
Article:
Zhou, J., Xiang, Y., Zhang, X. et al. (2025) Optimal self-consumption scheduling of highway electric vehicle charging station based on multi-agent deep reinforcement learning. Renewable Energy, 238. 121982. ISSN: 0960-1481

https://doi.org/10.1016/j.renene.2024.121982

Optimal Self-Consumption Scheduling of Highway Electric Vehicle Charging Station Based on Multi-Agent Deep Reinforcement Learning

Jianshu Zhou[a], Yue Xiang[a*], Xin Zhang[b], Zhou Sun[c], Xuefei Liu[d], Junyong Liu[a]

a. College of Electrical Engineering, Sichuan University, Chengdu, China.
b. Department of Automatic Control and Systems Engineering, University of Sheffield, Sheffield, United Kingdom
c. Sichuan Road and Bridge Construction Group Co., LTD, Chengdu, China.
d. State Grid Hebei Electric Power Company Economic and Technological Research Institute, Shijiazhuang, China.
*Corresponding author: xiang@scu.edu.cn

Abstract: Due to the randomness of renewable energy and electric vehicles (EVs) in highway charging stations, it is difficult to ensure the consistency of renewable energy supply and EVs demand. Considering the randomness of EVs charging and renewable energy power generation, an optimal self-consumption scheduling of a highway EV charging station based on multi-agent deep reinforcement learning (MADRL) is proposed to realize the economy, self-consumption, low-carbon operation and ensure reliability of power supply. In day-ahead, the traffic flow prediction model based on the CNN-BiLSTM and the queuing model based on user psychology are built to predict the charging load. The 24-hour optimal charging price is obtained by solving the incentive price optimization model and guides the orderly charging of EVs. In intra-day, considering the prediction errors of day-ahead and the diversity of charging levels, an optimal scheduling based on the MADRL is proposed. Regarding the multi-objective scheduling of the highway charging station, the multi-objective nonlinear and non-convex problem is transformed into multi-agent Markov game model. Finally, the effectiveness and optimality of the proposed method are verified on a highway charging station The results show that the proposed method can realize the economy, self-consumption and low-carbon operation of the charging station.

Keywords: highway EV charging station; day-ahead and intra-day optimization; traffic flow prediction; multi-agent deep reinforcement learning; self- consumption

1 Introduction

With the proliferation of electric vehicles (EVs), their high charging demands will have a profound impact on the operation of the distribution power networks and the electricity market [1-4]. At the same time, the development of renewable energy power generation policies and the automobile market will further promote the growth of charging demand [5-7].

Renewable energy charging stations use photovoltaic (PV), wind power (WP) and other power generation systems to charge EV. Renewable energy charging stations can be divided into urban and suburban charging stations (highway charging stations, etc.). At present, charging stations in urban areas are mainly charging stations equipped with PV generation and energy storage systems (ESS), and the PV generation is influenced by the weather and environment with large randomness [8-10]. Charging stations are configured with ESS to store and release energy, improve the consumption of PV generation. There have been some studies on the optimal scheduling of urban charging stations. Reference [11] proposed an energy management strategy for PV-ESS-charging stations based on the time of use prices. A comprehensive income optimization model of charging stations was established to make full use of the difference between peak and valley electricity prices as well as the electricity generated by PV generators. Moreover, reference [12] believed that the current deployment of renewable energy charging stations is not enough because of the lack of evaluation on economic and environmental benefits, and put forward an optimization strategy for renewable energy charging stations based on economic environment analysis. However,

due to the limitation of the surrounding environment, urban charging stations are not difficult to install large power generation systems, and can only be supported by power generation systems of small installed capacity such as PV, which is difficult to achieve self-consumption and low-carbon in the charging station.

Compared with urban charging stations, suburban charging stations (highway charging stations, etc.) are located in remote areas without a perfect power system, and the reliability of the power supply is low, which is similar to isolated islands. Taking highway charging stations as an example, highway charging stations usually have enough capacity and a safe distance. Therefore, there are conditions for installing wind turbines, solving the problems of single energy source, small installed capacity and poor power supply reliability of charging stations. There have been some studies on the highway charging stations. To further achieve carbon emission reduction, reference [13] proposed a new planning method for highway charging stations and a low-carbon facility planning framework for charging systems, including the construction of charging stations, decommissioning of gas stations, and installation of PV systems. Considering the environmental impact of highway slope, wind speed, and passenger number, reference [14] proposed an optimal location and scale planning method for highway EV charging stations based on the PV and ESS. Based on the Floyd and Monte Carlo algorithm, reference [15] proposed a collaborative planning method for highway charging stations and renewable energy systems. However, the current studies mainly focus on the site planning and capacity allocation of renewable energy charging stations on highways, and there are few studies on the scheduling of renewable energy charging stations on highways [16,17]. At the same time, the existing studies on highway renewable energy charging stations use model-driven (Monte Carlo simulation, etc.) methods to forecast load. However, it is difficult to simulate the load change characteristics with this prediction method. With more convenient access to historical data on highways and charging stations, the limitations of data-driven methods to predict the behavior or load of EVs on highways have been broken.

Due to the uncertainty of renewable energy, if the charging station equipped with WP and PV system does not have an effective scheduling method, it will cause derivative problems of low energy utilization efficiency. At the same time, compared with the regular travel of urban EV users, the randomness of EV stops on charging stations is greater, which will aggravate these derivative problems. In addition, in the previous study of highway scheduling [18,19], it is easy to ignore that parking is prohibited on the highway. Different from urban areas and other places, the highway has the particularity of banning parking, and the highway charging station can't be congested, and it is necessary to ensure the parking space of EVs with charging demand in the station. Therefore, a reasonable and effective scheduling method is needed to guide the orderly charging of EVs, achieve self-consumption of energy, and prevent the congestion of the charging station.

In terms of solving the optimal scheduling problem, the EV charging on highways is actually formulated as a nonlinear and non-convex problem, and it is difficult to fully cope with the complex problem by using traditional optimization algorithms and heuristic algorithms. In contrast, deep reinforcement learning (DRL) is an intelligent algorithm that does not need to establish an environment model and can interact with the environment to find the optimal control strategy that can achieve long-term rewards and enhance adaptability and robustness [20, 21]. At present, some studies are using the DRL to solve problems related to EVs' integration in the power grid. Reference [22] used the deep deterministic policy gradient (DDPG) to learn charging strategies and solve EV charging problems based on incentive and time-varying demand response. Reference [23] modeled the EV charging control model as a Markov decision process (MDP) from the user perspective. In the system model, considering the dynamic energy price and time-varying charging demand, the goal is to minimize charging costs and meet charging demand of EV users. In terms of the methodology, the DDPG algorithm was combined with transfer learning to control EV charging. However, a charging station is composed of charging piles with different charging power levels, and it is difficult for an agent to solve the scheduling problem of charging piles with different charging power levels. It is necessary to decompose the complex control problem into a multi-agent cooperation problem. It known that multi-agent deep reinforcement learning (MADRL) applies the ideas of the DRL to multi-agent systems [20]. The

MADRL can organize multi-agent to carry out autonomous learning and realize the cooperative solution to the scheduling problem of charging piles with different charging power levels through the interaction between multiple agents.

Because highway charging stations are generally located in areas with weak power construction, the reliability of the power supply is poor and the demand for energy self-consumption is greater. Due to the double uncertainty of renewable energy generation (WP and PV) and highway EVs charging, the difficulty of energy self-consumption has increased. At the same time, because the highway has the particularity of prohibiting parking, it is necessary to ensure the parking space of EVs with charging demand in the station. Therefore, aiming at the poor power supply reliability of highway charging stations and the low energy utilization efficiency caused by the randomness of source and load, an optimal day-ahead and intra-day self-consumption scheduling for a highway EV charging station based on the MADRL is proposed. Taking the highway charging station as the background, the PV, WP, and ESS are configured in the charging station. By using the CNN-BiLSTM network and queuing theory, the traffic flow and load prediction model is established. According to the prediction, the day-ahead optimal scheduling model guided by price incentives is established. In intra-day scheduling, the MADRL is used to control the power of the charging piles and ESS to achieve self-consumption, economic and low-carbon of the highway EV charging station.

The main contributions of the paper are as follows.

1) To accurately describe the charging behavior at a highway charging station, a multi-model traffic flow prediction method based on CNN-BiLSTM and a load modeling method based on user psychological $M/M/N_{sum}/C$ queuing theory were proposed. Aim at the prediction error superposition caused by the traditional single-step rolling prediction, the single-step rolling prediction is transformed into a multi-model traffic flow prediction with 24 CNN-BiLSTM. Then, the user response probability model of the Sigmoid function and $M/M/N_{sum}/C$ queuing theory are integrated to accurately model the charging station load.

2) To deal with the mismatch of the time dimension of source and load and take into account the global optimality and flexibility of scheduling, a self-consumption scheduling method based on day-ahead optimization guided by price incentive and intra-day optimization based on flexible power adjustment is proposed. Different from the traditional day-ahead charging plan, the proposed method formulates charging subsidies of the intra-day in the day-ahead stage, guides users to charge in an orderly manner, and indirectly transfers the load of the charging station to make the load consistent with the new energy power distribution. Considering the errors between the day-before price guidance and the intra-day actual situation, the charging piles and ESS are real-time scheduled to realize the re-transfer of the charging station load, providing a second guarantee for the consistency of the charging load and the power of renewable energy.

3) To improve the solving ability of the intra-day optimization model under the double uncertainty of EVs charging and renewable energy generation, the complex multi-objective nonlinear and non-convex scheduling is transformed into a multi-agent Markov game model (MGM). The multi-objective optimization is transformed into a multi-agent cooperative problem of multiple charging piles and ESS, to achieve dimensionality reduction of the optimization problem. The MATD3 algorithm, which is both efficient and stable, is used to solve the multi-agent MGM of multiple charging and ESS.

The rest of this paper is organized as follows. Section 2 presents the traffic flow and load prediction model of highway charging station. Section 3 presents the day-ahead and intra-day self-consumption scheduling. In section 4, the effectiveness and optimality of the proposed method are verified by test cases. Finally, conclusions and future work are provided in section 5.

