occurs in PEMFC due to limitations in mass transport. It is caused by the depletion of reactants near the catalyst layer, leading to reduced electrochemical reaction rates. Various factors can influence this form of loss, including the rate of reactant flow, the thickness and porosity of the electrode and membrane layers, and the solubility of the reactants in the electrolyte. [37]. The reactants, hydrogen, and oxygen, are consumed at the cathode and anode, respectively. The depletion of reactants in the reaction zone causes a concentration gradient that reduces the rate of the reaction. The concentration loss is proportional to the current density, and it can be expressed by the following equation:

$$\eta_{conc} = \left(\frac{RT}{nF}\right)\left(1 + \frac{1}{\alpha}\right) ln \frac{j_L}{j_L - j_{cell}} \tag{15}$$

where j_L represents the limiting current density.

Finally, the cell voltage of a PEMFC that considers activation loss, ohmic loss, and concentration loss is calculated as follows:

$$V_{cell} = E_{nersnt} - \eta_{act} - \eta_{ohmic} - \eta_{conc} \tag{16}$$

The polarization curve of the PEMFC model corresponding to the test run of FCHEVs is presented in Figure 14. The operation was mainly carried out at low current densities, and the curve exhibits a similar field pattern to the experimental data. Moreover, the power density as a function of current density shows analogous results to those obtained in the experiments.

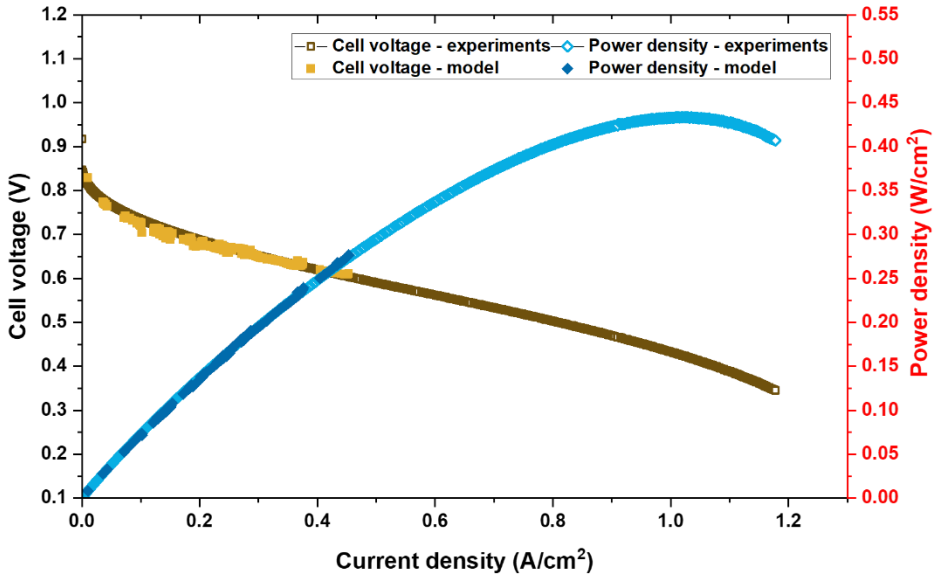


Figure 14. Polarization curve of the semi-empirical PEMFC model

Chapter 3. Energy-efficient Driving Considering Power Distribution of FCHEVs using DDPG.

3.1. Introduction

This chapter outlines the development of an energy-efficient driving assistance system that utilizes the vehicle model developed in the previous chapter. The objective of the system is to provide speed scenarios that minimize fuel consumption while the vehicle is driving on a test road with varying gradients. Furthermore, the system provides the power distribution strategy for the FCHEVs based on the required power for driving. The system considers both the internal state of the vehicle and external geographic information. To develop this system, the DDPG method, which is a type of actor-critic algorithm, is employed.

DDPG is a kind of reinforcement learning algorithm suitable for training policies in scenarios with continuous action spaces. It is a combination of deep learning and policy gradient methods, where the policy is learned directly from the state and action space of an environment. The DDPG algorithm comprises two neural networks: an actor and a critic network. The actor network receives the current state of the environment as input and produces the best action for the agent to execute. The critic network, on the other hand, evaluates the action taken by the actor network by estimating the Q-value of the current state and action [38].

The actor-critic approach has several advantages over other RL algorithms like Q-learning, in which the agent learns the optimal policy by estimating the optimal Q-value function. In contrast, the actor-critic approach can learn a policy directly without estimating the Q-values. This approach is especially useful when the action

space is continuous since estimating the Q-value for each possible action becomes impractical. Additionally, the actor-critic approach enables the agent to learn from its own experiences and enhance its performance over time by adjusting the policy and Q-value estimation. This makes DDPG a popular choice for developing complex control systems, such as autonomous driving, robotics, and game playing [39-41].

One key feature of DDPG is the use of a replay buffer, which stores the agent's experiences in a memory buffer. This buffer is then sampled randomly to train the neural networks, ensuring that the agent learns from a diverse set of experiences. An additional significant aspect involves the employment of target networks. These networks are duplicates of the actor and critic networks, serving to produce target Q-values and target actions throughout the training process. By using target networks, the agent's estimates of the Q-value and policy become more stable and less prone to oscillations [42]. DDPG is a model-free reinforcement learning method that combines the strengths of DQN, a type of value iteration, and policy iteration. A taxonomy of reinforcement learning algorithms is illustrated in Figure 15.

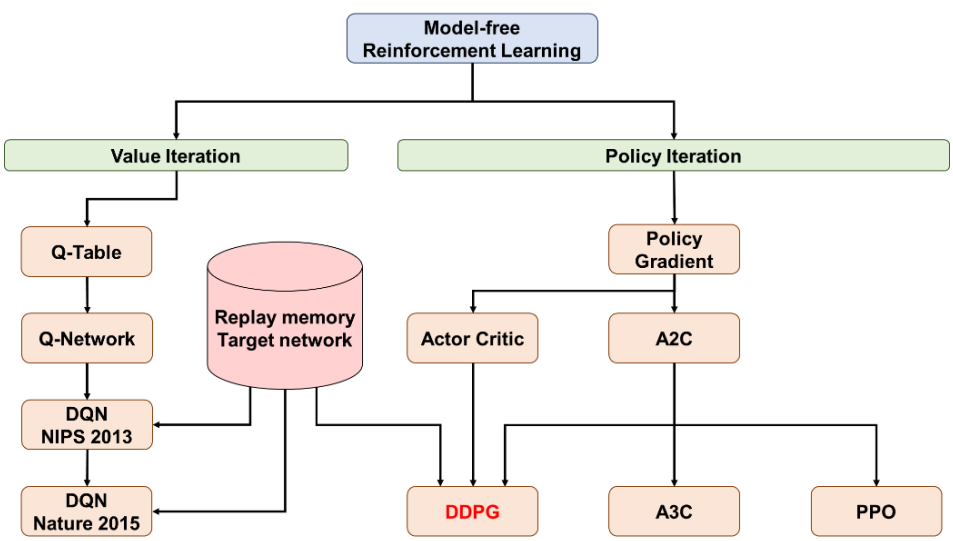


Figure 15. Classification for model-free reinforcement learning.

3.2. Learning Process for DDPG Agents

In this section, a suitable reward function is proposed for DDPG to enable energy-efficient driving and power distribution of FCHEVs. Due to constraints on vehicle specifications such as fuel cell power, battery power, and motor power, as well as steep slopes of roads, the range of actions for the vehicle is inevitably restricted. Consequently, the action space is updated at each step based on the vehicle's state and location.

3.2.1. DDPG Algorithm Principle

In this study, a distance-based approach is employed in the context of vehicle driving scenarios. The reason for this is that it is relatively straightforward to obtain topographical data regarding the road based on the vehicle's location, and there is minimal room for error. Within this framework, the DDPG algorithm is utilized to construct a system that generates actions for actions at each incremental distance step.

The DDPG algorithm framework implemented for driving FCHEVs is illustrated in Figure 16. This experience replay buffer, denoted as $\mathcal{R}$, represents a collection of prior experiences. During the operation of the vehicle model developed in the previous chapter, the internal and external states s of the vehicle, actions a, the next state s' after performing the action in the current state, the associated reward r, and an indicator variable, d, denoting whether the state is terminal are all stored in a replay memory. Subsequently, data is extracted from this replay buffer and employed for training the deep neural network. The replay buffer accumulates

data as the driving progresses step by step, and the oldest data is replaced with the most recent data. As there exists a potential risk of overfitting if only recent data is used for training, data samples for learning are randomly selected from the replay memory.

The neural networks involved in the DDPG algorithm are trained by extracting mini-batches of data from the replay memory and updating the parameters of both the actor and critic networks [43]. The distinguishing feature of DDPG, in contrast to other policy-based techniques, is that it uses a deterministic policy. Specifically, the actor function accepts the state as input and outputs a policy π whose probability distribution function corresponds to the Dirac delta function of a particular value. The deterministic target policy allows the Q-value, the action-value function, to be expressed by the Bellman equation as:

$$\pi: S \rightarrow p(a_t|s_t) = \delta(a_t - a_t^*) \tag{17}$$

$$Q_\phi(s_t, a_t) = \mathbb{E}_{r_t, s_{t+1} \sim E}\left[r(s_t, a_t) + \gamma Q_\phi\left(s_{t+1}, \mu_\theta(s_{t+1})\right)\right] \tag{18}$$

where S is the state space, a_t^* is the target action, γ is the discounting factor, and Q_ϕ is the Q-value generated by the critic networks, which serves to evaluate the values of states and actions. The actor networks are trained using the policy gradient approach. The gradient of the objective function J with respect to the parameter θ of the actor networks is given by the following expression:

$$\begin{aligned} \nabla_\theta J &\approx \mathbb{E}_{s_t \sim \rho^\beta}\left[\nabla_\theta Q(s, a|\phi)|_{s=s_t, a=\mu(s_t)}\right] \\ &= \mathbb{E}_{s_t \sim \rho^\beta}\left[\nabla_a Q(s, a|\phi)|_{s=s_t, a=\mu(s_t)} \nabla_\theta \mu(s|\theta)|_{s=s_t}\right] \end{aligned} \tag{19}$$

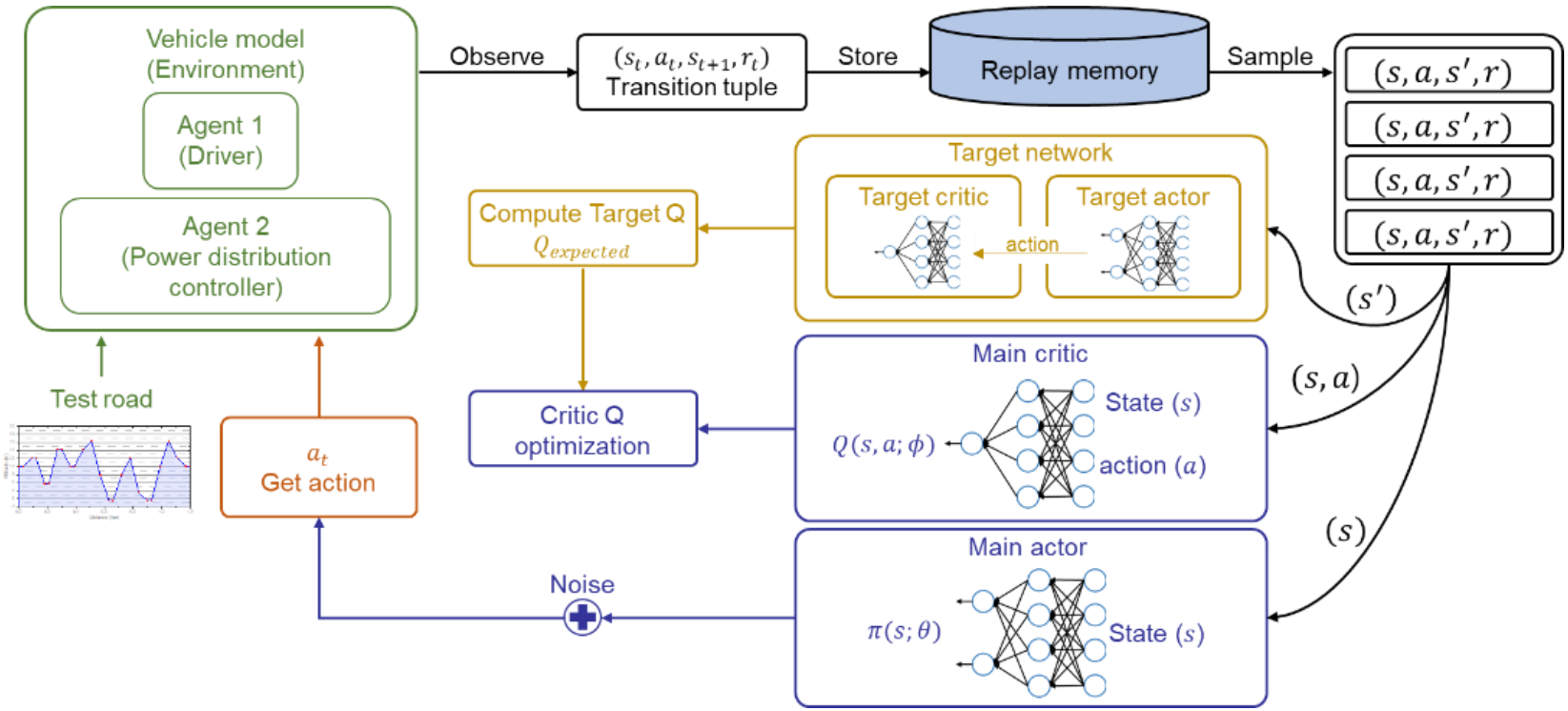


Figure 16. Framework for energy-efficient driving and power distribution of FCHEVs utilizing the DDPG methodology

Critic networks are action-value functions composed of parameters ϕ. They are trained using stochastic gradient descent to minimize the loss function of Mean Squared Bellman Error (MSBE), as expressed by the following equation:

$$L(\phi, \mathcal{R}) = \mathbb{E}_{(s,a,r,s',d)\sim\mathcal{R}}\left[\left(Q_\phi(s_t, a_t) - y_t\right)^2\right] \tag{20}$$

$$y_t = r(s_t, a_t) + \gamma(1-d)Q_{\phi_{targ}}\left(s_{t+1}, \mu_{\theta_{targ}}(s_{t+1})\right) \tag{21}$$

The DDPG algorithm is an off-policy algorithm and therefore utilizes target actor networks and target critic networks. At each step of the algorithm, the Q-networks are updated and their parameters are modified. However, this can lead to convergence issues since the parameters of the updated Q-networks are employed when calculating the target value. In order to address the convergence issues caused by the use of the parameters of the updated Q-networks when calculating the target value, target networks with parameters copied from the actor and critic networks are used instead to calculate the target value. To update the weights of the target networks, a soft target update is used. Specifically, the soft target update is represented as follows:

$$\phi_{targ} \leftarrow \tau\phi + (1-\tau)\phi_{targ} \tag{22}$$

$$\theta_{targ} \leftarrow \tau\theta + (1-\tau)\theta_{targ} \tag{23}$$

where τ regulates the rate at which the weights of the target networks are updated. Exploration is a crucial aspect of the learning process as it allows the agent to encounter diverse states and derive an optimal solution. To facilitate exploration

in the continuous action space of DDPG, random noise is added to the actor policy. The policy of the actor with noise can be represented as:

$$a_t = \mu_\theta(s_t|\theta_t) + \mathcal{N} \tag{24}$$

3.2.2. Development of the Agents for Speed Control and Power Distribution

In the proposed system, the DDPG algorithm is utilized, using the internal and external states of the vehicle as input values. The state used for learning is composed of battery SOC, vehicle speed v_t, and road gradient. The internal controller of the vehicle receives two actions from the learned model. The first action is the target speed for the vehicle in the next time step, which represents the desired acceleration. The second action corresponds to the battery's output power. A positive value signifies power assistance, while a negative value indicates charging. When decelerating, coasting, or driving downhill, the vehicle uses regenerative braking, whereas in other cases, it is charged with the output of the fuel cell. As actions are executed, rewards are calculated.

Table 4. Configuration of neural networks

Configuration	Actor networks	Critic networks
Input data	State s_t	State s_t, Action a_t (Concatenate layer)
Data scaling	Min-Max scaler	
Input layer	Node: 64 (Relu)	
Hidden layer 1	Node: 64 (Relu)	
Hidden layer 2	Node: 64 (Relu)	
Output layer	2 (tanh)	1 (-)

Table 4 and Table 5 detail the specific configuration of the actor networks and critic networks, as well as the hyperparameters utilized in the learning process. In order to train the agent, the vehicle model navigates through a test road, which is depicted in Figure 17. To enhance the agent's generality, the test road is comprised of a variety of gradients, with eight distinct types included. The learning process was conducted employing an Intel Core i7-8700 CPU, a NVIDIA GeForce RTX 3070 Ti GPU, 64 GB of RAM, and TensorFlow 2.5.0.

3.2.3. Adaptive Action Space

The vehicle driving system has constraints for driving due to the specification limits of each component of the vehicle powertrain. The constraints on acceleration and battery power policies, output by the Agent, undergo flexible changes that depend on the current state of the vehicle and the road conditions. When a policy proposed within a fixed action space is implemented, it may result in exceeding the constraints due to the characteristics of the Backward-looking simulator. Thus, in order to effectively train the DDPG agents within the constraints, a dynamically adaptive action space is necessary. This section presents an approach for developing such an action space that flexibly adjusts to the current vehicle state, allowing for optimal performance within the specified constraints. By considering the states at each step, the proposed action space can effectively regulate the agent's output policy and maintain the vehicle's performance within the prescribed boundaries, even when facing changing road conditions or varying vehicle states.

Table 5. Hyperparameters of learning process

Hyperparameters	Values
Experience memory buffer size	$[10^3,\ 10^4,\ 10^5]$
Learning rate of actor networks	10^{-5}
Learning rate of critic networks	10^{-5}
Discount factor γ	0.99
Update parameter for target networks τ	$[10^{-1},\ 10^{-2},\ 10^{-3},\ 10^{-4}]$
Batch size	32
Optimizer	Adam

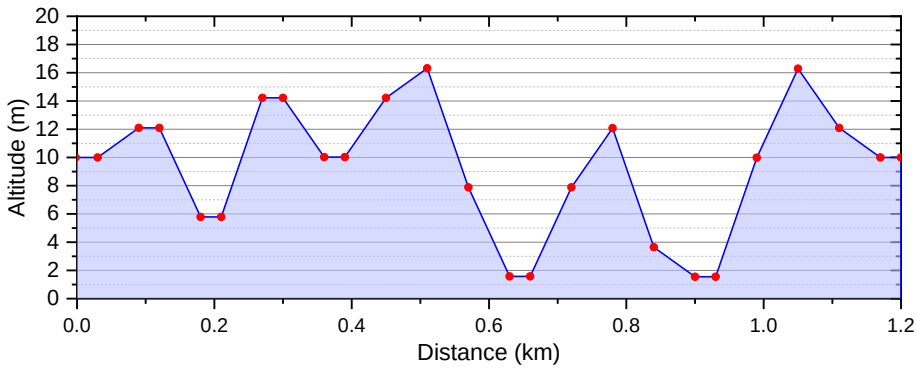


Figure 17. Test Road as an environment for Learning Agents through Vehicle Navigation

The adaptive action space is established according to the specification constraints of each component of the vehicle. The powertrain dynamics are constrained as follows:

$$T_{mot,min}(\omega_{mot}) \leq T_{mot}(t) \leq T_{mot,max}(\omega_{mot}) \tag{26}$$

$$I_{bat,min}(SOC) \leq I_{bat}(t) \leq I_{bat,max}(SOC) \tag{27}$$

$$P_{bat,min} \leq P_{bat}(t) \leq P_{bat,max} \tag{28}$$

$$v_{min} \leq v(t) \leq v_{max} \tag{29}$$

$$SOC_{min} \leq SOC(t) \leq SOC_{max} \tag{30}$$

In the reinforcement learning process, in which the vehicle drives on the test road, the action space for the speed change Δv is restricted to prevent it from exceeding the vehicle's speed limit and the motor's limit torque. Similarly, the action space for changing the ΔSOC is restricted to ensure that it does not exceed the limit current or limit power of the battery during power distribution. The updated action space at each step is defined as follows:

$$A_t = clip\big(A(\mu_\theta(s_t) + \mathcal{N}), A_{low}, A_{high}\big) \tag{31}$$

$$\begin{aligned} A_{low} &= [\Delta v_{min}(s_t), \Delta SOC_{min}(s_t)] \\ &= \big[max\big(\Delta v_{v,min}, \Delta v_{mot,min}\big), max\big(\Delta SOC_{I,min}, \Delta SOC_{p,min}\big)\big] \end{aligned} \tag{32}$$

$$\begin{aligned} A_{high} &= [\Delta v_{max}(s_t), \Delta SOC_{max}(s_t)] \\ &= \big[min\big(\Delta v_{v,max}, \Delta v_{mot,max}\big), min\big(\Delta SOC_{I,max}, \Delta SOC_{p,max}\big)\big] \end{aligned} \tag{33}$$

$$\Delta v_v = v_{lim} - v(t) \tag{34}$$

$$\Delta v_{mot} = \frac{\Delta d}{m_{veh}}\left(\frac{T_{mot,lim} r_f \eta_f}{R_{tire}} - F_{loss}\right) \tag{35}$$

$$\Delta SOC_I = -\frac{I_{lim}\Delta t}{q_{cap}} \tag{36}$$

$$\Delta SOC_p = -\frac{V - \sqrt{V^2 - 4RP_{bat}}}{2R}\left(\frac{\Delta t}{q_{cap}}\right) \tag{37}$$

where A_t is the physical action value in the vehicle model according to the action of actor networks at step t, A_{low} is minimum value of the action space, A_{high} is maximum value of the action space, Δv_v and Δv_{mot} represents the limit on speed change based on the vehicle's speed constraints and the motor's limit torque respectively, and $\Delta \mathrm{SOC_I}$ and ΔSOC_p represents the limit on SOC change based on the limit current of the battery and the limit power of the battery respectively.

