electronics

MDPI

Article

Federated Deep Reinforcement Learning for Energy Scheduling in Privacy-Sensitive PV-EV Charging Networks

Yongguang Zhao, Xinni Li, Yongqing Zheng * and Wei Guo

School of Software, Shandong University, Jinan 250101, China
* Correspondence: 13335135335@163.com

Abstract

The large-scale adoption of electric vehicles (EVs) improves transport sustainability but creates severe peak-time stress on distribution grids. In PV-assisted charging networks, station operators must jointly decide retail charging prices and energy-storage dispatch under uncertain demand and generation conditions. This paper develops a distributed federated deep reinforcement learning framework for multi-station scheduling, where each station trains a local soft actor–critic (SAC) policy and only model parameters are exchanged with a global aggregator. To better adapt prices to local supply–demand conditions, we introduce a sales-factor-based correction mechanism that links the announced price to demand pressure and storage status. The objective combines station revenue, operating expenses, and user-discomfort-related penalties under operational constraints. Simulation results on a five-station setting show stable convergence and consistent gains over benchmark methods, with profit improvements of 3.90–39.00%. The framework keeps raw operational data local and supports collaborative optimization across stations.

Keywords: electric vehicles; solar energy storage charging station; energy scheduling; game theory; deep reinforcement learning

1. Introduction

Driven by the rapid evolution of smart-grid technologies for power-and-energy management, traditional electric-vehicle (EV) charging stations are transitioning into photovoltaic-storage (PV-storage) charging stations that integrate PV generation and energy-storage systems (ESS). To further improve the profitability of such stations, devising rational dynamic-pricing and energy-scheduling schemes is essential. Existing research on dynamic pricing and energy scheduling for PV-storage charging stations is dominated by model-based approaches, which demand an accurate system model that faithfully captures the highly dynamic environment—encompassing the station's charging/discharging operations, EV users' charging behaviour, and PV-generation forecasts. Given this ever-changing context, accurately modeling the station's operating conditions is an arduous task. Moreover, as the system model scales up, the associated surge in computational time prevents model-based methods from efficiently leveraging large data sets.

To overcome these challenges, recent studies have adopted model-free reinforcement-learning (RL) and deep-RL (DRL) methods to optimize energy scheduling and pricing policies for individual or multiple PV-storage charging stations. However, because these centralized approaches must process vast amounts of data aggregated on a central server, they significantly increase computational latency. Furthermore, the data can be maliciously exploited by adversaries, compromising the privacy of station operations and leading

Check for updates

Academic Editor: Shiho Kim

Received: 24 January 2026
Revised: 11 February 2026
Accepted: 24 February 2026
Published: 28 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

to erroneous calculations of charging/discharging schedules and pricing plans [1]. In recent years, federated learning (FL) and federated RL (FRL) have emerged to mitigate computational complexity and strengthen privacy protection. By enabling distributed training without sharing raw local data, these techniques safeguard data privacy while simultaneously reducing computational burden.

Table 1 compares this paper with representative prior work. Existing studies mostly optimize single-station operations from either the energy-scheduling or pricing perspective, and partially consider EV-user satisfaction and uncertainty. However, research addressing multi-station scenarios with distributed energy resources remains insufficient. Although RL or DRL methods have been applied to complex multi-station environments, privacy protection is still largely overlooked. A comprehensive solution that simultaneously enables multi-site coordination and data privacy is lacking.

Table 1. Comparison between representative related work and our work.

Comparison Dimension	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	Our Work
Energy Scheduling	✓	✗	✗	✓	✓	✗	✓	✓	✓	✓
Pricing Strategy	✓	✓	✓	✓	✓	✓	✓	✓	✗	✓
Uncertainty Handling	✓	✗	✓	✗	✓	✗	✗	✓	✗	✓
Distributed Energy Resources	✓	✗	✓	✗	✗	✗	✗	✗	✓	✓
EV User Satisfaction	✓	✓	✗	✓	✓	✗	✓	✓	✗	✓
Multi-Charging-Station Operation	✗	✓	✗	✗	✗	✗	✗	✗	✓	✓
RL/DRL Methodology	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓
Privacy Protection	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓

To address these gaps, this paper proposes, to the best of our knowledge, a hierarchical distributed FRL-based DRL model that performs dynamic pricing and energy scheduling for multiple PV-storage charging stations while preserving their privacy. In the model, each station's DRL agent computes profitable selling prices and optimal ESS charge/discharge quantities, minimizing both station operating costs and EV-user discomfort costs, while keeping each station's private data secure. This paper consists of three parts: (1) A joint DRL algorithm that simultaneously schedules profitable selling prices and ESS charge/discharge actions for multiple PV-storage charging stations. (2) An action space and reward function designed for the DRL agents, where the action includes a sales factor set according to the station's operating conditions, and the reward function accounts for station profit, operating costs, and EV-user discomfort costs. (3) An FRL framework that protects the data privacy of PV-storage charging stations during pricing and energy-scheduling processes. The main contributions of this paper are as follows:

1. A hierarchical distributed FRL-based DRL model is proposed to maximize the profit of PV-storage charging stations under dynamic pricing while minimizing both station operating costs and EV-user discomfort costs, all while preserving the stations' data privacy.
2. State, action, and reward spaces are designed for the DRL-based local agents (LAs) within each PV-storage charging station, and an FRL framework is introduced in which all LAs and the global aggregator (GA) iteratively exchange their neural network models until every LA converges to its optimal local neural network model. Because only the neural network models—not the raw station data—are exchanged between LAs and the GA, the stations' data reduce raw-data exposure.
3. A sales factor that reflects the operating conditions of a PV-storage charging station (such as total energy demand and the remaining energy in the ESS) is proposed.

The actual selling price depends on the value of this factor, allowing station operators to adaptively adjust the sales factor to cope with potential future energy shortages.

4. Simulation results demonstrate that the proposed LAs successfully compute profitable selling prices for PV-storage charging stations and perform economical ESS charging/discharging while safeguarding station data privacy. Moreover, the results verify that the proposed method outperforms existing approaches in terms of profit and training stability.

2. Related Work

To optimize the economic benefits of PV-storage charging stations and establish an efficient energy-trading system among multi-stakeholder stations, researchers have begun exploring privacy-preserving energy-scheduling strategies between PV-storage charging stations to improve overall station revenue while ensuring data privacy.

Current research on PV-storage charging stations primarily focuses on planning and operational optimization. For instance, Ref. [11] proposed a multi-objective optimization method for site selection and charger allocation to improve overall efficiency. Ref. [12] considered queuing time and charging cost in siting decisions, aiming to reduce EV charging expenses and alleviate range anxiety. From the operational side, Ref. [13] developed a chance-constrained optimization method to minimize charging costs while accounting for uncertainties in load and PV output. Ref. [14] analyzed the impact of cyberattacks on station operations and their potential risks to the power grid. Ref. [15] introduced a two-stage optimization approach, with the first stage handling station planning and the second managing EV scheduling. Ref. [16] presented a coordinated optimization method for distribution networks and PV-storage charging stations, considering the coupling of energy and reserves to ease the burden of large-scale integration. However, most of these studies focus on individual station planning and optimization, lacking exploration of energy interactions among stations, and thus fail to fully enhance overall energy utilization efficiency.

To increase the profitability of PV-storage charging stations and guide EV charging behaviour, dynamic pricing strategies have become a major research focus. Ref. [17] proposed an optimization method based on zonal dynamic electricity pricing, assigning different tariffs according to load characteristics to improve local charging efficiency. Ref. [3] introduced a dynamic pricing model that shifts EV loads to off-peak hours, reducing overlap with residential demand and easing grid pressure. Ref. [4] developed a distributed pricing algorithm that generated individual selling-price signals for each station to lower user charging costs while maintaining grid stability. Ref. [18] formulated a bi-level optimization model for integrated transportation and power networks, enabling stations to profit while reducing operating costs. Ref. [19] presented a robust bi-level framework considering operational uncertainty, where the upper level determines pricing and station configuration, and the lower-level models EV user behaviour. Ref. [5] uses a queuing model to jointly optimize admission control, pricing, and scheduling to maximize station profit. Ref. [6] reduces service drop rates using a dual-mode charging strategy. Ref. [20] proposed a four-stage, multi-time-scale optimization algorithm covering day-ahead scheduling, pricing and incentives, hour-ahead adjustments, and real-time control. While these methods improve economic performance, most rely on centralized optimization and lack privacy-preserving mechanisms, raising concerns about user data leakage.