2 Traffic flow and load prediction model of highway charging station

2.1 Traffic flow prediction of a charging station based on the CNN-BiLSTM

Traffic flow prediction is to solve the full parking of highway charging station, and is the premise of the

charging station load prediction. Unlike other charging stations, highway charging stations need to ensure that EV s with low battery power (which makes it difficult to complete the next journey) have parking spaces, and cannot park on the highway. Therefore, the traffic flow peak of the highway charging station is predicted and the peak transfer is carried out before the day.

The convolutional neural network (CNN) adopts the method of local connection and weight sharing, which processes the original data by high-dimensional mapping and effectively extracts data features. The bidirectional Long short-term memory (BiLSTM) network is an improved network based on the LSTM [24]. The BiLSTM network structure is shown in Fig. 1. The BiLSTM network enables recursive feedback of past and future hidden layer states, to further explore the internal relationship between past and future traffic flow data, and further improve model prediction accuracy and feature data utilization.

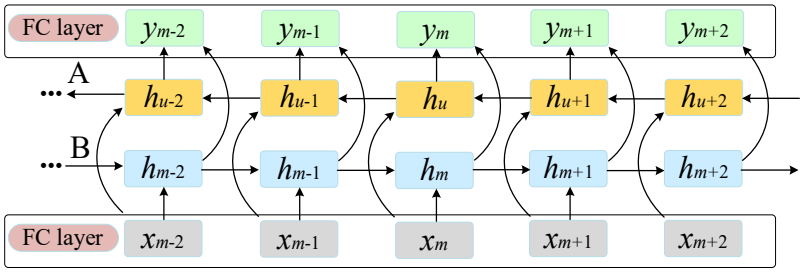


Fig. 1. The structure of the BiLSTM

Considering that the traffic flow data contains abundant temporal and spatial information, the paper adopts the CNN-BiLSTM network to forecast the traffic flow. First, the traffic flow data is normalized. Secondly, the day-ahead traffic flow on day m is predicted (24 points), but the prediction dimension is higher. Rolling and multi-step prediction will superimpose the error of prediction results, and the prediction accuracy is difficult to guarantee. Since the traffic flow is regular every day, the 0~23 points are divided into 24 groups, each group includes the data of the day m-1 to day m-d ($0^{[m-1,\ldots,m-d]}$, …, $23^{[m-1,\ldots,m-d]}$), and set up 24 CNN-BiLSTM networks. 24 sets of data were input into the corresponding CNN-BiLSTM network to train and output the traffic flow prediction results $Y_{tra}(t)$. Traffic flow prediction based on the CNN-BiLSTM is shown in Fig. 2. y^0 and x^0 are the output and input of the first CNN-BiLSTM network, y^1 and x^1 are the output and input of the second CNN-BiLSTM network, and so on. 0^m represents the 0 o 'clock on the m day, 0^{m-1} represents the 0 o 'clock on the m-1 day, and so on.

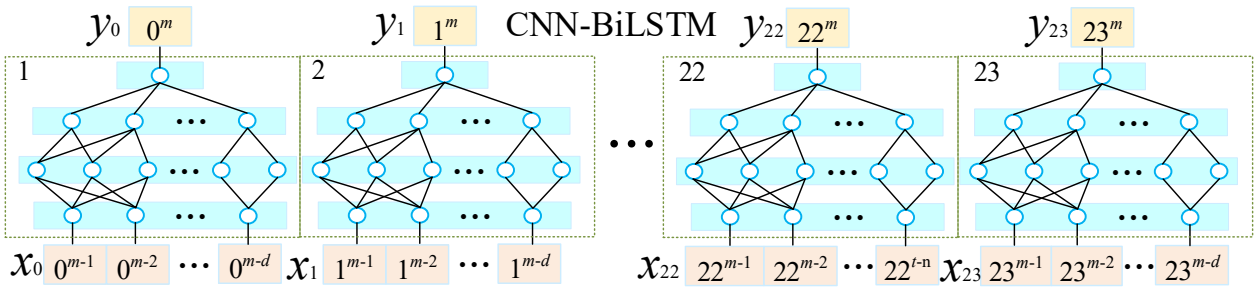


Fig.2. Traffic flow prediction based on the CNN-BiLSTM

2.2 Load prediction of a charging station based on the user psychological M/M/N_{sum}/C queuing theory

Charging station load forecasting is the premise of charging station scheduling. The scheduling system in a charging station adjusts the service charge price of the next day in advance according to the predicted load of the charging station to guide the charging of EVs. According to the set arrival ratio, the traffic flow data of each period

is converted into the arrival rate of EVs in the charging station, and then the load prediction model of the charging station is built through the queuing model.

The queuing model adopts the user psychological M/M/N_{sum}/C queuing theory. The arrival time interval of charging users follows the Poisson distribution with parameter λ, the number of charging piles is N_{sum}, and the charging time also follows the negative exponential distribution with parameter μ, and each charging pile is independent of each other. The total space capacity of the charging station (the sum of charging slots and parking Spaces) is C, and the charging rule of the queuing model of the charging station follows the first-come-first-served (FCFS) principle. The traditional M/M/N_{sum}/C queuing theory does not consider the user's psychology. Therefore, consider the following two scenarios:

1) Users may enter the queue because of the length of the queue and the charging fee at this time. The longer the queue, the smaller the probability of the user entering the queue, and conversely, the greater the probability. The higher the charging fee, the smaller the probability of the user entering the queue, and conversely, the greater the probability.

2) Users may choose to leave the queue because of the length of the queue and their emotional problems.

The M/M/N_{sum}/C queuing model of user psychology is shown in Fig. 3. The probability of the user joining the queue λ_k and the probability of leaving the queue μ_k:

$$\lambda_k = \begin{cases} \lambda P_{j,t}, & k < N_{\text{sum}} \\ \lambda P_{j,t} e^{-(k-N_{\text{sum}})\beta}, & N_{\text{sum}} \le k < C \\ 0, & k = C \end{cases} \tag{1}$$

$$\mu_k = \begin{cases} k\mu, & k \le N_{\text{sum}} \\ N_{\text{sum}}\mu + \beta_k, & N_{\text{sum}} < k \le C \end{cases} \tag{2}$$

$$\beta_k = \delta \ln(k - N_{\text{sum}} + 1),\ \delta \ge 0 \tag{3}$$

$$P_{j,t} = \mathrm{F}(z_{j,t}) = \frac{1}{1 + e^{-z_{j,t}}} \tag{4}$$

$$z_{j,t} = x_{j,t} + b_j \tag{5}$$

$$\lambda = \frac{N_{\text{avg}}}{100} Y_{tra}(t) \tag{6}$$

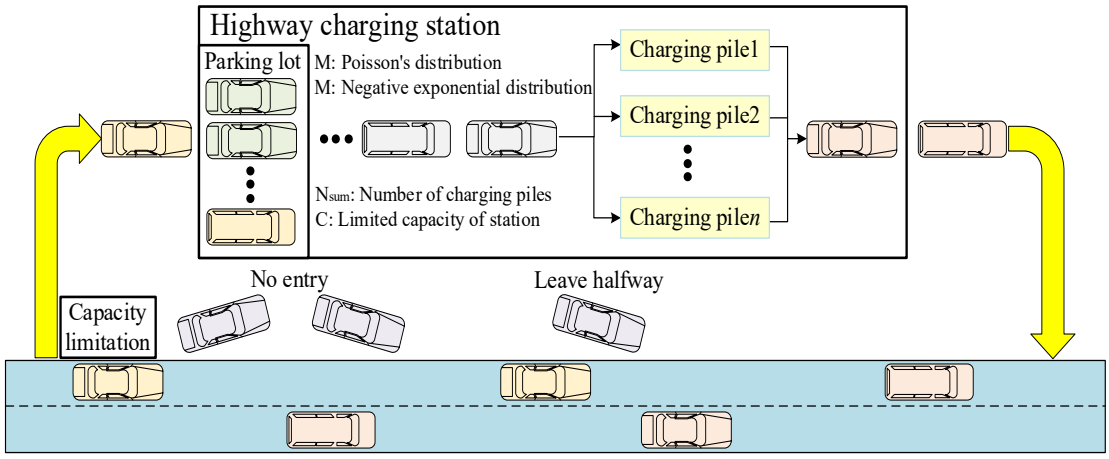


Fig.3. Queuing model of highway charging station

Where k is the length of the queue (k=0, 1, 2, …, N). β is the sensitivity parameter, which represents the decrease rate of probability ($\beta \geq 0$). As the queue grows longer, the probability of the user choosing to queue decreases rapidly, so the exponential function of base e is chosen in the setting of λ_k. β_k is the probability that a user in the queue will leave midway because they feel impatient. $P_{j,t}$ is the response probability of EV users to price, and the Sigmoid function is used as its response probability model. Considering that different users have different sensitivity to price, $z_{j,t}$ in the Sigmoid function is composed of stimulus price and price sensitivity. $x_{j,t}$ is the subsidized price, b_j is price sensitivity, and N_{avg} is the average number of a charging station per 100 vehicles.

The average number of charging piles in use is [25,26]:

$$N_{ch} = \sum_{k=0}^{N_{\text{sum}}} k \frac{\left(\lambda P_{j,t}\right)^k}{k!\mu^k} p_0 + N_{\text{sum}} \sum_{k=N_{\text{sum}}+1}^{C} \frac{e^{\frac{-(k-N_{\text{sum}})(k-N_{\text{sum}}-1)}{2}\beta}}{N_{\text{sum}}! \prod_{i=1}^{k-N_{\text{sum}}} [N_{\text{sum}} + \frac{\delta \ln(i+1)}{\mu}]} p_0 \tag{7}$$

In order to simplify load calculation, using the average power of all charging piles to calculate, the charging load of the charging station can be obtained:

$$P_{\text{cs}}(t) = N_{\text{ch}}(t) P_{\text{rated,avg}} \tag{8}$$

Where $P_{\text{rated,avg}}$ is the rated power of the charging piles.

3 Day-ahead and intra-day self-consumption scheduling

The structure of the renewable energy charging station on highways is shown in Fig.4. The PV and WP generations can deliver electricity to EVs, and an energy storage system can store excess energy and provide electricity to the grid when prices are high. The charging station can sell and purchase electricity to the grid, meeting the charging demand for EVs at the same time and increasing the overall revenue of the charging station.