Subsequent to receiving the policy from the actor networks, the adaptive action space is computed prior to executing the corresponding action on the vehicle model by the agent. The algorithm of the DDPG method with adaptive action space is summarized on Algorithm 1.

Algorithm 1 DDPG algorithm with adaptive action space

Initialize critic network $Q(s,a|\phi)$ and actor policy network $\mu(s|\theta)$ with random weights ϕ and θ
Initialize target networks $Q'(s,a|\phi_{targ})$ and $\mu'(s|\theta_{targ})$ with weights $\phi_{targ} \leftarrow \phi$ and $\theta_{targ} \leftarrow \theta$
Initialize replay buffer $\mathcal{R}$

for episode = 1 to M **do**
    Initialize a random process $\mathcal{N}$ for action exploration
    Receive initial observation state s_1

    while t < T **do**
        Select action $a_t = \mu(s_t|\theta) + \mathcal{N}_t$ according to the current policy and exploration noise
        Set action space $A_{low}, A_{high} = f(s_t)$ according to the function of the adaptive action space
        Execute action a_t and observe reward r_t and new state s_{t+1}
        Store transition $(s_t,\ a_t,\ r_t,\ s_{t+1},\ d)$ in replay buffer $\mathcal{R}$
        Sample a random mini-batch of N transitions $(s_i,\ a_i,\ r_i,\ s_{i+1},\ d)$ from $\mathcal{R}$
        Set $y_i = r_i + \gamma(1-d)Q'(s_{i+1}, \mu'(s_{i+1}|\theta_{targ})|\phi_{targ})$
        Update critic network by minimizing the loss:

$$L = \frac{1}{N}\sum_i \left(y_i - Q(s_i, a_i|\phi)\right)^2$$

        Update actor policy network μ by maximizing the Q-value:

$$\nabla_\theta J \approx= \frac{1}{N}\sum_i \nabla_a Q(s,a|\phi)|_{s=s_i, a=\mu(s_i)} \nabla_\theta \mu(s|\theta)|_{s=s_i}$$

        Update target networks:

$$\phi_{targ} \leftarrow \tau\phi + (1-\tau)\phi_{targ}$$
$$\theta_{targ} \leftarrow \tau\theta + (1-\tau)\theta_{targ}$$

    end while
end for

3.3. Results and Discussion

This section details the development of a driving system for FCHEV using the aforementioned DDPG algorithm. First, the reward function, which is the main element of reinforcement learning, is established. Subsequently, parameter tuning is executed to enable DDPG learning under optimal conditions. Finally, the efficiency of the developed driving system is assessed by evaluating its optimality in terms of energy consumption.

3.3.1. Development of the Reward Function

The reward function is a crucial element in reinforcement learning as it serves as an evaluation metric for an agent's behavior. It takes the current state and action as inputs and produces a corresponding reward value. The agent's goal is to choose actions that maximize the cumulative expected reward over a period of time, which leads to learning and improvement in performance. The primary objective of the proposed system is to discover optimal vehicle speed scenarios and power distribution strategies that minimize fuel consumption. To achieve this goal, the reward function includes three key components: the fuel consumption rate, driving time, and SOC deviation of the battery. Table 6 displays different combinations of these key components in the reward function. The reward function of Case 1 is the fuel consumption rate, and Case 2 multiplies the fuel consumption rate with the driving time to calculate the reward. Cases 3 and 4 incorporate a term for driving time into the reward functions of Cases 1 and 2, respectively, with ω_T as a

coefficient for speed control. Cases 5 and 6 introduce a term for SOC deviation from the SOC reference, which is set to 0.6, where ω_{soc} is the coefficient that controls the degree to which the SOC reference is followed. Cases 7 and 8 include a term for the rate of change in SOC in the reward functions of Cases 5 and 6, respectively, and γ is a correction coefficient between SOC change rate and fuel consumption rate according to battery and fuel cell output power.

Figure 18 shows the driving results of the DDPG agent for the reward function of each case. In case 1, the term for driving time is not included in the reward function, resulting in the vehicle reducing its speed at the beginning of driving and maintaining a low speed to minimize only fuel consumption rate. Despite controlling the state of charge (SOC) for achieving the minimum fuel consumption rate, the battery output is significantly high due to the absence of a battery-related controller. In cases 2, 3, and 4, the driving time term is added, which leads to an increase in speed to reduce the overall driving time. Speed control is performed to maximize the reward by minimizing fuel consumption based on the road gradient. However, due to the absence of a term for battery control, the battery SOC decreases almost linearly.

Table 6. Cases of the reward function combining key components

Cases	Reward function
1	$r = -\dot{m}_t$
2	$r = -m_t = -\dot{m}_t \Delta T_t$
3	$r = -(\dot{m}_t + \omega_T \Delta T_t)$
4	$r = -(m_t + \omega_T \Delta T_t)$
5	$r = -(\dot{m}_t + \omega_T \Delta T_t + \omega_{soc} \lvert SOC - SOC_{ref} \rvert)$
6	$r = -(m_t + \omega_T \Delta T_t + \omega_{soc} \lvert SOC - SOC_{ref} \rvert)$
7	$r = -(\dot{m}_t + \gamma \dot{SOC} + \omega_T \Delta T_t + \omega_{soc} \lvert SOC - SOC_{ref} \rvert)$
8	$r = -(m_t + \gamma \dot{SOC} + \omega_T \Delta T_t + \omega_{soc} \lvert SOC - SOC_{ref} \rvert)$

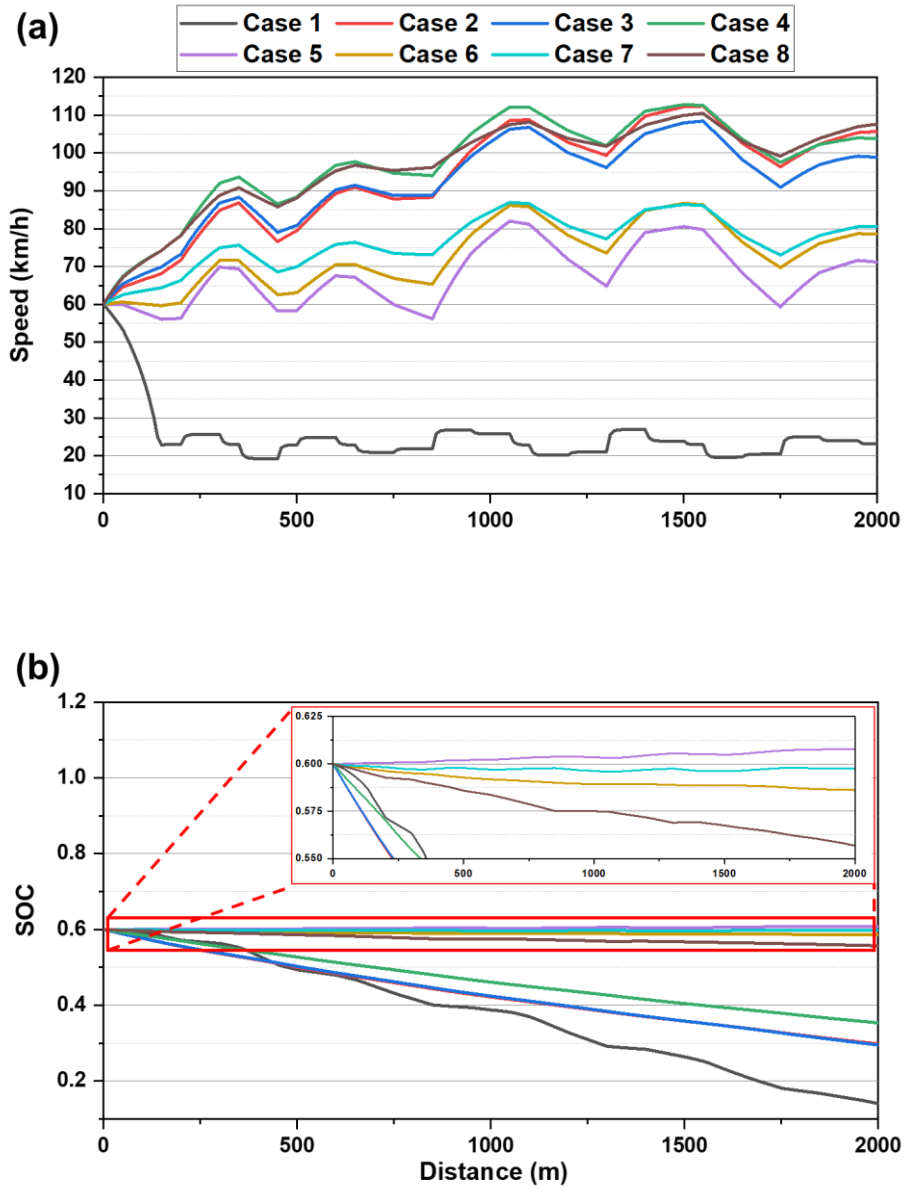


Figure 18. (a) Speed profile and (b) SOC trajectory according to reward function of DDPG algorithm

Therefore, the reward functions of cases 1, 2, 3, and 4 are not used for DDPG agent learning due to the absence of terms for speed control and power distribution control, respectively. Cases 5, 6, 7, and 8 perform control at an appropriate speed based on

the coefficient ω_T and maintain the battery SOC near the SOC reference while distributing power. The equivalent fuel consumption is calculated by applying an equivalent coefficient to the final SOC under the same total driving time to compare the specific impact of the reward function. The equivalent coefficient is calculated using dynamic programming (DP). DP is a global optimization method that is calculated using the Bellman equation, which is represented as:

$$J_t^*(s_t) = E[r_t + J_{t+1}^*(s_{t+1})] \tag{38}$$

DP results are commonly used as benchmarks for assessing optimality [44-46]. Optimization of the speed profile and power distribution for minimum fuel consumption is performed using sequential DP. The equivalent coefficient is calculated by figuring out the relationship between final SOC state and fuel consumption [47]. Figure 19 shows fuel consumption according to final SOC. The equivalent coefficient is obtained using the linear relationship between final SOC and fuel consumption. The fuel consumption according to reward function using the calculated equivalent coefficient is shown in Table 7. The reward function of Case 5 with the smallest fuel consumption is applied to the DDPG algorithm in this study. Figure 20 displays the action values in each step of Case 5. The actions indicate the variations in both speed and SOC. This demonstrates that the action space is limited according to the vehicle's specifications at each step, and an action is selected within that space. The Q value with reward of Case 5 can be updated as follows:

$$\begin{aligned} &Q_\phi(s_t, a_t) \\ &= \mathbb{E}_{r_t, s_{t+1} \sim E}\left[-(\dot{m}_t + \omega_T \Delta T_t + \omega_{soc}\left|SOC - SOC_{ref}\right|) + \gamma Q_\phi(s_{t+1}, \mu_\theta)\right] \end{aligned} \tag{25}$$

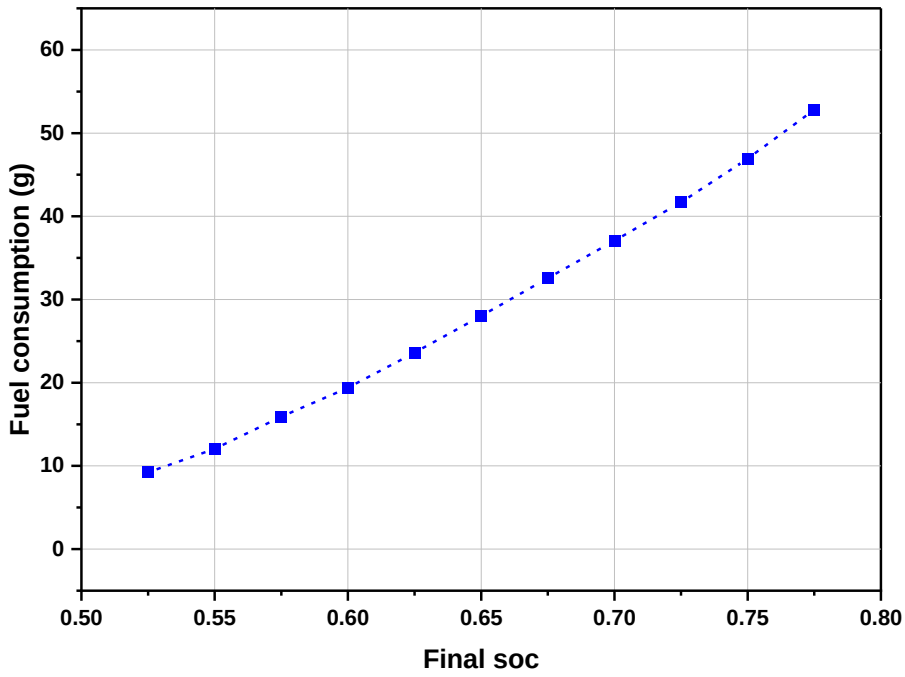
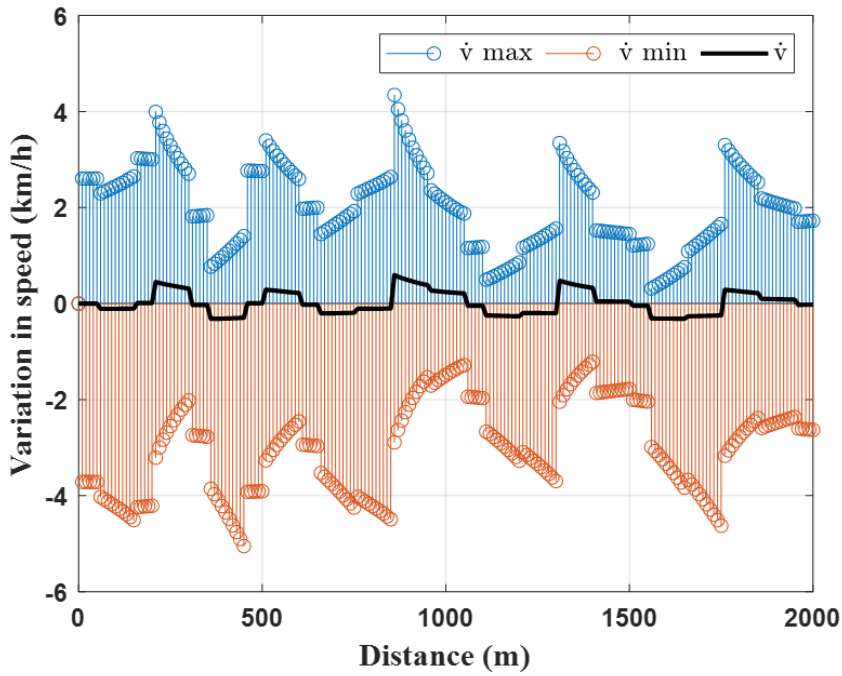


Figure 19. Fuel consumption according to final soc using sequential dynamic programming for speed control and power distribution.

Table 7. Fuel consumption based on reward function using equivalent coefficient

Cases	Driving time	Fuel consumption	Final SOC	Equivalent coefficient	Equivalent fuel consumption
5	107.58 s	20.19 g	0.6079		20.19 g
6	107.05 s	10.85 g	0.5367	168.85	22.87 g
7	106.87 s	17.31 g	0.5736		23.10 g
8	107.95 s	19.46 g	0.5940		21.81 g



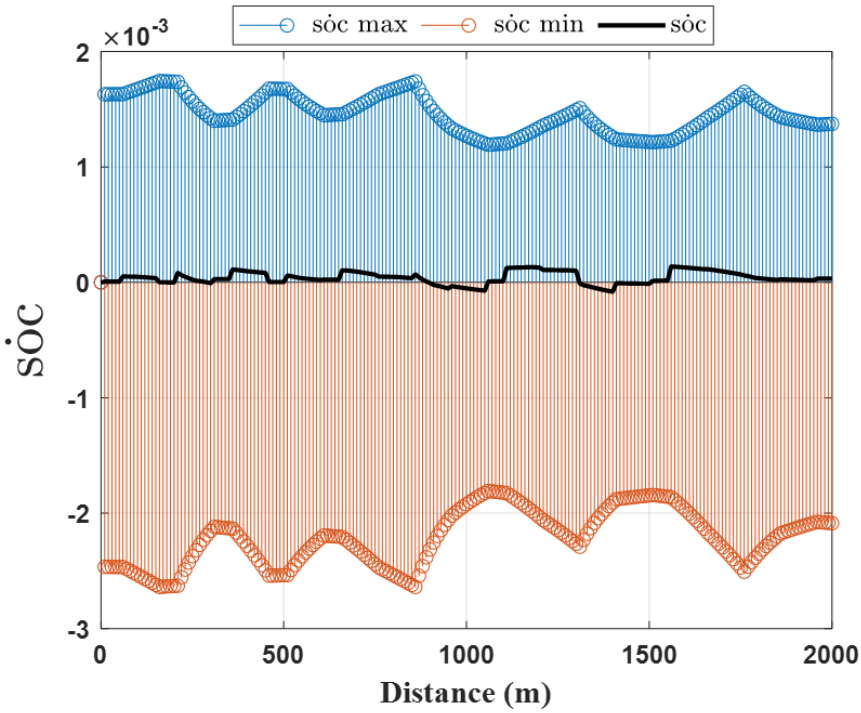
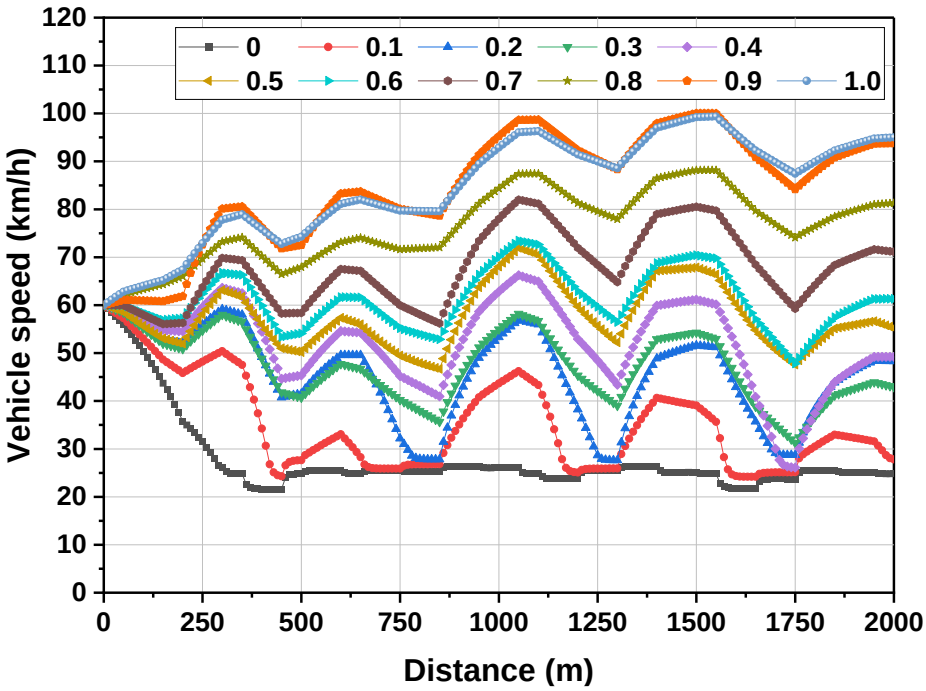


Figure 20. Actions selected within the constrained action space according to vehicle specifications in Case 5 (variation in speed and SOC)

3.3.2. Analysis of Coefficients in Reward Function

The selected reward function mentioned above incorporates two coefficients: ω_T, which is related to time, and ω_{soc}, which is related to the SOC of the battery. Figure 21 illustrates the relationship between vehicle speed, total driving time, and the time-related coefficients, ω_T. The time-related coefficient allows control over the total driving time within the vehicle's speed limit while accommodating the user's desired driving duration through adjustments in the speed spectrum to optimize energy-efficient driving. Moreover, Figure 22 presents the SOC history and equivalent fuel consumption based on coefficients related to SOC deviation. When the coefficient value is small, the impact of SOC deviation on the overall cost is minimal, resulting in rapid battery consumption to reduce fuel usage. However, with coefficient values of 5 or higher, the battery undergoes appropriate charging and discharging cycles to achieve optimal energy efficiency in driving. If the coefficient value becomes excessively large, the battery's SOC is maintained consistently at the reference value of 0.6, diminishing the role of the hybrid system in energy optimization. Notably, when the coefficient value was set to 10, it was observed that the least amount of fuel was consumed due to the implementation of proper charging and discharging strategies.



