



Robust real-time energy management for a hydrogen refueling station using generative adversarial imitation learning

Truong Hoang Bao Huy^a, Nguyen Thanh Minh Duy^b, Pham Van Phu^a, Tien-Dat Le^b, Seongkeun Park^{c,*}, Daehee Kim^{a,b,*}

^a Department of Future Convergence Technology, Soochunhyang University, Asan-si, Chuncheongnam-do 31538, South Korea

^b Department of Mobility Convergence Security, Soochunhyang University, Asan-si, Chuncheongnam-do 31538, South Korea

^c Department of Smart Automobile, Soochunhyang University, Asan-si, Chuncheongnam-do 31538, South Korea

HIGHLIGHTS

- A robust real-time energy management for hydrogen refueling station is developed.
- Hydrogen chains consider both power-to-hydrogen and hydrogen-to-power pathways.
- GAIL is proposed to guide the DRL agent to mimic expert scheduling strategy.
- Total profit increases by up to 29% when applying the proposed GAIL.
- GAIL outperforms the other algorithms by significant margins.

ARTICLE INFO

Keywords:

Energy management
Fuel cell electric vehicle
Generative adversarial imitation learning
Hydrogen refueling station
Hydrogen storage system

ABSTRACT

As the demand for hydrogen fuel increases with the rise of fuel-cell electric vehicles (FCEVs), the energy management of hydrogen refueling stations (HRSs) is crucial for operational efficiency and environmental sustainability. Although previous studies have applied various energy management methods to HRSs, the application of data-driven approaches for real-time optimization remains very limited. This study addresses this gap by proposing a novel energy management model for optimal real-time energy scheduling of on-grid HRSs using generative adversarial imitation learning (GAIL). The proposed algorithm aims to mimic expert demonstrations to enhance decision-making. Initially, expert trajectories are constructed by collecting state-action pairs, achieved by solving a deterministic energy scheduling model using historical data and a mixed integer linear programming (MILP) solver. These expert trajectories are then used to train the GAIL algorithm. Through adversarial training involving policy and discriminator networks, GAIL accurately simulates expert behavior, enabling strategic decisions regarding power-to-hydrogen conversion, hydrogen-to-power conversion, and FCEV refueling to maximize system profit. The applicability and feasibility of the GAIL algorithm are evaluated across a wide range of scenarios. The results show that total profit increases by up to 29% with the application of the proposed GAIL algorithm. Compared to well-regarded deep reinforcement learning methods, GAIL demonstrates superior performance, proving its effectiveness in real-time energy scheduling of on-grid HRSs.

1. Introduction

1.1. Background and motivation

Global energy demand and environmental concerns have surged significantly due to industrialization and economic growth [1].

Consequently, governments worldwide have gradually committed to building resource-efficient and environmentally friendly societies [2]. Hydrogen, with its high energy density and environmental benefits, is emerging as an ideal alternative energy source [3]. In recent years, there has been a significant acceleration in the market adoption of fuel cell electric vehicles (FCEVs) [4]. By 2030, it is projected that one in twelve

* Corresponding authors.

E-mail addresses: trhbhuy@sch.ac.kr (T.H.B. Huy), ntmduy@sch.ac.kr (N.T.M. Duy), pvphu@sch.ac.kr (P.V. Phu), ltdat@sch.ac.kr (T.-D. Le), keiny@sch.ac.kr (S. Park), daeheekim@sch.ac.kr (D. Kim).

cars sold in Germany, Japan, California, and South Korea will be hydrogen-powered. Additionally, over 350,000 hydrogen-powered trucks are expected to be in operation, and numerous trains and ships can ferry passengers without releasing carbon dioxide into the environment [5]. Fuel cell electric vehicles (FCEVs), alongside battery electric vehicles (BEVs), play a crucial role in decarbonizing the transportation sector. FCEVs are particularly user-friendly due to their long-range refueling cycles and ability to travel long distances [6]. However, the widespread adoption of FCEVs is hindered by the global shortage of hydrogen refueling stations (HRSs). A promising solution is the integration of onsite hydrogen generation within HRSs, utilizing onsite electrolyzers [7]. When combined with renewable energy sources, these HRSs can provide a sustainable supply of green hydrogen for FCEVs. Moreover, hydrogen storage technology can effectively manage excess renewable energy, enhancing system profitability [8]. Despite these advantages, such systems face significant challenges due to the high variability of renewable power sources, energy prices, and hydrogen demand. Current methods do not adequately address these challenges, particularly in real-time scenarios. Therefore, this study proposes a robust real-time energy management tool to ensure the efficient operation of HRSs, addressing the complexities and dynamics of modern energy systems.

1.2. Literature review

In recent years, HRS energy management has attracted the attention of researchers due to its critical role in the transition towards a sustainable and low-carbon energy future. Najafi et al. [9] developed a stochastic-robust bi-level optimization model to integrate HRS into active distribution networks (ADNs). Their study aimed to minimize ADN costs while ensuring system profitability. The model demonstrated that ADNs could adjust their strategies during peak times and worst-case electricity pricing scenarios, proving the technical feasibility of integrating HRSs. Furthermore, Wu et al. [10] investigated a multistage stochastic programming model to devise an optimal strategy for an HRS under various uncertainties, including hydrogen demand, system imbalance price, day-ahead price, and secondary reserve price. The main limitation of these studies is the lack of integration of renewable energy sources in energy management strategies, which could result in less environmentally friendly strategies and more dependence on fluctuating conventional energy prices.

Hydrogen production from renewable energy sources in HRSs offers a viable approach to utilizing clean energy and reducing greenhouse gas emissions. The production and utilization of low-carbon hydrogen are crucial for achieving a sustainable society, prompting several studies on renewable energy-powered HRSs. The authors in [11] introduced a photovoltaic (PV)-powered energy system for HRSs, where hydrogen was produced using a proton exchange membrane electrolyzer powered by both PV and grid electricity. This study optimized the system's operation using a mixed-integer linear programming (MILP) model, focusing on reducing CO₂ emissions and maximizing electrolyzer capacity utilization. However, this study did not consider uncertainties related to renewable energy supply, electricity prices, and hydrogen demand. Similarly, Tostado-Vélez et al. [12] developed a multi-energy microgrid model that integrated electrical, natural gas, and hydrogen subsystems. Their research explored charging stations for various vehicles, enabling FCEVs and natural gas vehicles (NGVs) to refuel using medium-sized electric gas systems. While this model advanced the integration of multiple energy sources, it also neglected the impact of uncertainties. Without accounting for the uncertainties, the models may not accurately reflect real-world conditions, potentially compromising the reliability and economic performance of HRSs.

Renewable energy-powered HRSs face numerous uncertainties, including fluctuating energy outputs, electricity prices, and hydrogen demand. Addressing these uncertainties is crucial for optimizing the scheduling strategy to enhance the profitability of such stations. Various

mathematical optimization methods are frequently employed to manage forecast errors in day-ahead scheduling [13]. Shams et al. [14] presented a model for coordinating HRS and delivery systems in the retail sector by applying bi-level programming to solve the MILP problem. They utilized the chance-constrained method to address uncertainties in renewable generation and demand, revealing that hydrogen prices are influenced by station production capacities and power prices. Grüger et al. [15] proposed an intelligent strategy for operating HRSs with onsite electrolysis, focusing on cost optimization and wind energy utilization based on electricity prices, wind availability, and hydrogen demand forecasts. Despite the conservative nature of the strategy, it successfully enhances wind energy usage and reduces hydrogen production costs by accounting for imperfect forecasts and the nonlinear behavior of electrolyzers. Lakouraj et al. [16] introduced a microgrid scheduling model that optimizes HRSs, demand response (DR), energy storage systems (ESS), and multi-market participation to minimize operational costs. Their risk-constrained stochastic scheduling model, which incorporates conservative parameters, effectively addresses uncertainties in market prices, renewable energy generation, loads, and hydrogen vehicle usage. Shoja et al. [17] developed a risk-averse optimization framework for integrating an EV charging park and HRS within a local multi-energy system. Utilizing power-to-hydrogen and DR programs, this framework enhances system flexibility and reduces costs. However, a significant limitation of these studies is the exclusion of fuel cells in hydrogen systems.

Recent studies have incorporated fuel cells into hydrogen systems to improve flexibility in energy management models. By converting hydrogen back into electricity, fuel cells can provide an additional revenue source for HRSs. Garcia-Torres et al. [18] developed a model-predictive control (MPC)-based dispatch methodology for microgrid operations, which included external agents and BEVs / FCEVs. This study optimized economic benefits by adjusting supply to meet external energy requests and incorporating operation and degradation costs into the objective function. The methodology utilized mixed-integer quadratic programming (MIQP) within a mixed-logic dynamic framework. However, the study did not model uncertainties. The authors in [19] presented a standalone hybrid refueling station supplying electricity, hydrogen, and natural gas to electric vehicles (EVs), FCEVs, and NGVs, leveraging wind and PV power for direct EV charging and versatile hydrogen production. This study addressed uncertainties in wind, PV power, and vehicle demands using stochastic optimization (SO) and information gap decision theory (IGDT) with risk-seeker and risk-averse approaches within an MILP framework. Mansour-Satloo et al. [20] introduced a low-carbon energy management model for a combined hydrogen, heat, and power microgrid, which included hydrogen fueling stations, EV parking, and fuel cell units for power and heat. Based on an IGDT-based robust approach, this model effectively managed uncertainties in wind and PV power, significantly reducing operational costs by 76.35%. Mobasseri et al. [21] introduced a robust hybrid energy management tool for multi-energy microgrids that handled demand and renewable generation uncertainties effectively. Their study utilized the Hong 2 m + 1 approach for renewables, IGDT for electric and heat demands, and scenario-based modeling for FCEV refueling. In a related study [22], the authors introduced a stochastic-interval model for optimizing PV-assisted HRSs, incorporating interval notation for renewable generation, energy pricing uncertainties, and scenario-based vehicle demand modeling. This detailed stochastic model for FCEVs included driving behavior and vehicle characteristics. Despite the significant advancements, these studies predominantly focused on modeling and forecasting uncertainties for day-ahead scheduling. A critical limitation is their inability to address real-time scheduling problems, as actual operating decisions may be re-optimized based on short-term forecasts of uncertainties. This highlights the need for methodologies capable of handling real-time adjustments to ensure optimal operation in dynamically changing environments [23].

Deep reinforcement learning (DRL) has recently emerged as a

significant area of research in artificial intelligence. In contrast to conventional methods, DRL algorithms are both data-based and model-free, continuously learning while interacting with the environment and finally making decisions. This characteristic makes DRL a robust framework for handling uncertainties in complex environments. Various studies have applied DRL algorithms to tackle the complexities of real-time energy management. Cao et al. [24] introduced a proximal policy optimization (PPO)-based scheduling method for HRS to address uncertainties in hydrogen demand. Qi et al. [25] developed an energy management optimization model for an on-grid HRS using an enhanced dueling double-deep Q-network (D3QN) algorithm integrated with NoisyNet. The NoisyNet-D3QN approach outperformed other DRL algorithms by improving exploration efficiency and action generalization. Additionally, Jiang et al. [26] proposed an optimal scheduling model for HRS based on transfer multi-agent RL, aiming to maximize revenue by catering to mobility and power systems via reserve demand response. This approach enhances decision-making and accelerates learning, with numerical studies indicating a potential revenue increase of up to 32.90% and a 34-fold reduction in computation time compared to conventional RL methods. Despite the promising applications, there are relatively few studies on DRL algorithms for the energy management of HRS. Although DRL is capable of solving complex optimization challenges in dynamic and uncertain environments, it still faces limitations due to the inherent nature of learning from trial and error, which can result in unstable performance [27].

Imitation learning (IL) [28] offers an alternative solution, which directly learns the optimal policy from expert demonstrations. In energy management problems where all information is available from historical data, a dedicated MILP solver can easily obtain global optimal solutions. In this way, the MILP solver can serve as the expert, and these optimal solutions can be used as expert demonstrations. Accordingly, IL leverages expert demonstrations to guide the learning process, significantly reducing the need for extensive trial-and-error exploration typical in reinforcement learning (RL). Consequently, IL can achieve more competent performance than RL. Behavior cloning (BC) [29] is a straightforward IL approach that treats the problem as a supervised learning task. The objective is to learn a mapping from states to actions by training on collected state-action pairs. However, the BC approach requires a large amount of data [30] and is particularly susceptible to compounding errors [31], where small prediction mistakes accumulate over time, causing the agent to deviate from the expert's trajectory. Inverse reinforcement learning (IRL) [32] is another IL approach that aims to learn the underlying reward function that the expert is optimizing. This reward function can then be used to derive a policy that replicates the expert behavior. However, IRL is computationally intensive due to the complexity of learning the reward function through optimization, making it challenging to apply to large or high-dimensional environments. Additionally, IRL focuses on learning a cost function that explains the expert behavior but does not directly guide the learner on how to act [33]. Recently, a third approach to IL has been proposed, namely generative adversarial imitation learning (GAIL) [33], which is inspired by generative adversarial networks (GANs) [34]. GAIL involves training a discriminator to differentiate between the expert's actions and the agent's actions while simultaneously training the agent to generate actions that the discriminator cannot distinguish from the expert's. GAIL can learn directly from expert demonstrations without needing a predefined reward function, making it particularly suitable for environments where designing a reward function is challenging or impractical. By iteratively refining the policy through adversarial training, GAIL can train agents to perform well even in scenarios not explicitly covered by expert demonstrations. Hence, GAIL can generalize better than BC. GAIL has been successfully applied in robotics [35], autonomous driving [36], and various other domains. However, it has not yet been applied to energy management systems.