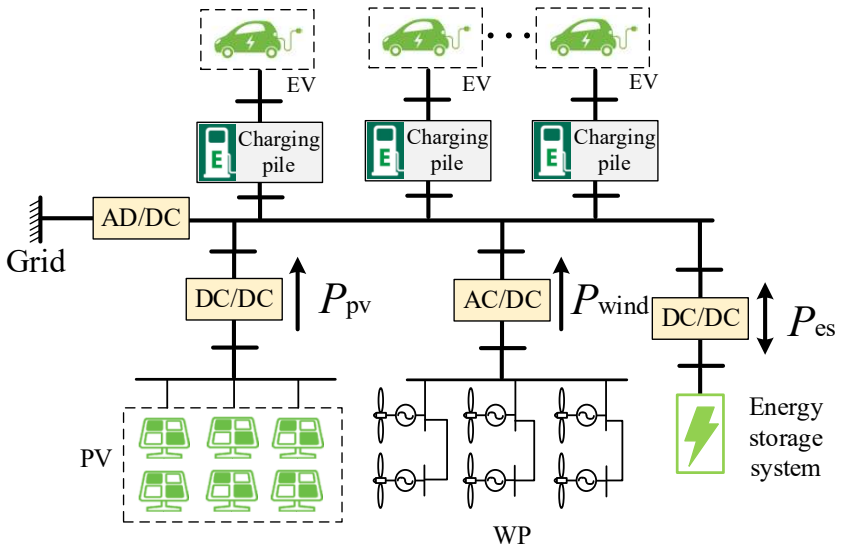


Fig.4. The structure of the renewable energy charging station on highways

Optimal day-ahead and intra-day self-consumption scheduling model is mainly divided into day-ahead optimal scheduling and day-ahead real-time optimal scheduling. The framework is shown in Fig. 5. Optimal day-ahead scheduling takes the 24-hour incentive price as the control variable. The CNN-BiLSTM algorithm and the queue theory were used to predict the 24-hour traffic flow and load, and the 24-hour optimization results were solved

according to the day-ahead optimization objective with an interval of 1 hour. The daily optimal scheduling takes the power of the charging piles and ESS as the control variables. The reward function is set within the day and the charging piles and ESS are regulated by multi-agent deep reinforcement learning with a time interval of 15 minutes.

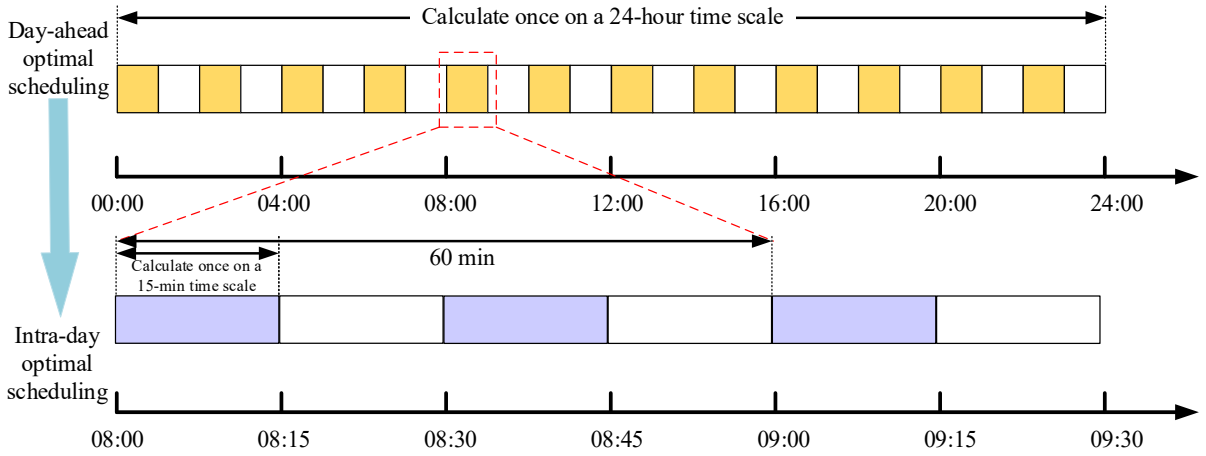


Fig.5. Day-ahead and intra-day scheduling framework

3.1 Optimal day-ahead self-consumption scheduling model

3.1.1 Objective function of day-ahead optimization

Considering the revenue, self-consumption, low-carbon and excess capacity penalty of a charging station, objective function of day-ahead optimization can be established as follows:

$$\max\left[K_{\text{sum}}+\varphi S_{\text{pw}}-\omega E_{\text{c}}-\psi Q_{\text{e,s}}\right] \tag{9}$$

where K_{sum} is the revenue of the charging station, and S_{pw} is the self-consumption rate. E_{c} is the carbon emission, and $Q_{\text{se,s}}$ is the excess capacity penalty. φ, ω and ψ are the coefficients of the objective function.

1) Revenue of the charging station

Revenue consists of charging revenue, revenue from electricity sales and electricity purchase cost of charging station.

$$K_{\text{sum}}=\sum_{t=0}^{T_{sum}}\left[\left(c_{\text{r}}(t)+c_{\text{s}}(t)+c_{\text{a}}(t)\right)P_{\text{cs}}(t)T_s+c_{\text{pl}}(t)P_{\text{out}}(t)T_s-c_{\text{r}}(t)P_{\text{g}}(t)T_s\right] \tag{10}$$

where $c_{\text{r}}(t)$ is the electricity price, $c_{\text{s}}(t)$ is the charging service fee and $c_{\text{a}}(t)$ is the incentive price at time t. T_{s} is the sampling time, and T_{sum} is the total time. $c_{\text{pl}}(t)$ is the electricity price sold to the grid at time t, and P_{out} is the power sold to the grid. $c_{\text{r}}(t)$ is the electricity price at time t, and $P_{\text{g}}(t)$ is the power input from the grid to the charging station at time t.

2) Self-consumption rate

$$S_{\text{pw}}=\frac{\sum_{t=0}^{T_{\text{sum}}}\left(P_{\text{wind}}(t)+P_{\text{pv}}(t)\right)}{\sum_{t=0}^{T_{\text{sum}}}\left(P_{\text{g}}(t)+P_{\text{wind}}(t)+P_{\text{pv}}(t)\right)} \tag{11}$$

where $P_{\text{pv}}(t)$ is the PV power at time t, $P_{\text{wind}}(t)$ is the WP at time t.

3) Carbon emission

The calculation formula for carbon emissions of electricity purchased from charging stations to the grid is as follows:

$$E_c = \sum_{t=0}^{T_{\text{sum}}} M_{\text{co}_2} \left(P_{\text{g}}(t) - P_{\text{out}}(t) \right) T_s \tag{12}$$

where M_{CO2} is the carbon dioxide emission per unit of electricity (taking the coal-fired machine as an example).

4) Excess capacity penalty

Highway charging stations need to ensure that there is a parking space for low-battery EVs, and that EVs cannot be parked on highways. Therefore, set excess capacity penalty of charging station:

$$e_{\text{s}}(t) = \begin{cases} 1, & k(t) = C(t) \\ 0.5, & 0.8C(t) \le k(t) = C(t) \end{cases} \tag{13}$$

$$Q_{\text{e,s}} = \sum_{t=0}^{T_{\text{sum}}} e_{\text{s}}(t) \tag{14}$$

Where $k(t)$ is the length of the queue (k=0, 1, 2, …, N), $C(t)$ is the total space capacity of the charging station (sum of the number of charging slots and parking spaces).

3.1.2 Constraints

1) Charging station power model

According to the WP, PV, charging power, and state of charge (SOC) of the ESS, the power balance formula of the charging station can be divided into three kinds. When the sum of the WP and PV power is greater than the total charging power, the charging power is provided by the WP and PV. When the difference between the sum of the WP and PV power and charging power is greater than the rated power of the ESS, the ESS is charged according to the rated power, and the remaining power is transmitted to the grid, otherwise, the charging power of the ESS is the difference between the sum of the WP and PV power and charging power. When the ESS is overcharged, the excess energy is directly transmitted to the grid.

$$\begin{cases} P_{\text{out}}(t) = P_{\text{cs}}(t) + P_{\text{es,rated}} - \left(P_{\text{wind}}(t) + P_{\text{pv}}(t) \right), & P_{\text{wind}}(t) + P_{\text{pv}}(t) - P_{\text{cs}}(t) > P_{\text{es,rated}} \,\&\, \text{SOC}(t) < SOC_{\text{es,max}} \\ P_{\text{es}}(t) = P_{\text{cs}}(t) - \left(P_{\text{wind}}(t) + P_{\text{pv}}(t) \right), & P_{\text{es,rated}} \ge P_{\text{wind}}(t) + P_{\text{pv}}(t) - P_{\text{cs}}(t) > 0 \,\&\, \text{SOC}(t) < SOC_{\text{es,max}} \\ P_{\text{out}}(t) = P_{\text{cs}}(t) - \left(P_{\text{wind}}(t) + P_{\text{pv}}(t) \right), & P_{\text{wind}}(t) + P_{\text{pv}}(t) - P_{\text{cs}}(t) > 0 \,\&\, \text{SOC}(t) = SOC_{\text{es,max}} \end{cases} \tag{15}$$

where $P_{\text{c,n}}(t)$ is the charging power of the n^{th} charging pile at time t, and $P_{\text{es,rated}}$ is the rated power of the battery energy storage system,

When the difference between the sum of the WP and PV power and charging power is 0 to -$P_{\text{es,rated}}$, the charging power is provided by the WP, PV, and ESS. When the ESS is over-discharged, the charging power is provided by the WP, PV, and power grid.

$$\begin{cases} P_{\text{es}}(t) = P_{\text{cs}}(t) - P_{\text{wind}}(t) + P_{\text{pv}}(t), & 0 > P_{\text{wind}}(t) + P_{\text{pv}}(t) - P_{\text{cs}}(t) > -P_{\text{es,rated}} \,\&\, \text{SOC}(t) > SOC_{\text{es,min}} \\ P_{\text{out}}(t) = P_{\text{cs}}(t) - P_{\text{wind}}(t) + P_{\text{pv}}(t), & 0 > P_{\text{wind}}(t) + P_{\text{pv}}(t) - P_{\text{cs}}(t) > -P_{\text{es,rated}} \,\&\, \text{SOC}(t) = SOC_{\text{es,max}} \end{cases} \tag{16}$$