To overcome the limitations of centralized approaches, recent studies have turned to model-free methods, such as deep reinforcement learning (DRL), for energy scheduling and dynamic pricing in PV-storage charging stations. For example, Ref. [21] combined Q-learning with Artificial Neural Networks (ANN) and Recurrent Neural Networks (RNN), achieving higher forecasting accuracy than traditional methods. Ref. [7] applied both model-

based dynamic programming and model-free methods like Q-learning and actor–critic to optimize pricing strategies. Ref. [8] introduced a multi-agent RL framework with utility and station agents to coordinate pricing. Ref. [9] proposed an online RL algorithm that considers stochastic EV behavior for joint scheduling. Ref. [10] developed a MADRL approach to optimize energy usage and costs across multiple stations through energy sharing. While these studies highlight the potential of DRL in dynamic and uncertain environments, most still focus on single-station scenarios and lack effective mechanisms for cooperative optimization and privacy protection in multi-station networks.

To further reduce computational complexity and enhance privacy protection, Federated Learning (FL) and Federated Reinforcement Learning (FRL) have gradually become emerging directions for energy-scheduling research in PV-storage charging stations. These approaches achieve optimal scheduling strategies through distributed model training without sharing local data. For example, Ref. [22] proposed a DRL-based FRL scheme for multiple industrial-IoT applications to meet their quality-of-service requirements. Ref. [23] presented an FL-based hybrid RL framework to improve network throughput and reduce handover costs. In Ref. [24], FRL was applied to edge caching to manage large-scale network issues by building shared prediction models. Ref. [25] designed a weighted-attention FRL framework to improve training performance and address local-model quality imbalance in device-to-device-assisted heterogeneous cooperative edge caching within mobile networks. These studies demonstrate the potential of FL in protecting data privacy and reducing computational cost, yet most focus on computer networks and communications rather than in-depth optimization for PV-storage charging-station energy scheduling.

Existing research has made significant advances in energy scheduling, dynamic pricing, and privacy protection for PV-storage charging stations. Optimization-based methods have effectively improved planning and scheduling efficiency, and intelligent strategies, such as DRL, have enhanced adaptive learning capabilities, while FL and FRL technologies have provided new solutions for data-privacy preservation. Nevertheless, these approaches still exhibit certain limitations. Centralized optimization entails high privacy risks, single-station optimization struggles to accommodate coordinated scheduling among multiple PV-storage charging stations, and current privacy-preserving techniques impose substantial computational overhead that compromises real-time performance. Therefore, future work should further integrate distributed deep reinforcement learning with federated reinforcement learning to optimize energy scheduling of PV-storage charging stations, strengthen privacy protection, and reduce computational costs, thereby achieving a more efficient and secure energy-trading mechanism.

3. Overview of Relevant Basic Theories

3.1. Photovoltaic-Storage Charging Station System

A typical PV-storage charging station integrates photovoltaic (PV) generation, an energy storage system (ESS), and electric vehicle (EV) charging infrastructure. The ESS plays a critical role in balancing the intermittent nature of solar energy and the continuous demand of EV charging. The operational status of the ESS is characterized by the State of Energy (*SOE*), which reflects the remaining electricity capacity. To maintain system stability and economic efficiency, the energy scheduling must optimize the power flow among the PV system, ESS, and the power grid under time-of-use (TOU) electricity prices.

3.2. Soft Actor–Critic (SAC) Algorithm

Deep reinforcement learning (DRL) optimizes energy scheduling through iterative interactions between an intelligent agent and the environment. This paper adopts the soft actor–critic (SAC) algorithm, an off-policy actor-critic framework based on the maximum

entropy principle. Unlike traditional DRL, SAC aims to maximize both the expected return and the policy's entropy:

$$J(\pi) = \sum_{t=0}^{T} \mathbb{E}_{(s_t,a_t)\sim\rho_\pi}[r(s_t,a_t) + \alpha\mathcal{H}(\pi(\cdot|s_t))] \quad (1)$$

where $\mathcal{H}(\cdot)$ denotes the entropy and α is the temperature parameter controlling the exploration-exploitation trade-off. The actor network outputs the pricing and scheduling policy, while the critic network evaluates the state-action value to guide policy improvement.

3.3. Federated Reinforcement Learning (FRL)

Federated Learning (FL) is a distributed machine learning paradigm that enables multiple entities to collaboratively train a model without sharing their raw private data. In the context of multi-station energy scheduling, Federated Reinforcement Learning (FRL) combines FL with DRL. The architecture consists of several Local Agents (LAs) and a Global Agent (GA). Each LA_n trains its local model parameters θ_n using its own operational data (e.g., local PV output and EV demand). The GA then aggregates these parameters to update the global model θ_G:

$$\theta_G = \frac{1}{N}\sum_{n=1}^{N}\theta_n. \quad (2)$$

This process ensures that private station data remain localized while benefiting from the collective knowledge of the entire network.

3.4. Game Theory in Dynamic Pricing

The energy scheduling and pricing problem among multiple charging stations can be modeled as a non-cooperative game, where each station acts as a rational player seeking to maximize its own profit. According to the Nash equilibrium theory, no station can unilaterally improve its profit by changing its pricing strategy when other stations' strategies are fixed. The pricing mechanism in this study is inspired by the supply–demand ratio in economics, where the selling price is adaptively adjusted based on the station's energy scarcity and user demand.

4. Problem Definition

The system model of this paper is shown in Figure 1, comprising a central-server PV-storage charging-station aggregator and PV-storage charging stations individually managed by their operators. The considered scenario involves multiple PV-storage charging stations scheduling their ESS charge/discharge quantities and setting selling prices when supplying EV users. Energy sharing and operational-data sharing among stations are prohibited. All stations purchase electricity from the grid under a Time-of-Use (TOU) tariff. Each station is equipped with an SAC-based energy-scheduling system, where the LA determines selling prices and ESS charge/discharge strategies every 30 min. PV generation output, total EV energy demand, and fixed station consumption are forecast. Every LA aims to maximize profit under dynamic pricing while respecting operational constraints and maintaining EV-user satisfaction by properly scheduling ESS charge/discharge. However, our framework prevents raw data sharing but does not, by itself, guarantee resistance to gradient/model-inversion attacks. Secure aggregation and differential privacy are left for future integration.

To train all LA neural network models in a distributed manner and protect the privacy of PV-storage charging station data, this section proposes an FRL framework in which the GA and all LA use the SAC algorithm to cooperatively schedule the selling prices and ESS charge/discharge of multiple PV-storage charging stations. As shown in Figure 2, the FRL framework has two levels. At the first level, each LA n uses the SAC algorithm to train its

local neural network model θ_n under diverse environments (including weather conditions, TOU prices, and predicted total EV energy demand). Then, each LA sends its trained model to the GA located at the PV-storage charging station aggregator. At the second level, the GA aggregates all received local neural network models to generate a global neural network model θ_G for all LA, and the resulting global model is broadcast to all LA, which use it to retrain their local models. Since only models—not data—are exchanged between the GA and the LA, the privacy of PV-storage charging station data is ensured.

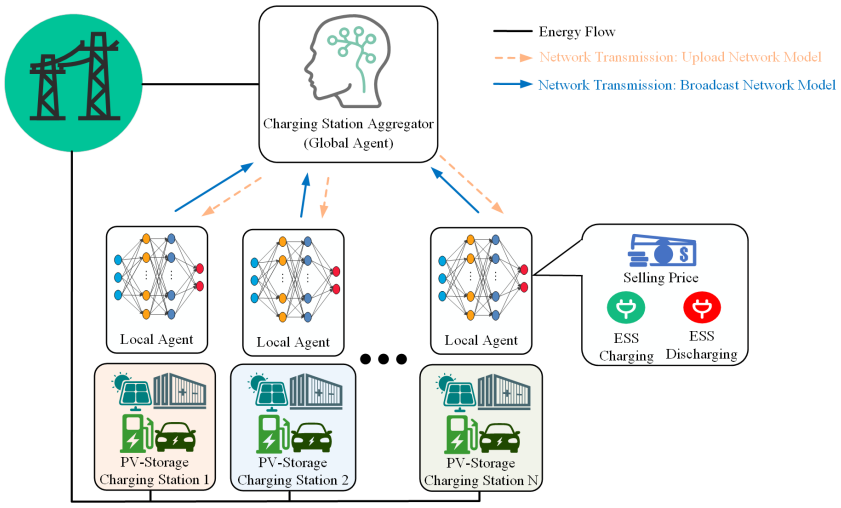


Figure 1. Energy scheduling and dynamic pricing model of PV-storage charging station under privacy protection.

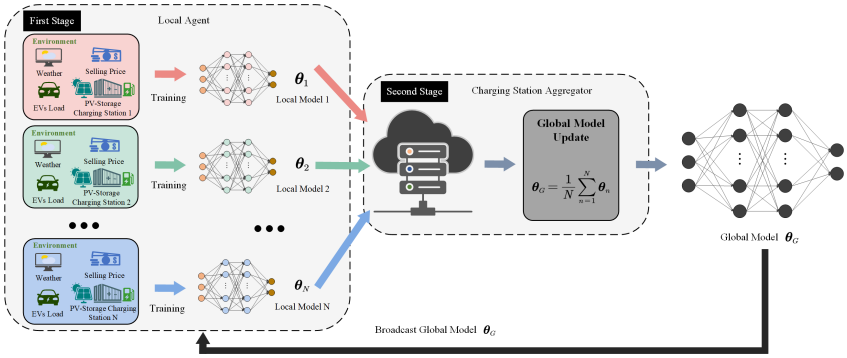


Figure 2. Interactive training process of LA and GA in FRL framework.