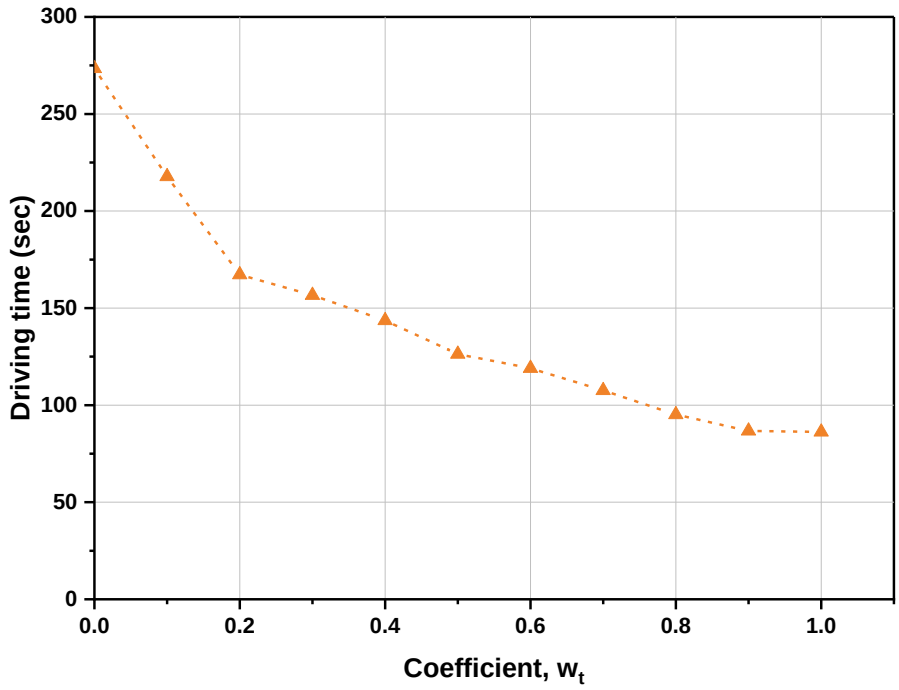


Figure 21. Vehicle speed and driving time according to time-related coefficient, ω_T in reward function

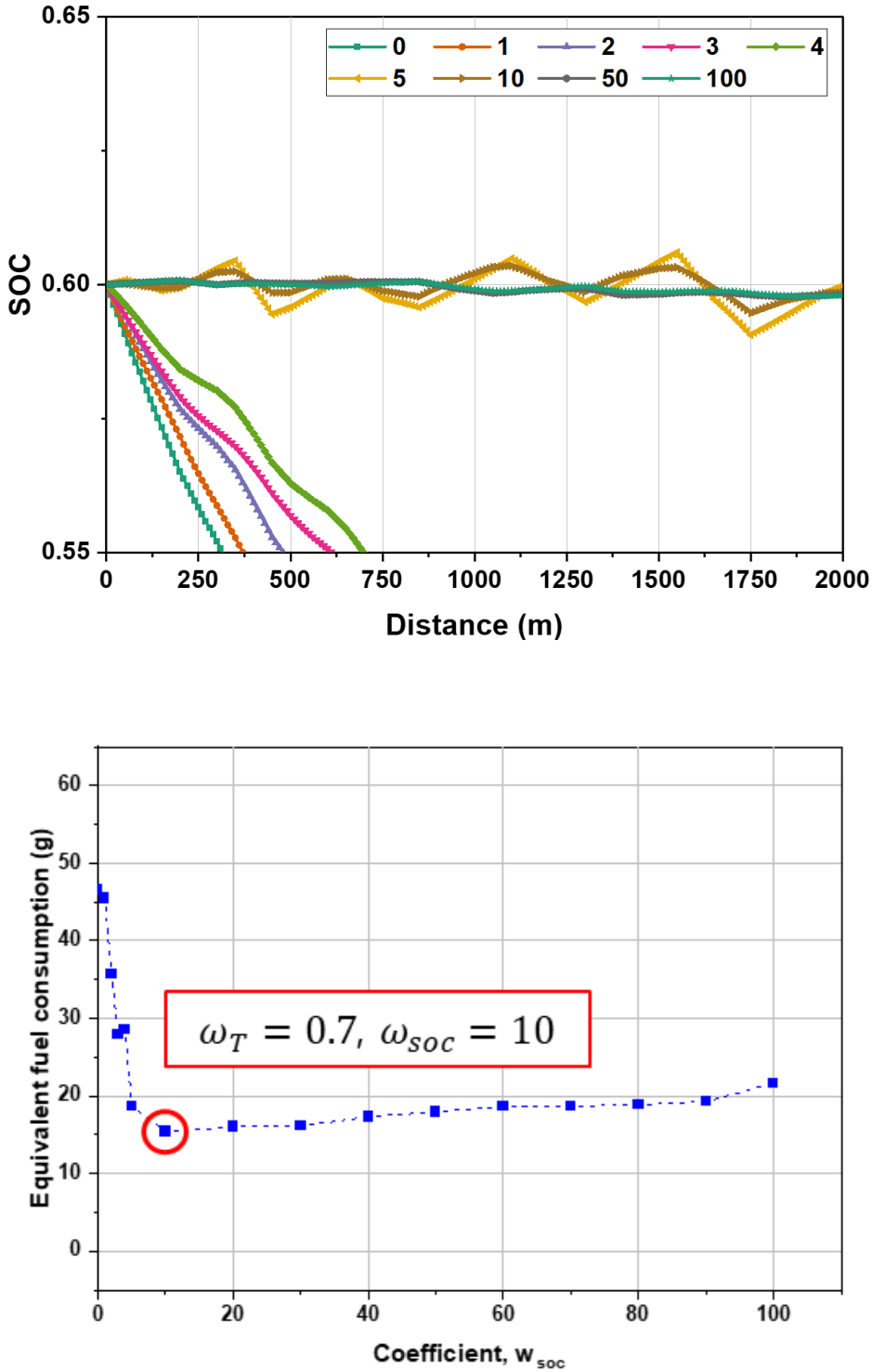


Figure 22. SOC history during driving and equivalent fuel consumption according to SOC deviation-related coefficient, ω_{SOC} in reward function

3.3.3. Parameter Tuning in DDPG Algorithm

The optimal values for the parameters of the DDPG algorithm are determined to improve its performance. The target update parameter τ is an important hyperparameter of the DDPG algorithm. It is used to control the update rate of the target network in the algorithm. Specifically, the target network, which is a copy of the original network used to calculate the target values, is updated gradually over time using the tau hyperparameter. Choosing an appropriate value for the tau hyperparameter is important to achieve a balance between the convergence rate and training stability in the DDPG algorithm. Figure 23 shows the return history according to the target update parameter τ. In the case of an update parameter of 0.1, convergence occurs at around the 600th episode, but thereafter, the

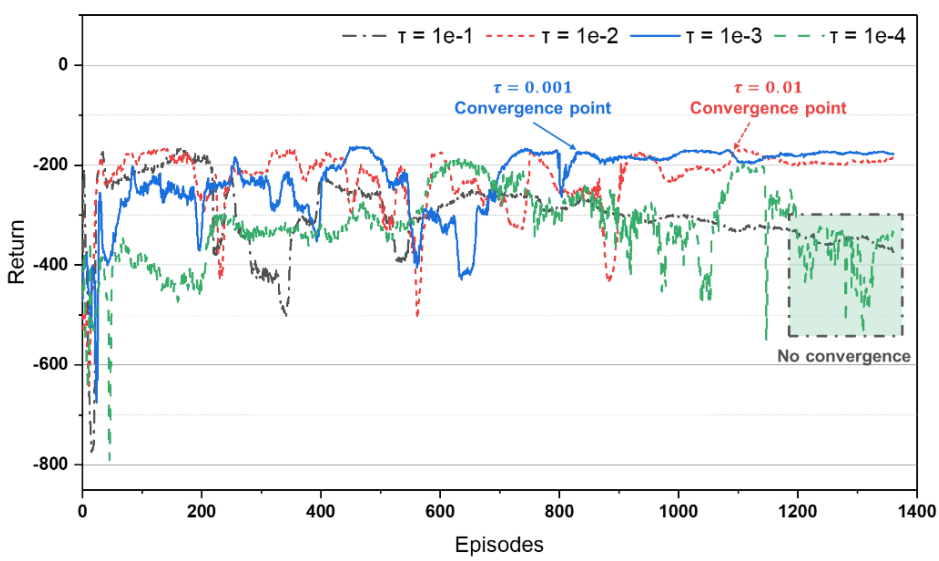


Figure 23. Return history according to update parameter τ with experience replay buffer size of 10^4

return decreases gradually over time. This is because frequent updates of the target networks result in the target values that the learning network aims to follow changing frequently, leading to instability in the learning direction. Convergence initiates approximately at episode 1100 for an update parameter of 0.01 and at episode 800 for an update parameter of 0.001. Despite having comparable return values in both cases, the latter has been verified to achieve faster convergence. For update parameter values of 0.0001, It is evident that the learning rate is relatively sluggish, as the small number of updates hinders the target networks from keeping pace with the latest trained networks. As a consequence, the learning process becomes unstable, leading to non-convergence.

Experience memory buffer size also has implications for the performance of DDPG. Research has shown that increasing the buffer size can lead to improved performance in some tasks, especially those that require a long-term memory. However, the optimal buffer size may vary depending on the task, and a larger buffer size may not always lead to better performance. Therefore, it is necessary to determine an appropriate buffer size that is suitable for the driving system under consideration. Figure 24 shows the return history according to the experience replay buffer size. The state of 200 steps (the number of steps per episode) per episode is stored in replay memory. The process of replacing the oldest memory with new memory begins at 5 episodes, 50 episodes, and 500 episodes for replay buffer sizes of 10^3, 10^4, and 10^5, respectively. When the buffer size is 10^3, the return value is greatly reduced in the initial learning process. This is due to the small size of the buffer, which limits the model's ability to reuse previous experience and therefore hinders its capacity to learn various situations. As learning progresses, the model performs more Exploitation, but only the latest data is used for learning, resulting in overfitting. Eventually, despite the model reaching convergence at 300 episodes, the performance is gradually declining. The results of the buffer size of

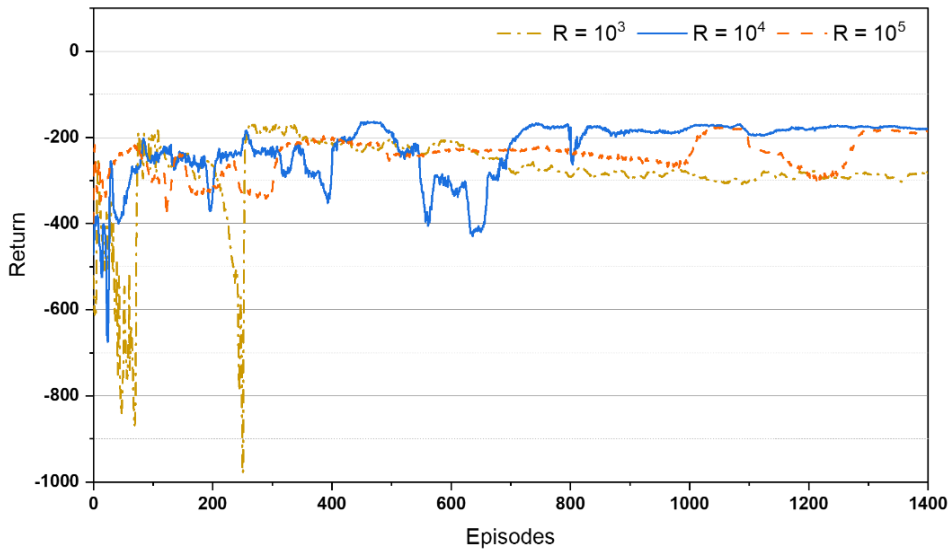


Figure 24. Return history according to experience replay buffer size with update parameter of 0.001

10^4 are the same as those of the update parameter 0.001 as illustrated in Figure 23. In the case of a buffer size of 10^5, the model exhibits stable learning at the initial stage of training due to the availability of a diverse set of experiences. However, from episodes 1000 to 1200, the model exhibits a lack of convergence and degrades performance during training. This can be attributed to the continual use of outdated or irrelevant data that is no longer useful for learning. In conclusion, the update parameter of 0.001 and the buffer size of 10^4 are the suitable choice for the specific environment considered in this study due to its fast convergence speed and stable learning stability.

Figure 25 illustrates the velocity profile and SOC trajectory of FCHEV per episode during the DDPG learning process. In the first episode, the random values assigned to the networks cause the model to exhibit biased outcomes such as continuous deceleration and battery output. As learning progresses, the speed of the

vehicle increases, and the SOC also operates near the reference SOC. Exploration of the model and non-convergence of learning cause the model to perform various experiences. Since the rate of the exploration decreases and the DDPG agent reaches convergence value in the latter part of learning, the vehicle continues similar driving and power distribution.

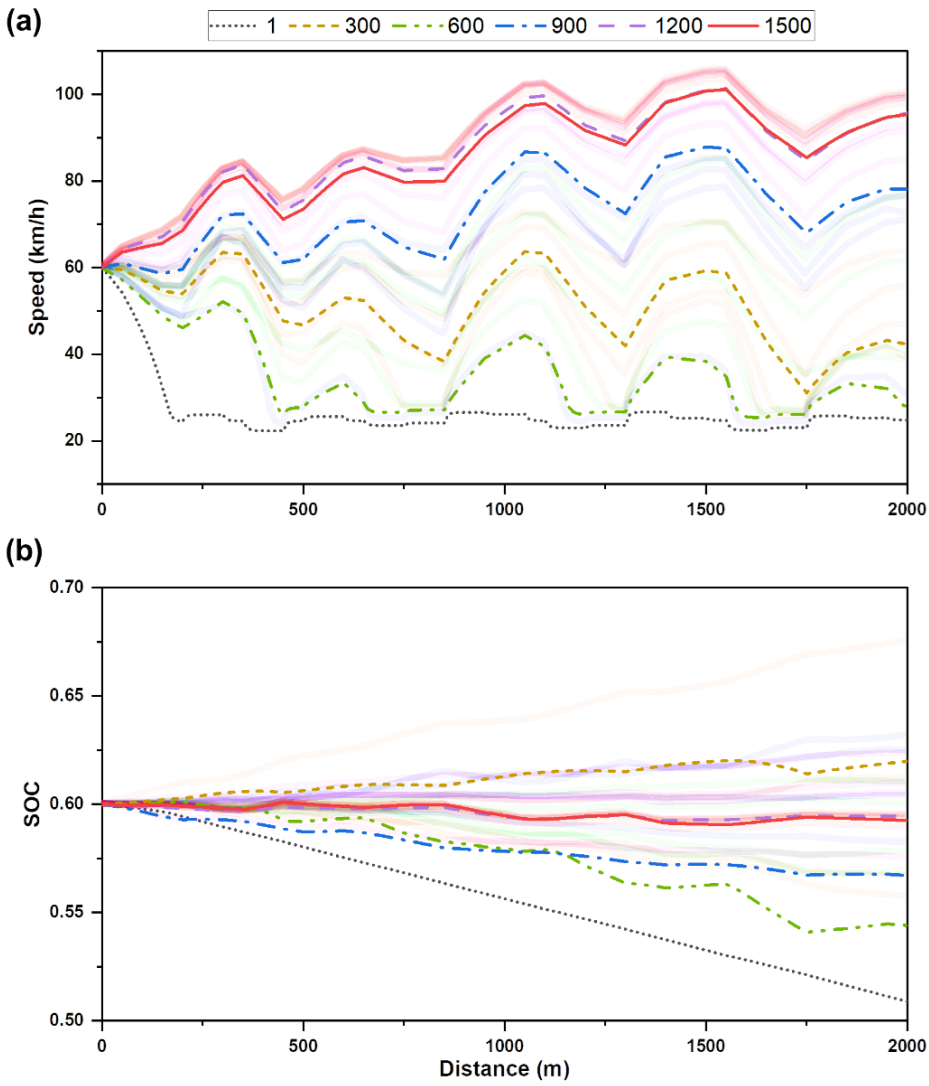


Figure 25. (a) Speed profile and (b) SOC trajectory of FCHEV across episodes for DDPG learning process

3.3.4. Optimality Verification of Speed Control and Power Distribution

The optimality of speed control and power distribution of the developed driving system are evaluated, and its performance is compared with conventional control methods used in actual driving controllers. The energy efficiency of speed control of the proposed system is tested by DP and cruise control to the system. Cruise control is an actual speed control method where the vehicle maintains a constant speed at a specific velocity. In addition, the power distribution of the proposed system is evaluated in comparison with the DP and the rule-based strategy developed as a heuristic. The rule-based strategy is developed by referencing the driving simulation tool, Autonomie, developed by Argonne National Laboratory. The strategy is summarized in Figure 26. The operation of the fuel cell is determined based on the SOC and required power, and the operation mode of the hybrid system is determined by the battery power, which depends on the SOC. The operation mode includes regeneration, normal, and power assist.

Figure 27 displays the vehicle driving speed and SOC for each control method, all having the same travel time. The cruise control strategy accelerates the vehicle to 100m, regardless of the road gradient, and then maintains a constant speed until 100m before the end of the journey, where it drives with a constant acceleration to reach the final speed. The proposed method of control shows a similar driving trend to the DP-based speed control, with a smaller speed deviation. Although the proposed method usually starts acceleration later, the deceleration start point is similar to that of DP.

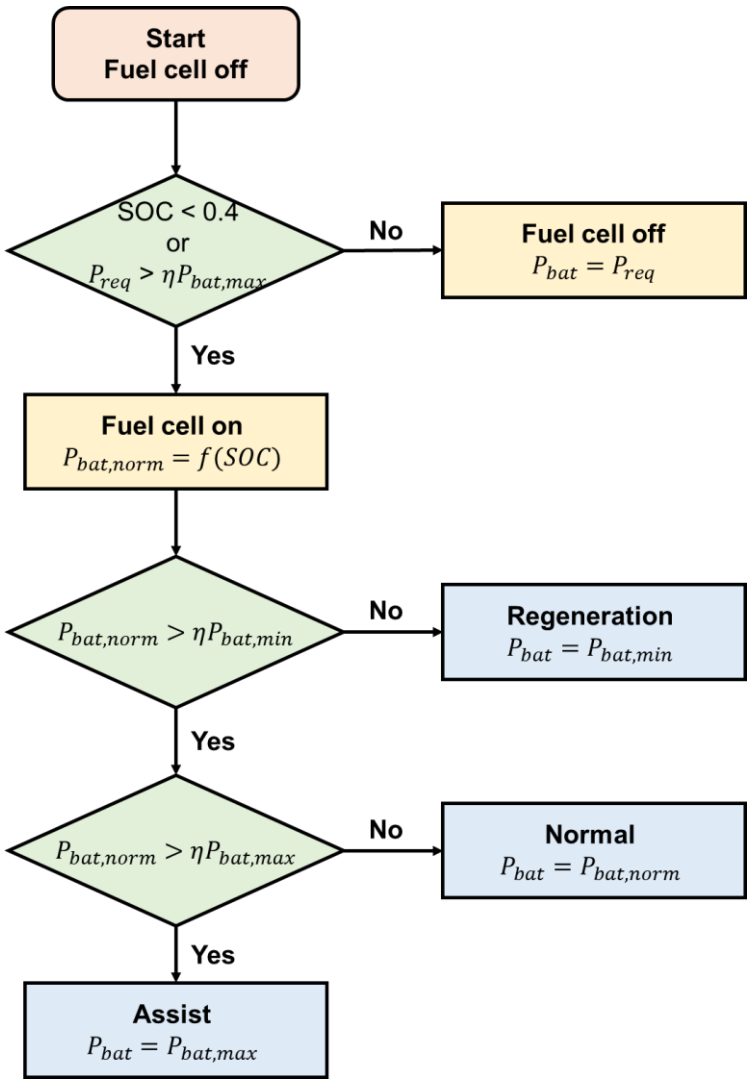


Figure 26. Flow chart summarizing the rule-based strategy for power distribution

The power distribution in a rule-based strategy is determined by referring to Figure 26, which displays the SOC and required power for charging and discharging. The SOC trajectory of DDPG-based power distribution displays a charge/discharge pattern that is similar to the optimal result. However, as the input and output power of the battery is relatively small, it is apparent that battery intervention is less active compared to DP.

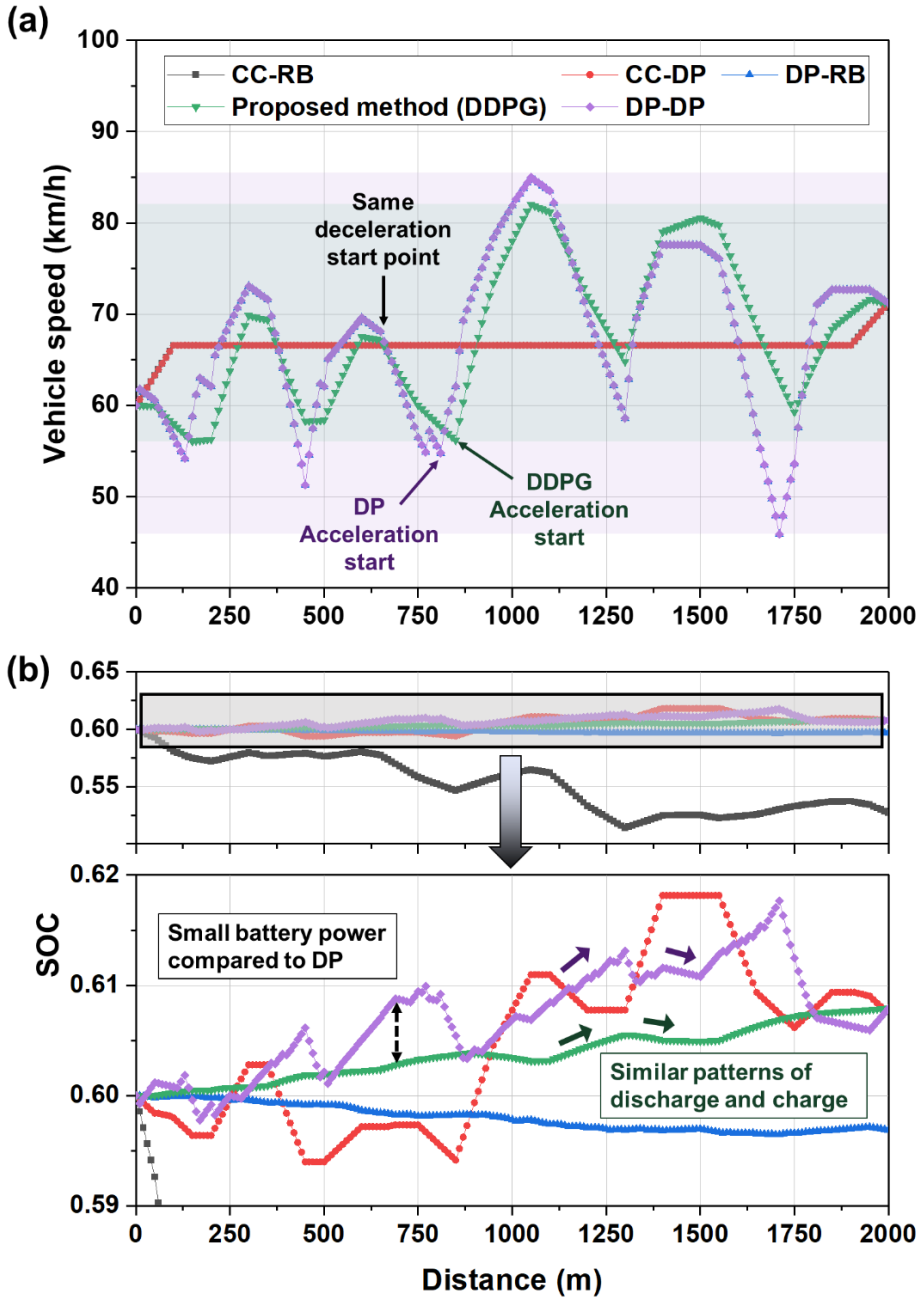


Figure 27. (a) Vehicle speed according to cruise control, dynamic programming and DDPG and (b) SOC trajectory according to rule-based strategy, dynamic programming and DDPG

Table 8 presents the fuel consumption results of each control method. To ensure a fair comparison of fuel consumption under equivalent SOC conditions, an equivalent coefficient calculated in Table 7 is used. Since this problem requires two optimization problems sequentially, the method applied with cruise control and rule-based strategy reduces the optimality. Consequently, a significant drop in performance is observed. Furthermore, superior performance outcomes are observed when DP is employed for speed control rather than power distribution, highlighting the greater energy impact of road gradient-based speed control. The DDPG-based proposed method achieves an optimality of 97.11%, which is superior to the result obtained by optimizing only one of speed control or power distribution.