When the sum of the rated power of the WP, PV and ESS is less than the total charging power, the charging power is first provided by the WP, PV and ESS, and power grid. When the ESS is over-discharged, the charging power is provided by the WP, PV, and power grid.

$$\begin{cases} P_{\text{cs}}(t) = P_{\text{wind}}(t) + P_{\text{pv}}(t) + P_{\text{es,rated}}(t) + P_{\text{g}}(t), & P_{\text{wind}}(t) + P_{\text{pv}}(t) + P_{\text{es,rated}} < P_{\text{cs}}(t) \,\&\, \text{SOC}(t) > SOC_{\text{es,min}} \\ P_{\text{cs}}(t) = P_{\text{wind}}(t) + P_{\text{pv}}(t) + P_{\text{g}}(t), & P_{\text{wind}}(t) + P_{\text{pv}}(t) + P_{\text{es,rated}} < P_{\text{cs}}(t) \,\&\, \text{SOC}(t) = SOC_{\text{es,max}} \end{cases} \tag{17}$$

2) Constraints on EV power and SOC:

$$\begin{cases} P_{\text{p1,min}} \le P_{\text{ev\_}i}(t) \le P_{\text{p1,rated}}, & i \in R_{\text{p1}} \\ P_{\text{p2,min}} \le P_{\text{ev\_}i}(t) \le P_{\text{p2,rated}}, & i \in R_{\text{p2}} \end{cases} \tag{18}$$

$$\begin{cases} \dfrac{d_i w_i}{C_i} \le SOC_{i,\text{out}}(t) \le 100\% \end{cases} \tag{19}$$

where $P_{\text{ev\_}i}(t)$ is the power of the i^{th} EV at time t. The charging pile types of highway charging stations can generally be divided into ultra-fast charging (Liquid-cooled charging piles) and fast charging piles (Air-cooled charging piles), so the power limits of EVs can be divided into two categories. $P_{\text{p1,min}}$ and $P_{\text{p2,min}}$ is the minimum charging power of ultra-fast charging and fast charging piles. $P_{\text{p1,rated}}$ and $P_{\text{p2,rated}}$ are the rated power of ultra-fast charging and fast charging piles. R_{p1}, and R_{p2} are collections of EVs connected to ultra-fast charging and fast charging piles respectively. $SOC_{i,\text{out}}$ is the driving capacity of the i^{th} EV. C_i is the battery capacity of the i^{th} EV, d_i is the expected travel mileage of the next journey of the i^{th} EV, and w_i represents the power consumption per kilometer of the EV. The SOC constraint of the EV means that the outgoing power needs to complete the estimated travel distance when off-grid.

3) Charging time constraints:

$$\frac{d_i w_i}{C_i P_{\text{p2,rated}}} \le t_{\text{ev\_}i} \le T_{\text{ev,max}} \tag{20}$$

where $t_{\text{ev\_}i}(t)$ is the charging time of the i^{th} EV, $T_{\text{ev,max}}$ is the longest charging time of the EV, and the reasonable longest charging time is determined according to the historical charging time of a highway charging station.

4) Constraints on ESS:

$$-P_{\text{es,rated}} \le P_{\text{es}}(t) \le P_{\text{es,rated}} \tag{21}$$

$$SOC_{\text{es,min}} \le SOC_{\text{es}}(t) \le SOC_{\text{es,max}} \tag{22}$$

$$SOC_{\text{es}}(t+1) = SOC_{\text{es}}(t) - \eta_{\text{es}} \frac{P_{\text{es}}(t) T_s}{C_{\text{es}}} \tag{23}$$

where $P_{\text{es}}(t)$ and $P_{\text{es,rated}}$ are the power and rated power of the ESS at time t. $SOC_{\text{es,max}}$ and $SOC_{\text{es,min}}$ are the SOC maximum and minimum limit of the ESS respectively. $SOC_{\text{es}}(t)$ and $SOC_{\text{es}}(t+1)$ are the SOC of the ESS at time t and $t+1$ respectively. η_{es} is the efficiency of the ESS, and C_{es} is the capacity of the ESS.

3.2 Optimal intra-day self-consumption scheduling based on the MADRL

The optimal intra-day self-consumption scheduling based on the MADRL is to model the charging piles and ESS as the agents and to model the highway EV charging station as a multi-agent environment. By interacting with the environment, the multiple agents constantly learn by using the feedback information, improves the decision-making ability under the change of environmental state, and finally realizes the optimal intra-day scheduling of the charging station. The optimal intra-day scheduling based on MADRL is shown in Fig. 6. The intra-day optimal control problem of the charging station is transformed into the MGM, and the MGM problem is solved by the multi-agent twin delayed deep deterministic policy gradient (MATD3) algorithm.

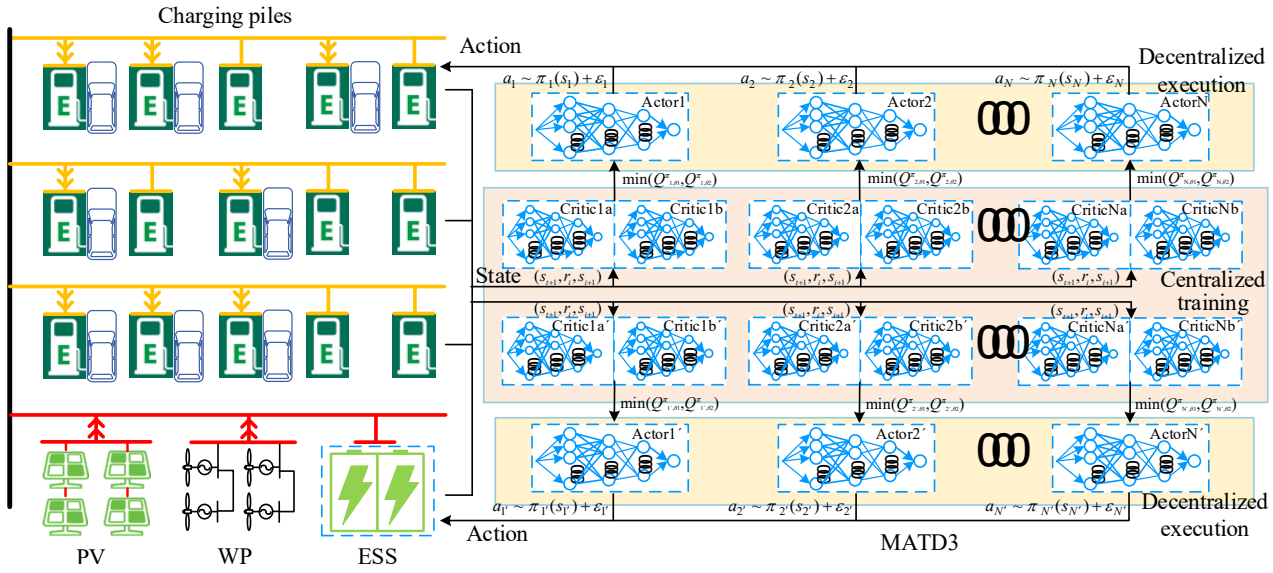


Fig. 6. The intra-day optimal scheduling based on MADRL

3.2.1 Markov game model

MADRL is a machine learning algorithm that combines deep learning methods, reinforcement learning methods, and multi-agent systems. MADRL can organize multi-agents to carry out autonomous learning and realize cooperative problem-solving through the interaction between agents.

For the charging scheduling problem of EVs and ESS, the state at each time is only related to the previous state and action of the agent, which conforms to the MDP. Moreover, it is a complex multi-objective optimization problem to control multiple charging piles and ESS in charging stations. Therefore, the multi-objective optimization problem is transformed into a game process among multiple agents to reduce the complexity of the solution. In the multi-agent system, the MDP is extended to the MGM, and the intraday optimal control problem of the charging station is transformed into the MGM. Usually represented by a set of multivariate groups, $\{S_1 \ldots S_i \ldots S_n, A_1 \ldots A_i \ldots A_n, P, R_1 \ldots R_i \ldots R_n, R_g, \gamma\}$. Among them, $S_1 \ldots S_i \ldots S_{n'}$ represents the state space observed by n′ agents, $A_1 \ldots A_i \ldots A_{n'}$ represents the action space of n′ agents, $R_1 \ldots R_i \ldots R_{n'}$ represents the reward space of n′ agents, R_g represents the global reward for all agents, P represents the state transition probability, and γ represents the reward discount coefficient. The MGM in the optimal intra-day self-consumption scheduling problem is shown below.

1) Agent set

If an agent controls charging piles and the ESS of a charging station, the calculation amount will be greatly increased. To simplify the calculation, according to the classification of object types, the agents of MGM are divided into three categories: the agent of fast pile charging pile I_{fc}, the agent of ultra-fast pile charging pile I_{sf}, and the agent of the ESS I_{ess}. The set of agents is $I=\{I_{fc}, I_{sf}, I_{ess}\}$.