1.3. Research gaps

A literature review revealed that past studies suffer from several significant limitations:

- Many studies have focused on HRSs without incorporating fuel-cell models. Consequently, these studies have overlooked the potential for selling excess energy back to the grid, a strategy that could significantly increase overall profitability. This highlights the need for more comprehensive models that include fuel-cell integration and the associated benefits of energy trading.
- Traditionally, day-ahead scheduling relies heavily on forecasting and modeling uncertain parameters. Inaccurate predictions or missing uncertainty distribution information can result in suboptimal solutions. Furthermore, pre-defined schedules may need to be adjusted or rescheduled based on real-time information, which these traditional methods cannot handle. As a result, these methods often lack the necessary flexibility to adapt to real-time changes in operating conditions, leading to inefficiencies and potential economic losses. This inflexibility underscores the need for more effective energy management strategies to accommodate the uncertainties in energy management problems.
- The application of data-driven methods to HRS energy management remains limited. Only a few recent studies have explored the use of DRL algorithms for this purpose. However, DRL algorithms often face challenges such as instability and slow convergence, primarily due to their reliance on trial-and-error learning processes. Furthermore, despite the potential advantages, there are currently no studies that apply GAIL in the context of HRS energy management. This highlights an opportunity to leverage GAIL capabilities to develop more stable, efficient, and effective energy management solutions that can overcome the limitations of existing DRL approaches.

1.4. Research contributions

Inspired by the above motivations and research gaps, this study develops a Generative Adversarial Imitation Learning (GAIL) algorithm for the energy management of hydrogen refueling stations (HRS). In contrast to traditional RL algorithms that can struggle with the complex and dynamic nature of the environment, GAIL leverages adversarial training to imitate optimal policy behaviors effectively. The proposed algorithm combines the strengths of GANs and IL to address the energy scheduling problem associated with hydrogen production, storage, and refueling operations. GAIL can effectively handle the associated uncertainties related to energy pricing, renewable generation, and FCEV demand. Additionally, GAIL enhances performance by delivering more reliable and cost-effective energy management solutions. Notably, this is the first study to develop and apply GAIL in the context of HRS energy management. The main contributions of this study are as follows:

- This study proposes a comprehensive model for an on-grid HRS incorporating various components such as a solar PV panel, an onsite electrolyzer, a fuel cell, a compressor for FCEV refueling, and a hydrogen storage tank. The model also includes a hydrogen-to-power process via fuel cells, allowing the sale of excess energy to the grid and generating additional revenue. The optimization problem is formulated as a MILP problem with the objective of maximizing total profit while ensuring a reliable hydrogen supply. A dedicated MILP solver resolves the optimization problem, collects state-action pairs, and constructs expert trajectories. This expert knowledge is then utilized to train a policy through GAIL.
- Our study develops a robust energy management framework for HRS that effectively handles uncertainties in energy pricing, renewable generation, and FCEV demand. By leveraging GAIL, we employ adversarial training of policy and discriminator networks to train a DRL agent. This agent emulates the optimal decisions of the MILP

solver for the operations of the electrolyzer, fuel cell, and FCEV refueling. The adversarial training process ensures that the actions of the agent closely resemble expert demonstrations, resulting in a more stable and efficient training process.

- The efficiency of the proposed GAIL algorithm is validated through comprehensive simulations using real-world data and compared with traditional DRL algorithms. The results demonstrate that GAIL significantly improves total profitability by up to 29% compared to the day-ahead MILP, presenting a highly effective and economically viable solution. The results achieved by GAIL closely approach the theoretical minimum for HRS models, indicating that the agent effectively learns and mimics the expert behaviors of the MILP solver. This capability allows GAIL to make near-optimal decisions based on real-time information, ensuring highly efficient energy management. Additionally, GAIL significantly outperforms standard DRL algorithms, confirming its superior performance and adaptability for real-time energy management. Consequently, GAIL emphasizes its potential to enhance the flexibility, robustness, and overall performance of HRS energy management systems.

The remainder of this paper is organized as follows. Section 2 introduces the mathematical model of the HRS. The methodology is described in detail in Section 3. Section 4 presents the simulation results, and the conclusions are presented in Section 5.

2. Mathematical model

Fig. 1 presents a schematic of the proposed on-grid HRS, which includes PV panels and a hydrogen chain established using an onsite electrolyzer, fuel cell, compressor for FCEV refueling, and a hydrogen storage system (HSS). Electricity powered by the utility grid and PV panels can produce hydrogen via water electrolysis. The produced hydrogen is then stored in HSS, from which the hydrogen chain can either refuel FCEVs via the compressor or generate electricity via the

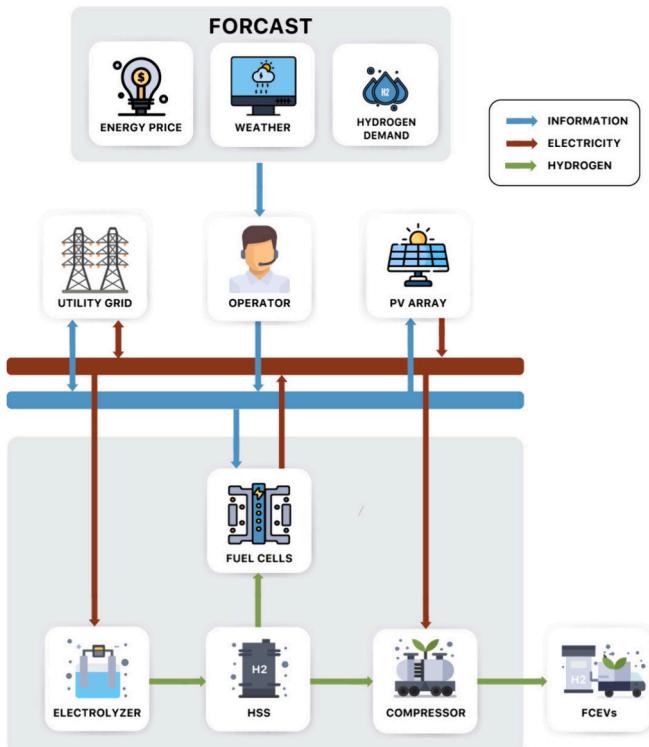


Fig. 1. Schematic diagram of the proposed HRS containing PV panels and a hydrogen chain established using an electrolyzer, fuel cells, a compressor for refueling FCEVs, and HSS.

fuel cell. Both power-to-hydrogen and hydrogen-to-power pathways are considered, allowing the HRS to buy or sell energy flexibly and generate revenue from the utility grid. The HRS is centrally operated by a system operator who receives forecast information on energy prices, weather, and hydrogen demand and then performs an energy scheduling strategy to maximize system profits.

2.1. System modeling

Energy scheduling for the HRS is conducted over a 24-h timeframe ($T = 24$) with a time step of 1 h ($\Delta\tau = 1$). The following subsections provide the mathematical formulations for each system component and the optimal objective function.

2.1.1. Utility grid modeling

The power exchanged with the utility grid is limited due to contractual agreements or physical constraints of the grid, as specified in (1) and (2). Constraint (3) ensures the complementary nature of the buying and selling processes. The utility grid is modeled as follows [37]:

$$0 \leq p_t^{\text{Grid.buy}} \leq u_t^{\text{Grid.buy}} \cdot \bar{p}^{\text{Grid.buy}}; \quad \forall t = 1, 2, \dots, T \quad (1)$$

$$0 \leq p_t^{\text{Grid.sell}} \leq u_t^{\text{Grid.sell}} \cdot \bar{p}^{\text{Grid.sell}}; \quad \forall t = 1, 2, \dots, T \quad (2)$$

$$0 \leq u_t^{\text{Grid.buy}} + u_t^{\text{Grid.sell}} \leq 1; \quad \forall t = 1, 2, \dots, T \quad (3)$$

where $p_t^{\text{Grid.buy}}$ and $p_t^{\text{Grid.sell}}$ are the buying and selling power with the utility grid at time step t , respectively; $\bar{p}^{\text{Grid.buy}}$ and $\bar{p}^{\text{Grid.sell}}$ are the power limits that can be bought and sold with the utility grid; and $u_t^{\text{Grid.buy}}$ and $u_t^{\text{Grid.sell}}$ are the binary variables, denoting the buying and selling modes at time step t , respectively.

2.1.2. PV generator modeling

The available PV generation is a function of solar irradiation and ambient temperature, which can be expressed as follows [37]:

$$\hat{p}_t^{\text{PV}} = \bar{p}^{\text{PV}} \cdot [0.25 \cdot v_t + 0.03 \cdot v_t \cdot \theta_t + (1.01 - 1.13 \cdot \eta^{\text{PV}}) \cdot v_t^2]; \quad \forall t = 1, 2, \dots, T \quad (4)$$

where \bar{p}^{PV} is the rated power of PV panels, η^{PV} is the conversion efficiency of PV panels, v_t and θ_t are the solar irradiance and ambient temperature at time step t , respectively.

The PV power output ought to be restricted by the rated power as follows [38]:

$$p_t^{\text{PV}} = \begin{cases} \bar{p}^{\text{PV}} & \text{if } \hat{p}_t^{\text{PV}} > \bar{p}^{\text{PV}} \\ \hat{p}_t^{\text{PV}} & \text{otherwise} \end{cases}; \quad \forall t = 1, 2, \dots, T \quad (5)$$

where p_t^{PV} is the PV power output at time step t .

2.1.3. Hydrogen chain modeling

Hydrogen can be produced through water electrolysis using advanced technologies such as polymer electrolyte membranes (PEM), solid oxide, or alkaline electrolyzers. The hydrogen gas is stored in tanks for later use in refueling FCEVs or generating electricity through fuel cells [22]. Water electrolyzers and fuel cells perform opposite reactions involving hydrogen and oxygen. **Table 1** lists the fundamental electrochemical reactions that occur in each system. An electrolyzer splits

Table 1
Fundamental electrochemical reactions in water electrolysis and fuel cells.

	Electrolyzer	Fuel cell
Anode	$2H_2O(l) \rightarrow O_2(g) + 4H^+(aq) + 4e^-$	$H_2(g) \rightarrow 2H^+(aq) + 2e^-$
Cathode	$4H^+(aq) + 4e^- \rightarrow 2H_2(g)$	$O_2(g) + 4H^+(aq) + 4e^- \rightarrow 2H_2O(l)$

water into hydrogen and oxygen using electrical energy, an endothermic process requiring energy input. In an electrolyzer, water at the anode side is oxidized to produce oxygen gas, protons, and electrons. The electrons travel through an external circuit to the cathode side, where

$$SOP_t^{HSS} = SOP_{t-1}^{HSS} + \gamma \frac{R \cdot \theta^{HSS}}{\nu^{HSS} \cdot \zeta} \cdot (g_t^{EZ} - g_t^{FC} - g_t^{FCEV}) \cdot \Delta\tau - \frac{\psi}{100} SOP_{t-1}^{HSS}; \quad \forall t = 1, 2, \dots, T \quad (12)$$

they reduce protons to produce hydrogen gas. Conversely, fuel cells generate electrical energy from hydrogen and oxygen through an exothermic reaction, which releases energy. In a fuel cell, hydrogen at the anode side is oxidized, producing protons and electrons. The electrons travel through an external circuit to the cathode side, performing work along the way (e.g., powering an electric motor). At the cathode, oxygen reacts with protons and electrons to produce water. In essence, an electrolyzer uses electrical energy to facilitate the reaction between water and oxygen to produce hydrogen gas, while a fuel cell decomposes hydrogen to initiate the flow of electricity from the anode to the cathode.