Table 8. Fuel consumption according to control methods

Case		**Fuel consumption**	**Final SOC**	**Equivalent fuel consumption**	**Optimality**
Speed control	**Power distribution**				
Cruise control	Rule-based	18.18 g	0.5272	31.81 g	61.63 %
Cruise control	DP	23.37 g	0.6079	23.37 g	83.89 %
DP	Rule-based	20.06 g	0.5969	21.92 g	89.44 %
Proposed algorithm (DDPG)		20.19 g	0.6079	20.19 g	97.11 %
DP	DP	19.60 g	0.6079	19.60 g	100 %

3.3.5. Characteristic Analysis of Roads

The slope of the road used for reinforcement learning and model testing has a great influence on the driving model as a state for learning as well as vehicle demand power. In this section, the DDPG (Deep Deterministic Policy Gradient) model undergoes training using roads with varying slopes, and subsequently, the model is evaluated on real roads with diverse slope characteristics. Figure 28 presents the altitude and grade of the roads for learning. Three types of roads are employed for training the driving model: roads with gentle slopes, roads with steep slopes, and roads that combine both characteristics. Figure 29 shows the normalized elevation and grade of the real road for testing the model. Specifically, the highway section corresponds to the initial 4 kilometers of the Seocho Interchange in Korea, while the flat road and hill road are derived from the Mojave proving ground roads.

Figure 30 presents the comparative analysis of equivalent fuel consumption when employing models trained with the DDPG algorithm on the test road. The integrated road-trained model exhibited the lowest fuel consumption across all test roads. Furthermore, the model trained on gentle roads demonstrated lower fuel consumption on flat roads compared to the model trained on steep roads. Conversely, the steep road-trained model displayed reduced fuel consumption in hilly terrains. Figure 31 shows vehicle acceleration and battery power, which are actions of the driving model, according to the grade of the test road. Irrespective of the learning road, the test road's gradient and vehicle acceleration demonstrate similar trends. However, there are disparities in battery power depending on the learning road. The gentle-trained model, predominantly trained on mild slopes, exhibited minimal charge and discharge, thus failing to leverage the advantages of hybrid functionality.

Conversely, the steep-trained model, unnecessarily induced continuous battery charging, resulting in increased fuel consumption.

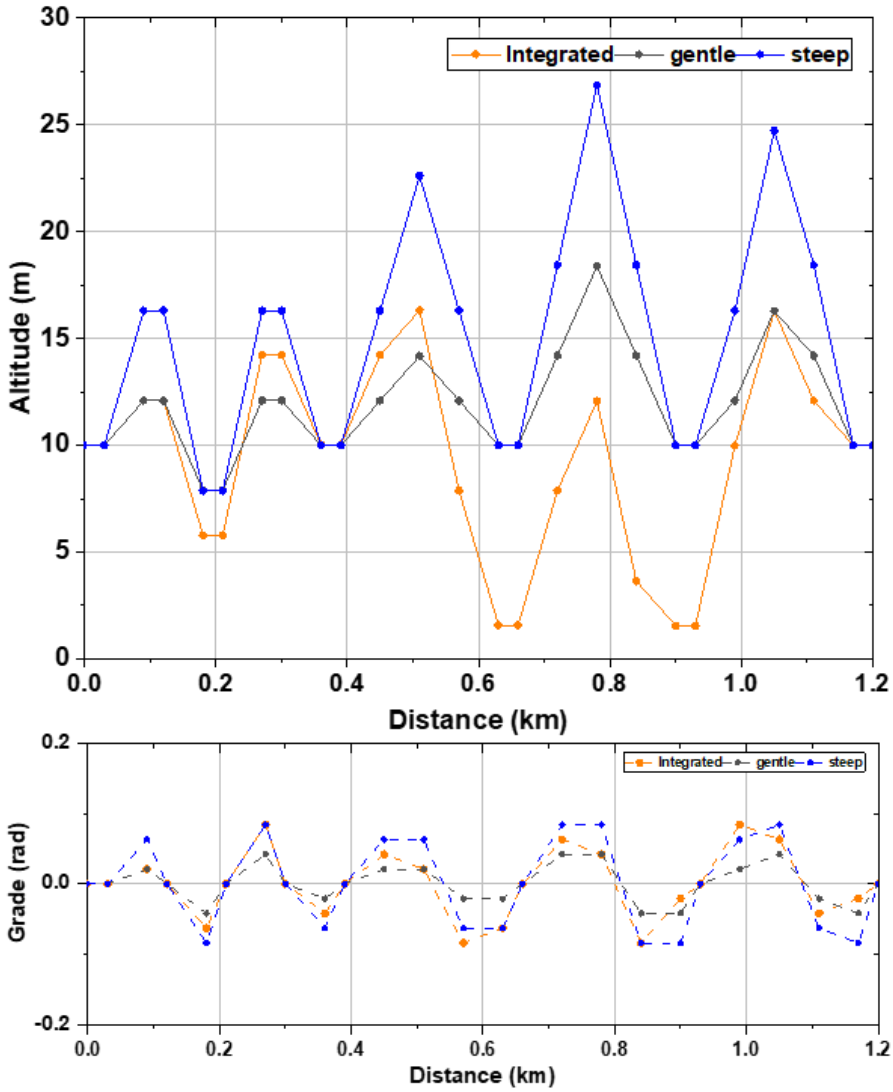


Figure 28. Altitude and grade of the road for DDPG learning

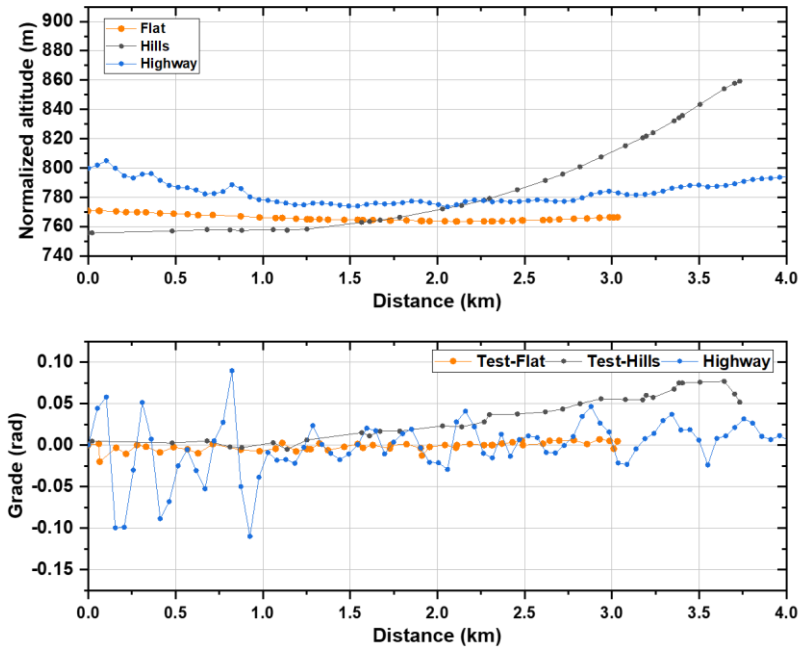


Figure 29. Normalized altitude and grade of the road for model test

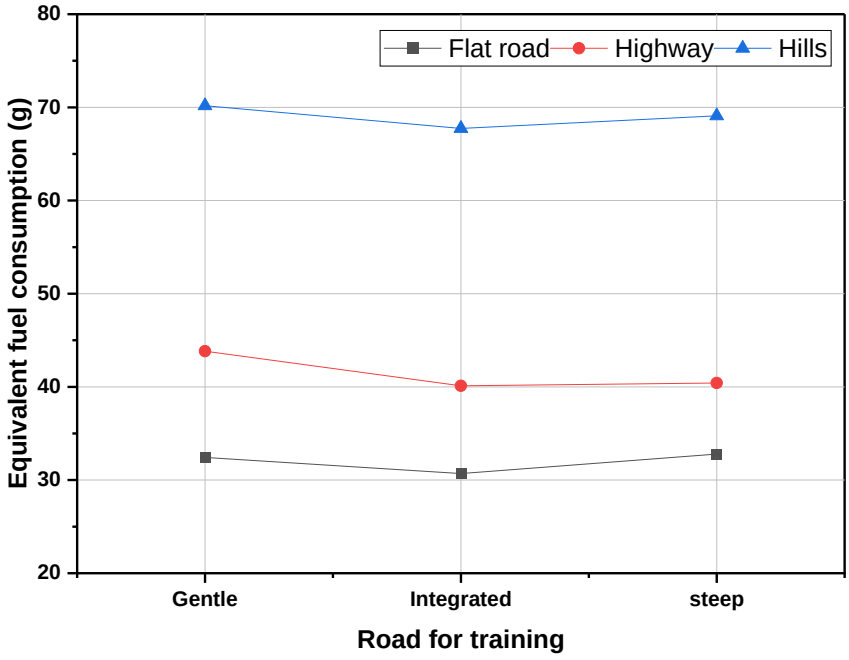


Figure 30. Equivalent fuel consumption according to train road

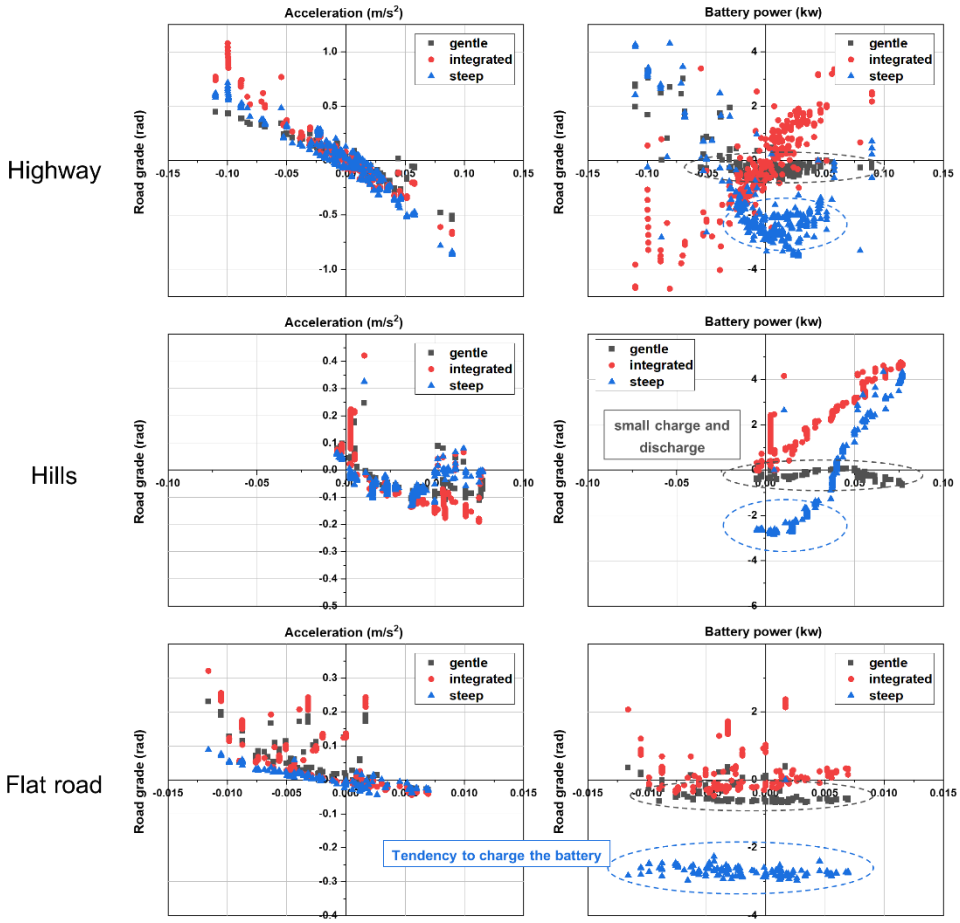


Figure 31. Actions of the driving model according to the grade of the test road: Acceleration and battery power

Chapter 4. Development of the DDPG Agent considering Moisture-related Errors in Interiors of PEMFC

4.1. Introduction

This chapter presents the development of a diagnostic system for moisture-related errors, specifically flooding and drying, in the interiors of PEMFC. To analyze the impact of flooding and drying on performance, a single PEMFC is fabricated and experimentally tested. In order to induce flooding and drying in the interiors of PEMFCs, extreme experimental parameters including temperature and relative humidity are set and the corresponding physical responses are recorded. These time-series experimental data are then used to construct a diagnostic system for flooding and drying, utilizing the LSTM and bagging methods. Finally, the performance of the diagnostic system is evaluated to assess its effectiveness in accurately diagnosing these moisture-related errors in PEMFCs.

Furthermore, an approximation model for performance degradation is developed through voltage pattern analysis during flooding and drying. It is then applied to the vehicle model. When flooding or drying is diagnosed, the vehicle's performance decreases gradually according to the corresponding degradation model. In the process of DDPG learning, the agent achieves energy-efficient driving while considering degradation caused by moisture-related errors through referencing the diagnostic system results.

4.2. Experiments for Inducing Flooding and Drying

The long-term operation of the fuel cell was carried out with the conditions outlined in Table 9. These operating conditions have been widely used in previous studies investigating the water management of PEMFC [48, 49]. The relative humidity ($\varphi = p_w/p_{sat}$) is determined by the temperature of the MEA and the humidifier, following the psychrometric chart. If the relative humidity is over 100%, flooding typically occurs, while if it is below 50%, drying occurs.

Table 9. Experimental conditions for inducing flooding and drying

Relative humidity (%)	**Cell temperature (℃)**	**Current density (A/cm^2)**
50	40	0.4
		0.8
	60	0.4
		0.8
	80	0.4
		0.8
100	40	0.4
		0.8
	60	0.4
		0.8
	80	0.4
		0.8

Peculiarities in the voltage and impedance modulus of a transparent flow channel PEMFC were identified through a comparison of its video and electrochemical data. In Figure 32, the flow channel experiences rapid liquid water accumulation during the flooding phenomenon. In the case of drying, only a few water droplets are observed, remaining stagnant in the GDL.

Figure 33 illustrates the Impedance modulus in the flooding. In the case of flooding, distinct anomalies are clearly observed in the data at lower frequencies (~10Hz). A fixed frequency of 10Hz is selected as the measurement frequency for impedance in the flooding state. On the other hand, in the drying state, the determination of Ohmic resistance is based on the High-Frequency Resistance (HFR) region, which corresponds to a semicircular region in the Nyquist plot as shown in Figure 34. Therefore, a fixed frequency of 1000Hz is selected for impedance measurements. Figure 35 depicts the voltage and impedance modulus during flooding and drying. Voltage sharply drops during flooding, while there is no variation in impedance modulus. As for drying, the voltage gradually decreases, and the impedance modulus increases. These tendencies are observed at relative humidity of 50% and 100%.

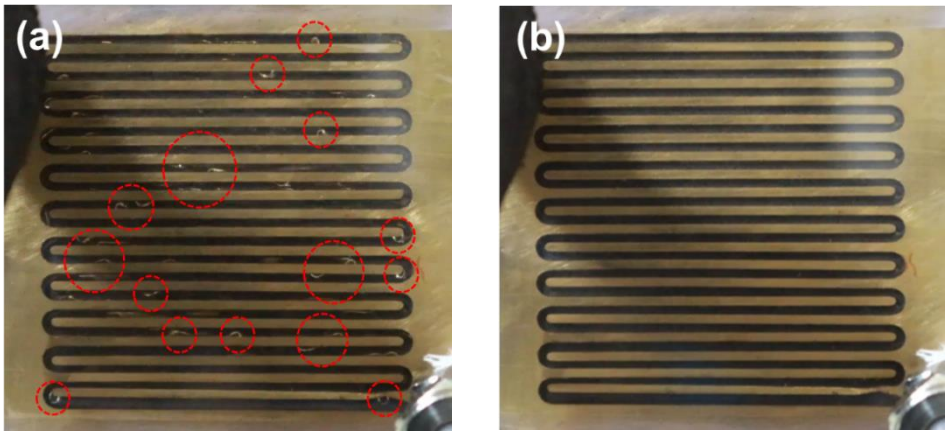


Figure 32. Transparent PEMFC during (a) flooding and (b) drying

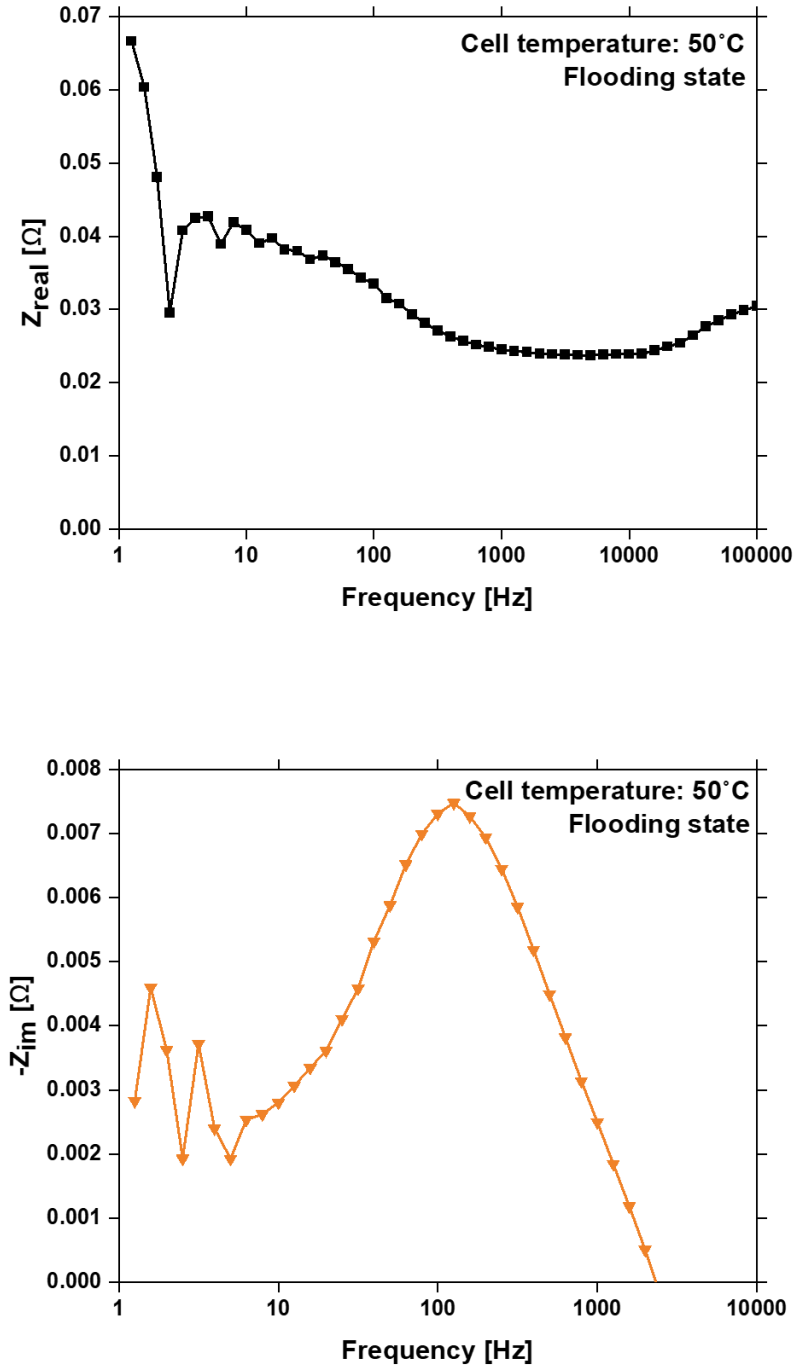


Figure 33. Real and imaginary values of impedance vary with the frequency of alternating current

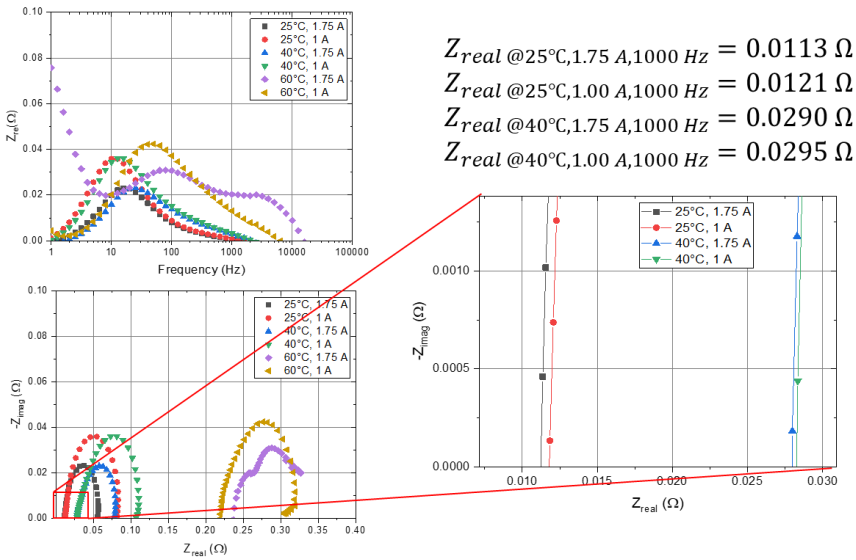


Figure 34. Real value of impedance according to frequency and Nyquist plot during drying

Polarization curves and Nyquist plots according to relative humidity for the PEMFC are shown in Figure 36. The cell temperature is maintained at 60 °C for relative humidity of 100% and 120%, and at 80 °C for a relative humidity of 50%. Results show that performance deteriorated more significantly at a relative humidity of 120% due to the increased likelihood of flooding. Performance declined rapidly at a relative humidity of 50% due to drying. Recorded data included current density, cell voltage, cell temperature, relative humidity of the anode and cathode, and impedance modulus. The impedance modulus at a fixed high frequency of 4000 Hz is employed as an indicator to determine the operational status of the PEMFC, which includes normal, flooding, and drying status. This impedance modulus is also included in the criteria to evaluate the occurrence of flooding and drying. Nevertheless, it is excluded from the deep learning models due to its prolonged measurement time.