2) State space

For the renewable energy charging station on highways, the information provided by the environment to the agents is the purchased power, sold power, PV power, PV, charging power of charging piles, SOC of EVs and output power and SOC of the ESS at time t. The state space of the model at time t is thus defined as follows,

$$S_i(t)=\left[P_g(t),P_{out}(t),P_{pv}(t),P_{wind}(t),P_{c,i}(t),P_{es}(t),k(t),SOC_{ev\_i}(t),SOC_{es}(t)\right] \tag{24}$$

3) Action space

The action space $A(t)$ is the set of decisions made by the agent based on the state S and strategy π in the current environment. The action space $A(t)$ is composed of the action set $a_{fc}(t)$ of the fast charging piles, the action set $a_{sf}(t)$ of the ultra-fast charging piles, and the action set $a_{ess}(t)$ of the ESS in the station. The action space $A(t)$ is as follows.

$$A(t)=\begin{cases} a_{\mathrm{fc}}(t)=[P_{\mathrm{p1},1}(t),\mathrm{K}\ ,P_{\mathrm{p1},j}(t)],P_{\mathrm{p1,min}}\le P_{\mathrm{p1},j}(t)\le P_{\mathrm{p1,rated}} \\ a_{\mathrm{sf}}(t)=[P_{\mathrm{p2},1}(t),\mathrm{K}\ ,P_{\mathrm{p2},k}(t)],P_{\mathrm{p2,min}}\le P_{\mathrm{p2},k}(t)\le P_{\mathrm{p2,rated}} \\ a_{\mathrm{ess}}(t)=P_{\mathrm{es}}(t),-P_{\mathrm{es,rated}}(t)\le P_{\mathrm{es}}(t)\le P_{\mathrm{es,rated}}(t) \end{cases} \tag{25}$$

Where I_{ess} controls the ESS power P_{es} in the station through a_{ess} (t); The action set $I_{\mathrm{fc}}(t)$ of the fast charging piles consists of the power $P_{\mathrm{p1},1}(t),\ldots, P_{\mathrm{p1},j}(t)$, and I_{fc} controls the power of m fast charging piles through $a_{\mathrm{fc}}(t)$. The action set $a_{\mathrm{sf}}(t)$ of the ultra-fast charging piles consists of the power $P_{\mathrm{p2},1}(t),\ldots, P_{\mathrm{p2},k}(t)$, and the agent I_{sf} controls the power of n ultra-fast charging piles through $a_{\mathrm{sf}}(t)$.

4) Reward space

The goal of intra-day optimal scheduling is to improve real-time revenue, self-consumption, low-carbon and excess capacity penalty of a charging station. The objective function of intra-day scheduling is as follows.

$$\max J_{\mathrm{RL}}=\sum_{t=0}^{T_{\mathrm{sum}}}K'_{\mathrm{sum}}+\varphi' S_{\mathrm{pw}}-\omega' E_{\mathrm{c}}-\psi' Q_{\mathrm{e,s}}-\theta' C_{\mathrm{loss,s}} \tag{26}$$

Where φ', ω' and ψ' are the coefficients of the objective function of intra-day scheduling.

In the day-ahead, the total charging load predicted by the load model is used to calculate the charging station revenue, while in the intra-day stage, the real-time revenue of the charging station is calculated by directly collecting the real-time power of all charging piles and superimposing the total charging power. The real-time revenue formula for charging stations is as follows.

$$K'_{\mathrm{sum}}(t)=\sum_{n=0}^{N_{\mathrm{sum}}}\left(c_{\mathrm{r}}(t)+c_{\mathrm{s}}(t)+c_{\mathrm{a}}(t)\right)P_{\mathrm{c},n}(t)+c_{\mathrm{pl}}(t)P_{\mathrm{out}}(t)-c_r(t)P_{\mathrm{g}}(t) \tag{27}$$

The ESS in the charging station has undergone multiple charge and discharge cycles, which may cause overcharge and over-discharge. If the cost of battery charging and discharging and the cost of battery life loss caused by overcharging and discharging are considered, the $C_{\mathrm{loss,s}}$ of battery loss cost is,

$$C_{\mathrm{loss}}(t)=\begin{cases} c_{\mathrm{over}}P_{\mathrm{over}}(t)T+c_{\mathrm{es}}P_{\mathrm{es}}(t)T_{\mathrm{s}}, & SOC_{\mathrm{es}}(t)\le SOC_{\mathrm{es,min}}\ or\ SOC_{\mathrm{es,max}}<SOC_{\mathrm{es}}(t) \\ c_{\mathrm{es}}P_{\mathrm{es}}(t)T, & SOC_{\mathrm{es,min}}\le SOC_{\mathrm{es}}(t)\le SOC_{\mathrm{es,max}} \end{cases} \tag{28}$$

$$C_{\mathrm{loss,s}}(t)=\sum_{t=0}^{T_{\mathrm{sum}}}C_{\mathrm{loss}}(t) \tag{29}$$

where $P_{\mathrm{over}}(t)$ is the overcharge and over-discharge power of the ESS at time t, $P_{\mathrm{es}}(t)$ is the power of the ESS at time t, c_{es} is the unit charge and discharge cost of the battery, and c_{over} is the penalty cost of the unit overcharge and over-discharge power of the battery. $SOC_{\mathrm{es,min}}$ and $SOC_{\mathrm{es,max}}$ are the maximum and minimum of the SOC of the ESS, respectively, and SOC_{es} are the SOC of the ESS at time t.

The reward function evaluates the charging station environment and gets an immediate reward, which influences the agent's choice of action. According to the intra-day optimization objective function (26), the reward space of the model at time t is defined as follows.

$$\begin{cases} r_{\mathrm{fc}}(t)=r_{\mathrm{sf}}(t)=\alpha_1\left[\sum_{n=0}^{N_{\mathrm{sum}}}\left(c_{\mathrm{r}}(t)+c_{\mathrm{s}}(t)+c_{\mathrm{a}}(t)\right)P_{\mathrm{c},n}(t)+c_{\mathrm{pl}}(t)P_{\mathrm{out}}(t)-c_r(t)P_{\mathrm{g}}(t)\right] \\ +\alpha_2\dfrac{P_{\mathrm{wind}}(t)+P_{\mathrm{pv}}(t)}{P_{\mathrm{g}}(t)+P_{\mathrm{wind}}(t)+P_{\mathrm{pv}}(t)}-\alpha_3 M_{\mathrm{co_2}}\left(P_{\mathrm{g}}(t)-P_{\mathrm{out}}(t)\right)-\alpha_4 e_{\mathrm{s}}(t), \\ r_{\mathrm{es}}=\beta_1\left[\sum_{n=0}^{N_{\mathrm{sum}}}\left(c_{\mathrm{r}}(t)+c_{\mathrm{s}}(t)+c_{\mathrm{a}}(t)\right)P_{\mathrm{c},n}(t)+c_{\mathrm{pl}}(t)P_{\mathrm{out}}(t)-c_r(t)P_{\mathrm{g}}(t)\right] \\ +\beta_2\dfrac{P_{\mathrm{wind}}(t)+P_{\mathrm{pv}}(t)}{P_{\mathrm{g}}(t)+P_{\mathrm{wind}}(t)+P_{\mathrm{pv}}(t)}-\beta_3 C_{\mathrm{loss}}(t) \end{cases} \tag{30}$$

where α_1, α_2, α_3, β_1, β_2, β_3 are the coefficients of the reward function. $r_{\mathrm{ess}}(t)$ is the reward function of ESS agent I_{ess}, which consists of real-time revenue, self-consumption rate, and ESS loss cost of the charging station. $r_{\mathrm{fc}}(t)$ is the reward function of the agent I_{fc}, which consists of real-time revenue, self-consumption rate, carbon emission, and excess capacity penalty of the charging station. $r_{\mathrm{sf}}(t)$ is the reward function of the agent I_{sf}.

3.2 Intra-day optimization solution based on the MATD3

As a new MADRL algorithm for solving continuous problems, the MATD3 is an improvement based on the multi-agent deep deterministic policy gradient (MADDPG) [29,30]. The MATD3 is designed to solve the overestimation bias of the Q value, high variance accumulation, and unstable learning process of the MADDPG algorithm based on value learning. Therefore, the MATD3 algorithm was used to solve the MGM model of intra-day optimal scheduling.

During the training of the MATD3, the agent $I=\{I_{\mathrm{fc}}; I_{\mathrm{sf}}; I_{\mathrm{ess}}\}$ observes the current state, selects the action $A=\{a_{\mathrm{fc}}(t), a_{\mathrm{sf}}(t), a_{\mathrm{ess}}(t)\}$ from the action space based on the policy π, and obtains the immediate reward $\{r_{\mathrm{fc}}(t), r_{\mathrm{sf}}(t), r_{\mathrm{ess}}(t)\}$ based on the reward function (30). The formula for calculating the cumulative reward R_{cum} from the period t=0 to the end of the agent learning is as follows.

$$R_{\mathrm{cum}}(t)=\sum_{t=0}^{T_{\mathrm{sum}}}\gamma R\left(S',\pi_{\phi,\mathrm{fc}}(S')+\omega,\pi_{\phi,\mathrm{sf}}(S')+\omega,\pi_{\phi,\mathrm{ess}}(S')+\omega\right) \tag{31}$$

Where γ is the discount factor, which determines the impact of future rewards on cumulative rewards. $\pi_{\phi,\mathrm{fc}}$, $\pi_{\phi,\mathrm{sf}}$ and $\pi_{\phi,\mathrm{ess}}$ are the action strategies of fast charging piles, ultra-fast charging piles and ESS, respectively. ω is Gaussian noise.

The goal of the MATD3 algorithm is to find the optimal policy π so that the agent expects the maximum cumulative reward. The state-action value function evaluates the action $A(t)$ of the charging pile and the ESS agent based on the charging station information (a set of multiple groups) $\{S(t), a_{\mathrm{fc}}(t), a_{\mathrm{sf}}(t), a_{\mathrm{ess}}(t), P, r_{\mathrm{fc}}(t), r_{\mathrm{sf}}(t), r_{\mathrm{ess}}(t), S'(t), \gamma\}$. Seek to minimize the difference between the state-action value function Q value and the target Q value. The Bellman equation of the state-action value function is as follows.

$$Q(S(t),A(t))=R_i(t)+\gamma E_\pi\left[R_{\mathrm{cum}}(t)\,|\,S(t),A(t)\right] \tag{32}$$

To solve the overestimation bias of Q value in the deterministic policy gradient algorithm, the MATD3 algorithm takes the minimum value of the two evaluators when calculating the target value Q. The target value Q approximated by the charging piles and ESS agent estimated value network is as follows.

$$y_i=R_i+\gamma\min_{n=1,2}Q^{\pi}_{i,\theta'_n}(S',\pi_{\phi,\mathrm{fc}}(S')+\omega,\pi_{\phi,\mathrm{sf}}(S')+\omega,\pi_{\phi,\mathrm{ess}}(S')+\omega) \tag{33}$$

Based on the policy π, action $A(t)=\{a_{\mathrm{fc}}(t), a_{\mathrm{sf}}(t), a_{\mathrm{ess}}(t)\}$ is selected from the action space. The calculation formula of action (power) of the charging piles and ESS agent in the policy network is as follows.

$$\begin{cases} a_{\text{fc}} = \pi_{\phi,\text{fc}}(S') + \omega \sim N(\pi_{\phi,\text{fc}}(S'), \sigma^2) \\ a_{\text{sf}} = \pi_{\phi,\text{sf}}(S') + \omega \sim N(\pi_{\phi,\text{sf}}(S'), \sigma^2) \\ a_{\text{ess}} = \pi_{\phi,\text{ess}}(S') + \omega \sim N(\pi_{\phi,\text{ess}}(S'), \sigma^2) \end{cases} \tag{34}$$

Where agents observe state variables such as PV power, WP, SOC of the charging station and makes the optimal action (optimal power) by the strategies $\pi_{\phi,\text{fc}}$, $\pi_{\phi,\text{sf}}$, and $\pi_{\phi,\text{ess}}$. $N(0,\sigma)$ is the positive distribution noise; c is the noise range.