In this study, the mathematical modeling of the hydrogen chain is simplified in energy management problems using linearized equations, as proposed in [16]. The hydrogen mass flows in the electrolyzer and fuel cell are defined by (6) and (7), respectively [8]. Moreover, the operations of the electrolyzer and fuel cell are constrained by their maximum and minimum powers, as specified in (8) and (9) [39]. It is also assumed that the electrolyzer and fuel cell do not operate simultaneously, as indicated in (10).

$$g_t^{EZ} = \eta^{EZ} \cdot \frac{p_t^{EZ}}{LHV}; \quad \forall t = 1, 2, \dots, T \quad (6)$$

$$g_t^{FC} = \frac{p_t^{FC}}{LHV \cdot \eta^{FC}}; \quad \forall t = 1, 2, \dots, T \quad (7)$$

$$u_t^{EZ} \cdot p_t^{EZ} \leq p_t^{EZ} \leq u_t^{EZ} \cdot \bar{p}^{EZ}; \quad \forall t = 1, 2, \dots, T \quad (8)$$

$$u_t^{FC} \cdot p_t^{FC} \leq p_t^{FC} \leq u_t^{FC} \cdot \bar{p}^{FC}; \quad \forall t = 1, 2, \dots, T \quad (9)$$

$$0 \leq u_t^{EZ} + u_t^{FC} \leq 1; \quad \forall t = 1, 2, \dots, T \quad (10)$$

where g_t^{EZ} and g_t^{FC} are the hydrogen mass flows of the electrolyzer and fuel cell at time step t , respectively; p_t^{EZ} and p_t^{FC} are the powers of the electrolyzer and fuel cell at time step t , respectively; \underline{p}^{EZ} and \bar{p}^{EZ} are the minimum and maximum powers of the electrolyzer, respectively; \underline{p}^{FC} and \bar{p}^{FC} are the minimum and maximum powers of the fuel cell, respectively; u_t^{EZ} and u_t^{FC} are the binary variables, denoting the operation modes of the electrolyzer and fuel cell at time step t , respectively; η^{EZ} and η^{FC} are the efficiencies of the electrolyzer and fuel cell, respectively; and LHV is hydrogen lower heating value.

The amount of refueled hydrogen must not exceed the expected FCEV refueling demand, as shown in (11) [22].

$$g_t^{FCEV} \leq u_t^{FCEV} \cdot \bar{g}_t^{FCEV}; \quad \forall t = 1, 2, \dots, T \quad (11)$$

where g_t^{FCEV} is the amount of hydrogen refueled at time step t , \bar{g}_t^{FCEV} is the expected demand of the FCEV at time step t , and u_t^{FCEV} is a binary variable denoting the operation mode of the compressor for FCEV refueling at time step t .

In this study, a steady-state model is applied to represent the state of pressure (SOP) in the HSS, as presented in (12) [16]. This model accounts for gas dissipation and equates the inflow and outflow mass flows to the hydrogen flows through the electrolyzer and compressor/fuel cell.

Moreover, the SOP of the HSS is limited by constraint (13) [22]. Finally, the SOPs at the initial and final time steps of the scheduling horizon are established according to constraint (14) [22].

$$\underline{SOP}^{HSS} \leq SOP_t^{HSS} \leq \overline{SOP}^{HSS}; \quad \forall t = 1, 2, \dots, T \quad (13)$$

$$SOP_{\Gamma(1)}^{HSS} = SOP_{\Gamma(end)}^{HSS} = \overline{SOP}^{HSS} \quad (14)$$

where SOP_t^{HSS} is the SOP of the HSS at time step t ; \underline{SOP}^{HSS} and \overline{SOP}^{HSS} are the minimum and maximum SOPs of the HSS, respectively; γ is the conversion factor, R is the gas constant, θ^{HSS} is the HSS temperature, ν^{HSS} is the gas volume, ζ is the mass molar of hydrogen, and ψ is the dissipation factor.

2.1.4. Power balance

In the proposed system, the power balance should be maintained at any time step as follows [22]:

$$p_t^{Grid,buy} + p_t^{PV} + P_t^{FC} = P_t^{Grid,sell} + P_t^{EZ} + p_t^{Comp}; \quad \forall t = 1, 2, \dots, T \quad (15)$$

where p_t^{Comp} is the power consumption of the compression stages, which can be calculated based on the amount of hydrogen refueled as follows [22]:

$$p_t^{Comp} = \frac{g_t^{FCEV} \cdot \zeta^{Comp}}{\Delta\tau \cdot \eta^{Comp}}; \quad \forall t = 1, 2, \dots, T \quad (16)$$

where ζ^{Comp} and η^{Comp} are the energy consumption and efficiency of the compressor, respectively.

2.1.5. System profit

The total profit of an HRS can be expressed as follows [22]:

$$F_t = F_t^{FCEV} - F_t^{Grid} - F_t^{EZ} - F_t^{FC}; \quad \forall t = 1, 2, \dots, T \quad (17)$$

In (17), the first term represents the FCEV refueling revenue, which is given by (17) [22]:

$$F_t^{FCEV} = \lambda^{FCEV} \cdot g_t^{FCEV}; \quad \forall t = 1, 2, \dots, T \quad (18)$$

where λ^{FCEV} is the hydrogen price.

The second term in (17) represents the cost of the energy exchange with the main grid [22].

$$F_t^{Grid} = \Delta\tau \cdot (\lambda_t^{Grid,buy} \cdot p_t^{Grid,buy} - \lambda_t^{Grid,sell} \cdot p_t^{Grid,sell}); \quad \forall t = 1, 2, \dots, T \quad (19)$$

where $\lambda_t^{Grid,buy}$ and $\lambda_t^{Grid,sell}$ are the buying and selling prices at time step t , respectively.

The costs related to the operation and maintenance of the electrolyzer and fuel cell are given in the last two terms in (17) as follows [39]:

$$F_t^i = \Delta\tau \cdot \left(\frac{k^i \cdot \bar{p}^i}{T^i} \cdot u_t^i + p_t^i \cdot \mu^i \right); \quad \forall i \in \{EZ, FC\}; t = 1, 2, \dots, T \quad (20)$$

where k^i , μ^i , and T^i are capital cost, operation and maintenance cost, and number of life hours of the electrolyzer/fuel cell, respectively; p_t^i and u_t^i are the power and operation mode of electrolyzer/fuel cell at time step t , respectively.

2.2. Objective function

In this study, the objective of the energy scheduling problem is to maximize the system profit of an HRS by optimally scheduling the operations of the electrolyzer, fuel cell, compressor, and energy trading with the utility grid. Mathematically, the energy scheduling problem can be formulated as follows:

$$\mathbf{P1} : \min_{x_t} F = \sum_{t=1}^T F_t \quad (21)$$

s.t. : (1) – (20)

where the vector of decision variables can be given in (22):

$$x = \left\{ \begin{array}{l} p_t^{Grid.buy}, p_t^{Grid.sell}, p_t^{EZ}, p_t^{FC}, p_t^{Comp}, g_t^{EZ}, g_t^{FC}, g_t^{FCEV}, \\ SOP_t^{HSS}, u_t^{Grid.buy}, u_t^{Grid.sell}, u_t^{EZ}, u_t^{FC}, u_t^{FCEV} \end{array} \right\}; \quad \forall t = 1, 2, \dots, T \quad (22)$$

Remark: Modeling binary variables directly in MILP offers accurate and interpretable solutions that naturally enforce operational constraints. This approach makes it straightforward to model scenarios where certain actions, such as selling and buying, cannot occur simultaneously. Although solving MILP problems with many binary variables can be computationally intensive, it ensures that the operational constraints are strictly adhered to. Alternatively, treating these binary variables as continuous can transform the problem into a Linear Programming (LP) problem. This transformation typically makes the problem easier and faster to solve using standard LP solvers. Relaxing binary variables to be continuous can significantly reduce computational complexity and improve scalability, making it feasible to solve larger problems more efficiently. However, this approach requires careful handling during post-processing and constraint enforcement to ensure that the final solution is both feasible and optimal. In medium-scale problems, such as HRS energy management, the binary approach provides a clear and straightforward framework for system modeling. In future studies, we plan to explore the relaxation of binary variables to continuous for large-scale problems to take advantage of the computational efficiencies this approach offers.

3. Proposed methodology

3.1. Markov decision process (MDP) formulation

Energy scheduling for an HRS is essentially a sequential decision-making problem. Thus, it can be mathematically formulated as a Markov decision process (MDP). Within this framework, the agent observes the current state s_t at each time step t , takes an action a_t based on a policy $\pi(a|s)$, earns a reward r_t based on this action, and transitions to the next state s_{t+1} . The agent aims to maximize the cumulative reward by strategically adapting its policy. The MDP for energy scheduling of an HRS can be expressed as follows:

State: The state is represented by a vector that includes all observable information at a given time step, including the time step (t), electricity buying price ($\lambda_t^{Grid.buy}$), PV generation (p_t^{PV}), expected FCEV demand (\bar{g}_t^{FCEV}), and the SOP of the HSS at the previous time step (SOP_{t-1}^{HSS}). The buying and selling prices typically show a correlation, which is represented by the buying price ($\lambda_t^{Grid.buy}$) within the state vector. Therefore, the state vector can be expressed as (23):

$$s_t = [t, \lambda_t^{Grid.buy}, p_t^{PV}, \bar{g}_t^{FCEV}, SOP_{t-1}^{HSS}] \quad (23)$$

Action: The agent generates actions to control the hydrogen mass flow in the electrolyzer, fuel cell, and compressor for FCEV refueling, based on a policy and the observed states at the current time step. The action vector is expressed as:

$$a_t = [g_t^{EZ}, g_t^{FC}, g_t^{FCEV}] \quad (24)$$

Transition model: $T(s_t, a_t, s_{t+1})$ is a transition model that defines the next state $s_{t+1} \in S$ given the current state $s_t \in S$ and the action $a_t \in A$. The transition model for the SOP of the HSS is given by (12), and other information in the state, including the time step, energy price, PV generation, and expected FCEV demand, are observed in real-time.

Reward function: Reward r_t gives the immediate reward received after transitioning from state s_t to state s_{t+1} owing to action a_t . The objective of the agent is to find a policy π that maximizes the expected sum of discounted rewards over a horizon of T time steps as follows:

$$\max J(\pi) = E_\pi \left[\sum_{t=1}^T \gamma^{t-1} r_t \right] \quad (25)$$

where Π is the collection of all allowable policies, $\gamma \in [0,1]$ represents the discount factor, and E_π represents the expected value under policy π . We do not define the reward function since an agent in GAIL can learn a policy from expert demonstrations without the explicit reward function.

3.2. The proposed real-time optimal energy scheduling using GAIL

This section introduces real-time energy scheduling for an HRS based on GAIL. GAIL is an advanced framework that combines IL with adversarial training, which was introduced by GANs [34]. This enables agents to learn complex behaviors by imitating expert demonstrations. The key idea behind GAIL is to train a policy that generates behavior indistinguishable from that of an expert without requiring explicit access to the expert policy or rewards. First, state-action pairs are collected sequentially to build expert trajectories using the MILP solver to solve problem **P1** using historical data. Second, the expert trajectories generated by the MILP solver are fed into the discriminator of GAIL for

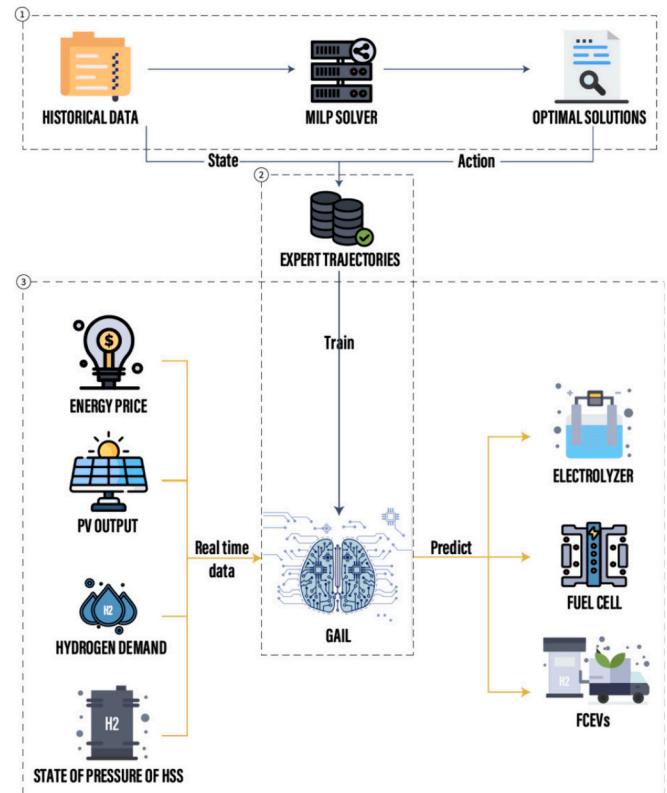


Fig. 2. Overall framework of GAIL algorithm for HRS energy management, consisting of three stages: ① Generation of expert trajectories; ② GAIL training; ③ Real-time energy scheduling with learned policy.

training. Finally, the reward output from the discriminator is used to direct the agent towards actions that yield a better performance to maximize the total profit of the HRS. Fig. 2 depicts the overall framework of the proposed methodology consisting of three stages: ① Generation of expert trajectories; ② GAIL training; ③ Real-time energy scheduling with the learned policy.