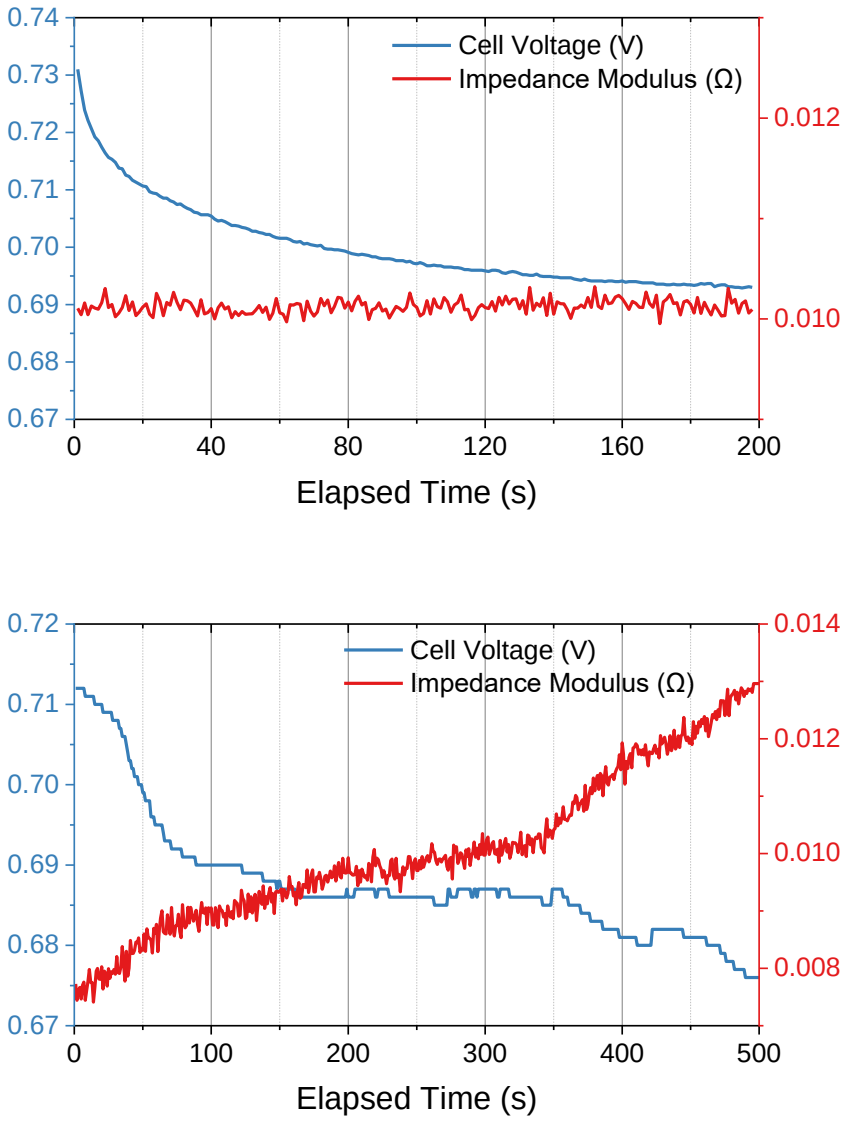
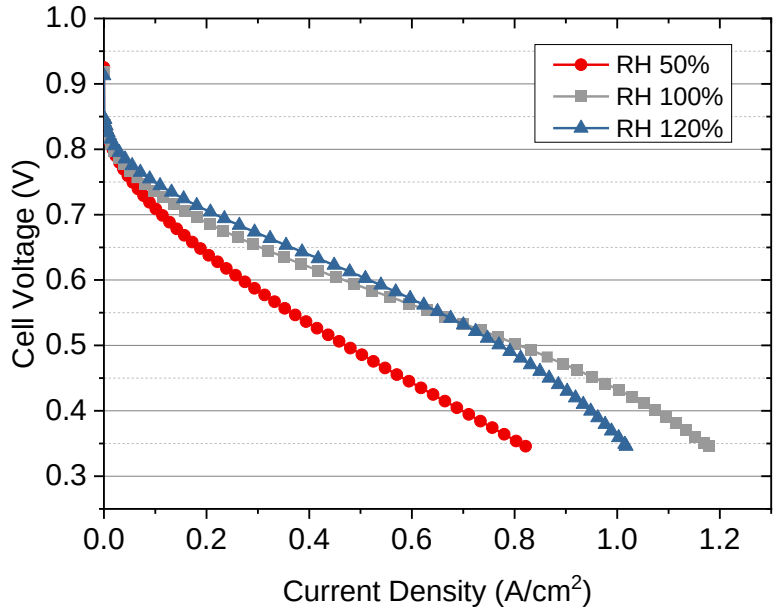


Figure 35. Voltage and impedance modulus during (a) flooding and (b) drying



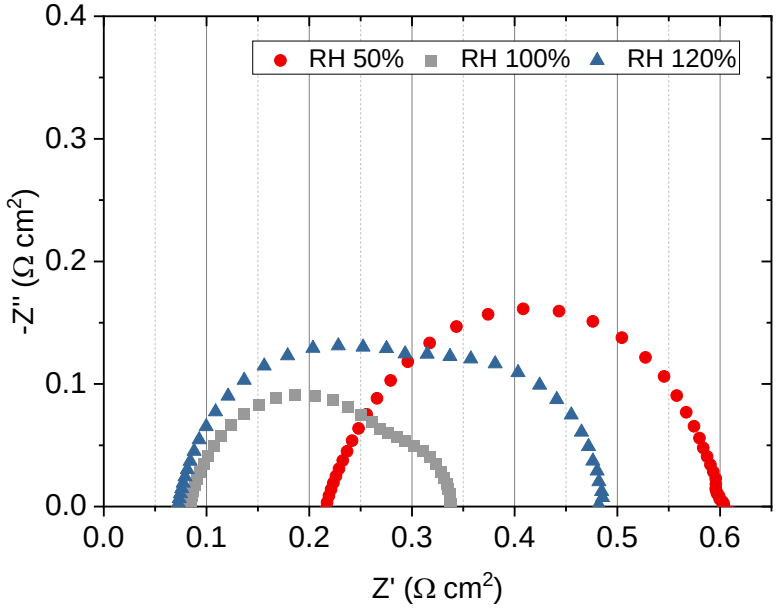


Figure 36. Polarization curves and Nyquist plot according to relative humidity in PEMFC experiments

4.3. Development of Diagnosis System using LSTM

4.3.1. Calculation of LSTM

The present study aimed to develop a diagnosis system for fuel cell operation by utilizing LSTM, a type of Recurrent Neural Network (RNN), to model the sequential correlation from past to present states in the time-series data. LSTM is known for its ability to detect long-term dependencies in data, which makes it suitable for this task. This method was developed to solve the vanishing gradient problem commonly encountered in Vanilla RNN [50, 51]. This problem arises when the time step increases, and the gradient decreases, leading to the loss of information from the past hidden state. However, LSTM incorporates several gates in the cell state that regulate the flow of information, reducing the occurrence of this issue [52]. Figure 37 represents the structure of the LSTM cell. The forget gate in LSTM employs a logistic activation function to generate outputs that range between 0 and 1. These outputs are then subjected to an element-by-element multiplication operation, which selectively erases part of the long-term memory. Additionally, a new memory component is added to the input gate to improve the model's performance. The long-term state beyond the input gate is then passed to the hyperbolic tangent ($tanh$) function and filtered by the output gate, which generates a short-term state and the final output of the cell. By employing these mechanisms, LSTM can effectively detect long-term dependencies in sequential data and address the vanishing gradient problem associated with Vanilla RNN. The equations governing the LSTM cell for a single step can be expressed as follows:

$$f_t = \sigma\left(W_{xf}^T x_t + W_{hf}^T h_{t-1} + b_f\right) \tag{38}$$

$$i_t = \sigma\left(W_{xi}^T x_t + W_{hi}^T \mathrm{h}_{\mathrm{t-1}} + b_i\right) \tag{39}$$

$$o_t = \sigma\left(W_{xo}^T x_t + W_{ho}^T h_{t-1} + b_o\right) \tag{40}$$

$$g_t = \tanh\left(W_{xg}^T x_t + W_{hg}^T h_{t-1} + b_g\right) \tag{41}$$

$$c_t = f_t \circ c_{t-1} + i_t \circ g_t \tag{42}$$

$$y_t = h_t = o_t \circ \tanh(c_t) \tag{43}$$

where f_t (forget gate) controls which information from the previous cell state and the current input should be removed, i_t (input gate) controls how much new information should be added to the current cell state, o_t (output gate) determines which part of the cell state should be output as the output of the LSTM, g_t (candidate value) is a new piece of information calculated by the activation function, which is used to update the cell state, c_t (cell state) is the long-term memory of the LSTM, which is updated based on the input and forget gates, the candidate value, and the previous cell state, y_t (output) is the short-term memory generated based on the current cell state and output gate, and h_t (hidden state) is the output of the LSTM at each time step and is derived from the cell state through an activation function.

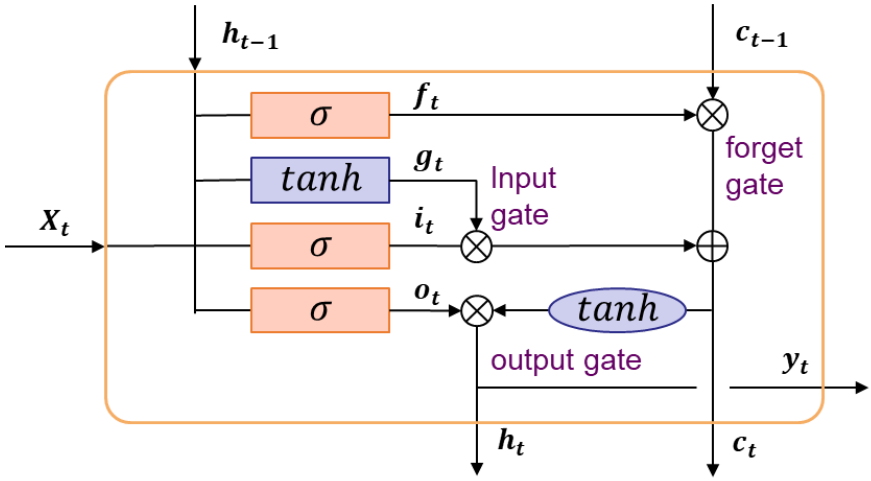


Figure 37. Structure of the LSTM cell

The architecture of the LSTM is depicted in Figure 38. The input sequence is processed by the LSTM cells based on the time step t. The LSTM layers, except for the last one, are sequence-to-sequence networks which are designed to process sequential input data and generate sequential output data. These networks consist of multiple LSTM layers, where each layer receives an input sequence and outputs a sequence that is passed to the next layer. The final LSTM layer is a sequence-to-vector network where a sequence of inputs is processed and transformed into a single output vector. It produces a fixed-length vector by taking the final output state of the LSTM cells at time step T. This vector output is forwarded to the subsequent dense layer.

The input features for the proposed system are selected from the experimental results obtained from full-scale tests. However, in order to make the system applicable in real-time, experimental data that required several minutes to collect, such as impedance modulus, were eliminated from the input features. In addition, it is observed during the experimental phase that flooding and drying can be confirmed

by monitoring voltage changes while keeping the current constant. However, in real vehicle operation where the current is constantly changing, adding voltage as an input feature may result in inaccurate diagnosis. Therefore, voltage is also excluded from the input features. The selected features for the system were the current density, relative humidity inside the flow channel of the anode and cathode, and cell temperature. The LSTM networks produces an estimation of the operating state of the PEMFC as output. This is achieved by utilizing the softmax function to determine the probabilities of each class, which include the states of flooding, normal, and drying, based on the last calculation results of the dense layer. The state of the PEMFC is determined by the model selecting the class with the highest probability. The duration of the time step t and the window size T in this study were 1 s and 45 s, respectively. The diagnosis of the PEMFC's state was conducted by analyzing the data from the previous 45 s to the current time.

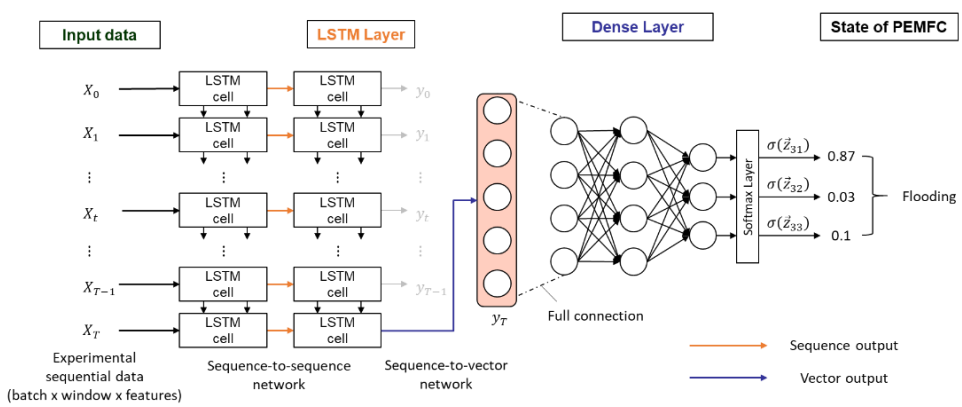


Figure 38. Architecture of the deep LSTM model [53]

4.3.2. Bagging Ensemble Method

The Bagging method, also known as Bootstrap Aggregating, is a widely-used ensemble learning technique that aims to improve the accuracy and stability of neural network models [53-57]. Bagging entails generating multiple bootstrapped datasets from the original training set and training an individual model on each of these datasets. The predictions of these models are then aggregated using a voting or averaging mechanism to produce the final prediction. Figure 39 illustrates the process of the Bagging method. The Bagging method involves the following steps [58]:

1. Randomly sample the training dataset with replacement to create multiple bootstrap samples with data balancing.

2. Train a separate model on each bootstrap sample.

3. Use each model to make predictions on the test dataset.

4. Aggregate the predictions of all models to produce the final prediction.

The aggregation can be done in several ways, including majority voting, weighted voting, and averaging. Majority voting involves selecting the most common prediction among all models, while weighted voting assigns weights to each model's prediction based on its performance on the training set. Averaging involves taking the average of all models' predictions. Table 10 presents the parameters of the Bagging method used in this study. Table 11 displays the configuration of the LSTM models, which serve as weak classifiers in this study.

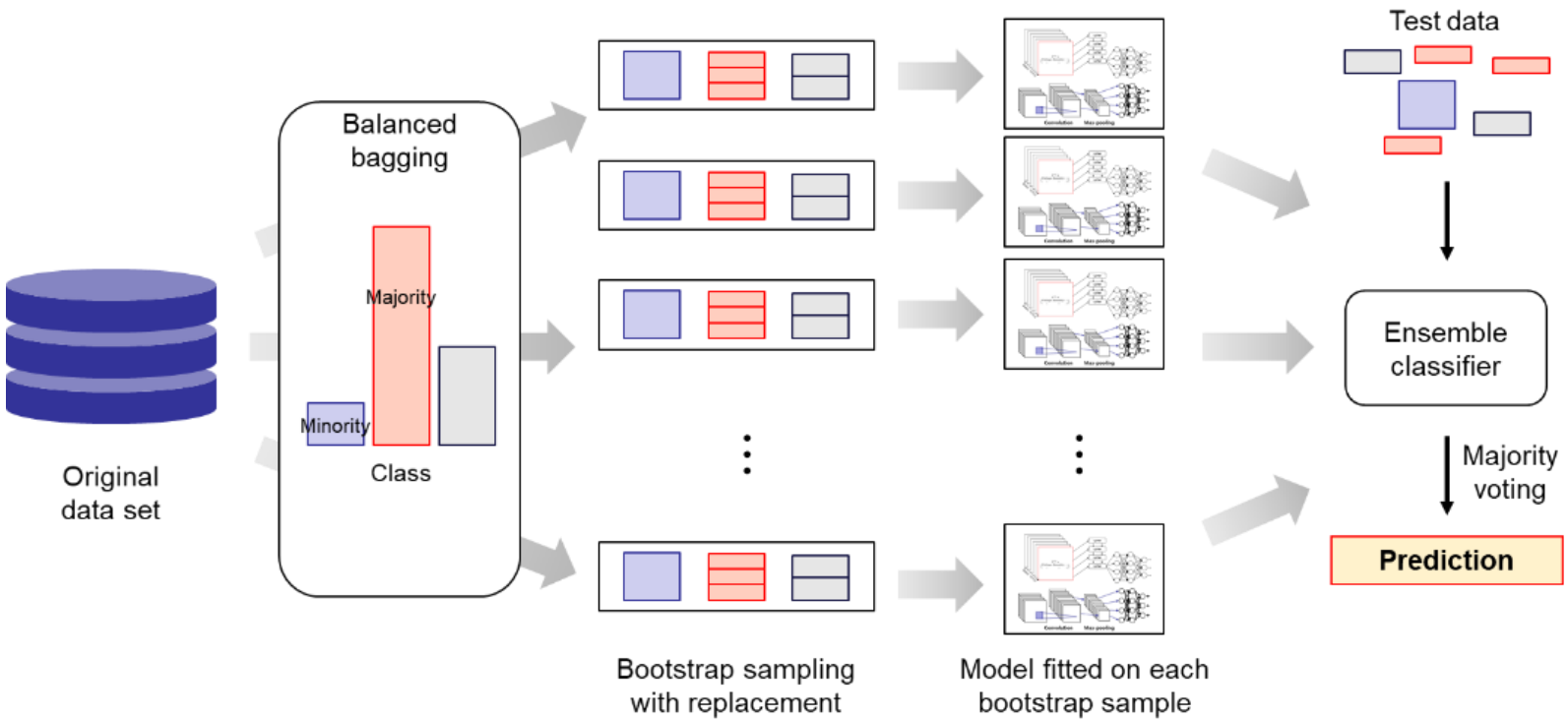


Figure 39. Process of the Bagging ensemble method [53]

Table 10. Parameters of the Bagging methods based on LSTM

Parameters	Values
Number of samples	9
Size of each dataset	10000 (balanced)
Aggregation	Majority voting
Window size	45
Batch size	512
Learning rate	1e-3
Optimizer	Adam

Table 11. Configuration of the LSTM for weak classifiers in Bagging method

Configuration	LSTM networks
Input data	$[j_{cell}, \varphi_{an}, \varphi_{ca}, T_{cell}]$
Data scaling	Min-Max scaler
LSTM layer 1	Unit: 128 (Sequence to sequence)
LSTM layer 2	Unit: 128 (Sequence to sequence)
LSTM layer 3	Unit: 64 (Sequence to vector)
Dense layer	Node: 32 (elu)
Output layer	3 (softmax)

4.3.3. Diagnosis Results of PEMFC State

Figure 40 depicts the diagnostic accuracy of LSTM-based network model and the bagging ensemble method. The individual models demonstrate an accuracy range of 82-83%, with an average of 82.6%. Nevertheless, the collective accuracy of the

ensemble method, which integrates the outputs of each model through a majority vote, shows a marked improvement, reaching 88.1% accuracy. The confusion matrix of the bagging model is presented in Figure 41. The primary reason for model misclassification is the misdiagnosis of normal PEMFC states as either flooding or drying. Nonetheless, the model exhibits a high degree of accuracy in the majority of cases, and in particular, demonstrates a detection rate of 97.39% for flooding cases.