To reduce the high variance of the target value when updating the actor, the MATD3 algorithm uses the regularization technique of target policy smoothing. The policy gradient of agent i is as follows.

$$\nabla_{\theta_{\phi,i}} J(\phi) \approx N^{-1} \sum \nabla_{\theta} \pi_{\phi} \nabla_{\phi}(S) Q_{i,\theta_1}^{\pi}(S, A_1, \text{K}\ A_N)|_{A_i=\pi_{\phi}(s_i)} \tag{35}$$

Finally, two target network parameters are updated by soft update policy:

$$\begin{cases} \theta'_{i,n} \leftarrow \tau\theta_{i,n} + (1-\tau)\theta'_{i,n}, n = 1,2 \\ \phi' \leftarrow \tau\phi + (1-\tau)\phi' \end{cases} \tag{36}$$

The MATD3 is a centralized training and decentralized execution mode. During the training period, the two evaluation networks of the charging piles and ESS agent observe the observable data signals in the charging station to evaluate the action (power) of the policy network, and through continuous interactive learning, the charging piles and ESS agent can learn the optimal intraday control policy. In the decentralized execution stage, the actor of each agent performs decentralized tasks according to the policy, and the charging piles and ESS agent can quickly and real-time schedule the power of the charging piles or ESS after obtaining the state of the charging piles or ESS. Intraday scheduling algorithm of the MATD3 is shown in Table 1.

Table 1 Intraday scheduling algorithm of the MATD3

Algorithm1: Centralized training

Initialize: the two critic networks $Q^{\pi}_{i,\theta 1}$, $Q^{\pi}_{i,\theta 2}$, the network parameters $\theta_{i,1}$, $\theta_{i,2}$, ϕ_i of the actor network for each agent i, the experience buffer B, the model parameter of the charging station;
Assign network parameters to corresponding target network parameters;
Input: States of the charging station such as the length of the queue, WP, and PV power;
Output: Actions of the charging piles and ESS $A(t)$=$\{a_{\text{fc}}(t), a_{\text{sf}}(t), a_{\text{ess}}(t)\}$;
for episode=1, 2, …, M **do**
  Initialize a random process for action exploration;
  Receive initial observation state;
  for t=1,2…T_{sum} **do**
    Agents of charging piles and ESS select random action $A(t)$= {1:$a_{\text{fc}}(t)$, 2:$a_{\text{sf}}(t)$, 3: $a_{\text{ess}}(t)$}, and the noise is dynamically adjusted to explore the current deterministic policy;
    Execute the action $A(t)$, schedule power of charging piles and ESS, observe states of charging station $S(t)$, and calculate new reward $\{r_{\text{fc}}(t), r_{\text{sf}}(t), r_{\text{ess}}(t)\}$;
    Put the experience transition ($S(t)$, $a_{\text{fc}}(t)$, $a_{\text{sf}}(t)$, $a_{\text{ess}}(t)$, $r_{\text{fc}}(t)$, $r_{\text{sf}}(t)$, $r_{\text{ess}}(t)$, $S'(t)$) into the experience buffer B;
    for agent i=1 to 3 **do**
      a minibatch $\{S(t), a_{\text{fc}}(t), a_{\text{sf}}(t), a_{\text{ess}}(t), P, r_{\text{fc}}(t), r_{\text{sf}}(t), r_{\text{ess}}(t), S'(t), \gamma\}$ is sampled from the experience buffer for the training of the policy network and value network.
      Update target value

$$y_i = R_i + \gamma \min_{o=1,2} Q^{\pi}_{i,\theta'_o}(S', \pi_{\phi,1}(S'_1) + \varepsilon, ..., \pi_{\phi,n'}(S'_{n'}) + \varepsilon)$$

      Update critic network parameter
      if t mod d **then**
        Update action parameters by policy gradient

$$\nabla_{\theta_{\phi,i}} J(\phi) \approx N^{-1} \sum \nabla_{\theta} \pi_{\phi} \nabla_{\phi}(S) Q_{i,\theta_1}^{\pi}(S, A_1, \mathrm{K}\ A_{n'}) \big|_{A_i = \pi_{\phi}(S_i)}$$

Update the target value network and policy network parameters,

$$\begin{cases} \theta'_{i,o} \leftarrow \tau\theta_{i,o} + (1-\tau)\theta'_{i,o}, o = 1,2 \\ \phi' \leftarrow \tau\phi + (1-\tau)\phi' \end{cases}$$

end if

end for

end for

end for

Algorithm2: Decentralized scheduling

Input: Real-time states of the charging station such as the length of the queue, WP, and PV power;

Output: Real-time actions of the charging piles and ESS $\{a_{fc}(t), a_{sf}(t), a_{ess}(t)\}$;

for t=1,2…T_{sum} **do**

for agent i=1 to 3 **do**

Agent i {1: I_{fc}, 2: I_{sf}, 3: I_{ess}} observe state $S(t)$;

Execute the action $A(t)$ ={1:$a_{fc}(t)$, 2:$a_{sf}(t)$, 3: $a_{ess}(t)$}, and schedule power of charging piles and ESS;

end for

end for

4 Case study

4.1 Parameter configuration

Take a highway charging station as the test case, which is with 20 charging piles, including a wind farm, a photovoltaic power generation, and an ESS. The 20 charging piles are subdivided into 16 ultra-fast charging piles and 4 fast charging piles. By using the data of EVs entering a highway charging station and the power data of a wind farm and a photovoltaic power generation, the effectiveness and feasibility of the proposed method are verified. The highway charging station layout is shown in Fig. 7, the system parameters are shown in Table 2. The power of a typical PV and wind farm are used for the data in the experiments. The 24-hour PV and WP are shown in Fig. 8.

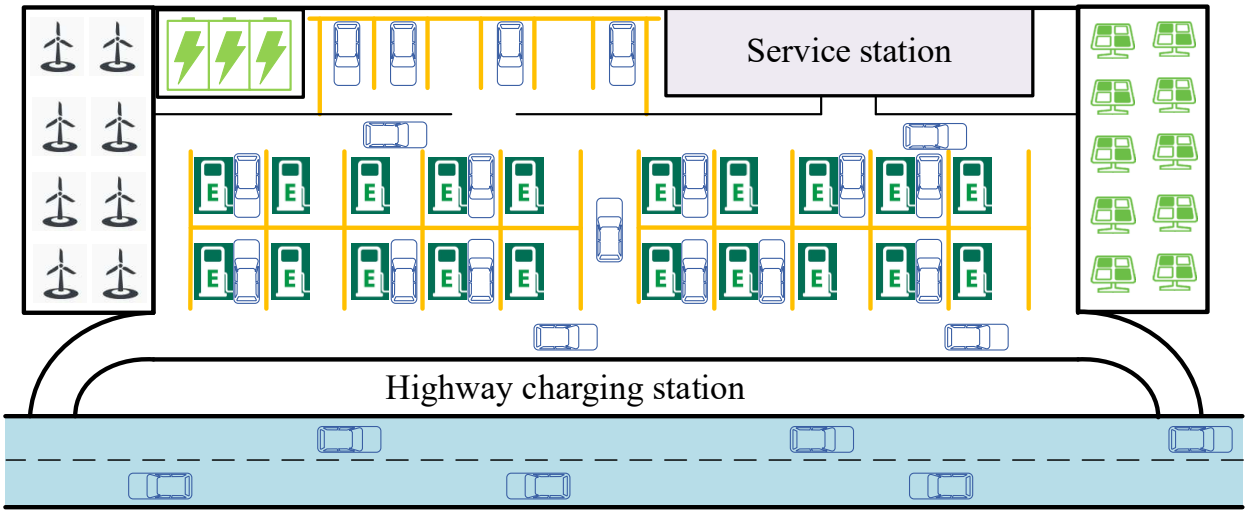


Fig.7. The highway charging station layout

Table.2. system parameters

Parameters	Values	Parameters	Values
N_{sum}	20	$T_{ev,max}$ (h)	2
$P_{p1,min}/P_{p1,rated}/P_{p2,min}/P_{p2,rated}$ (kW)	40/120/80/200	$P_{es,rated}$ (kW)	3000
Capacity of the ESS (MW•h)	6	C_i (kW•h)	100
Initial SOC of ESS	0	c_{es} (CNY/kW•h)	0.1
M_{CO2} (kg/kW•h)	0.785	c_{over} (CNY/kW•h)	0.15
$SOC_{es,max}/SOC_{es,min}$	0.95/0	c_r (CNY/kW•h)	0.505
η_{es}	0.96	c_s (CNY/kW•h)	0.2
T_s/min(day-ahead)	60	T_{sum}/h	23
T_s/min(intra-day)	15	$P_{rated,avg}$ (kW)	136

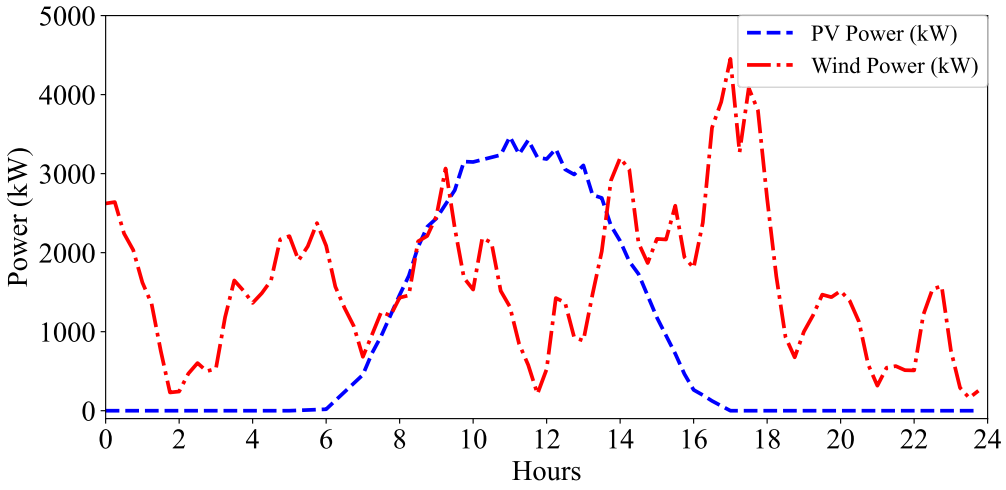


Fig.8. 24-hour PV and WP

4.2 Analysis of training results

The study uses the traffic flow of an actual highway as experimental data. The input of the prediction model is the traffic flow data for the past four days, recorded every hour for 24 hours, and the output is the traffic flow of day t+1. The study compares the prediction results of the LSTM, BiLSTM, and CNN-BiLSTM models and presents the results in Fig. 9.