3.2.1. Generation of expert trajectories

The initial step involves the generation of expert trajectories. This crucial phase is designed to create a comprehensive dataset that models the optimal operational strategies and serves as a benchmark for learning and simulation. To accomplish this, optimization problem **P1** is formulated as a deterministic MILP model that can be solved explicitly using accurately known input data. This approach leverages a wealth of scenarios derived from accurate input information obtained through historical data collection and sophisticated scenario generation approaches. Each scenario is considered a scheduling cycle, encompassing a sequence of T timesteps representing the operational data for one scheduling cycle. The data include the real-time electricity price ($\lambda_{\Delta t}^{\text{Grid.buy}}, \lambda_{2\Delta t}^{\text{Grid.buy}}, \dots, \lambda_T^{\text{Grid.buy}}$), solar irradiation ($v_{\Delta t}, v_{2\Delta t}, \dots, v_T$), and FCEV refueling demand ($\bar{g}_{\Delta t}^{\text{FCEV}}, \bar{g}_{2\Delta t}^{\text{FCEV}}, \dots, \bar{g}_T^{\text{FCEV}}$). For each scenario, the collected data serves as inputs to the formulated problem (problem **P1**), which is tackled using a specialized MILP solver. Thus, problem **P1** is conceptualized as a day-ahead scheduling problem, with the MILP solver tasked with finding the optimal values for the decision variables. Thus, the solution is deemed the best possible solution under the assumption of the availability of complete information and accurate input data. It is important to note that this idealized condition is not achievable in real-world applications, as it requires having the perfect information for an entire scenario in advance.

After resolving problem **P1**, the input data and optimal solutions for each scenario enable the construction of an expert dataset comprising T state-action pairs $D = \{s_t, a_t\}_{t=1}^T$, where the state variables (s_t) and action variables (a_t) are explicitly defined and extracted for the time steps of a scheduling cycle, as in (23) and (24). This dataset captures the essence of expert decision-making in optimal energy scheduling scenarios. This approach is systematically applied across all selected scenarios, and the resulting state-action pairs from each scenario are merged to form a comprehensive expert dataset. By resolving problem **P1** across N scenarios, each scenario consists of T time steps, and expert trajectories of NT state-action pairs $D = \{s_i, a_i\}_{i=1}^{NT}$ are produced. These expert trajectories form the foundation for the subsequent stages of the GAIL algorithm, providing expert knowledge for training and fine-tuning learning models.

3.2.2. GAIL training

GAIL involves two primary components: a generator G (or the policy of the agent) and discriminator D . The training process consists of

alternating between updating the discriminator and the generator (policy) based on feedback from the discriminator in a min-max game, where the generator generates samples that are indistinguishable from the expert demonstrations, and the discriminator seeks to differentiate between the samples generated by the generator and those generated by expert demonstrations. Through iterative training, the generator learns to generate samples sufficiently similar to expert demonstrations, from which the discriminator network cannot distinguish the difference between the two samples. The objective of GAIL is to define a saddle point (π, D) of the following expression:

$$\minmax_{\theta} \mathbb{E}_{\pi_\theta} [\log D_\omega(s, a)] + \mathbb{E}_{\tau_E} [\log(1 - D_\omega(s, a))] - \lambda H(\pi_\theta) \quad (26)$$

where π_θ is the policy parameterized by θ , D_ω is the discriminator parameterized by ω , τ_E is the set of expert trajectories, \mathbb{E}_{π_θ} and \mathbb{E}_{τ_E} are the mathematical expectation of the learner samples and the expert demonstration, respectively, $H(\pi_\theta) \triangleq \mathbb{E}_{\pi_\theta} [-\log \pi_\theta(a|s)]$ represents the causal entropy of the policy π_θ weighted by a regularization parameter λ .

Fig. 3 depicts the training process of GAIL. The gradient steps are taken to update the parameters of discriminator network D_ω and policy network π_θ alternatively until both networks converge. Specifically, the agent (learner) obtains expert trajectories as inputs. Then, the agent interacts with the environment based on the policy π_θ and generates a trajectory that contains a sequence of state-action pairs as $\tau_i = \{(s_1, a_1), (s_2, a_2), \dots, (s_n, a_n)\}$. The discriminator in GAIL outputs a value $D(s, a)$ for the given state-action pair (s, a) and determines whether it originates from expert trajectories or is generated by the generator policy by interacting with the environment. Thus, the trajectory $\tau_i \sim \pi_\theta$ and the expert trajectory τ_E are used to update the discriminator with the Adam optimizer, which maximizes the following function:

$$\mathbb{E}_{\tau_i} [\log D_\omega(s, a)] + \mathbb{E}_{\tau_E} [\log(1 - D_\omega(s, a))] \quad (27)$$

The objective of the generator is to minimize the following function:

$$\mathbb{E}_{\tau_i} [\log D_\omega(s, a)] - \lambda H(\pi_\theta) \quad (28)$$

The discriminator network can be interpreted as a reward function that provides a learning signal to the policy $r(s, a) = -\log D_\omega(s, a)$. In this study, the parameter θ of the policy network π_θ can be updated using proximal policy optimization (PPO) as follows:

$$\mathbb{E}_{\tau_i} [\nabla_\theta \log \pi_\theta(a|s) Q(s, a)] - \lambda \nabla_\theta H(\pi_\theta) \quad (29)$$

Thus, the generator policy is encouraged to produce high-quality trajectories that mimic the expert demonstrations. Further details of the PPO algorithm can be found in [40].

3.2.3. Real-time energy scheduling with learned policy

Upon completion of the GAIL training phase, the learned policy can be applied to energy scheduling tasks for the HRS. This process is designed to unfold into two primary stages at each time step, ensuring a

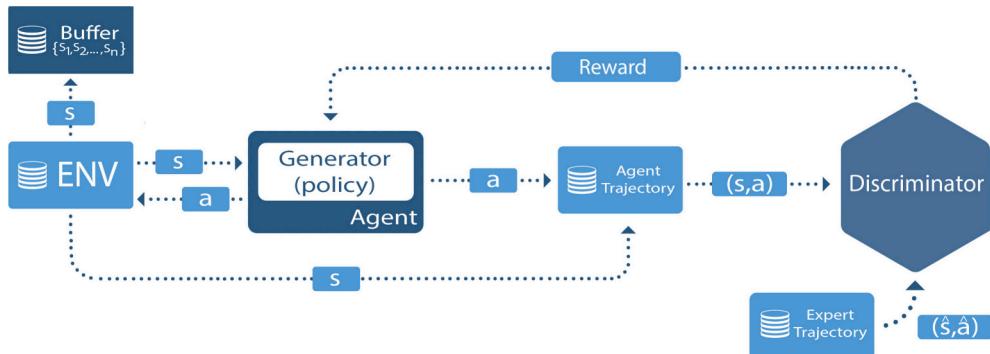


Fig. 3. The training process of GAIL involves training an agent to imitate expert behavior. During this process, the generator (agent) and discriminator are trained adversarially until the actions of the agent become indistinguishable from those of the expert.

dynamic and responsive energy management system that adapts to real-time operational conditions. The first phase of energy scheduling involves determining the hydrogen mass flow in the electrolyzer, fuel cell, and compressor for FCEV refueling based on the learned policy, π_θ^* . To facilitate this, real-time operational data are gathered, and a state vector is formulated for each time step t , as shown in (23). The state vector contains current conditions and operational status of the hydrogen refueling station. Utilizing this state vector, the learned policy π_θ^* process to define the predicted action is as follows: $\hat{a}_t = \pi_\theta^*(s_t)$ outlines the operational directives for the electrolyzer, fuel cell, and FCEV refueling. The expected action essentially encompasses the inflow and outflow mass flows of the HRS, as shown in (24). Following the action determination, the SOP of the HSS at time step t is ascertained using (12), which considers the optimal inflow and outflow of the hydrogen mass to satisfy the FCEV refueling demand efficiently. Concurrently, the power exchanged with the utility grid is determined following (15) to ensure the energy balance in the HRS.

With all decision variables explicitly defined, including the power levels for the electrolyzer, fuel cell, FCEV refueling, and utility grid, the system operator dispatches these values as setpoint signals. These signals command the operational behaviors of the corresponding components within the HRS, ensuring operation according to the optimal schedule defined by the GAIL model. This process is repeated at each time step throughout the energy scheduling cycle, providing a systematic and intelligent approach to energy management. By leveraging the desired efficiency of the learned policy, the HRS can dynamically adjust its operations to real-time data and conditions, optimize energy utilization, maximize total profit, and ensure the reliable provision of hydrogen fuel.

4. Simulation results

In this section, the feasibility of the developed methodology is validated for the benchmark HRS described in Section 2. All simulations are conducted using Python on a Mac Studio with an Apple M1 Max chip and 32 GB of RAM. Energy scheduling is performed over a 24-h period with a time step of 1 h.

4.1. Simulation setup

4.1.1. Input data

The HRS under study comprises a PV panel and a hydrogen chain connected to the utility grid. Table 2 lists the data related to the hydrogen chain, including the electrolyzer, fuel cell, compressor, and hydrogen storage tank, which are obtained from [22]. The power exchanged between the utility grid and the HRS is limited to 500 kW. In the proposed simulations, the datasets used as inputs for training and testing include the time-series data of the hourly electricity prices ($\lambda_t^{Grid.buy}$), PV generation (p_t^{PV}), and FCEV demand (g_t^{FCEV}), as shown in Fig. 4. These data are collected from real data or generated using different approaches. These datasets closely resemble real-world scenarios, ensuring simulation applicability and reliability. Gaussian noise

is injected into a nominal scenario based on historical data from the PJM FE Ohio system on November 2nd, 2019, to simulate electricity price uncertainty. The Gaussian noise level of the electricity price is set to 10%. The selling price ($\lambda_t^{Grid.sell}$) is assumed to be 0.8 times the buying price ($\lambda_t^{Grid.buy}$). The expected power output of a PV array of 250 kW is defined based on the solar irradiance and temperature extracted from the European Commission database in Madrid (Spain) from January 1st, 2015, to December 31st, 2018.

Because of the limited availability of FCEV demand data, the scenario-based generation approach is applied using the probability distribution functions of two essential factors of an HRS: (1) the total number of refueling events in one scheduling scenario and (2) the amount of hydrogen refueled for a single refueling event. Table 3 presents the data related to the distribution of FCEV refueling demand. Fig. 5 depicts the percentage variation in FCEV refueling demand, presented as a percentage of the total number of refueling events occurring within a day [41]. The data in Table 3 and Fig. 5 are provided by the National Renewable Energy Laboratory (NREL) [42]. Based on these data, a Gaussian distribution is used to generate 821 scenarios for hourly FCEV demand profiles, as shown in Fig. 4(c). Moreover, a refueling price (λ^{FCEV}) of 4 \$/kg is considered.

4.1.2. Data analysis

This subsection provides a concise analysis of the input data used in HRS energy management. The boxplots in Fig. 6 illustrate the distribution of energy price, PV generation, and FCEV demand across different hours of the day. As shown in Fig. 6(a), energy prices exhibit significant variability throughout the day, with peaks at 7:00 and 18:00, and lower prices observed between 14:00 and 16:00. In Fig. 6(b), a clear pattern is evident for PV generation, which begins around 6:00, peaks between 11:00 and 15:00, and tapers off by 18:00. The highest generation occurs at midday, reflecting the availability of sunlight and aligning with the natural daylight cycle. Fig. 6(c) shows that FCEV demand follows a distinct daily cycle, with low demand during the early morning hours (midnight to 6:00) and late evening (after 20:00). During these periods, occasional spikes in usage are observed. Demand starts to increase significantly from 7:00, peaking between 12:00 and 17:00. This pattern suggests that FCEV demand is closely tied to typical daily activities, with higher usage during daytime hours when people are commuting or traveling for work, and lower usage during nighttime when fewer trips are made.

The frequency distribution of energy price, PV generation, and FCEV demand is shown in Fig. 7. In Fig. 7(a), most of the data points for energy prices are concentrated at lower levels, primarily below 0.05 \$/kWh, with a small tail extending towards higher prices, indicating occasional spikes. The histogram for PV generation in Fig. 7(b) exhibits a bimodal distribution. Many data points are clustered around zero, reflecting numerous instances of minimal or no power generation, likely during nighttime or periods of low sunlight. There is also a significant peak around 250 kW, indicating high PV generation during peak sunlight hours. As shown in Fig. 7(c), the histogram for FCEV demand displays a right-skewed distribution, indicating that low to moderate demand is common, with high demand occurring less frequently. The demand peaks around 2–4 kg, with fewer instances of very high demand.