The ESS charging and discharging power set of the PV-storage charging station is expressed as follows:

$$\mathbf{E}_t^{\mathrm{ch/dch}} = \left[E_{1,t}^{\mathrm{ch/dch}}, E_{2,t}^{\mathrm{ch/dch}}, \ldots, E_{n,t}^{\mathrm{ch/dch}}\right] \tag{3}$$

where $E_{n,t}^{\mathrm{ch/dch}}$ represents the charge/discharge amount of PV-storage charging station n during time period t. The set of maximum charge/discharge amounts for the PV-storage charging station is denoted as follows:

$$\bar{\mathbf{E}}_t^{\mathrm{ch/dch}} = \left[\bar{E}_1^{\mathrm{ch/dch}}, \bar{E}_2^{\mathrm{ch/dch}}, \ldots, \bar{E}_n^{\mathrm{ch/dch}}, \ldots, \bar{E}_N^{\mathrm{ch/dch}}\right]. \tag{4}$$

The set of minimum charge/discharge amounts for the PV-storage charging station is denoted as follows:

$$\underline{\mathbf{E}}_t^{\mathrm{ch/dch}} = \left[\underline{E}_1^{\mathrm{ch/dch}}, \underline{E}_2^{\mathrm{ch/dch}}, \ldots, \underline{E}_n^{\mathrm{ch/dch}}, \ldots, \underline{E}_N^{\mathrm{ch/dch}}\right] \tag{5}$$

where $\bar{E}_N^{\mathrm{ch/dch}}$ and $\underline{E}_N^{\mathrm{ch/dch}}$ represent the maximum and minimum charge/discharge amounts of the ESS for PV-storage charging station n, respectively. In addition, there is the State of Energy (SOE) set for the ESS of the PV-storage charging station, which is denoted as follows:

$$\mathbf{SOE}_t = [\mathrm{SOE}_{1,t}, \mathrm{SOE}_{2,t}, \ldots, \mathrm{SOE}_{n,t}, \ldots, \mathrm{SOE}_{N,t}] \tag{6}$$

where $\mathrm{SOE}_{n,t}$ is the SOE of PV-storage charging station n during time period t. The set of maximum SOE capacities for the ESS of the PV-storage charging station is denoted as follows:

$$\overline{\mathbf{SOE}} = \left[\overline{\mathrm{SOE}}_1, \overline{\mathrm{SOE}}_2, \ldots, \overline{\mathrm{SOE}}_n, \ldots, \overline{\mathrm{SOE}}_N\right]. \tag{7}$$

The set of minimum SOE capacities for the ESS of the PV-storage charging station is denoted as follows:

$$\underline{\mathbf{SOE}} = [\underline{\mathrm{SOE}}_1, \underline{\mathrm{SOE}}_2, \ldots, \underline{\mathrm{SOE}}_n, \ldots, \underline{\mathrm{SOE}}_N] \tag{8}$$

where $\overline{\mathrm{SOE}}_n$ and $\underline{\mathrm{SOE}}_n$ are the maximum and minimum SOE capacities of the ESS for PV-storage charging station n, respectively. $\widehat{\mathbf{E}}_t^{\mathrm{PV}} = \left[\widehat{E}_{1,t}^{\mathrm{PV}}, \widehat{E}_{2,t}^{\mathrm{PV}}, \ldots, \widehat{E}_{N,t}^{\mathrm{PV}}\right]$ represents the set of forecasted power generation of the PV system connected to the PV-storage charging station.The state of operation of the SOE is dynamic, expressed by the SOE of the ESS of the photovoltaic energy storage station n at time t, which is the $SOE_{n,t-1}$ of the previous period $t-1$, the charge quantity $E_{n,t}^{ch}$, the discharge quantity $E_{n,t}^{dch}$, and the predicted PV generation quantity $\widehat{E}_{n,t}^{PV}$ expressed, expressed as follows:

$$SOE_{n,t} = SOE_{n,t-1} + E_{n,t}^{\mathrm{ch}} - E_{n,t}^{\mathrm{dch}} + \widehat{E}_{n,t}^{\mathrm{PV}}. \tag{9}$$

The set of selling prices for the PV-storage charging station is denoted as follows:

$$\boldsymbol{\pi}_t^{\mathrm{sell}} = \left[\pi_{1,t}^{\mathrm{sell}}, \pi_{2,t}^{\mathrm{sell}}, \ldots, \pi_{n,t}^{\mathrm{sell}}, \ldots, \pi_{N,t}^{\mathrm{sell}}\right] \tag{10}$$

where $\pi_{n,t}^{\mathrm{sell}}$ represents the selling price of electricity for PV-storage charging station n when discharging to EV users during time period t.

The set of maximum selling price sets for the PV-storage charging station during time period t is denoted as follows:

$$\overline{\boldsymbol{\pi}}_t^{\mathrm{sell}} = \left[\overline{\pi}_{1,t}^{\mathrm{sell}}, \overline{\pi}_{2,t}^{\mathrm{sell}}, \ldots, \overline{\pi}_{N,t}^{\mathrm{sell}}\right]. \tag{11}$$

The set of minimum selling price sets for the PV-storage charging station during time period t is denoted as follows:

$$\boldsymbol{\pi}_t^{\mathrm{sell}} = \left[\pi_{1,t}^{\mathrm{sell}}, \pi_{2,t}^{\mathrm{sell}}, \ldots, \pi_{N,t}^{\mathrm{sell}}\right]. \tag{12}$$

The predicted energy consumption of the PV-storage charging station is divided into two categories. The set of predicted fixed energy consumption for the PV-storage charging station (for example, heating, ventilation, and air conditioning equipment):

$$\widehat{\mathbf{E}}_t^{\mathrm{Fixed}} = \left[\widehat{E}_{1,t}^{\mathrm{Fixed}}, \widehat{E}_{2,t}^{\mathrm{Fixed}}, \ldots, \widehat{E}_{N,t}^{\mathrm{Fixed}}\right]. \tag{13}$$

The set of predicted total energy demand for EVs users at the PV-storage charging station is denoted as follows:

$$\widehat{\mathbf{E}}_t^{\mathrm{EV}} = \left[\widehat{E}_{1,t}^{\mathrm{EV}}, \widehat{E}_{2,t}^{\mathrm{EV}}, \ldots, \widehat{E}_{N,t}^{\mathrm{EV}}\right]. \tag{14}$$

5. Dynamic Pricing and Energy Scheduling Strategy of DRL Based on FRL

5.1. FRL and DRL

The DRL method used in this subsection is the SAC algorithm. FRL is a distributed framework used to train LA neural network models and compute their optimal policies while protecting the private data of LAs. The FRL framework includes one GA deployed at the PV-storage charging station aggregator and multiple LAs distributed across local devices. During distributed training, the GA and LAs interact periodically to update the global and local neural network models of LAs.

In this subsection, we define a set of LAs $\mathcal{N} := \{1, 2, 3, \ldots, N\}$, where each $LA_n \in \mathcal{N}$ is trained using its private dataset D_n, and builds a local neural network model (i.e., the neural network weights of LA_n). After each LA_n completes local training, its trained neural network model θ_n is sent to the PV-storage charging station aggregator. The GA then aggregates and stores the neural network models from all LAs as a global batch $\chi = [\theta_1, \theta_2, \ldots, \theta_N]$. Next, the GA uses the local models in the global batch to construct a global model $\theta_G = f(\theta_1, \theta_2, \ldots, \theta_N)$, which represents the global neural network parameters shared across all LAs. Here, $f(\cdot)$ is defined as average aggregation: $f(\theta_1, \theta_2, \ldots, \theta_N) = \frac{1}{N} \sum_{n=1}^{N} \theta_n$. After that, the GA sends the newly constructed θ_G to all LAs. Each LA replaces its previous local model parameters with θ_G, and then improves its local model using new data to compute its optimal policy. Each GA–LA iteration is referred to as one communication round. The entire FRL process continues until all LAs obtain the required local models after several communication rounds. The theoretical proof of LA training-curve convergence can be found in Ref. [26].

In traditional centralized DRL algorithms, model training is performed at the PV-storage charging station aggregator to manage multiple local devices. The aggregator receives all training data from local devices and trains an optimal model for them. However, the large amount of data collected from local devices may significantly increase GA training time. In addition, centralized collection of all local data may cause privacy leakage. Compared with centralized methods, fully distributed DRL algorithms do not require data sharing at the PV-storage charging station aggregator. Each DRL agent is allowed to train a local model using only its own data, thereby reducing training time in the DRL process. However, in fully distributed DRL, some DRL agents may not have sufficient training data, which can lead to overfitting. As a result, the trained local models may be biased, further affecting the accuracy of DRL agent policies.