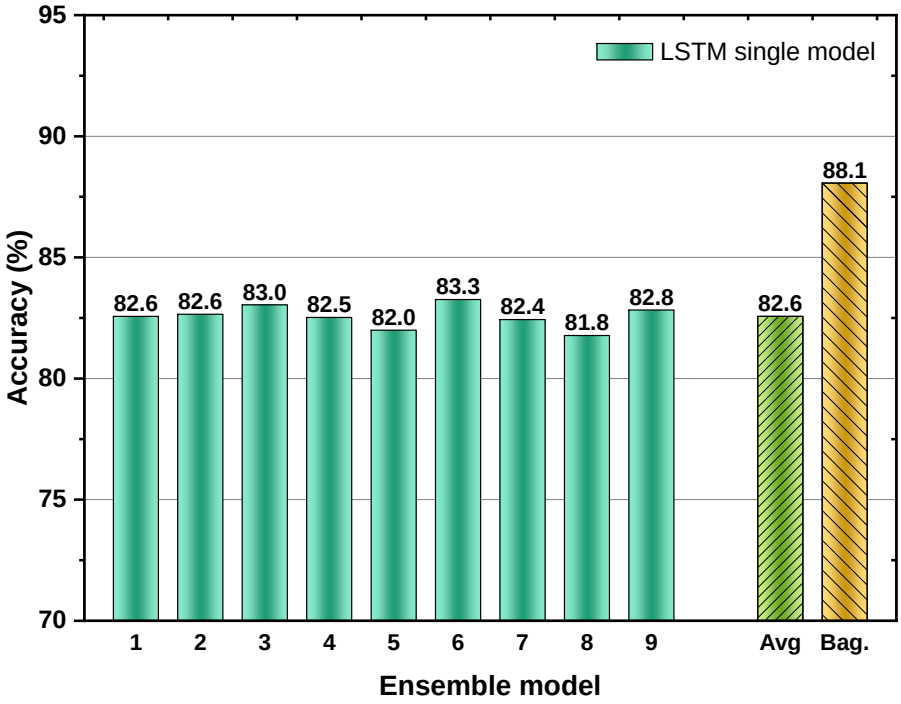


Figure 40. Diagnosis accuracy of nine LSTM models and bagging model

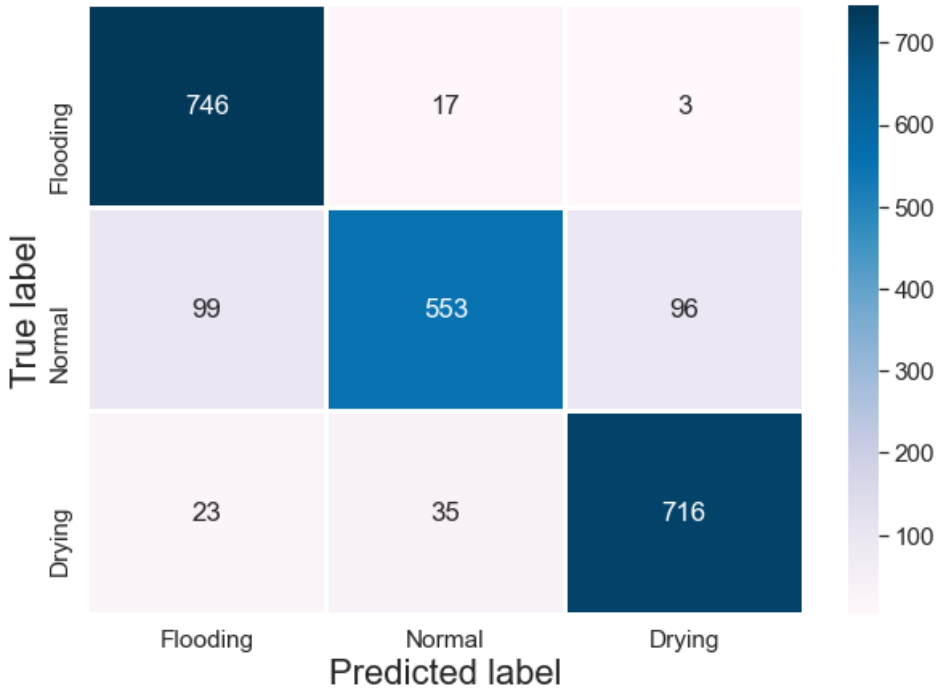


Figure 41. Confusion matrix for bagging model

4.4. DDPG Agent Training with Diagnosis of Flooding and Drying

4.4.1. Reward Function with Additional Cost for Moisture-related Errors

DDPG agent training is performed considering flooding and drying using the previously developed diagnostic model. Based on the results of the cell test, the magnitude of the voltage drop of the PEMFC is determined when flooding and drying occur. Additional fuel is consumed as the voltage drop causes the current

density to increase to meet the required power. DDPG agent should minimize flooding and drying to reduce fuel consumption. Also, even if the corresponding errors occur, it should swiftly return to the normal state. To accomplish this, cost term for flooding and drying is added to the reward function.

The cost associated with flooding and drying is computed by means of an LSTM-based bagging model and the voltage drop metrics associated with these phenomena. The probability of errors occurring, which are predicted by the diagnostic model, is used to determine the expected rate of additional fuel consumption. The cost including the additional fuel consumption rate is as follows:

$$\dot{m}_{fc} = -(\bar{\sigma}_F \dot{m}_F + \bar{\sigma}_N \dot{m}_N + \bar{\sigma}_D \dot{m}_D) \tag{44}$$

where $\bar{\sigma}_F$, $\bar{\sigma}_N$, and $\bar{\sigma}_D$ are the average probabilities for each state of the PEMFC predicted by the bagging models. The sum of probabilities for each state is equal to 1. $\dot{m}_F$, $\dot{m}_N$, and $\dot{m}_D$ are the expected fuel consumption rate caused by voltage drop due to each error. The driving system that considers the occurrences of flooding and drying is trained using the DDPG algorithm mentioned earlier by incorporating the associated costs into the reward function.

4.4.2. Training Results considering Flooding and Drying

A performance evaluation is conducted on the driving system that considers the occurrences of flooding and drying. The vehicle model equipped with the developed system is driven five consecutive times on the designated test road. The energy-

efficient driving assistance system that incorporates the diagnosis of flooding and drying is referred to as EDAS-D, while the system that does not consider it is denoted as EDAS. The results of applying the diagnosis model to the DDPG learning process are presented in Figure 42, which displays the vehicle speed, SOC, and PEMFC condition diagnosis outcomes. The progression of vehicle speed appears to be similar regardless of the inclusion of the diagnostic model. Both EDAS and EDAS-D models exhibit similar charging and discharging times, but there is a notable difference in power size. Specifically, the EDAS-D model employs the battery to assist with more power compared to the EDAS model without error diagnosis. Moreover, the EDAS

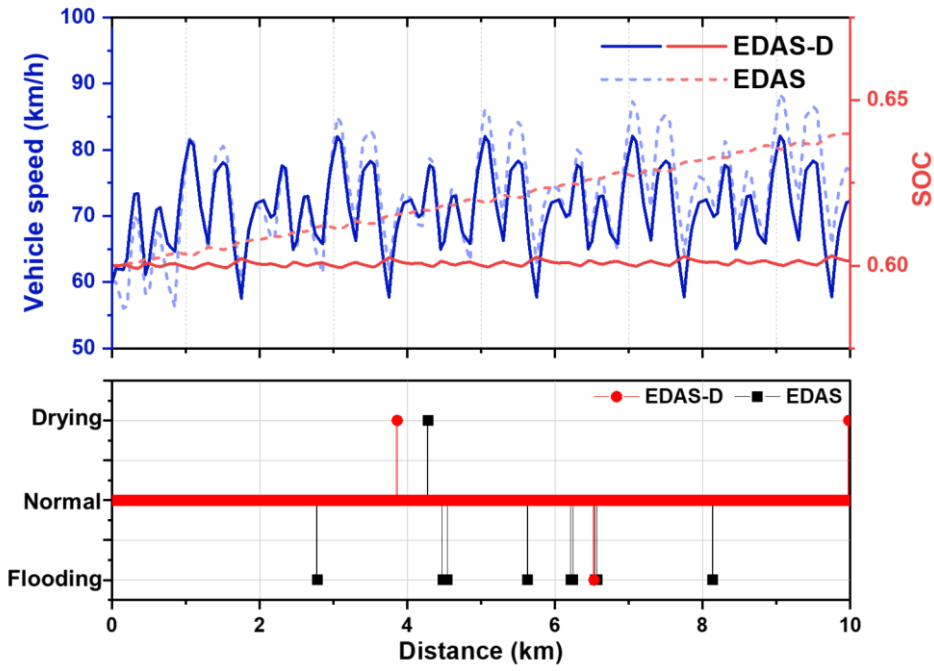


Figure 42. Vehicle speed, SOC and state of PEMFC according to the application of the diagnosis model of the DDPG algorithm.

model shows a higher frequency of short-duration occurrences of flooding and drying. Specifically, a significant portion of errors are related to the phenomenon of flooding. The voltage drop phenomenon when errors occur is presented in Figure 43. The voltage is initially reduced by approximately 0.8V, and as the error persists, the magnitude of the voltage drop gradually increases, up to a maximum of 2.33V. This underscores the significance of prompt action upon the occurrence of flooding and drying. The fuel consumption of the EDAS model is 92.14 g, while that of the EDAS-D model is 90.98 g, thereby reducing energy consumption by around 1.25%.

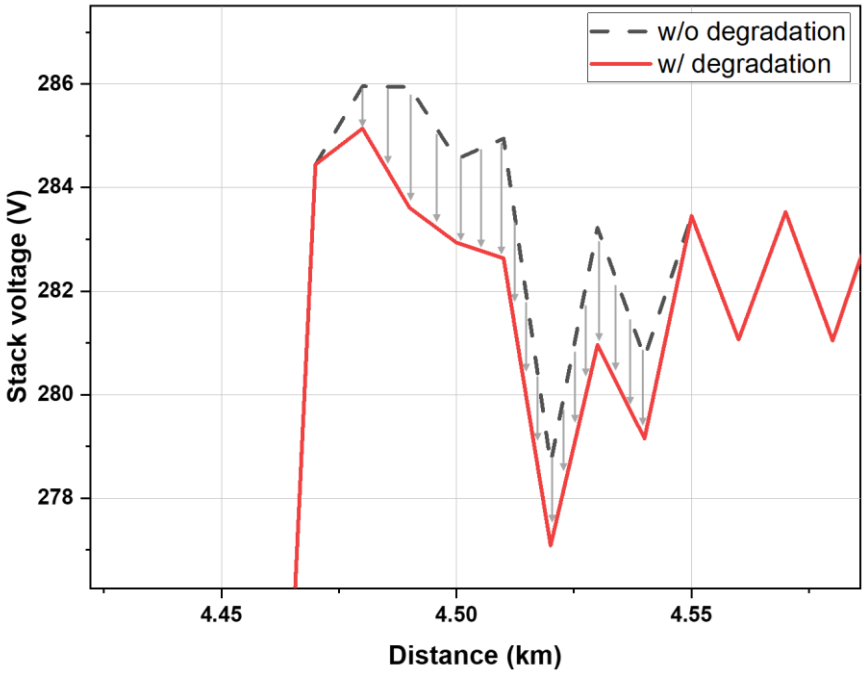


Figure 43. Voltage-drop during flooding and drying phenomena in PEMFC

Table 12 presents a comparative analysis of the effects of specific flooding and drying in each model during 20 consecutive drives on the test road. The EDAS model's longer duration of flooding and drying leads to an increase in total voltage-drop per step. Furthermore, it takes more time for EDAS to return to normal after an error occurs. The actions of the DDPG algorithm, i.e., vehicle acceleration and battery power, during flooding and drying are shown in Figure 44. The control of vehicle speed and power distribution in the EDAS model remains largely unchanged even if flooding and drying occur. Since the EDAS model independently determines the speed of the vehicle and the power control of the components only by the vehicle state and road gradient without considering the fuel cell state, it can be seen that flooding and drying occur in similar states. In contrast, the EDAS-D model exhibits a tendency to decelerate in case of flooding and to reduce the power allocated to the fuel cell. This reduces the water generated by low PEMFC loads, resulting in faster recovery from flooding. In the case of Drying, the vehicle accelerates slightly and the battery is slightly charged. A certain level of load is necessary on the PEMFC to provide water to the channel, but it should not be excessive because of low efficiency caused by voltage-drop.

Table 12. Comparative analysis of the effects of specific flooding and drying in EDAS and EDAS-D during 20 consecutive drives on the test road.

	EDAS	**EDAS-D**
Flooding duration (sec)	28.49	7.10
Drying duration (sec)	4.36	1.84
Total voltage-drop per step (V)	62.55	10.83
Average recovery time (sec)	0.7045	0.5956

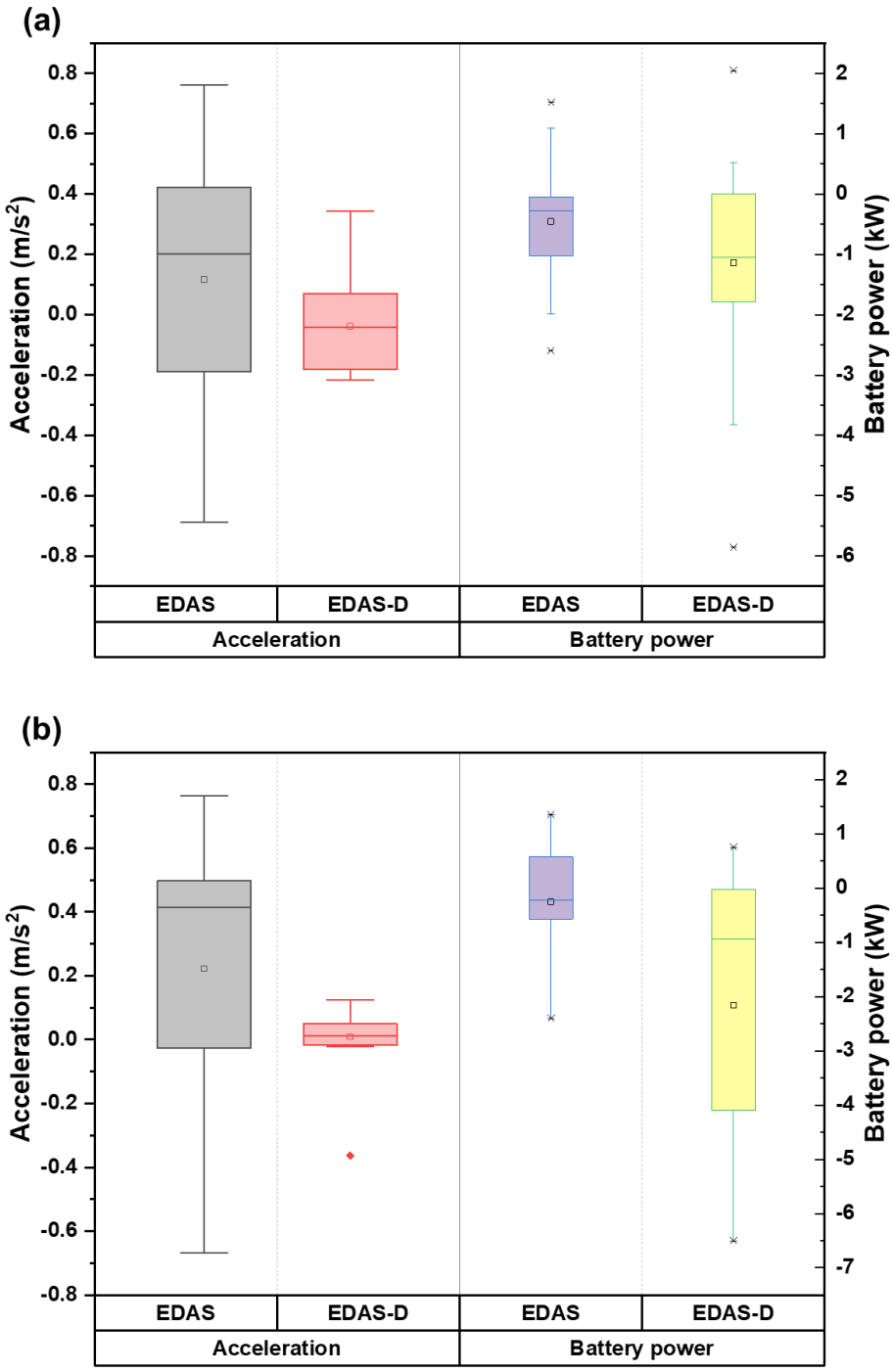


Figure 44. Box plots of DDPG algorithm actions: vehicle acceleration and battery power for (a) flooding and (b) drying

Chapter 5. Online-learning for Generality of the Driving System

5.1. Introduction

Previous chapters have demonstrated the impressive driving performance of the FCHEV's reinforcement learning-based system. However, in general, reinforcement learning has limitations in that training and testing are performed in the same environment [59]. If the environment is changed, the generalization capability of the agent is likely to be jeopardized due to reinforcement learning's dependency on the Bellman equation, which calculates the future Q value based on information about the future environment [60, 61]. Consequently, to enhance the agent's generality, perpetual online-learning is imperative in a novel environment. Online-learning is an extension of reinforcement learning that allows the agent to train and adapt continuously in real-time without access to the entire dataset. Online-learning algorithms use the reward signal to update the policy in real-time, improving the agent's decision-making ability. In this study, driving scenario is conducted on an unfamiliar road environment using a model generated through offline learning on the test road. Initially, the policy from the existing agent is delivered to the vehicle controller. During the course of driving, a mini-batch of acquired data is continuously collected, and online-learning is employed to enhance the agent's generalization capability.

5.2. Online-learning Process

This section describes the online-learning process of this study. Figure 45 shows a flowchart of the online learning process. In this process, the vehicle is driven on a new road environment that differs from the one used in the offline-learning phase. Initially, the vehicle adheres to the policy derived from the offline-trained model. Similar to the DDPG algorithm, the experience replay buffer is employed to store the experiences recorded during driving, including the state of the vehicle and action, the following state, and the reward in response to the action. When the replay memory surpasses the mini-batch size, a copy of the existing model is generated, and DDPG learning based on the copied model is performed using new driving data. It is important to note that, during the online-learning phase, the model being trained does not affect the driving and power distribution. The driving and learning procedures are conducted simultaneously, and when the driving distance reaches the update distance d_{seg} of the online-learning model, a test is performed to compare the existing model and the newly learned model. The evaluation of the newly learned model is carried out on the road where the vehicle is driven based on the policy of the existing model, and the performance of the existing model is determined from the driving results. The superior model of the two models is updated with a new driving model. This learning process is repeated until the vehicle reaches its final destination. Table 13 displays the parameters employed for online-learning in this study.

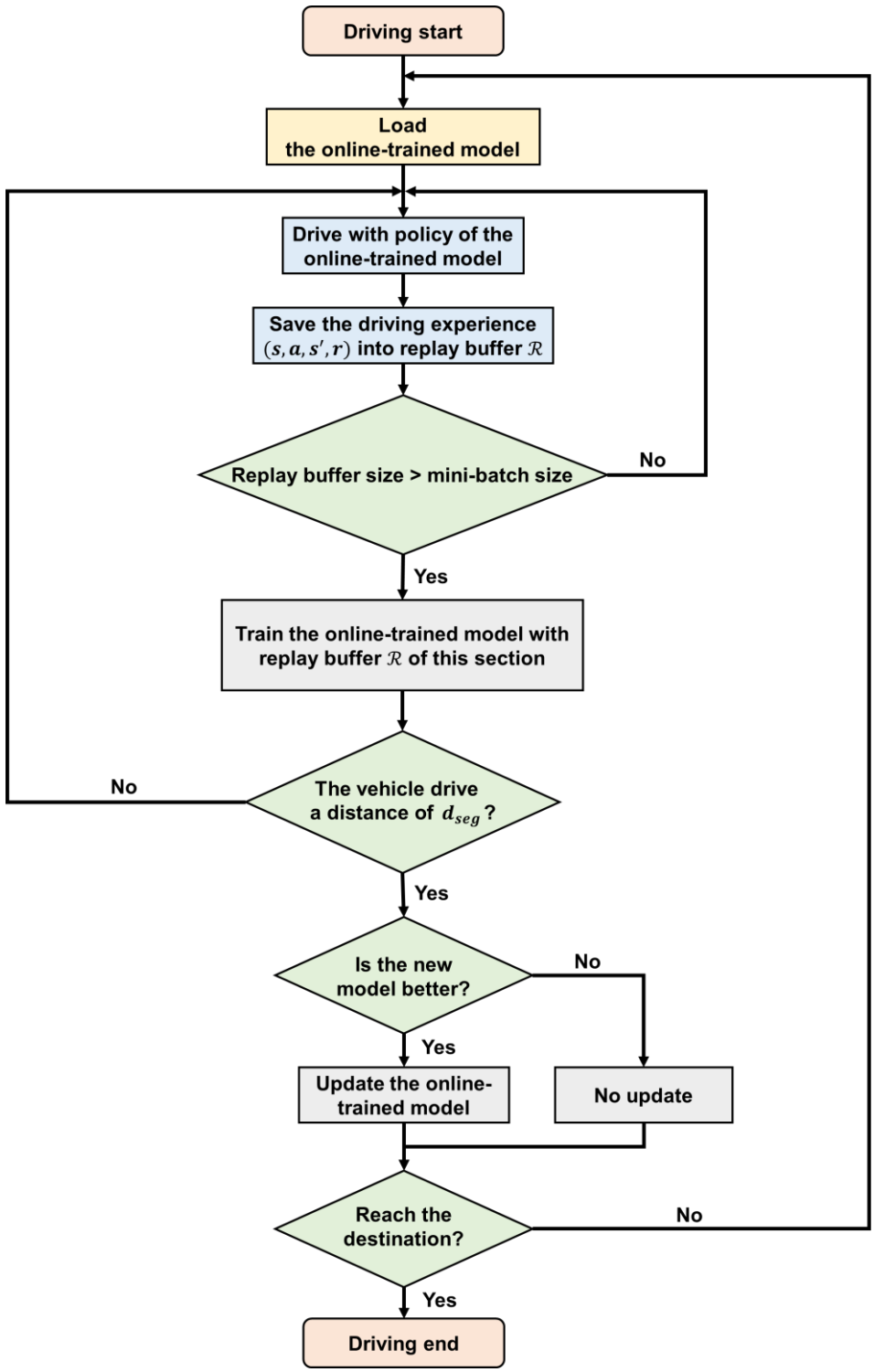


Figure 45. Flowchart of online learning process

Table 13. Hyperparameters of learning process for online-learning

Parameters	Values
Update distance of online-learning model d_{seg}	2 km
Experience memory buffer size	1000
Learning rate of actor networks	1e-5
Learning rate of critic networks	1e-5
Discount factor γ	0.99
Update parameter for target networks τ	0.001
Batch size	32
Optimizer	Adam

5.3. Online-learning for Enhancing Generality in DDPG under Environmental Changes

To assess and enhance the generality of the proposed vehicle driving system, online-learning is conducted. The previously learned DDPG agent, EDAS-D, is tested on a new road and compared with the continuously updated system that has undergone online-learning during driving. The road route for online-learning and its elevation profile based on distance are depicted in Figure 46. This route spans a total distance of 26.28 km, from Seocho IC to Suwonsingal IC in Korea, featuring a steep initial slope and subsequently gentle slopes, providing the agent with diverse experiences. To facilitate exploration by the agent in a completely different environment, noise is randomly sampled from a distribution with mean 0 and standard deviation of 0.1.