Fig. 8 is a heatmap that displays the correlation coefficients between state and action variables in HRS energy management. The correlation analysis justifies the selection of these state and action variables. The chosen state variables capture essential factors influencing energy management, including temporal, economic, renewable energy, and demand aspects. The action variables, representing the hydrogen mass flow of the electrolyzer, fuel cell, and FCEV refueling, are critical for making informed decisions about hydrogen production, power generation, and refueling. The strong correlation between state and action variables is crucial for effectively training the GAIL agent to make optimal decisions in HRS operations.

Table 2
Data of the proposed HRS.

Parameter	Value	Parameter	Value
p_{EZ}^{EZ}/\bar{p}^{EZ}	25/400 kW	v^{HSS}	200 m ³
p_{FC}^{FC}/\bar{p}^{FC}	25/400 kW	θ^{HSS}	313 K
$\eta^{EZ}/\eta^{FC}/\eta^{Comp}$	0.65/ 0.77/ 0.8 p.u.	γ	0.0001
k^{EZ}/k^{FC}	8.5/ 32 \$/kW	R	8.314 J/K mol
T^{EZ}/T^{FC}	10,000 h	ζ	0.002 kg/mol
μ^{EZ}/μ^{FC}	0.03 \$/kWh	ψ	0.0006%
		LHV	39.72 kWh/kg
		ζ^{Comp}	2.7 kWh/kg
		$SOP^{HSS}/\bar{SOP}^{HSS}$	2/10 bar

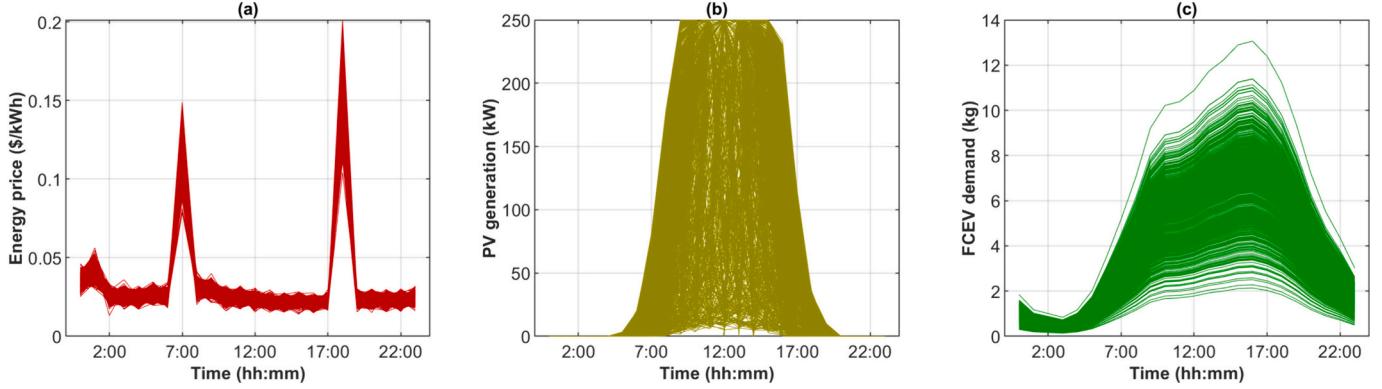


Fig. 4. Dataset for training and testing: (a) Energy price, (b) PV generation, and (c) FCEV demand. The dataset has a total of 821 scenarios, with 730 training scenarios and 91 testing scenarios.

Table 3

Data of FCEV refueling demand distribution.

	Mean	Standard deviation	Min	Max
Number of refueling events	30	5	–	–
Fueling amounts per event (kg)	2.93	1.9	0.7	6.95

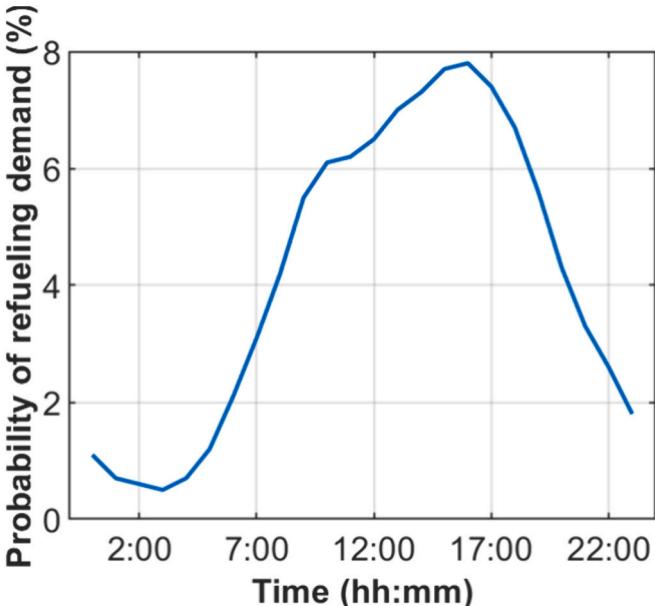


Fig. 5. Percentage variation of FCEV refueling demand during a day. Data is shown as a percentage of the total refueling events occurring over a day.

4.1.3. Network architecture, hyperparameter tuning, and training analysis

The proposed GAIL algorithm is implemented using the *Stable-Baselines3* [43] and *imitation* [44] libraries, with the PPO algorithm serving as the learner model. GAIL comprises a policy network (actor), value network (critic), and discriminator network. The policy network includes two linear layers with 256 neurons each, activated by ReLU functions. The final layer, a linear activation function, outputs action probabilities corresponding to the three possible actions in the environment. The critic network, similar to the actor network, also has two hidden layers but concludes with a single neuron in the output layer, which represents the estimated state value. The discriminator network, designed as a reward neural network, comprises two hidden layers with 32 neurons each, utilizing ReLU activation functions. The output layer of the discriminator network employs a sigmoid activation function to

produce a probability value that serves as a discounted reward.

To optimize model performance, extensive hyperparameter tuning for the GAIL algorithm is conducted using the *Optuna* hyperparameter optimization framework [45]. This framework allows for an efficient and systematic search for optimal hyperparameters. Each set of hyperparameters is evaluated based on the cumulative reward obtained in the test scenarios, with evaluations conducted every 600,000 timesteps and two evaluations per trial. A total of 100 trials are conducted, allowing *Optuna* to explore a broad range of hyperparameter combinations. **Table 4** summarizes the tuned hyperparameters for GAIL training and their optimal values as identified through the optimization process, with some hyperparameters kept at their default values as specified in the *imitation* library.

Fig. 9 illustrates the training curve of the tuned agent during the GAIL training process. Initially, the mean reward is low but increases steadily over the first training steps, indicating effective learning from the expert trajectories. As training progresses, the average reward rises rapidly before stabilizing, signifying significant improvement and convergence towards the expert policy. The learning curve without significant fluctuations, suggests stable training and well-tuned hyperparameters.

4.2. Real-time energy scheduling performance of the proposed GAIL

After offline training, the learned policy can provide scheduling decisions for the optimal operation of the HRS based on the real-time state. Accordingly, the developed GAIL is tested under 91 scenarios. **Table 5** presents the simulation results for all test scenarios obtained by the developed GAIL algorithm. **Table 5** also lists the results of day-ahead MILP, which is a practical approach to solving problem **P1** through forecasting. For clarity in comparison, the forecast v_t of an actual value v_t^* follows a normal distribution as $v_t \sim \mathcal{N}(v_t^*, \gamma v_t^*)$. Given the substantial stochastic variability, the parameter γ escalates steadily from 0.2 at $t = 1$ to 0.5 at $t = T$, as proposed in [46]. The results in **Table 5** show that the system profitability is higher when applying the scheduling strategy given by GAIL, leading to an improvement of 29%. In the base case, the FCEV demand is met by a maximum of 96%, whereas the proposed algorithm can satisfy 91% of the FCEV demand owing to the scale of the HRS; however, the proposed GAIL still yields higher profits than the day-ahead MILP. Instead of passively buying electricity from the utility grid to fill the HSS and fully meet the FCEV demand, the HRS flexibly buys or sells electricity at different times to maximize profits.

Fig. 10 depicts the revenue for five scenarios with increasing FCEV demand. Total profit is determined according to (17)–(20), which takes into account revenue from FCEV refueling and selling energy and deducts the costs of purchasing electricity and maintaining electrolyzers and fuel cells. From **Fig. 10**, there is a clear upward trend in the total profit and FCEV refueling revenue as the total FCEV demand increases.

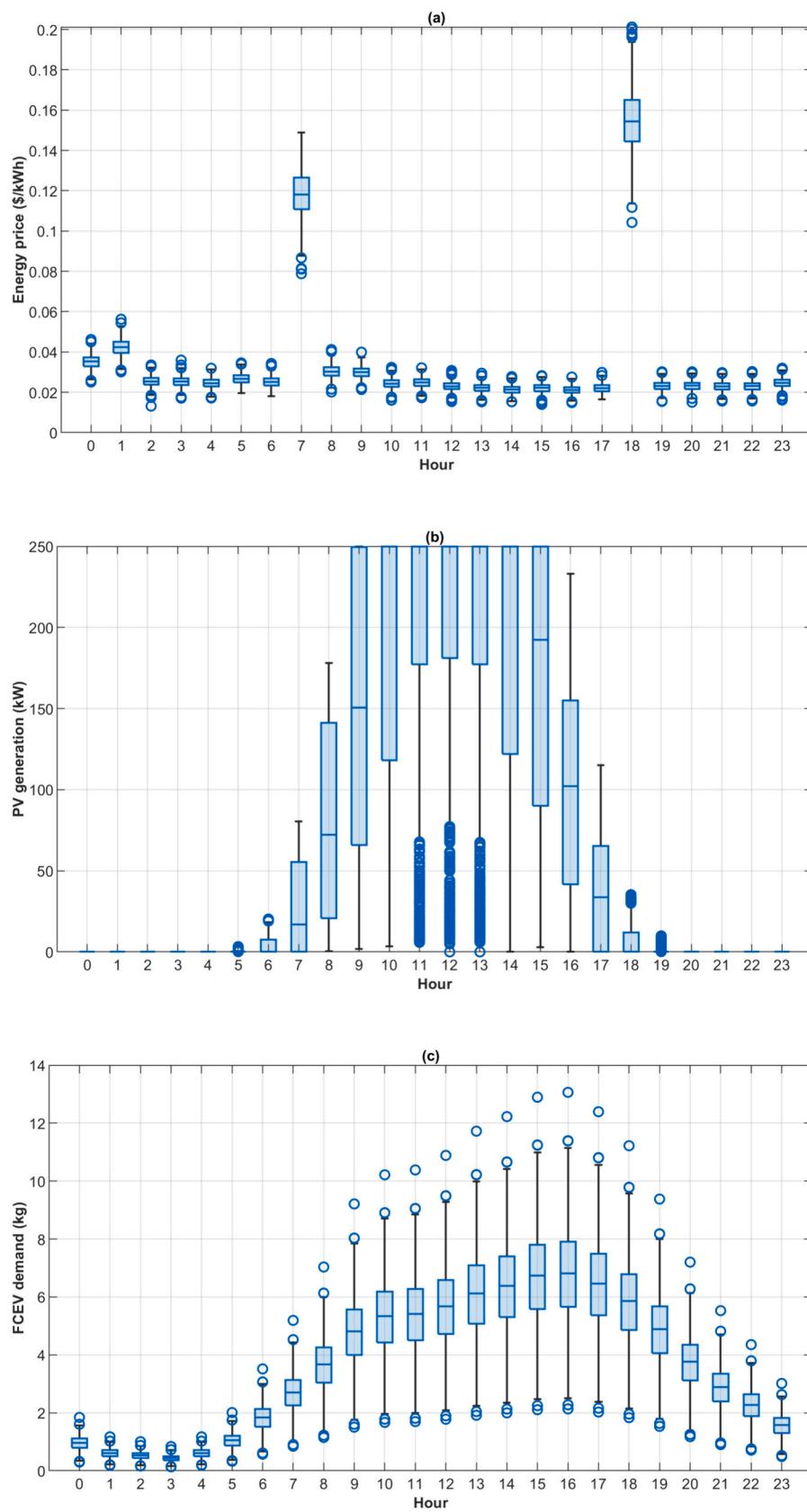


Fig. 6. Boxplot of hourly distribution: (a) Energy price, (b) PV generation, and (c) FCEV demand.