FRL is a promising solution to the above challenges in centralized and fully distributed multi-agent DRL. Under the FRL framework, all LAs are trained in a distributed manner, thereby reducing computational time during training. Moreover, the GA at the PV-storage charging station aggregator relies only on local neural network models and does not access private data, effectively preventing privacy leakage. Local neural network models are periodically updated through the global neural network model built by the PV-storage charging station aggregator. Such periodic global updates help improve the model performance of all LAs and further prevent overfitting of local models. In this subsection, all DRL-based LAs in the PV-storage charging station interact with the GA in the PV-storage charging station aggregator to improve their local models and find the optimal policies for all LAs (i.e., profitable selling prices and ESS charging/discharging power), while ensuring data privacy of the PV-storage charging station and reducing the training time of all LAs.

5.2. SAC-Based Local Intelligent Agent

5.2.1. State Space

For $t = 1, 2, \ldots, 48$, the state space of LA_n at time slot t, denoted by $S_{n,t}$, is defined as follows:

$$S_{n,t} = \left\{ \pi_t^{\mathrm{TOU}}, \mathrm{SOE}_{n,t-1}, E_{n,t-1}^{\mathrm{EV,tr}}, E_{n,t}^{\mathrm{EV,total}}, \hat{E}_{n,t}^{\mathrm{PV}} \right\} \tag{15}$$

where π_t^{TOU} is the TOU price for purchasing electricity from the grid at time slot t for the PV-storage charging station. $\mathrm{SOE}_{n,t-1}$ is the SOE of the ESS in station n at time slot $t-1$. $E_{n,t-1}^{\mathrm{EV,tr}}$ is the unmet EV energy demand transferred from time slot $t-1$, caused by insufficient ESS discharge in station n at $t-1$. $E_{n,t}^{\mathrm{EV,total}}$ is the sum of the total EV energy demand at time slot t and the unmet EV energy demand from time slot $t-1$. Finally, $\hat{E}_{n,t}^{\mathrm{PV}}$ is the predicted PV generation of station n at time slot t.

5.2.2. Action Space

Given a state, each LA in the PV-storage charging station computes two actions corresponding to ESS charging/discharging energy and selling price. The action space of LA_n at time slot t is defined as follows:

$$A_{n,t} = \left\{ E_{n,t}^{\mathrm{ch/dch}}, \pi_{n,t}^{\mathrm{sell}} \right\} \tag{16}$$

where $E_{n,t}^{\mathrm{ch/dch}}$ and $\pi_{n,t}^{\mathrm{sell}}$ denote the ESS charging/discharging energy and selling price of LA_n at time slot t, respectively. These two actions are interdependent, and the value of $\pi_{n,t}^{\mathrm{sell}}$ is determined according to the charging or discharging amount $E_{n,t}^{\mathrm{ch/dch}}$. This is because the selling price is positive only when the PV-storage charging station is in ESS discharging mode; otherwise, the selling price is zero. In addition, $\pi_{n,t}^{\mathrm{sell}}$ is a corrected action of the original action $\tilde{\pi}_{n,t}^{\mathrm{sell}}$, where $\tilde{\pi}_{n,t}^{\mathrm{sell}}$ is the actual output of the actor network. The corrected action is obtained by multiplying a selling factor, which reflects the capability of ESS charging/discharging energy relative to the total energy demand of the PV-storage charging station.

The derivation of the selling factor is inspired by the results in [27], where a pricing model among producers is built in various energy-sharing environments based on the supply–demand ratio in economics. At time slot t, the selling factor of station n is defined as the ratio of the total energy demand of the PV-storage charging station to the remaining ESS energy. The specific expression is as follows:

$$\lambda_{n,t} = \epsilon_n \frac{E_{n,t}^{\mathrm{EV,total}}}{\mathrm{SOE}_{n,t} - \mathrm{SOE}_n^{\mathrm{reg}}}. \tag{17}$$

According to the above formula and the ESS charging/discharging state of the PV-storage charging station, the action $\pi_{n,t}^{\mathrm{sell}}$ is expressed as follows:

$$\pi_{n,t}^{\mathrm{sell}} = \begin{cases} \lambda_{n,t}\, \tilde{\pi}_{n,t}^{\mathrm{sell}}, & \text{if ESS discharges,} \\ 0, & \text{if ESS charges} \end{cases} \tag{18}$$

where $\mathrm{SOE}_n^{\mathrm{reg}}$ and ϵ_n affect the selling price and cause its variation in (18). $\mathrm{SOE}_n^{\mathrm{reg}}$, the SOE regulation parameter of ESS in station n, is used to prevent unexpected future energy demand. $\mathrm{SOE}_n^{\mathrm{reg}}$ is represented as $\mathrm{SOE}_n^{\mathrm{reg}} = \max\{\mathrm{SOE}_n^{\mathrm{reg}}, \underline{\mathrm{SOE}}_n\}$. A larger $\mathrm{SOE}_n^{\mathrm{reg}}$ better addresses unexpected future energy demand, at the cost of limiting current ESS discharge capability and thereby increasing the selling price at the current time slot. The parameter ϵ_n denotes the sensitivity of the selling price to the ratio between total energy consumption and SOE in the PV-storage charging station. Similar to the effect of $\pi_{n,t}^{\mathrm{sell}}$ on $\mathrm{SOE}_n^{\mathrm{reg}}$, a larger

ϵ_n increases $\pi_{n,t}^{\text{sell}}$. These two parameters can be adaptively tuned by the station operator in different environments, e.g., by changing operating conditions and energy consumption of the PV-storage charging station.

5.2.3. Reward Function

The total reward of LA_n at time slot t, denoted by $R_{n,t}^{\text{total}}$, is composed of three reward terms: $R_{n,t}^{\text{total}} = R_{n,t}^{(1)} + R_{n,t}^{(2)} + R_{n,t}^{(3)}$. The three terms are: (1) $R_{n,t}^{(1)}$: the sum of profit from selling electricity to EV users and the cost of purchasing electricity from the grid; (2) $R_{n,t}^{(2)}$: the weighted sum of PV-storage charging station operating cost and EV user discomfort cost; and (3) $R_{n,t}^{(3)}$: the penalty for ESS over-charging and under-charging in the PV-storage charging station.

The specific definitions are as follows:

$$R_{n,t}^{(1)} = \pi_{n,t}^{\text{sell}} E_{n,t}^{\text{dch}} - \pi_t^{\text{TOU}} \left(E_{n,t}^{\text{ch}} + \hat{E}_{n,t}^{\text{Fixed}} \right), \tag{19}$$

$$R_{n,t}^{(2)} = -\left(\delta_1 C_{n,t}^{\text{EVCS}} + \delta_2 C_{n,t}^{\text{EV}} \right), \tag{20}$$

$$R_{n,t}^{(3)} = \begin{cases} -\mu_1 \left(\underline{\text{SOE}_n} - \text{SOE}_{n,t} \right), & \text{if } \text{SOE}_{n,t} < \underline{\text{SOE}_n}, \\ -\mu_2 \left(\text{SOE}_{n,t} - \overline{\text{SOE}_n} \right), & \text{if } \text{SOE}_{n,t} > \overline{\text{SOE}_n}, \\ 0, & \text{otherwise.} \end{cases} \tag{21}$$

In (19), the reward represents the profit of station n at time slot t, i.e., the difference between the ESS discharging revenue $\left(\pi_{n,t}^{\text{sell}} E_{n,t}^{\text{dch}} \right)$ and the cost associated with purchasing electricity from the grid and fixed station consumption $\left(\pi_t^{\text{TOU}} (E_{n,t}^{\text{ch}} + \hat{E}_{n,t}^{\text{Fixed}}) \right)$. Equation (20) is the weighted sum of negative cost terms, involving the PV-storage charging station operating cost $C_{n,t}^{\text{EVCS}}$ and EV user discomfort cost $C_{n,t}^{\text{EV}}$, where $\delta_1 + \delta_2 = 1$ and $\delta_1, \delta_2 > 0$. In (21), $R_{n,t}^{(3)}$ denotes the penalty cost for ESS over-charging and under-charging, represented by penalty factors μ_1 and μ_2, respectively.