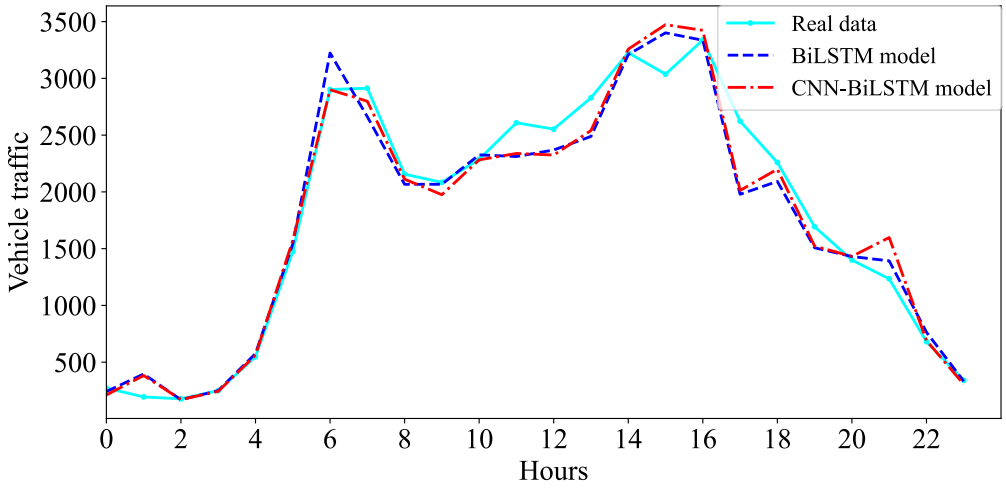
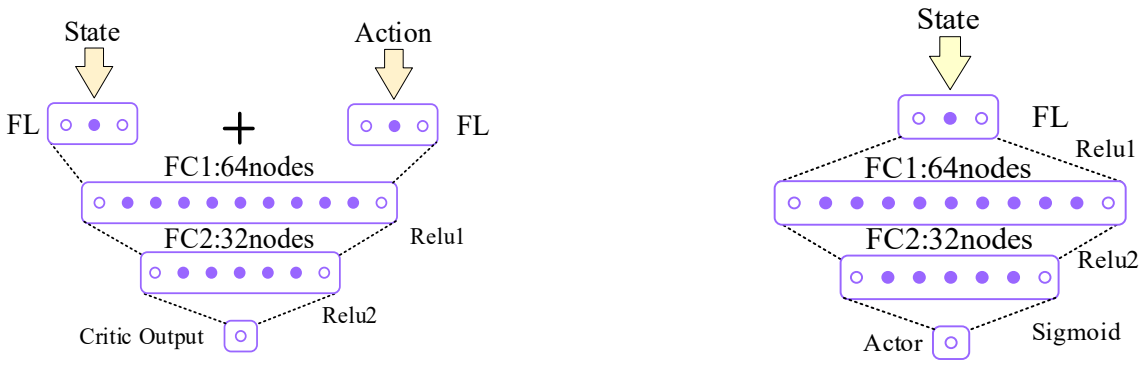


Fig.9. Prediction of vehicle traffic

After many experiments, the hyperparameter of algorithm training was determined (the TD3 and MATD3 share this hyperparameter), as shown in Table 3. The structural design of critic network and actor network of the algorithm is shown in Fig. 10.

Table 3 Hyperparameters of algorithm training

Hyperparameters	Values	Weights	Values
Training number	500	α_1	1
Batch number	512	α_2	10
Capacity of the experience buffer	1×10^6	α_3	1
Discount factor	0.99	β_1	1
Learning rate	0.01	β_2	10
Optimizer	Adam	β_3	1



a Structure of critic network

b Structure of actor network

Fig.10. Structure of the algorithm network

To ensure the training effect, the agent conducts 500 trial-and-error training. According to the objective function of intra-day scheduling, the global reward J_{RL} of the MATD3 and TD3 algorithms are shown in Fig. 11. As can be seen from Fig. 11, the MATD3 algorithm began to converge after the 100th generation, and the global reward was 7.44×10^5 in the 500th generation, which was significantly higher than the 7.28×10^5 of the TD3 algorithm. At the same time, the global reward fluctuation of the MATD3 algorithm is smaller than that of the TD3 algorithm after the 100th convergence. It shows that the MATD3 algorithm has better optimization effect and stability than the TD3 in the complex scheduling problem of the highway charging station.

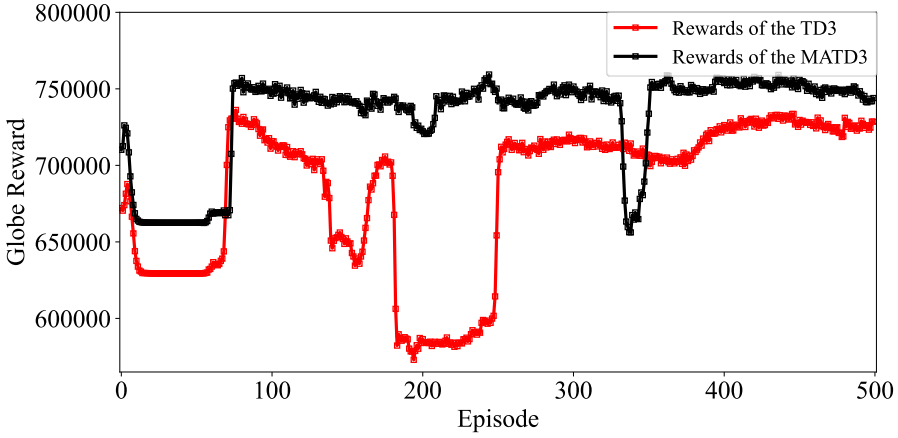


Fig. 11. Global reward

4.3 Analysis of simulation experiment

During the day-ahead stage, traffic flow and load for the next 24 hours are predicted. Optimization results for the next 24 hours are then generated based on the pre-day optimization objective, with a time scale of 1 hour. In the same environmental conditions, the results of a day-ahead experiment with price incentives are compared with the results of an experiment without price incentives.

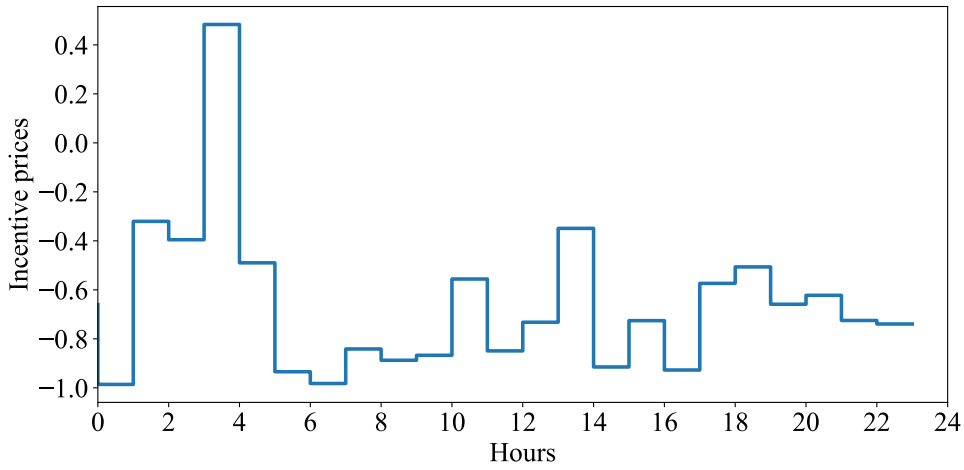
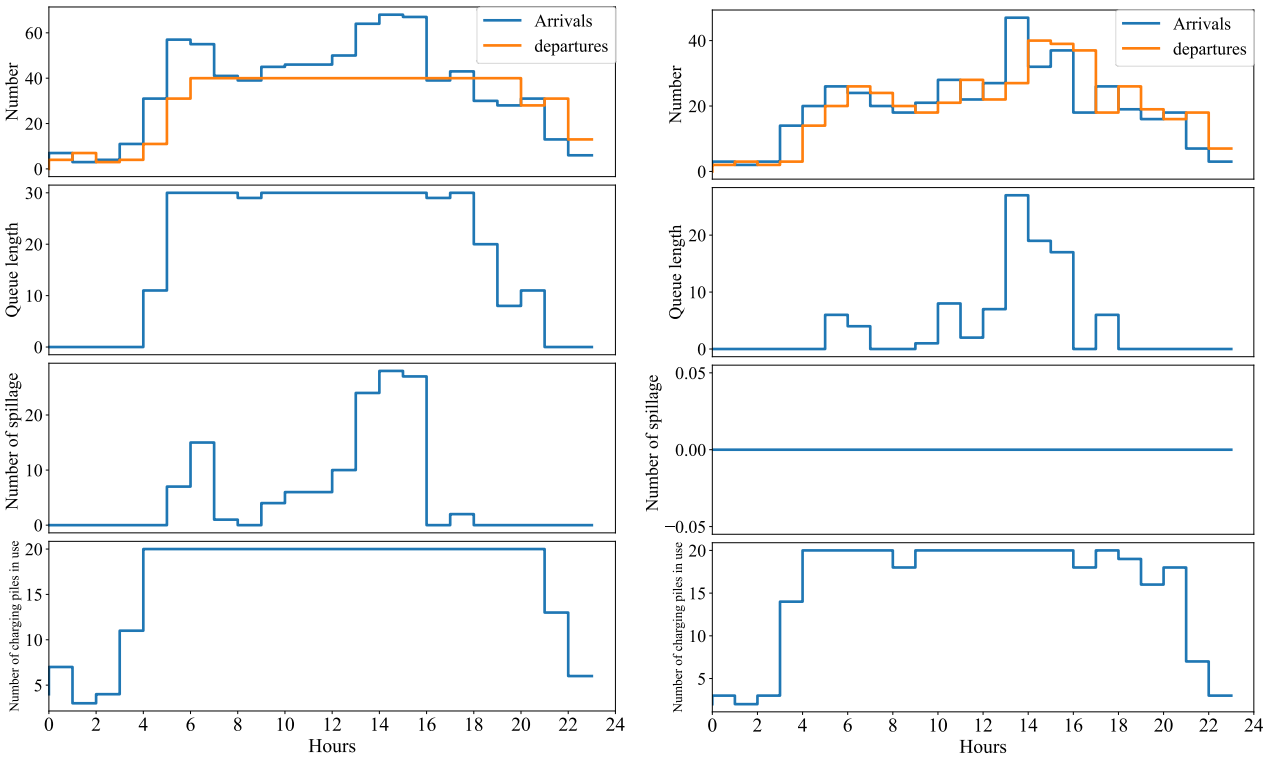