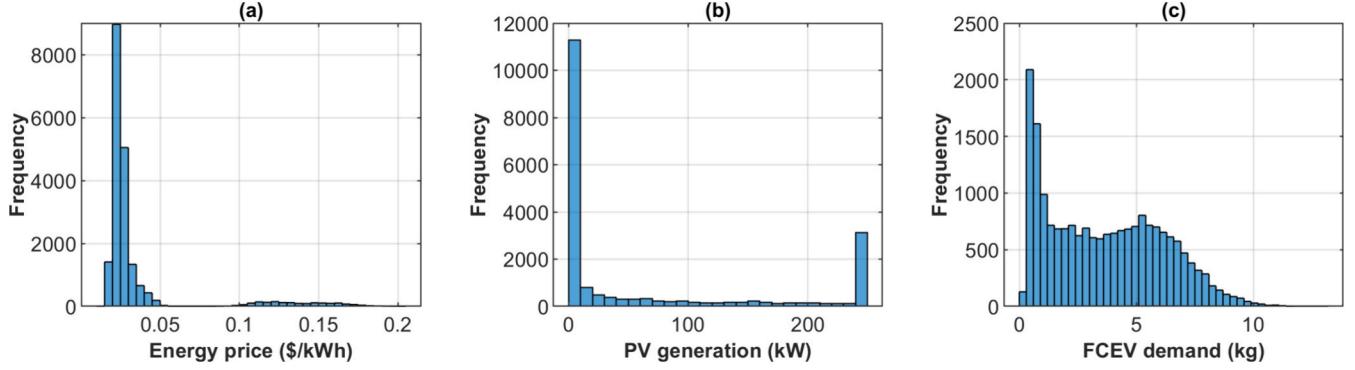


Fig. 7. Distribution histograms of input data: (a) Energy price, (b) PV generation, and (c) FCEV demand. The histograms for energy price and FCEV demand show highly skewed distributions, whereas the PV generation histogram shows a bimodal distribution.

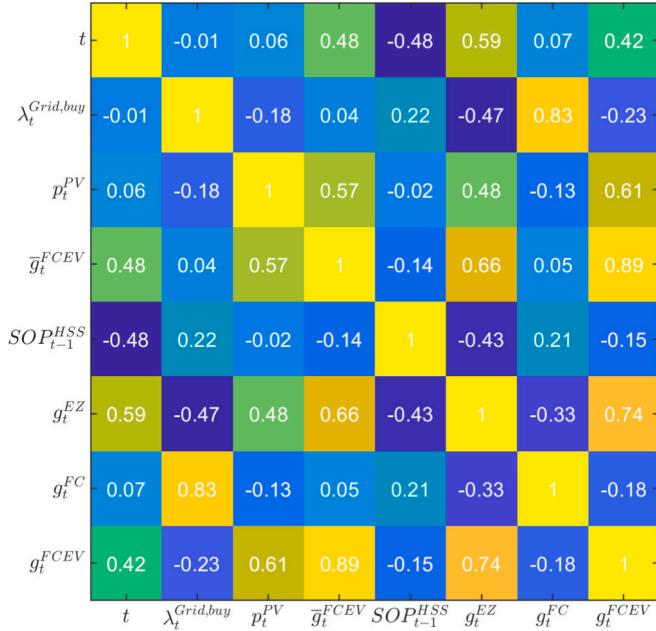


Fig. 8. Heatmap of the correlation matrix for state and action variables in HRS energy management (warmer colors indicate positive correlations and cooler colors indicate negative correlations).

Table 4
Search space and optimal values for the GAIL hyperparameters.

Hyperparameter	Search space	Optimal value
Batch Size	[8, 16, 32, 64, 128, 256, 512]	64
Number of Steps	[512, 1024, 2048]	2048
Gamma	[0.9, 0.95, 0.98, 0.99, 0.995, 0.999, 0.9999]	0.99
Learning Rate	[1e-5, 1] (log scale)	0.0003
Entropy Coefficient	[0.00000001, 0.1] (log scale)	0.01
Clip Range	[0.1, 0.2, 0.3, 0.4]	0.2
Number of Epochs	[1, 5, 10, 20]	10
GAE Lambda	[0.8, 0.9, 0.92, 0.95, 0.98, 0.99, 1.0]	0.95
Max Gradient Norm	[0.3, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 5]	0.5
Value Function Coefficient	[0, 1]	0.5
Demo Batch Size	-	1024
Generator Replay Buffer Capacity	-	512
Number of Discriminator Updates per Round	-	8

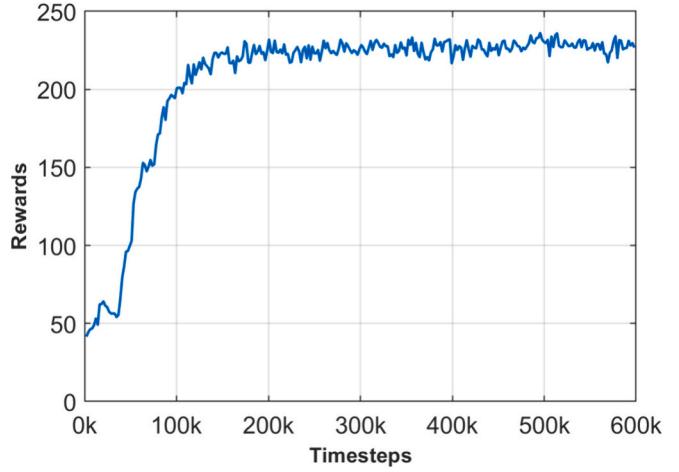


Fig. 9. Reward convergence curve in the GAIL training process. The mean reward starts low, increases steadily during initial training, and then rises rapidly before stabilizing, indicating effective learning and convergence towards the expert policy.

Table 5
Results of the proposed GAIL and day-ahead MILP for test scenarios.

Indicators	Day-ahead MILP	The proposed GAIL
Profit (\$)	16,277	20,997
FCEV demand satisfied (%)	96.39	91.70

However, the revenue from selling energy decreases significantly in scenarios with a higher FCEV demand. When FCEV demand is low, selling energy through a fuel cell is the main revenue stream, along with FCEV refueling. However, when the FCEV demand increases, most of the hydrogen in the storage tanks is devoted to meeting the FCEV refueling demand, significantly reducing hydrogen-to-power conversion of fuel cells. Therefore, the proposed GAIL can provide effective energy scheduling strategies for different scenarios, which can economically benefit HRS.

4.3. Energy scheduling via GAIL in a specific scenario

A scenario is selected randomly from the testing dataset to confirm the feasibility of the GAIL algorithm. Fig. 11 presents the data for the selected scenario, including the energy price, PV generation, and FCEV demand.

Fig. 12 shows the scheduling results for a random scenario, demonstrating HRS operations including the power-to-hydrogen function of

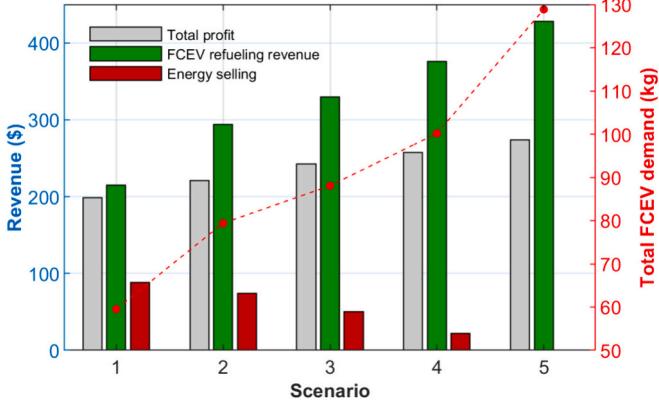


Fig. 10. Revenue from different sources for five scenarios with gradually increasing refueling demand. Higher FCEV demand increases total profit and refueling revenue but reduces energy selling revenue due to increased hydrogen allocation for refueling.

the electrolyzer, hydrogen consumption by the compressor to meet FCEV demand, and the hydrogen-to-power function of the fuel cell. Notably, the HRS is designed to prevent simultaneous operation of the electrolyzer and fuel cell.

Fig. 12(a) illustrates that the electrolyzer and compressor are appropriately scheduled to maintain the hydrogen tank pressure at an acceptable level to satisfy FCEV demand. Most of the stored hydrogen is consumed to meet peak FCEV demand from 8:00 to 17:00. However, due to high selling prices at 7:00 and 18:00, hydrogen is primarily used by the fuel cell to generate electricity for sale back to the grid, thereby generating profit. This demonstrates that an HRS can function as an electricity generator. The hydrogen stored in the tank can thus be managed to supply either FCEV demand or fuel cells, based on real-time data of refueling demand and energy prices, to maximize revenue.

Fig. 12(b) illustrates the operations of various components in the HRS, with positive values indicating energy consumption or HRS-to-grid flows, and negative values indicating energy generation or grid-to-HRS flows. The power exchanged with the grid, as well as the energy consumed/generated by the electrolyzer, fuel cell, and compressor, remain within the allowable range throughout the scheduling horizon. In conclusion, the proposed GAIL algorithm effectively coordinates HRS operations to optimize system profit while satisfying all operational constraints.

4.4. Performance evaluation and comparison

A comparative analysis of the proposed GAIL, DDPG [47], PPO [40], TRPO [48], and SAC [49] is conducted to further prove their effectiveness in the proposed GAIL in HRS energy management. In this study, the

performance of an algorithm in real-time energy scheduling is evaluated using a performance error for the n^{th} scenario, which is defined as follows:

$$J_{\text{gap}} = \frac{J_n^{\text{expert}} - J_n}{J_n^{\text{expert}}} \quad (30)$$

where J_n^{expert} and J_n are the profits achieved by the expert and proposed methods for the n^{th} scenario, respectively. Performance errors are used to assess the total costs accumulated across all test scenarios.

The expert approach addresses the problem using a day-ahead deterministic MILP optimization model, where all necessary forecast data for the entire scheduling cycle, including energy price, solar irradiation, and FCEV refueling demand, are assumed to be known precisely in advance. Consequently, the results obtained by this method represent the best possible solutions and serve as a baseline for comparison.

A comparison of the model performances of the GAIL and other algorithms is presented in Table 6. Among the considered algorithms, the GAIL significantly outperforms the other algorithms. Regarding the performance error, the GAIL achieves an average of 1.81% and a cumulative cost of 1.72% across all test scenarios. In contrast, DDPG and PPO are the second-best algorithms that achieve the closest performance to the GAIL algorithm, with 3.27% and 3.86% on average and 3.07% and 3.87% for the cumulative cost, respectively.

Moreover, the GAIL algorithm demonstrates stable performance with a low standard deviation of the performance error (only 1.09%). The findings in Table 6 are further validated by comparing the boxplots of performance errors for these algorithms, as shown in Fig. 13. The developed GAIL exhibits the lowest mean performance error, indicated by the central line in the boxplot. Additionally, the interquartile range, represented by the height of the box, is relatively narrow for the GAIL algorithm, indicating lower variability in the performance error. Therefore, the effective generalization ability and robustness of the proposed GAIL are confirmed across different scenarios. This indicates that the GAIL algorithm can effectively control the operation of an HRS, achieving near-optimal profit in a given scenario without violating any constraints.

4.5. Detailed analysis of energy scheduling strategies

In this section, a comprehensive comparison is conducted to evaluate the energy scheduling decisions provided by different methods. The same test scenario used in Section 4.3, with data depicted in Fig. 11, is chosen for consistency. Fig. 14 illustrates the actions of various algorithms, including TRPO, SAC, PPO, DDPG, and the proposed GAIL, compared to the expert policy. As mentioned, the decisions from the expert policy represent the best possible results, serving as a benchmark to assess the performance of other methods.

The electrolyzer is crucial for the efficient operation of an HRS, as it converts electricity into hydrogen for storage in tanks. Fig. 14(a)

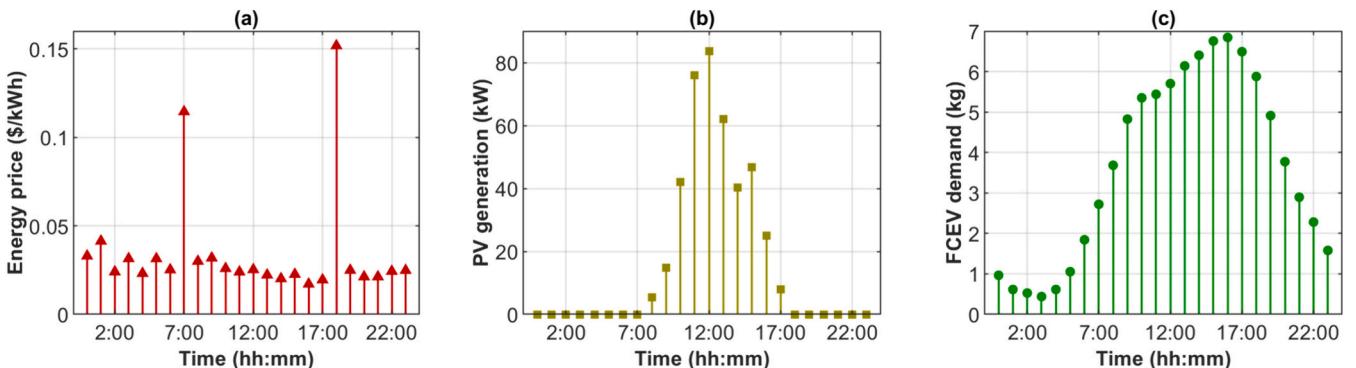


Fig. 11. Testing data for a selected scenario: (a) Energy price, (b) PV generation, and (c) FCEV demand.