In (20), $C_{n,t}^{\text{EVCS}}$ and $C_{n,t}^{\text{EV}}$ are calculated as follows:

$$C_{n,t}^{\text{EVCS}} = \beta_n \left(E_{n,t}^{\text{ch}} + E_{n,t}^{\text{dch}} \right), \tag{22}$$

$$C_{n,t}^{\text{EV}} = \begin{cases} \gamma_{n,1} \, \pi_{n,t}^{\text{sell}} E_{n,t}^{\text{EV,tr}}, & \text{if } E_{n,t}^{\text{EV,total}} > E_{n,t}^{\text{dch}}, \\ \gamma_{n,2} \, \pi_{n,t}^{\text{sell}}, & \text{if } E_{n,t}^{\text{EV,total}} = E_{n,t}^{\text{dch}}. \end{cases} \tag{23}$$

where $C_{n,t}^{\text{EVCS}}$ denotes the operating cost of ESS charging/discharging in the PV-storage charging station, and β_n is the operating-cost coefficient. $C_{n,t}^{\text{EV}}$ denotes the EV user discomfort cost of station n at time slot t, which is represented by the selling price $\pi_{n,t}^{\text{sell}}$ and unmet EV energy demand $E_{n,t}^{\text{EV,tr}}$ ($E_{n,t}^{\text{EV,tr}} = E_{n,t}^{\text{EV,total}} - E_{n,t}^{\text{dch}}$). $\gamma_{n,1}$ and $\gamma_{n,2}$ are penalty factors. The EV user discomfort cost $C_{n,t}^{\text{EV}}$ is directly related to the selling price and unmet energy demand.

5.3. Algorithm Process

This paper presents a privacy-preserving algorithm for scheduling ESS charge/discharge and selling prices across multiple PV-storage charging stations. To safeguard station data privacy, the algorithm integrates an FRL framework with the SAC method. All LA and the GA iteratively train their models through interaction, using only local and global models of LA to find the optimal policy—namely, the optimal ESS charge/discharge quantities and selling prices. All LA synchronously receive from and send to the GA their global and local models. The algorithm consists of three tasks: (1) initialization of all parameters to

update the local and global models of every LA and the GA; (2) model training to update each LA local model (i.e., the weights of its actor, value, and critic networks); and (3) the GA employs the FedSGD algorithm to update the global model for all LA. The detailed training procedure of the algorithm is as follows:

1. Initialize the weights of the actor network, critic network, value network, and target value network, along with the initialization of the experience replay pool β.
2. Initialize each LA's action probabilities, Q-values, state-value function, and target state-value function.
3. Initialize the ESS operation (charging or discharging), selling prices, and parameters associated with training LA.
4. Initialize the global neural network model θ_G and global batch data x for all LA (including the local models of all LA).
5. In each communication round and local training episode, every LA repeats the following training process to search for the optimal ESS charge/discharge strategy and selling price from time slot $t = 1$ to $t = 48$:
    (a) Receive the output of the actor network—that is, the actions of all LA at time slot t given the state and computed action probabilities.
    (b) Compute the ESS charge/discharge action $E_{n,t}^{ch/ch}$ and the selling-price action $\pi_{n,t}^{sell}$, obtain the reward $R_{n,t+1}$, and derive the next-time state $S_{n,t+1}$ from the state-transition function.
    (c) Store the sample data $[s_{n,t}, a_{n,t}, R_{n,t+1}, s_{n,t+1}]$ in the experience replay pool for subsequent sampling during training.
    (d) If it is an update round, each LA randomly samples $[s, a, r, s']$ from the experience replay pool β.
    (e) Each LA uses samples from the experience replay pool to estimate action probabilities, state values, and action-state Q-values.
    (f) Calculate the loss function of the value network and minimize it to update the network weights.
    (g) Estimate the target Q-value using the reward and the state value obtained from the target Value network.
    (h) Compute the loss functions of the actor network and critic network and minimize them to update the weights of these two networks.
    (i) In each training round, update the weights of the target value network using a smoothing constant δ with a soft update approach.
6. Each LA transmits the trained SAC-based neural network model to the GA, and the GA stores them in the global batch data x.
7. The GA calculates the global model for all LAs using all local models in x through the FedSGD method.
8. The GA broadcasts the computed global model to all LAs, and the LAs replace their local models with the newly received global model.

5.4. Structural Framework of LA Based on SAC

Figure 3 illustrates the structure of the SAC algorithm, where the LA is responsible for dynamic pricing and energy scheduling of the PV-storage charging station.

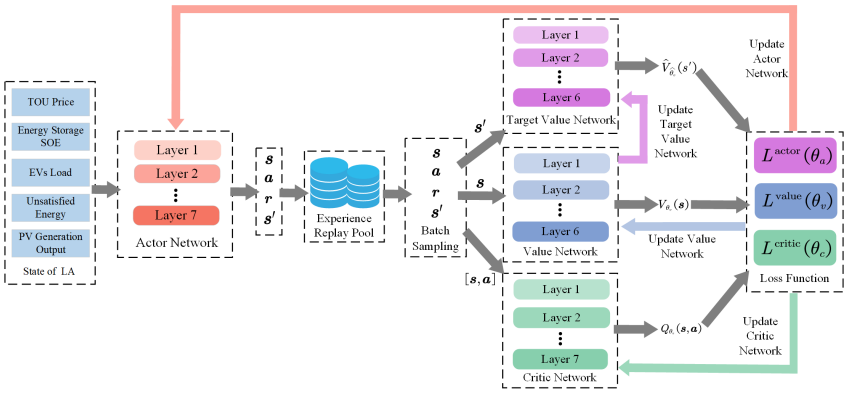
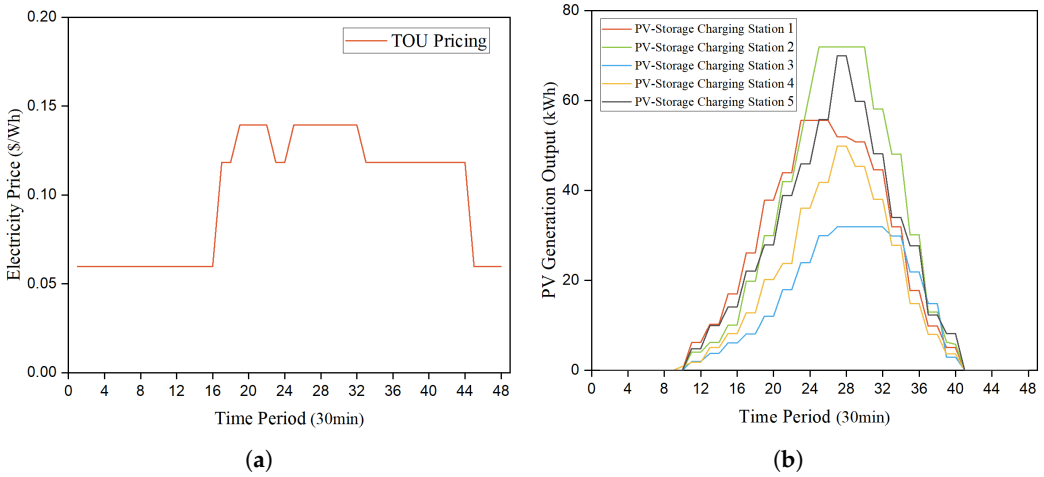


Figure 3. Structure of LA based on SAC.

6. Experiment

6.1. Experimental Setup

This chapter explores dynamic pricing and energy scheduling for five PV-storage charging stations under the TOU electricity prices shown in Figure 4a. The forecasted PV generation ($\widehat{E}_{n,t}^{PV}$) for each PV-storage charging station in each time period is depicted in Figure 4b, while the predicted total energy demand ($\widehat{E}_{n,t}^{EV}$) for EVs and fixed energy consumption ($\widehat{E}_{n,t}^{Fixed}$) are shown in Figure 5. The ESS capacity for each PV-storage charging station is set at 650 kWh, with the maximum charging amount $\overline{E}_n^{ch/dch}$ and minimum discharging amount $\underline{E}_n^{ch/dch}$ set to 350 kWh and 0 kWh, respectively. The parameters for adjusting the State of Energy (SOE), as well as the maximum and minimum SOE, along with the initial SOE, are set to SOE_n^{reg} = 10%, $\overline{SOE}_n$ = 100%, $\underline{SOE}_n$ = 10%, and SOE_n^{ini} = 50%, respectively. The maximum selling prices $\overline{\pi}_{n,t}^{sell}$ and minimum selling prices $\underline{\pi}_{n,t}^{sell}$ for PV-storage charging station n during time period t are set to 300% and 120% of the TOU price π_t^{TOU}, respectively, with the maximum and minimum sales factors $\lambda_{n,t}$ set to 1.5 and 1, respectively. For the five PV-storage charging stations, the weights in the reward function are uniformly set to δ_1 = 0.1, δ_2 = 0.9, δ_n = 0.002, $\gamma_{n,1}$ = 1, $\gamma_{n,2}$ = 50, and $\mu_1 = \mu_2 = 1.5$. The sensitivity parameter ϵ_n for the sales factor of the five PV-storage charging stations is set to 1.5.



(**a**) (**b**)

Figure 4. Time-of-use electricity price and photovoltaic forecast power generation: (**a**) Time-of-use (TOU) electricity price profile; (**b**) Forecasted PV generation of five charging stations.