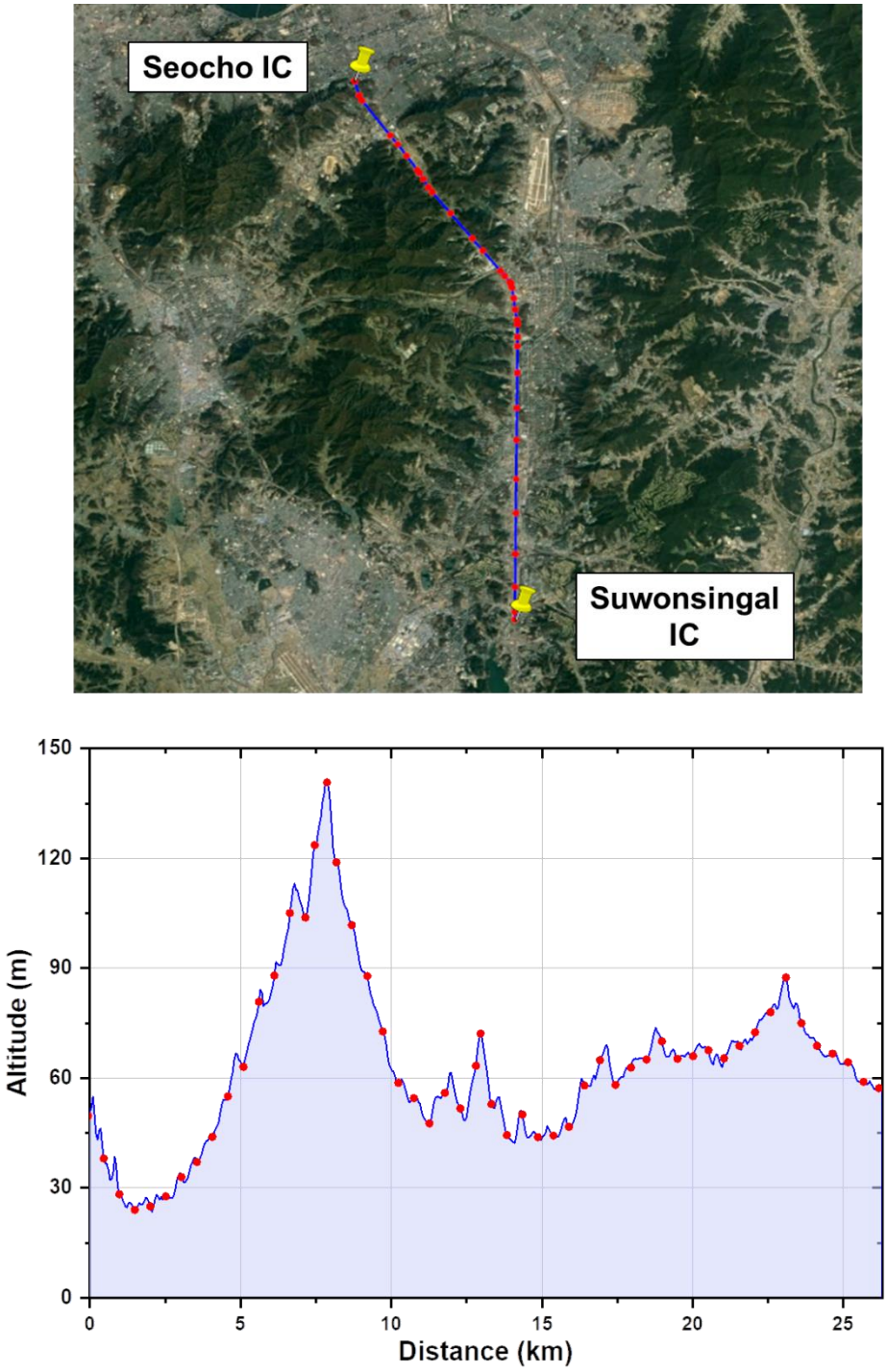


Figure 46. Road route and elevation profile for online-learning (Seocho IC to Suwonsingal IC)

Cumulative fuel consumption of two vehicles, one equipped with only the EDAS-D model without online-learning and the other with continuous updates based on the EDAS-D model (EDAS-online), is shown in Figure 47. The initial driving is performed using the EDAS-D model in both vehicles, and hence the cumulative fuel consumption is the same during the initial online-learning process. The model is updated a total of three times during the online-learning process, and a slight difference in cumulative fuel consumption is observed initially. However, as the learning progresses, the DDPG agent adapts to the new road, leading to a gradual widening of the gap in fuel consumption between the two models.

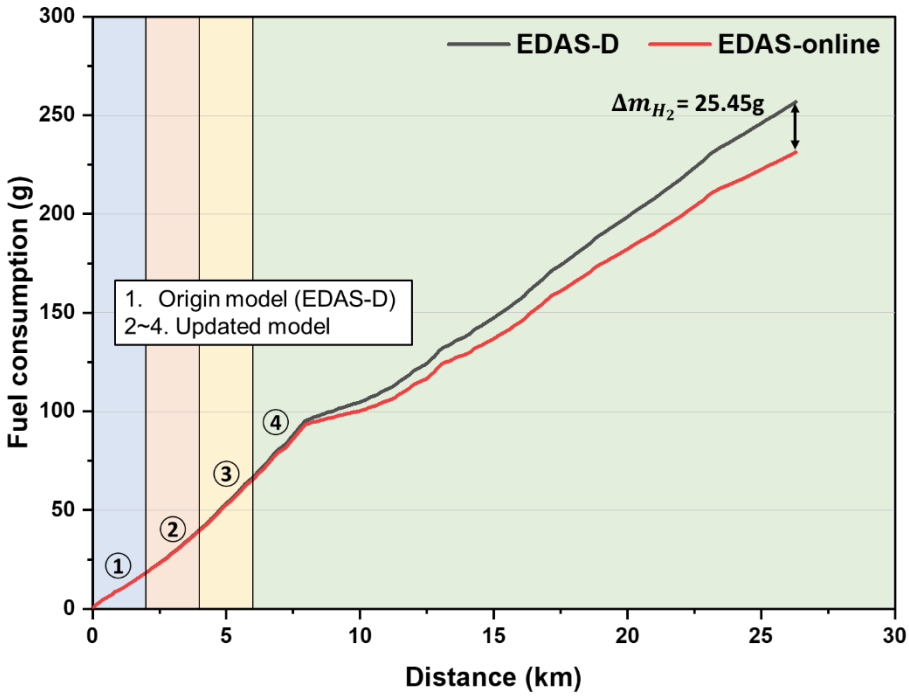


Figure 47. Cumulative fuel consumption of EDAS-D and EDAS-online

The graph presented in Figure 48 displays the vehicle speed, SOC, and PEMFC state of EDAS-D and EDAS-online. Both models exhibit a similar trend in terms of vehicle speed, with a slight decrease observed as the model undergoes updates. The SOC charging and discharging trends were found to be similar for both EDAS-D and EDAS-online models; however, the EDAS-online model exhibited a higher discharge rate. Table 14 presents the detailed driving results of EDAS-D and EDAS-online. The results reveal that online-learning lead to a 5.59% reduction in equivalent fuel consumption. The EDAS-online model exhibits more frequent flooding at 12.74 sec and less drying at 4.24 sec. However, due to the rapid voltage-drop associated with flooding, the total voltage-drop was greater for EDAS-online. While online-learning do not significantly prevent moisture-related errors, it is confirmed that the speed control and power distribution close to the optimum for the new road significantly contribute to the increase in energy efficiency.

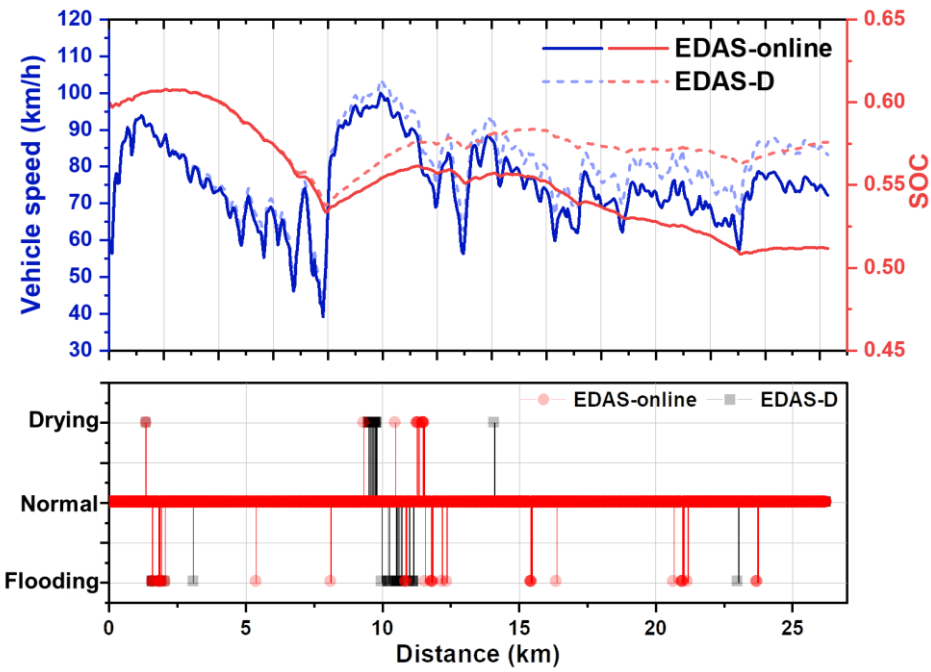


Figure 48. Vehicle speed, SOC and state of PEMFC of EDAS-D and EDAS-online

Table 14. Detailed driving results of EDAS-D and EDAS-online

	EDAS-D	**EDAS-online**
Equivalent fuel consumption (g)	260.85	246.26 (+5.59%)
Flooding duration (sec)	9.77	12.74
Drying duration (sec)	7.65	4.24
Total voltage-drop per step (V)	22.24	24.60

5.4. Analysis of Online Learning Performance according to Offline Base Models

In this section, an analysis of the performance of online learning based on the learning level of the offline model is conducted. The process involves uploading three different versions of the learning model to the driving controller when navigating a new road: the initial model with an unconverged return, an intermediate model in the midst of the learning process, and a final model with a converged return. The training history of the offline model is presented in Figure 49, where the red points indicate the return values of the base models used for analyzing the performance of online learning. In the process of online learning, the coefficients of the reward function are configured to match the values employed in the training of the offline model. The update point for the models during online learning is depicted in Figure 50. Table 15 provides information on the return value and update count of the base model for each episode. At the beginning of learning, the base model is updated in every update interval. The frequency of updates decreases until the midpoint of the episodes for the base model. Subsequently, the number of updates remains constant until the model reaches convergence.

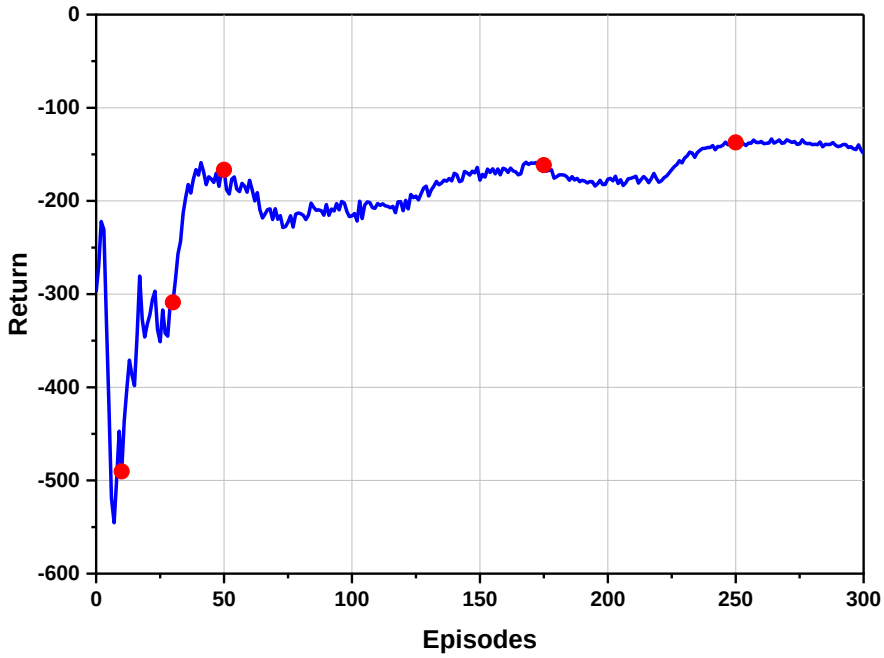


Figure 49. Return history of offline models

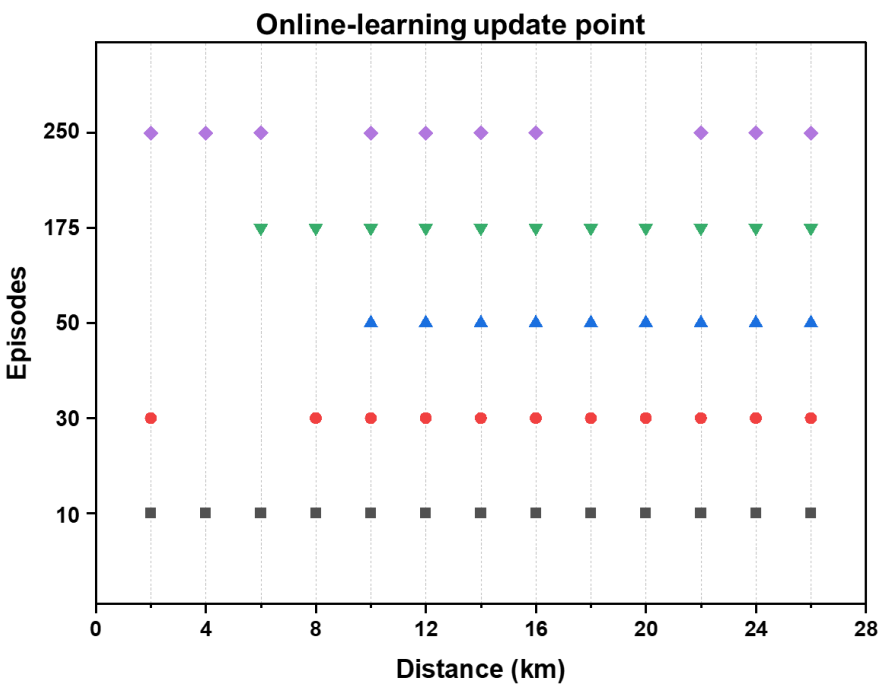


Figure 50. Online-learning update point according to offline models

Table 15. Episodes, return of base models and number of model update in online learning

Episodes	Return	Number of updates
10	-490.32	13
30	-308.92	11
50	-166.50	9
175	-161.62	11
250	-137.19	10

Figure 51 presents equivalent fuel consumption, flooding duration, drying duration, and improvements of the offline base model and the online model according to the learning episode. It is observed that when online learning is conducted using a highly trained model, a significant improvement in fuel consumption is achieved. This suggests that utilizing a well-learned model for online learning results in faster enhancements in new road environments. Additionally, when employing a less trained base model for online learning, there is a decrease in the duration of flooding. If the base model has received sufficient training, flooding incidents are less likely to occur, rendering online learning less impactful in reducing flooding occurrences. As for drying duration, it is evident that the effect of online learning is minimal, as the offline model already demonstrates satisfactory performance even in new environments.

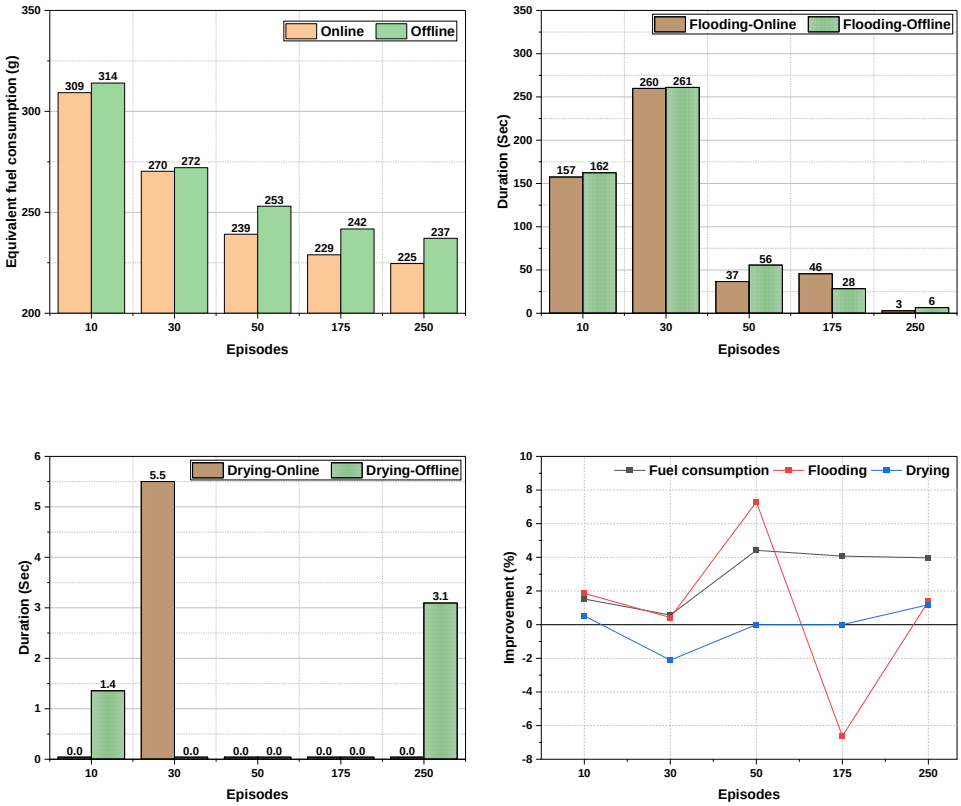


Figure 51. Equivalent fuel consumption, flooding duration, drying duration and improvement of Online models and Offline models

Chapter 6. Conclusion and Future works

6.1. Conclusion

This study developed an energy-efficient driving assistance system for autonomous FCHEVs using an AI model based on reinforcement learning. The study aimed to develop the energy management strategy of an FCHEVs by developing an AI model that presents a reference for speed control and power distribution. The vehicle was modeled using a backward-looking simulation method, and a semi-empirical model was developed based on data obtained by fabricating and testing the PEMFC. Based on the developed vehicle model, a reward function capable of controlling fuel consumption, driving time, and SOC was developed. The results of the study showed that the developed DDPG model with the developed reward function improved energy efficiency by 36.52% compared to the method combining cruise control and rule-based strategy that are actually used. It also achieved an optimality of 97.11%.

The driving system was additionally developed considering flooding and drying, which are chronic problems of PEMFC, the power source of FCHEV. The study intentionally caused moisture-related errors with the fabricated PEMFC to identify patterns when flooding and drying occurred. A model was trained to diagnose flooding and drying using LSTM based on the sequential data obtained from the error-induced experiments. The bagging ensemble method was used to improve the performance by integrating the output of multiple LSTM models, achieving an accuracy of 88.1%. The detection rate for flooding was particularly high at 97.39%.

Furthermore, a driving model was trained by adding an additional fuel consumption rate to the reward function using the average of the probability of flooding and drying, which is the output of the bagging ensemble model. The occurrence of flooding and drying was observed to be less frequent, and the recovery time for normalization of PEMFC was improved during such events. As a consequence, fuel consumption was reduced by 1.25% compared to the driving system without the diagnostic system.

The nature of the Bellman equation used in reinforcement learning leads to a deterioration in generality when the environment changes. In order to overcome this limitation, online learning was conducted while the vehicle model was driving on a new real road. This approach enables the reinforcement learning algorithm to adapt to changing environments and to continuously improve the performance of the driving assistance system in real-world scenarios. Although the online-learning model did not yield significant improvements in preventing flooding and drying, the developed methodology resulted in a fuel consumption reduction of approximately 5.59%. This outcome was attributed to the near-optimal speed control and power distribution achieved through the reinforcement learning algorithm. Despite the limited impact on moisture-related stability, the developed model still demonstrates the potential for improving the energy efficiency of FCHEVs.

This study proposed a novel methodology that addresses both speed control and power distribution using the DDPG algorithm, which achieves high optimality for solving these two nonlinear problems. By constructing dual neural networks, the reinforcement learning model can ensure not only energy efficiency but also moisture-related stability of the PEMFC. This methodology represents a significant contribution as it is the first to integrate these two functions using the DDPG algorithm. Additionally, the incorporation of online learning further enhances the effectiveness and robustness of the system under diverse driving conditions and

environments. Overall, the study provided insights for the development of energy-efficient driving assistance systems that can manage moisture-related issues in FCHEVs.

6.2. Future works

Future research intends to develop a more improved and realistic energy-efficient driving system by utilizing the expanded V2I-based information. In addition to geographical information on the road, V2I communication also allows vehicles to access surrounding vehicle information and traffic light information. The energy-efficient driving assistance system developed based on this information has great potential.

Interaction with surrounding vehicles is a critical aspect of driving in real-world scenarios. The state of the surrounding vehicles has a direct impact on the speed control limitations of a vehicle. Factors such as the distance between vehicles, the speed of the preceding vehicle, and the safety distance, impose speed and acceleration constraints, thereby restricting the vehicle's operation to within the allowable control range. And Traffic signals play a crucial role in achieving energy efficiency while driving. Vehicle idling induced by stopping at traffic signals results in wasteful energy consumption. The operation of traffic lights is influenced by both time and the location of vehicles, as the decision to continue driving or halt is made according to the traffic light's current status when the vehicle reaches the intersection. Consequently, the traffic light is represented as a function of both the position of the vehicle and the duration of the driving time. In the event that the traffic light is red and the vehicle arrives at the intersection, a penalty is imposed to encourage the

driving system to avoid coming to a halt at the signal. Drawing on the V2I-based information, a prospective driving system is envisaged to be constructed in the future. The system is intended to integrate more precise constraining conditions, utilizing real-time data on nearby vehicle positions, as well as current speed and acceleration. Additionally, appropriate incentives will be established in line with the prevailing traffic signal conditions, leading to a more practical driving system.