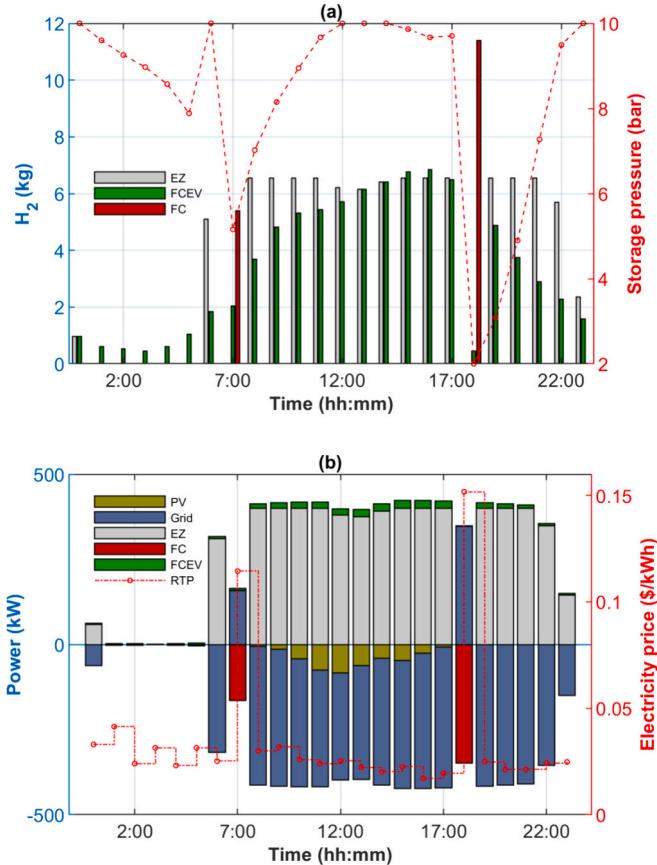


Fig. 12. Scheduling results obtained by GAIL method: (a) Inflows and outflows of hydrogen chain, (b) Operations of various components in the HRS. A positive value indicates energy consumption or HRS-to-grid flows, while a negative value indicates energy generation or grid-to-HRS flows.

Table 6
Comparisons of the different algorithms for test scenarios.

Algorithm	Performance error for each scenario (%)		Performance error for cumulative cost (%)
	Average	Std.	
TRPO	4.29	1.62	4.28
SAC	4.33	1.74	4.40
PPO	3.86	1.56	3.87
DDPG	3.27	2.19	3.07
The proposed GAIL	1.81	1.09	1.72

visually depicts the scheduling decisions of the electrolyzers. It can be observed from Fig. 14(a) that the decisions made by agents trained with different algorithms deviate from the expert decisions, which can be attributed to inevitable errors in model training. Despite these deviations, the agent trained with GAIL exhibits behavior that more closely aligns with the expert policy compared to other algorithms. The difference is most pronounced between 1:00 and 6:00, where GAIL demonstrates high accuracy in capturing the expert policy, while the other algorithms show significant deviations. Similar conclusions can be drawn from the fuel cell scheduling decisions shown in Fig. 14(b), where the actions of GAIL are almost identical to those of the expert policy. Although other algorithms capture the general trend of the expert decisions, they still exhibit notable discrepancies from the optimal decisions. The schedules for FCEV refueling provided by different algorithms are relatively consistent across most time slots, as shown in Fig. 14(c). However, the most significant difference occurs at 18:00,

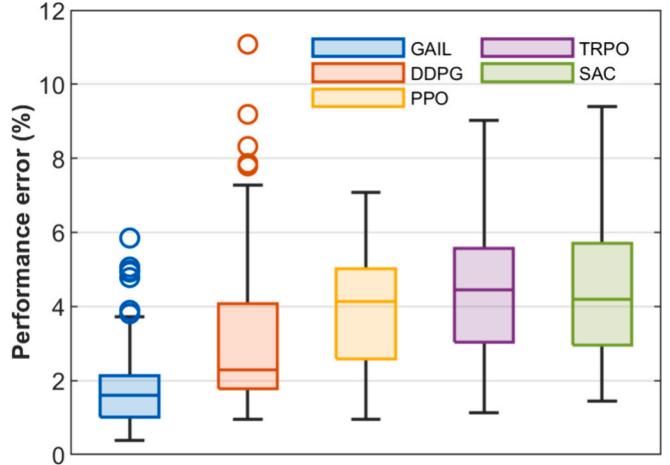


Fig. 13. Comparison of performance errors of five algorithms across 91 test scenarios, presented in a boxplot. The horizontal line indicates the average value for each algorithm. GAIL (blue) exhibits the lowest median performance error and smallest interquartile range, indicating consistent and superior performance. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

when GAIL closely follows the expert policy, while other methods are less effective during this period. The SOP of the HSS over the scheduling horizon is depicted in Fig. 14(d), indicating that GAIL effectively learns and mimics the optimal operations of the hydrogen chain, making it the most accurate algorithm among those compared.

For a quantitative comparison, the proposed GAIL yields a total profit of \$215.04 for the selected scenario. In contrast, the total profits achieved by TRPO, SAC, PPO, and DDPG are \$210.36, \$208.58, \$211.29, and \$213.08, respectively. Compared to the expert policy, which achieves a total profit of \$216.91, all five algorithms demonstrate effectiveness with acceptable gaps, highlighting their good generalization ability. Notably, the proposed GAIL achieves the highest profit among the data-driven approaches, which is relatively close to the theoretically optimal result of the expert policy, with a gap of only 0.86%. In summary, the comparison highlights the superior performance of the proposed GAIL in mimicking the expert policy, particularly regarding scheduling accuracy and cost efficiency. The consistent and accurate performance of GAIL across different components of the HRS underscores its effectiveness in HRS energy management.

5. Conclusion

This study developed a new real-time energy scheduling model based on the GAIL algorithm for HRS energy management. The proposed algorithm uses adversarial training in policy and discriminator networks to mimic expert decision-making processes by directly learning operational strategies from expert demonstrations. To achieve this, expert trajectories are constructed through a sequential collection of state-action pairs, derived by solving a pre-defined optimization problem using a MILP solver with historical data. These expert trajectories are then fed into the discriminator in GAIL during the training phase. The reward output of the discriminator guides the agent towards actions that enhance performance. Extensive simulations were conducted on a benchmark HRS to evaluate the feasibility of the proposed method. The model also enables hydrogen-to-power conversion via a fuel cell to sell energy, serving as an optimal supplement to FCEV refueling, particularly as a primary economic venture during periods with few refueling events. The GAIL algorithm increases system profit by closely imitating expert operational strategies, with a total profit increase of 29% compared to the day-ahead MILP strategy. Comparisons between GAIL and other DRL algorithms show that GAIL surpasses DDPG, PPO, TRPO, and SAC,

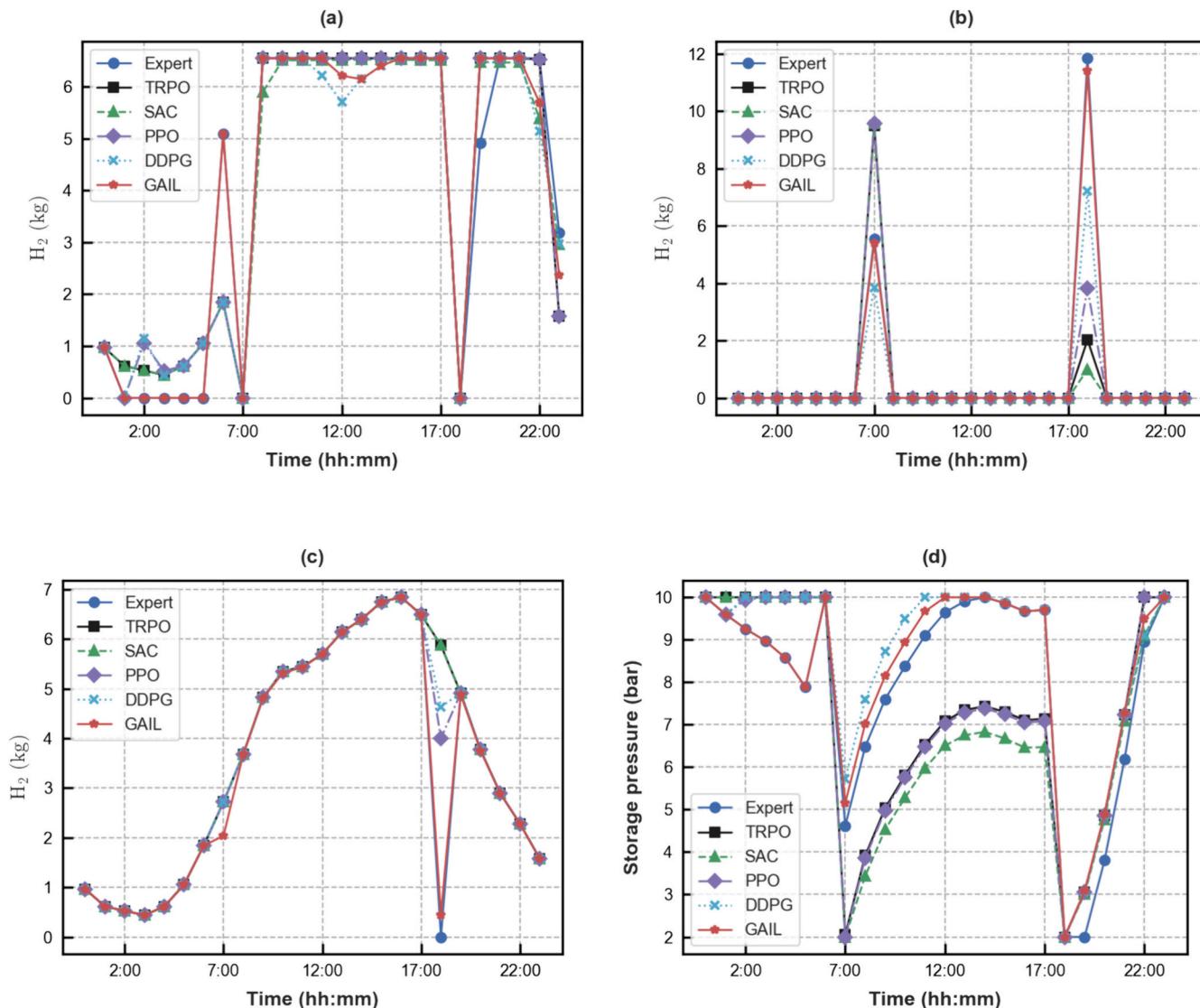


Fig. 14. Comparison of energy scheduling strategies between the proposed GAIL, the expert policy, and other DRL algorithms in a specific scenario: (a) actions of the electrolyzer, (b) actions of the fuel cell, (c) actions of FCEV refueling, and (d) storage pressure in the hydrogen tank. The proposed GAIL algorithm (red line) closely aligns with the expert policy (blue line) across various actions, demonstrating superior performance and reliability. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

demonstrating its superior performance and adaptability in new, unseen scenarios. Thus, the proposed GAIL algorithm shows great potential as a real-time energy management tool for HRS.

Despite the effectiveness of the proposed GAIL, this study has certain limitations. This study does not consider the integration and coordination of HRS with the active distribution network. Additionally, the focus is solely on the economic aspects of HRS, overlooking technical aspects such as battery degradation and grid component failures to simplify the model. Future research should aim to expand the energy management model to include coordination with the distribution network. Furthermore, the generalized performance and robustness of GAIL need to be enhanced to adapt to more complex and large-scale systems.

CRediT authorship contribution statement

Truong Hoang Bao Huy: Writing – original draft, Validation, Methodology, Investigation, Conceptualization. **Nguyen Thanh Minh Duy:** Writing – original draft, Visualization, Validation, Software. **Pham Van Phu:** Validation, Software, Investigation, Data curation. **Tien-Dat Le:** Visualization, Resources, Methodology, Formal analysis.

Seongkeun Park: Writing – original draft, Supervision, Investigation, Funding acquisition, Formal analysis. **Daehee Kim:** Writing – original draft, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This research was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (MOE) (2021RIS-004), supported by Institute for

Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. 2022-0-01197, Convergence security core talent training business (Soon-ChunHyang University)), supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (RS-2023-00245084), and supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2023-00277255). This work was also supported by the Soonchunhyang Research Fund.