Table 2 provides the neural network structure of the SAC-based LA, using the ADAM optimization method to compute the optimal weights of the actor network, critic network, and value network. The capacity of the experience replay pool is set to 50,000, and the

batch size for sampling during training is 48. The soft update parameter δ for updating the target Value network is set to 0.2, and the temperature parameter ζ for the entropy term is set to one. In the FRL framework, all five LAs participate in each communication round, and it is set that after every 500 training episodes, each LA communicates with the GA once within 3500 training episodes. The simulation is conducted using PyTorch 1.1.0 in Python 3.7.0.

To evaluate the performance of the algorithm proposed in this paper, the following three comparison algorithms were selected for analysis:

1. SAC algorithm without a sales factor: This algorithm uses the SAC method for dynamic pricing and energy scheduling of the PV-storage charging station, just like the one proposed in this paper, but without considering the impact of the sales factor.
2. Actor–critic (AC) algorithm: This classic reinforcement learning AC algorithm is used for decision-making. It employs a separate actor to select actions and a critic to estimate the value function, in order to optimize the profit of the PV-storage charging station.
3. Advantage actor–critic (A2C) algorithm [28]: This algorithm introduces the "advantage function" based on the AC algorithm, improving the stability and efficiency of the algorithm's training by reducing variance.

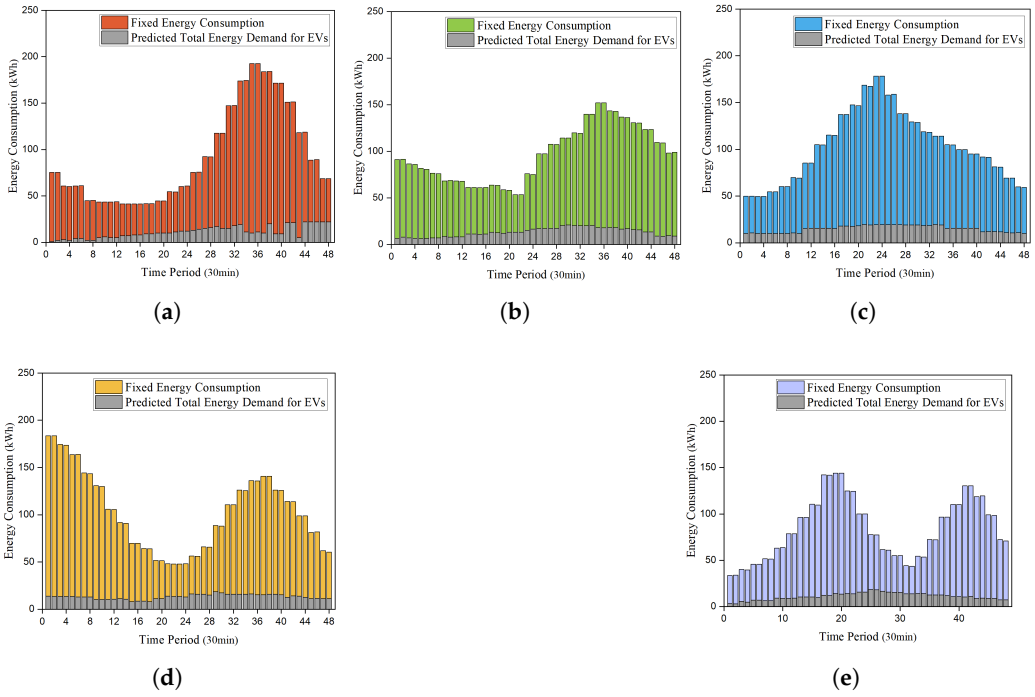


Figure 5. Total energy demand and fixed energy consumption of EVs predicted by five PV-storage charging stations: (**a**–**e**) show the forecasted EV demand and fixed load for Stations 1–5, respectively.

Table 2. Neural network structure of LA based on SAC.

Neural Network	Hidden Layer	Number of Neurons (1–7 Layers)							Activation Function	Optimizer	Learning Rate
Actor Network	7	512	512	256	256	128	128	128	Hyperbolic Tangen	ADAM	1×10^{-6}
Critic Network	7	512	512	256	256	128	128	128	Leaky-Relu	ADAM	5×10^{-6}
Value Network	6	512	256	256	256	128	128	-	Leaky-Relu	ADAM	5×10^{-6}
Target Value network	6	512	256	256	256	128	128	-	Leaky-Relu	-	-

6.2. Experimental Results and Analysis

This section demonstrates the effectiveness of the proposed method through performance comparison results with the comparison algorithms. Figure 6a–e shows the selling prices of each PV-storage charging station under different values of the SOE regulation parameter $\mathrm{SOE}_n^{\mathrm{reg}}$, i.e., 10%, 20%, and 30%. As $\mathrm{SOE}_n^{\mathrm{reg}}$ increases, each PV-storage charging station tends to store more energy to better cope with possible future energy-supply shortages, which reduces ESS discharging energy and ultimately raises the selling price. This trend can be observed in Figure 6, and all PV-storage charging stations have higher selling prices under larger $\mathrm{SOE}_n^{\mathrm{reg}}$.

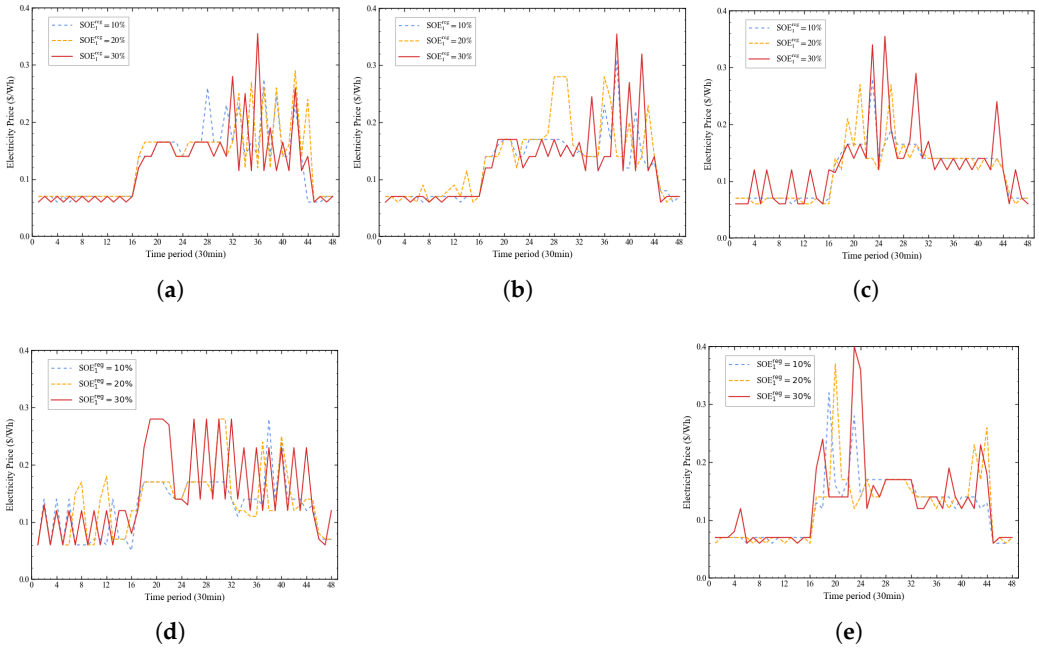


Figure 6. Electricity sales prices of five photovoltaic-storage charging stations under different SOE_n^{reg} conditions: (**a**–**e**) present the selling price at Stations 1–5 under different SOE_n^{reg} settings, respectively.

Figure 7 compares the average SOE of ESSs in five PV-storage charging stations over one day under different $\mathrm{SOE}_n^{\mathrm{reg}}$ settings. From the figure, it can be observed that increasing $\mathrm{SOE}_n^{\mathrm{reg}}$ increases the ESS SOE of PV-storage charging stations. In general, each PV-storage charging station n can adaptively tune ϵ_n and $\mathrm{SOE}_n^{\mathrm{reg}}$ in the selling factor, so as to achieve higher profit from the operator's perspective (increase ϵ_n and $\mathrm{SOE}_n^{\mathrm{reg}}$) or provide cheaper and greater energy supply from the EV users' perspective (decrease ϵ_n and $\mathrm{SOE}_n^{\mathrm{reg}}$).

This subsection also proposes a performance metric to quantify the similarity between the relative changes in station selling price and total EV energy demand across all N PV-storage charging stations and $T-1$ time slots, denoted by DOS:

$$\mathrm{DOS} = \frac{\sum_{n=1}^{N}\sum_{t=2}^{T}\left|\Delta\pi_{n,t}^{\mathrm{sell}} - \Delta E_{n,t}^{\mathrm{EV,total}}\right|}{N(T-1)} \times 100(\%), \tag{24}$$

$$\Delta\pi_{n,t}^{\mathrm{sell}} = \frac{\pi_{n,t}^{\mathrm{sell}} - \pi_{n,t-1}^{\mathrm{sell}}}{\pi_{n,t-1}^{\mathrm{sell}}}, \qquad \Delta E_{n,t}^{\mathrm{EV,total}} = \frac{E_{n,t}^{\mathrm{EV,total}} - E_{n,t-1}^{\mathrm{EV,total}}}{E_{n,t-1}^{\mathrm{EV,total}}} \tag{25}$$

where $\Delta\pi_{n,t}^{\mathrm{sell}}$ and $\Delta E_{n,t}^{\mathrm{EV,total}}$ are the relative deviations of the selling price and total EV energy demand at time slot t from those at time slot $t-1$, respectively. A smaller DOS value indicates higher similarity between the variations in selling price and total EV energy

demand, meaning that the selling price computed in this subsection can more effectively reflect total EV energy demand. Experimental results show that the proposed dynamic pricing method reduces DOS by 8.7% compared with the fixed TOU pricing method.