Bibliography

[1] Weiskopf SR, Rubenstein MA, Crozier LG, Gaichas S, Griffis R, Halofsky JE, et al. Climate change effects on biodiversity, ecosystems, ecosystem services, and natural resource management in the United States. Sci Total Environ 2020;733. https://doi.org/10.1016/j.scitotenv.2020.137782.

[2] Schleussner CF, Rogelj J, Schaeffer M, Lissner T, Licker R, Fischer EM, et al. Science and policy characteristics of the Paris Agreement temperature goal. Nat Clim Chang 2016;6:827–35. https://doi.org/10.1038/nclimate3096.

[3] United Nations Climate Change, "Nationally Determined Contributions Registry." https://unfccc.int/NDCREG

[4] Berkeley earth, "Global Temperature Report for 2022." https://berkeleyearth.org/global-temperature-report-for-2022/

[5] Our World in Data, "Emissions by sector." https://ourworldindata.org/emissions-by-sector

[6] Statista, " Size of the global autonomous vehicle market in 2021 and 2022, with a forecast through 2030." https://www.statista.com/statistics/1224515/av-market-size-worldwide-forecast/

[7] Han J, Shen D, Jeong J, Russo M Di, Kim N, Grave JJ, et al. Energy Impact of Connecting Multiple Signalized Intersections to Energy-Efficient Driving: Simulation and Experimental Results. IEEE Control Syst Lett 2023;7:1297–302. https://doi.org/10.1109/lcsys.2023.3234808.

[8] Zhou H, Xu W, Chen J, Wang W. Evolutionary V2X Technologies Toward the Internet of Vehicles: Challenges and Opportunities. Proc IEEE 2020;108:308–23. https://doi.org/10.1109/JPROC.2019.2961937.

[9] Yu P, Li M, Wang Y, Chen Z. Fuel Cell Hybrid Electric Vehicles: A Review of Topologies and Energy Management Strategies. World Electr Veh J 2022;13. https://doi.org/10.3390/wevj13090172.

[10] Wang X, Park S, Han K. Energy-Efficient Speed Planner for Connected and Automated Electric Vehicles on Sloped Roads. IEEE Access 2022;10:34654–64. https://doi.org/10.1109/ACCESS.2022.3162871.

[11] Lee H, Kim K, Kim N, Cha SW. Energy efficient speed planning of electric vehicles for car-following scenario using model-based reinforcement learning. Appl Energy 2022;313:118460. https://doi.org/10.1016/j.apenergy.2021.118460.

[12] Ahmadi S, Bathaee SMT, Hosseinpour AH. Improving fuel economy and performance of a fuel-cell hybrid electric vehicle (fuel-cell, battery, and ultra-capacitor) using optimized energy management strategy. Energy Convers Manag 2018;160:74–84. https://doi.org/10.1016/j.enconman.2018.01.020.

[13] Lin X, Xu X, Lin H. Predictive-ECMS based degradation protective control strategy for a fuel cell hybrid electric vehicle considering uphill condition. ETransportation 2022;12:100168. https://doi.org/10.1016/j.etran.2022.100168.

[14] Zheng C, Li W, Li W, Xu K, Peng L, Cha SW. A Deep Reinforcement Learning-Based Energy Management Strategy for Fuel Cell Hybrid Buses. Int J Precis Eng Manuf - Green Technol 2022;9:885–97. https://doi.org/10.1007/s40684-021-00403-x.

[15] Zhou J, Liu J, Xue Y, Liao Y. Total travel costs minimization strategy of a dual-stack fuel cell logistics truck enhanced with artificial potential field and deep reinforcement learning. Energy 2022;239:121866. https://doi.org/10.1016/j.energy.2021.121866.

[16] Kim Y, Figueroa-Santos M, Prakash N, Baek S, Siegel JB, Rizzo DM. Co-optimization of speed trajectory and power management for a fuel-cell/battery electric vehicle. Appl Energy 2020;260:114254. https://doi.org/10.1016/j.apenergy.2019.114254.

[17] Nie Z, Jia Y, Wang W, Chen Z, Outbib R. Co-optimization of speed planning and energy management for intelligent fuel cell hybrid vehicle considering complex traffic conditions. Energy 2022;247:123476. https://doi.org/10.1016/j.energy.2022.123476.

[18] Afra M, Nazari M, Kayhani MH, Sharifpur M, Meyer JP. 3D experimental visualization of water flooding in proton exchange membrane fuel cells. Energy 2019;175:967–77. https://doi.org/10.1016/j.energy.2019.03.128.

[19] Ijaodola OS, El- Hassan Z, Ogungbemi E, Khatib FN, Wilberforce T, Thompson J, et al. Energy efficiency improvements by investigating the

water flooding management on proton exchange membrane fuel cell (PEMFC). Energy 2019;179:246–67. https://doi.org/10.1016/j.energy.2019.04.074.

[20] Huang Z, Jian Q, Zhao J. Experimental study on improving the dynamic characteristics of open-cathode PEMFC stack with dead-end anode by condensation and circulation of hydrogen. Int J Hydrogen Energy 2020;45:19858–68. https://doi.org/10.1016/j.ijhydene.2020.05.108.

[21] Laribi S, Mammar K, Sahli Y, Koussa K. Air supply temperature impact on the PEMFC impedance. J Energy Storage 2018;17:327–35. https://doi.org/10.1016/j.est.2018.03.020.

[22] Izhan Noor Azam AM, Choon PM, Masdar MS, Zainoodin AM, Husaini T. Performance and water transport behaviour in Polymer Electrolyte Membrane fuel cells. Int J Hydrogen Energy 2022. https://doi.org/10.1016/j.ijhydene.2021.12.146.

[23] Yang Z, Du Q, Jia Z, Yang C, Jiao K. Effects of operating conditions on water and heat management by a transient multi-dimensional PEMFC system model. Energy 2019;183:462–76. https://doi.org/10.1016/j.energy.2019.06.148.

[24] Chen K, Laghrouche S, Djerdir A. Degradation model of proton exchange membrane fuel cell based on a novel hybrid method. Appl Energy 2019;252:113439. https://doi.org/10.1016/j.apenergy.2019.113439.

[25] Guilbert D, Gaillard A, N'Diaye A, Djerdir A. Fault-tolerant control for PEMFC and its DC/DC converter. IYCE 2015 - Proc 2015 5th Int Youth Conf Energy 2015:8–13. https://doi.org/10.1109/IYCE.2015.7180758. [26] Lee H, Kim N, Cha SW. Model-Based Reinforcement Learning for Eco-Driving Control of Electric Vehicles. IEEE Access 2020;8:202886–96. https://doi.org/10.1109/ACCESS.2020.3036719.

[27] Ou M, Zhang R, Shao Z, Li B, Yang D, Ming P, et al. A novel approach based on semi-empirical model for degradation prediction of fuel cells. J Power Sources 2021;488:229435. https://doi.org/10.1016/j.jpowsour.2020.229435.

[28] James SS, Anderson SR, Lio M Da. Longitudinal Vehicle Dynamics: A Comparison of Physical and Data-Driven Models under Large-Scale Real-

World Driving Conditions. IEEE Access 2020;8:73714–29. https://doi.org/10.1109/ACCESS.2020.2988592.

[29] E J, Zhang B, Zeng Y, Wen M, Wei K, Huang Z, et al. Effects analysis on active equalization control of lithium-ion batteries based on intelligent estimation of the state-of-charge. Energy 2022;238:121822. https://doi.org/10.1016/j.energy.2021.121822.

[30] Ruan S, Ma Y, Yang N, Xiang C, Li X. Real-time energy-saving control for HEVs in car-following scenario with a double explicit MPC approach. Energy 2022;247:123265. https://doi.org/10.1016/j.energy.2022.123265.

[31] Wang Y, Wu K, Zhao H, Li J, Sheng X, Yin Y, et al. Degradation prediction of proton exchange membrane fuel cell stack using semi-empirical and data-driven methods. Energy AI 2023;11:100205. https://doi.org/10.1016/j.egyai.2022.100205.

[32] El-Fergany AA, Hasanien HM, Agwa AM. Semi-empirical PEM fuel cells model using whale optimization algorithm. Energy Convers Manag 2019;201:112197. https://doi.org/10.1016/j.enconman.2019.112197.

[33] Akimoto Y, Okajima K. Simple on-board fault-detection method for proton exchange membrane fuel cell stacks using by semi-empirical curve fitting. Appl Energy 2021;303:117654. https://doi.org/10.1016/j.apenergy.2021.117654.

[34] R. O'Hayre, S. W. Cha, W. Colella, F. B. Prinz, "Fuel Cell Fundamentals (3rd edition)", Wiley, 2016

[35] Kim MS, Kim J, Kim SY, Chu CH, Rho KH, Kim M, et al. Parametric study on the performance of electrochemical hydrogen compressors. Renew Energy 2022;199:1176–88. https://doi.org/10.1016/j.renene.2022.09.081.

[36] Dutta S, Shimpalee S, Van Zee JW. Numerical prediction of mass-exchange between cathode and anode channels in a PEM fuel cell. Int J Heat Mass Transf 2001;44:2029–42. https://doi.org/10.1016/S0017-9310(00)00257-X.

[37] Chugh S, Chaudhari C, Sonkar K, Sharma A, Kapur GS, Ramakumar SSV. Experimental and modelling studies of low temperature PEMFC performance. Int J Hydrogen Energy 2020;45:8866–74. https://doi.org/10.1016/j.ijhydene.2020.01.019.

[38] Hou Y, Liu L, Wei Q, Xu X, Chen C. A novel DDPG method with prioritized experience replay. 2017 IEEE Int Conf Syst Man, Cybern SMC 2017 2017;2017-Janua:316–21. https://doi.org/10.1109/SMC.2017.8122622.

[39] Chen Y, Han W, Zhu Q, Liu Y, Zhao J. Target-driven obstacle avoidance algorithm based on DDPG for connected autonomous vehicles. EURASIP J Adv Signal Process 2022;2022. https://doi.org/10.1186/s13634-022-00872-5.

[40] Fernandez-Fernandez R, Aggravi M, Giordano PR, Victores JG, Pacchierotti C. Neural Style Transfer with Twin-Delayed DDPG for Shared Control of Robotic Manipulators. Proc - IEEE Int Conf Robot Autom 2022:4073–9. https://doi.org/10.1109/ICRA46639.2022.9812245.

[41] Gong C, He Q, Bai Y, Hou X, Fan G, Liu Y. Wide-Sense Stationary Policy Optimization With Bellman Residual on Video Games. Proc - IEEE Int Conf Multimed Expo 2021. https://doi.org/10.1109/ICME51207.2021.9428293.

[42] Yang J, Peng G. DDPG with Meta-Learning-Based Experience Replay Separation for Robot Trajectory Planning. 2021 7th Int Conf Control Autom Robot ICCAR 2021 2021:46–51. https://doi.org/10.1109/ICCAR52225.2021.9463493.

[43] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. 4th Int Conf Learn Represent ICLR 2016 - Conf Track Proc 2016.

[44] Kim K, Lee H, Song C, Kang C, Cha SW. Optimization of speed trajectory for eco-driving considering road characteristics. 2018 IEEE Veh Power Propuls Conf VPPC 2018 - Proc 2019:0–4. https://doi.org/10.1109/VPPC.2018.8605038.

[45] Wang J, Wang J, Wang Q, Zeng X. Control rules extraction and parameters optimization of energy management for bus series-parallel AMT hybrid powertrain. J Franklin Inst 2018;355:2283–312. https://doi.org/10.1016/j.jfranklin.2017.12.039.

[46] Zhang Y, Wei C, Liu Y, Chen Z, Hou Z, Xu N. A novel optimal power management strategy for plug-in hybrid electric vehicle with improved adaptability to traffic conditions. J Power Sources 2021;489:229512. https://doi.org/10.1016/j.jpowsour.2021.229512.

[47] Zheng CH, Oh CE, Park YI, Cha SW. Fuel economy evaluation of fuel cell hybrid vehicles based on equivalent fuel consumption. Int J Hydrogen Energy 2012;37:1790–6. https://doi.org/10.1016/j.ijhydene.2011.09.147.

[48] Choi H, Kim J, Kwon O, Yoo H, Kim H, Cha H, et al. Observation of flooding-induced performance enhancement in PEMFCs. Int J Hydrogen Energy 2022;47:6259–68. https://doi.org/10.1016/j.ijhydene.2021.11.234.

[49] Ren P, Pei P, Li Y, Wu Z, Chen D, Huang S, et al. Diagnosis of water failures in proton exchange membrane fuel cell with zero-phase ohmic resistance and fixed-low-frequency impedance. Appl Energy 2019;239:785–92. https://doi.org/10.1016/j.apenergy.2019.01.235.

[50] Hochreiter S, Schmidhuber J ¨urgen. Long Short-Term Memory. Neural Comput 1997;9:1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

[51] Zhou J, Cui G, Hu S, Zhang Z, Yang C, Liu Z, et al. Graph neural networks: A review of methods and applications. AI Open 2020;1:57–81. https://doi.org/10.1016/j.aiopen.2021.01.001.

[52] G´eron A. Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: concepts, tools, and techniques to build intelligent systems. 2019.

[53] Kim K, Kim J, Choi H, Kwon O, Jang Y, Ryu S, et al. Pre-diagnosis of flooding and drying in proton exchange membrane fuel cells by bagging ensemble deep learning models using long short-term memory and convolutional neural networks. Energy 2023;266:126441. https://doi.org/10.1016/j.energy.2022.126441.

[54] Hong H, Liu J, Zhu AX. Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci Total Environ 2020;718:137231. https://doi.org/10.1016/j.scitotenv.2020.137231.

[55] Hong H, Liu J, Zhu AX. Modeling landslide susceptibility using LogitBoost alternating decision trees and forest by penalizing attributes with the bagging ensemble. Sci Total Environ 2020;718:137231. https://doi.org/10.1016/j.scitotenv.2020.137231.

[56] Mosavi A, Sajedi Hosseini F, Choubin B, Goodarzi M, Dineva AA, Rafiei Sardooi E. Ensemble Boosting and Bagging Based Machine Learning

Models for Groundwater Potential Prediction. Water Resour Manag 2021;35:23–37. https://doi.org/10.1007/s11269-020-02704-3.

[57] Lin J, Chen H, Li S, Liu Y, Li X, Yu B. Accurate prediction of potential druggable proteins based on genetic algorithm and Bagging-SVM ensemble classifier. Artif Intell Med 2019;98:35–47. https://doi.org/10.1016/j.artmed.2019.07.005.

[58] Skurichina M, Duin RPW. Bagging , Boosting and the Random Subspace Method for Linear Classifiers 2002:121–35.

[59] Zhang R, Zong Q, Zhang X, Dou L, Tian B. Game of Drones: Multi-UAV Pursuit-Evasion Game With Online Motion Planning by Deep Reinforcement Learning. IEEE Trans Neural Networks Learn Syst 2022:1–10. https://doi.org/10.1109/TNNLS.2022.3146976.

[60] Cobbe K, Klimov O, Hesse C, Kim T, Schulman J. Quantifying generalization in reinforcement learning. 36th Int Conf Mach Learn ICML 2019 2019;2019-June:2280–93.

[61] Hansen N, Wang X. Generalization in Reinforcement Learning by Soft Data Augmentation. Proc - IEEE Int Conf Robot Autom 2021;2021-May:13611–7. https://doi.org/10.1109/ICRA48506.2021.9561103.

국문 초록

최근 급격한 기후 변화로 인한 지구 온난화 문제는 전 세계적으로 큰 관심을 받고 있다. 차량 업계에서도 온실 가스 배출을 줄이기 위해 방안들을 모색하고 있다. 첫번째로는 차량-인프라 통신 (Vehicle-to-infrastructure, V2I) 기반으로 자율 주행 기술의 발전으로 최소의 연료를 소모하도록 에너지 효율 주행을 하는 것이다. 두번째는 친환경 연료를 사용하는 고분자 전해질 연료전지 (Polymer electrolyte membrane fuel cell, PEMFC)로 차량 에너지 동력원을 변경하고 하이브리드화하는 것이다. 하지만 이러한 방안들은 해결과제들을 가지고 있다.

먼저 연료전지 하이브리드 전기 자동차 (Fuel cell hybrid electric vehicles, FCHEVs)의 에너지 효율 주행을 하기 위해서 차량의 내, 외적인 정보를 기반으로 최적의 속도 컨트롤과 고분자 전해질 연료전지, 배터리간의 요구 동력 분배 전략 개발이 필수 적이다. 위 두 가지 개발방안은 최고의 성능을 위해서는 최적화 기법이 요구된다. 하지만 긴 계산 시간 때문에 최적화 기법은 실제 차량에 적용시키기 어렵기 때문에 적절한 컨트롤 전략이 필요하다. 또한 연료전지 하이브리드 전기 자동차의 에너지 동력원인 고분자 전해질 연료전지는 플러딩 (Flooding), 드라잉 (drying)이라는 치명적인 결함이 존재한다. 이러한 결함들은 연료전지가 극한의 상황에서 장시간 작동되었을 때 발생되기 때문에 주행 컨트롤과 배터리의 동력 어시스트로 통해 안정적인 조건에서 연료전지를 가동시켜야 한다. 따라서 본 논문에서는 에너지 효율 속도 컨트롤, 동력분배 전략, 고분자 전해질 연료전지의 수분 관련 결함을 고려한 연료전지 하이브리드 전기 자동차를 위한 주행 시스템을 제시한다.

먼저 후방향 시뮬레이션 (Backward-looking simulation)으로 연료전지 하이브리드 전기 자동차의 모델링을 수행한다. 또한 단일 연료전지 셀을 제작하고 실험 결과값을 참조하여 준 경험적 고분자 전해질 연료전지 모델을

개발한다. 개발된 차량 모델을 이용하여 강화학습의 일종인 심층 결정적 정책 경사법 (Deep deterministic policy gradient, DDPG)으로 주행 시스템을 학습시킨다. 이 시스템은 차량의 목표 속도와 연료전지, 배터리의 컨트롤러에게 동력 분배에 대한 참조 값을 제시한다. 매 스텝마다 파워트레인의 제한 조건을 넘지 않도록 심층 결정적 정책 경사법의 행동 공간 (Action space)을 갱신함으로써, 차량모델이 시스템에서 제시한 행동에 대한 수행 가능성을 높인다. 파라미터 최적화를 통해 심층 결정적 정책 경사법에 적합한 파라미터를 적용하여 모델의 성능을 개선한다. 또한 학습에 적용된 도로의 구배에 따른 모델의 연료소모와 작동점을 분석한다. 제안된 시스템은 글로벌 최적화 방법인 동적 계획법 (Dynamic programming, DP)과 비교하여 97.11 % 최적성을 보였으며, 크루즈 컨트롤 (Cruise control)과 규칙 기반 전략 (Rule-based strategy) 기반의 컨트롤보다 36.52 % 우수한 성능을 보였다.

에너지 효율 주행 시스템의 연료전지 하이브리드 차량에서 플러딩과 드라잉의 발생 여부를 파악하기 위해 해당 결함들을 진단하는 모델을 개발한다. 의도적으로 플러딩과 드라잉을 유발하는 실험을 수행하고, 각각 결함 발생시의 전기화학적인 데이터를 분석한다. 습득된 시계열 데이터를 기반으로 장단기 기억 신경망 (Long-short term memory, LSTM)과 배깅 앙상블 방법 (Bootstrap aggregation, Bagging)을 이용해서 진단 모델을 개발한다. 진단 시스템의 Flooding 과 drying 에 대한 진단율은 88.11 %을 달성했다. 진단 모델의 출력 값을 심층 결정적 정책 경사법의 보상 함수 (Reward function)에 포함시켜 고분자 전해질 연료전지의 수분 관련 결함을 고려한 에너지 효율 주행 시스템을 개발한다. 연료 전지 상태 진단이 통합된 주행 시스템은 플러딩과 드라잉에 대한 감소가 확인되었으며, 발생 후에 평균 0.5956 초 후에 정상화되었다. Error 회피로 인해 기존 주행 모델에 비해 연료 소모 효율이 약 1.25 % 개선되었다.

개발된 주행 시스템의 일반성을 확인하기위해 새로운 도로 환경에서 차량 모델을 주행 시킨다. 미래의 Q 값을 갱신하는 강화학습의 벨만 방정식

(Bellman equation)의 특성상 환경 (Environment) 변경되었을 때 최적성이 떨어진다. 따라서 성능 감소를 방지하기 위해 온라인 러닝 (Online-learning)을 수행한다. 또한 오프라인 러닝의 학습 수렴성에 대한 온라인 러닝의 효과를 검증한다. 온라인 러닝이 수행된 모델은 기존 오프라인 모델보다 5.59% 연료가 적게 소모되었다.

본 연구를 통해서 최적 속도 컨트롤과 동력 분배 전략의 목표 값을 동시에 제시하는 시스템을 단일 심층 결정적 정책 경사법을 이용하여 최초로 개발되었다. 본 강화학습 모델은 고분자 전해질 연료전지의 치명적인 결함인 플러딩과 드라잉의 발생을 줄였으며, 발생하더라도 빠르게 정상으로 돌아오도록 컨트롤하는 것이 확인되었다. 시스템의 일반화에 대한 우수성이 확인되었으며 온라인 러닝을 통해 성능을 개선시켰다. 이로써 본 연구에서 제안하는 고분자 전해질 연료전지의 안정성을 고려한 연료전지 하이브리드 차량의 에너지 효율 주행 시스템은 자동차산업의 주요 관심사인 자율주행의 친환경화 발전에 기여했으며 온라인 적용이 가능한 고성능 동력분배 전략 개발에 대한 방법론을 제시하였다.