References

- [1] Tian Z, Lv H, Zhou W, Zhang C, He P. Review on equipment configuration and operation process optimization of hydrogen refueling station. *Int J Hydron Energy* 2022;47:3033–53. <https://doi.org/10.1016/j.ijhydene.2021.10.238>.
- [2] Sen S, Ganguly S. Opportunities, barriers and issues with renewable energy development – a discussion. *Renew Sust Energ Rev* 2017;69:1170–81. <https://doi.org/10.1016/j.rser.2016.09.137>.
- [3] Rosen MA, Koohi-Fayegh S. The prospects for hydrogen as an energy carrier: an overview of hydrogen energy and hydrogen energy systems. *Energy Ecol Environ* 2016;1:10–29. <https://doi.org/10.1007/s40974-016-0005-z>.
- [4] Ade N, Wilhite B, Goyette H. An integrated approach for safer and economical design of hydrogen refueling stations. *Int J Hydron Energy* 2020;45:32713–29. <https://doi.org/10.1016/j.ijhydene.2020.08.232>.
- [5] Council H. Path to hydrogen competitiveness: A cost perspective. 2020.
- [6] Genovese M, Schlüter A, Scionti E, Piraino F, Corigliano O, Fragiacomo P. Power-to-hydrogen and hydrogen-to-X energy systems for the industry of the future in Europe. *Int J Hydron Energy* 2023;48:16545–68. <https://doi.org/10.1016/j.ijhydene.2023.01.194>.
- [7] Tebibel H. Methodology for multi-objective optimization of wind turbine/battery/electrolyzer system for decentralized clean hydrogen production using an adapted power management strategy for low wind speed conditions. *Energy Convers Manag* 2021;238:114125. <https://doi.org/10.1016/j.enconman.2021.114125>.
- [8] Tostado-Véliz M, Kamel S, Hasanien HM, Turky RA, Jurado F. A mixed-integer-linear-logical programming interval-based model for optimal scheduling of isolated microgrids with green hydrogen-based storage considering demand response. *Journal of Energy Storage* 2022;48:104028. <https://doi.org/10.1016/j.est.2022.104028>.
- [9] Najafi A, Homaei O, Jasinski M, Tsaoosoglou G, Leonowicz Z. Integrating hydrogen technology into active distribution networks: the case of private hydrogen refueling stations. *Energy* 2023;278:127939. <https://doi.org/10.1016/j.energy.2023.127939>.
- [10] Wu X, Zhao W, Li H, Liu B, Zhang Z, Wang X. Multi-stage stochastic programming based offering strategy for hydrogen fueling station in joint energy, reserve markets. *Renew Energy* 2021;180:605–15. <https://doi.org/10.1016/j.renene.2021.08.076>.
- [11] Aki H, Sugimoto I, Sugai T, Toda M, Kobayashi M, Ishida M. Optimal operation of a photovoltaic-generation-powered hydrogen production system at a hydrogen refueling station. *Int J Hydron Energy* 2018;43:14892–904. <https://doi.org/10.1016/j.ijhydene.2018.06.077>.
- [12] Tostado-Véliz M, Arévalo P, Jurado F. A comprehensive electrical-gas-hydrogen microgrid model for energy management applications. *Energy Convers Manag* 2021;228:113726. <https://doi.org/10.1016/j.enconman.2020.113726>.
- [13] Huy THB, Dinh HT, Kim D. Multi-objective framework for a home energy management system with the integration of solar energy and an electric vehicle using an augmented e-constraint method and lexicographic optimization. *Sustain Cities Soc* 2023;88:104289. <https://doi.org/10.1016/j.scs.2022.104289>.
- [14] Shams MH, Niaz H, Liu JJ. Energy management of hydrogen refueling stations in a distribution system: a bilevel chance-constrained approach. *J Power Sources* 2022;533:231400. <https://doi.org/10.1016/j.jpowsour.2022.231400>.
- [15] Grüger F, Hoch O, Hartmann J, Robinius M, Stolten D. Optimized electrolyzer operation: employing forecasts of wind energy availability, hydrogen demand, and electricity prices. *Int J Hydron Energy* 2019;44:4387–97. <https://doi.org/10.1016/j.ijhydene.2018.07.165>.
- [16] MansourLakouraj M, Niaz H, Liu JJ, Siano P, Anvari-Moghaddam A. Optimal risk-constrained stochastic scheduling of microgrids with hydrogen vehicles in real-time and day-ahead markets. *J Clean Prod* 2021;318:128452. <https://doi.org/10.1016/j.jclepro.2021.128452>.
- [17] Shoa JZM, Mirzaei MA, Seiedy H, Zare K. Sustainable energy supply of electric vehicle charging parks and hydrogen refueling stations integrated in local energy systems under a risk-averse optimization strategy. *Journal of Energy Storage* 2022;55:105633. <https://doi.org/10.1016/j.est.2022.105633>.
- [18] Garcia-Torres F, Vilaplana DG, Bordona C, Roncero-Sánchez P, Ridao MA. Optimal Management of Microgrids with External Agents Including Battery/fuel cell electric vehicles. *IEEE Transactions on Smart Grid* 2019;10:4299–308. <https://doi.org/10.1109/TSG.2018.2856524>.
- [19] Xu X, Hu W, Liu W, Wang D, Huang Q, Huang R, et al. Risk-based scheduling of an off-grid hybrid electricity/hydrogen/gas/ refueling station powered by renewable energy. *J Clean Prod* 2021;315:128155. <https://doi.org/10.1016/j.jclepro.2021.128155>.
- [20] Mansour-Saatloo A, Ebadi R, Mirzaei MA, Zare K, Mohammadi-Ivatloo B, Marzband M, et al. Multi-objective IGDT-based scheduling of low-carbon multi-energy microgrids integrated with hydrogen refueling stations and electric vehicle parking lots. *Sustain Cities Soc* 2021;74:103197. <https://doi.org/10.1016/j.scs.2021.103197>.
- [21] Mobasseri A, Tostado-Véliz M, Ghadimi AA, Reza Miveh M, Jurado F. Multi-energy microgrid optimal operation with integrated power to gas technology considering uncertainties. *J Clean Prod* 2022;333:130174. <https://doi.org/10.1016/j.jclepro.2021.130174>.
- [22] Tostado-Véliz M, Ghadimi AA, Miveh MR, Bayat M, Jurado F. Uncertainty-aware energy management strategies for PV-assisted refuelling stations with onsite hydrogen generation. *J Clean Prod* 2022;365:132869. <https://doi.org/10.1016/j.jclepro.2022.132869>.
- [23] Huy THB, Le T-D, Phu PV, Park S, Kim D. Real-time power scheduling for an isolated microgrid with renewable energy and energy storage system via a supervised-learning-based strategy. *Journal of Energy Storage* 2024;88:111506. <https://doi.org/10.1016/j.est.2024.111506>.
- [24] Cao B, Wu X, He M, Li X, He W, Chen S. Proximal-Policy-Optimization-based Intraday Scheduling of Hydrogen Fueling Station. 2022 IEEE 5th International Electrical and Energy Conference (CIEEC) 2022:4357–62. <https://doi.org/10.1109/CIEEC54735.2022.9846726>.
- [25] Qi Y, Xu X, Liu Y, Pan L, Liu J, Hu W. Intelligent energy management for an on-grid hydrogen refueling station based on dueling double deep Q network algorithm with NoisyNet. *Renew Energy* 2024;222:119885. <https://doi.org/10.1016/j.renene.2023.119885>.
- [26] Jiang Y, Liu J, Zheng H. Optimal scheduling of distributed hydrogen refueling stations for fuel supply and reserve demand service with evolutionary transfer multi-agent reinforcement learning. *Int J Hydron Energy* 2024;54:239–55. <https://doi.org/10.1016/j.ijhydene.2023.04.128>.
- [27] Huy THB, Truong Dinh H, Ngoc Vo D, Kim D. Real-time energy scheduling for home energy management systems with an energy storage system and electric vehicle based on a supervised-learning-based strategy. *Energy Convers Manag* 2023;292:117340. <https://doi.org/10.1016/j.enconman.2023.117340>.
- [28] Schaal S. Learning from Demonstration. *Advances in Neural Information Processing Systems9*. MIT Press; 1996.
- [29] Osa T, Pajarinen J, Neumann G, Bagnell JA, Abbeel P, Peters J. An algorithmic perspective on imitation learning. *FNT in Robotics* 2018;7:1–179. <https://doi.org/10.1561/2300000053>.
- [30] Ross S, Gordon G, Bagnell D. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings* 2011:627–35.
- [31] Ross S, Bagnell D. Efficient Reductions for Imitation Learning. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. JMLR Workshop and Conference Proceedings* 2010:661–8.
- [32] Ng AY, Russell SJ. Algorithms for Inverse Reinforcement Learning. *Proceedings of the Seventeenth International Conference on Machine Learning, San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2000. p. 663–70*.
- [33] Ho J, Ermon S. Generative Adversarial Imitation Learning. 2016. <https://doi.org/10.48550/arXiv.1606.03476>.
- [34] Goodfellow IJ, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, et al. Generative Adversarial Networks. 2014. <https://doi.org/10.48550/arXiv.1406.2661>.
- [35] Li R, Zou Z. Enhancing construction robot learning for collaborative and long-horizon tasks using generative adversarial imitation learning. *Adv Eng Inform* 2023;58:102140. <https://doi.org/10.1016/j.aei.2023.102140>.
- [36] Pham D-T, Tran T-N, Alami S, Duong VN. A generative adversarial imitation learning approach for realistic aircraft taxi-speed modeling. *IEEE Trans Intell Transp Syst* 2022;23:2509–22. <https://doi.org/10.1109/TITS.2021.3119073>.
- [37] Tostado-Véliz M, Kamel S, Hasanien HM, Turky RA, Jurado F. Uncertainty-aware day-ahead scheduling of microgrids considering response fatigue: an IGDT approach. *Appl Energy* 2022;310:118611. <https://doi.org/10.1016/j.apenergy.2022.118611>.
- [38] Tostado-Véliz M, Kamel S, Aymen F, Rezaee Jordhei A, Jurado F. A stochastic-IGDT model for energy management in isolated microgrids considering failures and demand response. *Appl Energy* 2022;317:119162. <https://doi.org/10.1016/j.apenergy.2022.119162>.
- [39] Tostado-Véliz M, Rezaee Jordhei A, Fernández-Lobato L, Jurado F. Robust energy management in isolated microgrids with hydrogen storage and demand response. *Appl Energy* 2023;345:121319. <https://doi.org/10.1016/j.apenergy.2023.121319>.
- [40] Schulman J, Wolski F, Dhariwal P, Radford A, Klimov O. Proximal Policy Optimization Algorithms. 2017. <https://doi.org/10.48550/arXiv.1707.06347>.
- [41] Mukherjee U, Marouf mashat A, Narayan A, Elkamel A, Fowler M. A stochastic programming approach for the planning and operation of a power to gas energy hub with multiple energy recovery pathways. *Energies* 2017;10:868. <https://doi.org/10.3390/en10070868>.
- [42] Kurtz JM, Sprik S, Saur G, Onorato S. Fuel Cell Electric Vehicle Driving and Fueling Behavior. Golden, CO (United States): National Renewable Energy Lab. (NREL); 2019. <https://doi.org/10.2172/1501674>.
- [43] Raffin A, Hill A, Gleave A, Kanervisto A, Ernestus M, Dormann N. Stable-Baselines3: reliable reinforcement learning implementations. *J Mach Learn Res* 2021;22:1–8.
- [44] Gleave A, Taufeeque M, Rocamonde J, Jenner E, Wang SH, Toyer S, et al. imitation: Clean Imitation Learning Implementations. 2022. <https://doi.org/10.48550/arXiv.2211.11972>.
- [45] Akiba T, Sano S, Yanase T, Ohta T, Koyama M. Optuna: A Next-generation Hyperparameter Optimization Framework. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, New York, NY,

- USA: Association for Computing Machinery. 2019. p. 2623–31. <https://doi.org/10.1145/3292500.3330701>.
- [46] Gao S, Lee R bin Z, Huang Z, Xiang C, Yu M, Tan KT, et al. A hybrid approach for home energy management with imitation learning and online optimization. IEEE Trans Industr Inform 2024;20:4527–39. <https://doi.org/10.1109/TII.2023.3324939>.
- [47] Lillicrap TP, Hunt JJ, Pritzel A, Heess N, Erez T, Tassa Y, et al. Continuous control with deep reinforcement learning. 2019. <https://doi.org/10.48550/arXiv.1509.02971>.
- [48] Schulman J, Levine S, Moritz P, Jordan MI, Abbeel P. Trust Region Policy Optimization. 2017. <https://doi.org/10.48550/arXiv.1502.05477>.
- [49] Haarnoja T, Zhou A, Abbeel P, Levine S. Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor. 2018. <https://doi.org/10.48550/arXiv.1801.01290>.