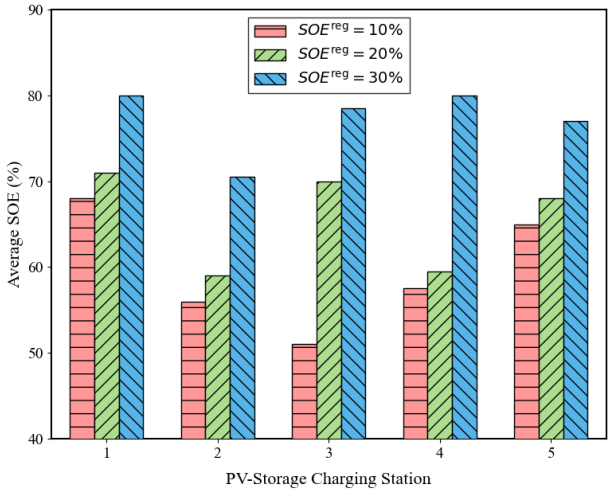


Figure 7. Average SOE of ESS for five photovoltaic-storage-charging stations under different SOE_n^{reg} conditions.

Figure 8 shows the increase in profit for the five PV-storage charging stations of this chapter's algorithm compared to the contrast algorithms, calculated as follows:

$$\frac{P_n - P_n^b}{P_n^b} \times 100(\%) \tag{26}$$

where P_n is the profit obtained by charging station n using the algorithm presented in this paper, and P_n^b represents the profits of charging station n using three comparison methods, including: (1) SAC algorithm without a sales factor and (2) Two existing DRL algorithms (actor–critic and advantage actor–critic (A2C) [28]). From Figure 8, it can be seen that the method proposed in this chapter yields higher profits for all PV-storage charging stations compared to the three comparison algorithms, with the relative profit increase for each station ranging from 3.90% to 39.00%. The method proposed in this chapter achieves more considerable profits than the SAC algorithm without a sales factor, mainly because the increase in the sales factor leads to higher selling prices when the energy demand at the PV-storage charging station is high. Compared with the actor–critic and A2C algorithms, the algorithm using SAC also achieves greater profits, as the entropy term of the soft Q-value in the SAC algorithm encourages the LA to explore the environment more effectively.

Figure 9 compares the performance of the algorithm in this chapter with the actor–critic algorithm, A2C algorithm, and SAC algorithm without using FRL in terms of reward training curves. It can be observed from the figure that the training curve of the SAC algorithm with FRL fluctuates significantly during the first 1000 rounds of iterative training, but then quickly converges to the optimal policy. In contrast, the training curves of the actor–critic and A2C algorithms with FRL fluctuate more throughout the training period and converge much slower than the SAC algorithm. Additionally, Figure 9 also shows that the training curve of the SAC algorithm without FRL performs worse than the SAC algorithm with FRL between the 1000th and 2000th rounds of iterative training. This result verifies that the FRL framework avoids unnecessary training processes and prevents training divergence by providing a global model for all LAs in each communication round. After 2500 iterations, both training curves of the SAC algorithm (with and without the FRL framework) converge to the optimal policy. From Figures 8 and 9, it can be concluded that the algorithm

in this chapter has better performance than the comparison algorithms and converges quickly without sharing private data. For operators and aggregators of PV-storage charging stations, reducing unnecessary training processes and adjusting selling prices and ESS charge/discharge quantities in a timely manner are crucial. Therefore, the algorithm in this chapter is suitable for managing the operation of multiple PV-storage charging stations with consideration of privacy issues and efficient training.

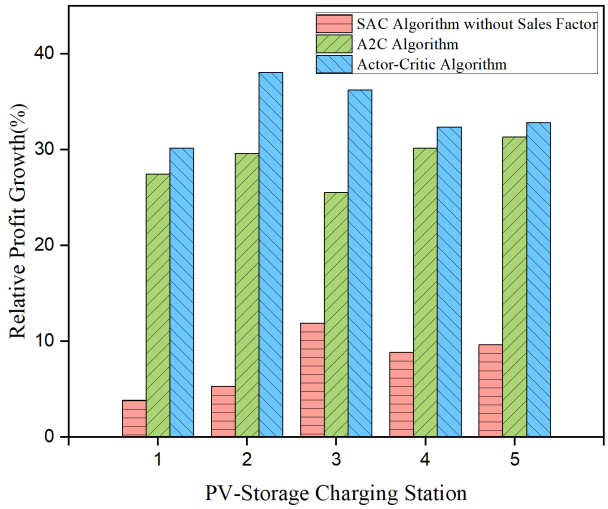


Figure 8. Comparison of the relative profit growth of PV-storage charging station under different algorithms.

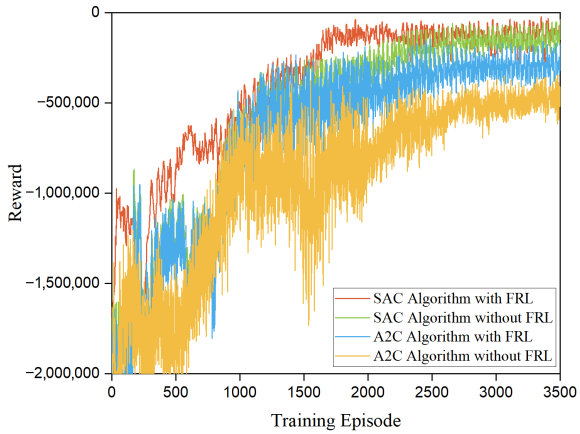


Figure 9. Convergence curve of training reward for different algorithms.

6.3. Robustness and Practical Overhead Analysis

To keep the empirical claims transparent, we report deterministic comparisons under a fixed experimental configuration and avoid unsupported over-claims on statistical significance. Under the reported setup, the proposed FRL-SAC framework consistently yields better profitability and convergence behavior than the compared baselines. Specifically, the reported profit gains are 3.90% over SAC (without sales factor), 39.00% over AC, and 30.10% over A2C. These improvements are consistent with the convergence trends shown in the reward curves, where the proposed policy reaches a stable high-reward regime earlier than AC/A2C-based alternatives.

At the same time, we explicitly acknowledge that the current manuscript does not yet include multi-seed statistics (e.g., mean ± standard deviation) or hypothesis-testing results. Therefore, conclusions in this version are stated as deterministic gains under the

reported setting. A full statistical extension with repeated runs, confidence intervals, and significance tests will be included in subsequent experiments.

Beyond economic performance, practical deployment also depends on communication and computation costs. In the proposed two-level architecture, each local agent performs on-site SAC updates and periodically exchanges model parameters with the global coordinator. This design avoids raw-data transfer and reduces direct data-exposure risk, while introducing synchronization latency and communication overhead. In general, shorter aggregation intervals improve cross-station knowledge sharing but increase network burden, and longer intervals reduce communication load but may slow global coordination. Our current results indicate that periodic aggregation improves learning stability compared with fully isolated local training, as summarized in Table 3, the proposed FRL-SAC with sales factor achieves the best deterministic performance. A detailed system-level benchmark (wall-clock training time, per-round payload, and latency under heterogeneous links) is left for future work.

Table 3. Deterministic performance comparison under the reported setup.

Method	Profit Gap vs. Proposed	Convergence Trend (Reported Curves)	Key Difference
Proposed FRL-SAC + Sales Factor	–	Faster approach to a stable high-reward regime	Federated SAC + adaptive sales-factor correction
SAC (without sales factor)	−3.90%	Stable but weaker in middle/late training stages	No adaptive sales-factor mechanism
AC	−39.00%	Slower and less stable convergence	Simpler actor–critic update dynamics
A2C	−30.10%	Better than AC, below SAC variants	Advantage estimator improves variance but remains suboptimal

7. Conclusions

This work presents a federated DRL approach for coordinated pricing and ESS scheduling in multi-station PV-EV charging systems. Each station trains a local SAC policy, while a global server periodically aggregates model parameters to enhance cross-station learning. The proposed sales-factor mechanism enables demand-aware price adjustment and helps balance profitability with service considerations. Experimental results show consistent profit improvements and stable convergence compared with representative DRL baselines.

To enhance the operational efficiency of PV-storage charging stations and mitigate their impact on the distribution network, future research should focus on coordinated optimization among the distribution grid, PV-storage systems, and EVs. Building upon the proposed federated deep reinforcement learning (FRL-DRL) framework, new algorithms can be developed to support more complex interactions while ensuring system reliability and data privacy. For example, integrating graph neural networks can capture spatial and topological relationships among charging stations to improve learning efficiency. Additionally, incorporating environmental constraints, such as carbon emissions and grid safety margins, can support greener and safer scheduling decisions. Lightweight distributed reinforcement learning architectures and user-side behavior modeling with incentive mechanisms are also promising directions for enabling real-time, user-station-grid collaborative energy scheduling. These advancements will drive the development of more intelligent, sustainable, and resilient PV-storage charging networks.

Author Contributions: Conceptualization, Y.Z. (Yongguang Zhao) and Y.Z. (Yongqing Zheng); Methodology, Y.Z. (Yongguang Zhao), X.L., Y.Z. (Yongqing Zheng) and W.G.; Validation, Y.Z. (Yongguang Zhao) and X.L.; Investigation, X.L.; Data curation, X.L. and W.G.; Writing—original draft, Y.Z. (Yongguang Zhao), X.L. and Y.Z. (Yongqing Zheng); Writing—review & editing, Y.Z. (Yongguang Zhao), Y.Z. (Yongqing Zheng) and W.G.; Supervision, Y.Z. (Yongguang Zhao). All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fraiji, Y.; Azzouz, L.B.; Trojet, W.; Saïdane, L.A. Cyber security issues of Internet of electric vehicles. In Proceedings of the 2018 IEEE Wireless Communications and Networking Conference (WCNC 2018), Barcelona, Spain, 15–18 April 2018; pp. 1–6. [CrossRef]
2. Luo, C.; Huang, Y.; Gupta, V. Stochastic Dynamic Pricing for EV Charging Stations with Renewable Integration and Energy Storage. *IEEE Trans. Smart Grid* **2018**, *9*, 1494–1505. [CrossRef]
3. Moghaddam, Z.; Ahmad, I.; Habibi, D.; Masoum, M.A. A coordinated dynamic pricing model for electric vehicle charging stations. *IEEE Trans. Transp. Electrif.* **2019**, *5*, 226–238. [CrossRef]
4. Rasheed, M.B.; Awais, M.; Alquthami, T.; Khan, I. An Optimal Scheduling and Distributed Pricing Mechanism for Multi-Region Electric Vehicle Charging in Smart Grid. *IEEE Access* **2020**, *8*, 40298–40312. [CrossRef]
5. Wang, S.; Bi, S.; Zhang, Y.J.; Huang, J. Electrical Vehicle Charging Station Profit Maximization: Admission, Pricing, and Online Scheduling. *arXiv* **2017**, arXiv:1705.02116. [CrossRef]
6. Zhang, Y.; You, P.; Cai, L. Optimal Charging Scheduling by Pricing for EV Charging Station with Dual Charging Modes. *IEEE Trans. Intell. Transp. Syst.* **2019**, *20*, 3386–3396. [CrossRef]
7. Fang, C.; Lu, H.; Hong, Y.; Liu, S.; Chang, J. Dynamic Pricing for Electric Vehicle Extreme Fast Charging. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 531–541. [CrossRef]
8. Moghaddam, V.; Yazdani, A.M.; Wang, H.; Parlevliet, D.; Shahnia, F. An Online Reinforcement Learning Approach for Dynamic Pricing of Electric Vehicle Charging Stations. *IEEE Access* **2020**, *8*, 130305–130313. [CrossRef]
9. Wang, S.; Bi, S.; Zhang, Y.A. Reinforcement Learning for Real-Time Pricing and Scheduling Control in EV Charging Stations. *IEEE Trans. Ind. Inform.* **2021**, *17*, 849–859. [CrossRef]
10. Shin, M.; Choi, D.; Kim, J. Cooperative Management for PV/ESS-Enabled Electric Vehicle Charging Stations: A Multiagent Deep Reinforcement Learning Approach. *IEEE Trans. Ind. Inform.* **2020**, *16*, 3493–3503. [CrossRef]
11. Zhang, Y.; Wang, Y.; Li, F.; Wu, B.; Chiang, Y.; Zhang, X. Efficient Deployment of Electric Vehicle Charging Infrastructure: Simultaneous Optimization of Charging Station Placement and Charging Pile Assignment. *IEEE Trans. Intell. Transp. Syst.* **2021**, *22*, 6654–6659. [CrossRef]
12. Lu, L.; Fei, H. Location Model of Electric Vehicle Charging Stations Considering Queuing Time and Charging Cost. *J. Transp. Inf. Saf.* **2023**, *41*, 154–162.
13. Li, D.; Zouma, A.; Liao, J.T.; Yang, H.T. An energy management strategy with renewable energy and energy storage system for a large electric vehicle charging station. *Etransportation* **2020**, *6*, 100076. [CrossRef]
14. Acharya, S.; Dvorkin, Y.; Pandzic, H.; Karri, R. Cybersecurity of Smart Electric Vehicle Charging: A Power Grid Perspective. *IEEE Access* **2020**, *8*, 214434–214453. [CrossRef]
15. Graber, G.; Calderaro, V.; Mancarella, P.; Galdi, V. Two-stage stochastic sizing and packetized energy scheduling of BEV charging stations with quality of service constraints. *Appl. Energy* **2020**, *260*, 114262. [CrossRef]
16. Liu, Y.; Zhang, S.; Guo, L.; Li, X.; Jia, X.; Wang, T.; Zhao, Z. Coordinated Optimal Scheduling Method for Distribution Network and PV-Storage-Charging Stations Considering Energy-Reserve Coupling. *Power Syst. Technol.* **2024**, *48*, 3175–3185. [CrossRef]
17. Deng, Y.; Li, J.; Lu, G.; Wang, H. Electric Vehicle Charging Optimization Strategy Guided by Zonal Dynamic Electricity Pricing Mechanism. *Power Syst. Prot. Control* **2024**, *52*, 33–44. [CrossRef]
18. Cui, Y.; Hu, Z.; Duan, X. Optimal Pricing of Public Electric Vehicle Charging Stations Considering Operations of Coupled Transportation and Power Systems. *IEEE Trans. Smart Grid* **2021**, *12*, 3278–3288. [CrossRef]
19. Zeng, B.; Dong, H.; Sioshansi, R.; Xu, F.; Zeng, M. Bilevel robust optimization of electric vehicle charging stations with distributed energy resources. *IEEE Trans. Ind. Appl.* **2020**, *56*, 5836–5847. [CrossRef]

20. Yan, Q.; Zhang, B.; Kezunovic, M. Optimized Operational Cost Reduction for an EV Charging Station Integrated with Battery Energy Storage and PV Generation. *IEEE Trans. Smart Grid* **2019**, *10*, 2096–2106. [CrossRef]
21. Dabbaghjamanesh, M.; Moeini, A.; Kavousi-Fard, A. Reinforcement Learning-Based Load Forecasting of Electric Vehicle Charging Station Using Q-Learning Technique. *IEEE Trans. Ind. Inform.* **2021**, *17*, 4229–4237. [CrossRef]
22. Messaoud, S.; Bradai, A.; Ahmed, O.B.; Quang, P.T.A.; Atri, M.; Hossain, M.S. Deep Federated Q-Learning-Based Network Slicing for Industrial IoT. *IEEE Trans. Ind. Inform.* **2021**, *17*, 5572–5582. [CrossRef]
23. Liu, Y.; Feng, G.; Sun, Y.; Qin, S.; Liang, Y. Device Association for RAN Slicing Based on Hybrid Federated Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 15731–15745. [CrossRef]
24. Wang, X.; Wang, C.; Li, X.; Leung, V.C.M.; Taleb, T. Federated Deep Reinforcement Learning for Internet of Things with Decentralized Cooperative Edge Caching. *IEEE Internet Things J.* **2020**, *7*, 9441–9455. [CrossRef]
25. Wang, X.; Li, R.; Wang, C.; Li, X.; Taleb, T.; Leung, V.C.M. Attention-Weighted Federated Deep Reinforcement Learning for Device-to-Device Assisted Heterogeneous Collaborative Edge Caching. *IEEE J. Sel. Areas Commun.* **2021**, *39*, 154–169. [CrossRef]
26. Li, X.; Huang, K.; Yang, W.; Wang, S.; Zhang, Z. On the convergence of fedavg on non-iid data. *arXiv* **2019**, arXiv:1907.02189.
27. Liu, N.; Yu, X.; Wang, C.; Li, C.; Ma, L.; Lei, J. Energy-sharing model with price-based demand response for microgrids of peer-to-peer prosumers. *IEEE Trans. Power Syst.* **2017**, *32*, 3569–3583. [CrossRef]
28. Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; Riedmiller, M.A. Deterministic Policy Gradient Algorithms. In Proceedings of the 31th International Conference on Machine Learning, (ICML 2014), Beijing, China, 21–26 June 2014; Volume 32, pp. 387–395.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.