Fig. 12. Incentive price



a Without the optimal price incentives

b With the optimal price incentives

Fig. 13. Number of arrivals, departures, queue, spillage and charging piles in use

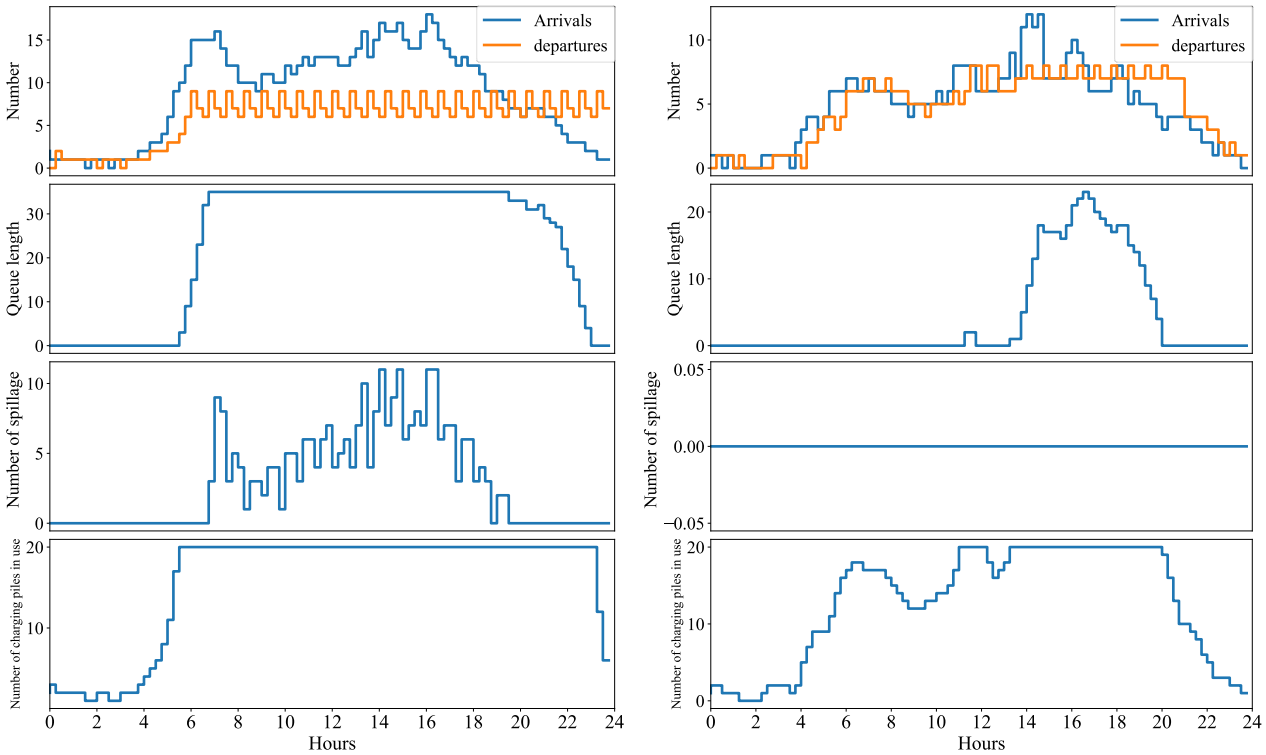
Table.4 Day-ahead evaluation indicators

Methods	K_{sum}/CNY	S_{pw}	E_c/kg	Number of spillage
Day-ahead scheduling	38004	85.83%	-4441	130
Day-ahead scheduling with the optimal price incentives	70471	88.04%	-6753	0

The incentive price is shown in Fig. 12. The number of arrivals, departures, queue, spillage and charging piles in use with or without price incentives are shown in Fig. 13. The evaluation indexes of day-ahead scheduling are shown in Table 4. It can be seen from Fig. 13 that during 5 to 19h, the number of spillage is greater than 0 or even more than 20 without price incentives, and the number of spillage is 0 with price incentives. Moreover, it can be seen from Table 4 that the number of spillage in 24 hours without price incentives is 130, and 130 EVs that need to be charged cannot enter the charging station, leading to full parking in charging stations and even parking in emergency lanes. Under the price incentives, the number of overflows in 24 hours is 0, and there is no situation that the highway charging station is full.

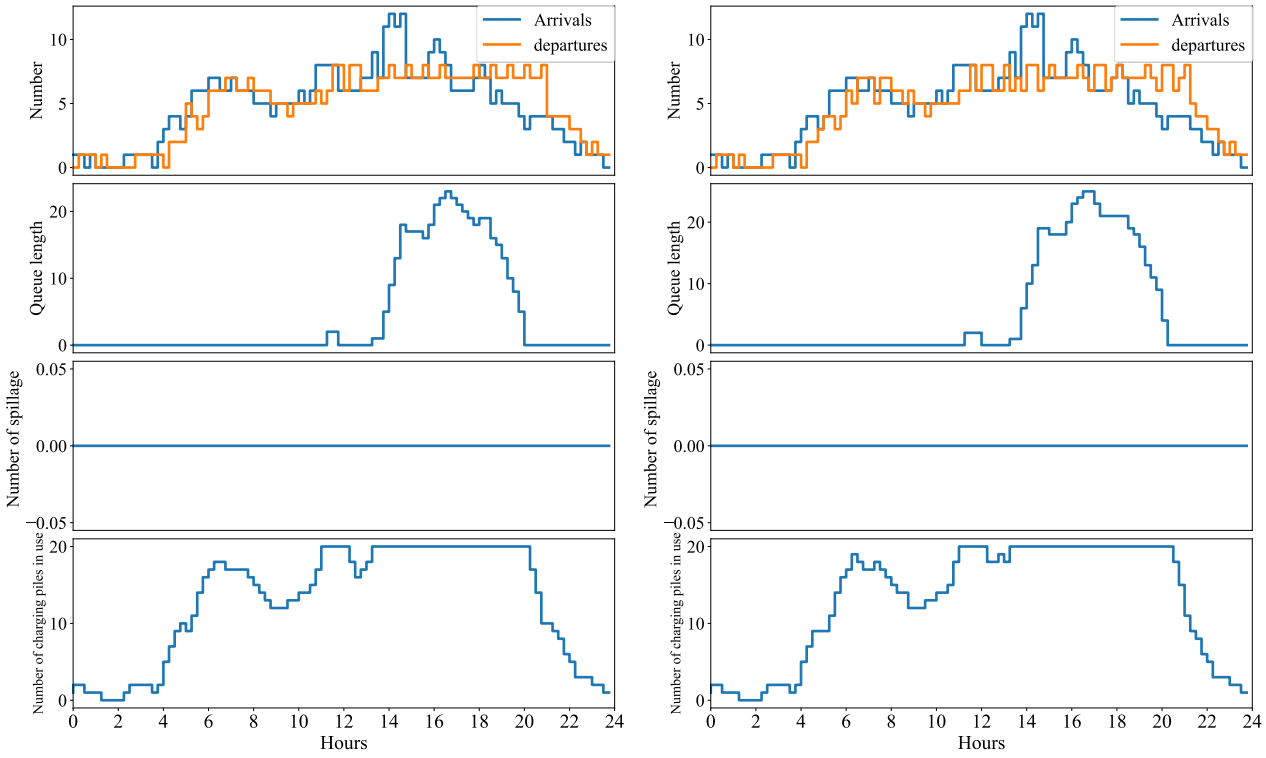
As can be seen from Table 4, Revenue K_{sum}=70471CNY and self-consumption rate S_{pw}=88.04% with price incentives are significantly higher than K_{sum}=38004CNY and self- consumption rate S_{pw}=85.83% with no price incentives, and the carbon emission E_c=-6753kg is 2312kg less than the E_c=-4441kg without price incentive. The results show that under the price incentives, there is no full parking of highway charging station, the revenue is increased, the energy self-cycling in the station is realized, and the carbon emission in the station is 0, providing clean energy for the power grid. If the energy loss is not considered, the external carbon emission is reduced by 7476kg.

In the intra-day stage, the MATD3 controls the power of the charging piles and ESS according to the set reward function (the time interval is 15 minutes). Run simulation experiments in the same environment variables. To verify the effectiveness and optimality of the proposed method, it is compared with the average allocation without the PV-WP-ES, average allocation with the PV-WP-ES, optimal method based on the day-ahead price incentives, and optimal method based on the day-ahead price incentive and TD3 algorithm (TD3 algorithm). The solution algorithm of the optimal method based on the day-ahead price incentives is the Whale optimization algorithm (WOA).



a Average allocation with the PV-WP-ES

b Optimal method based on the day-ahead price incentives



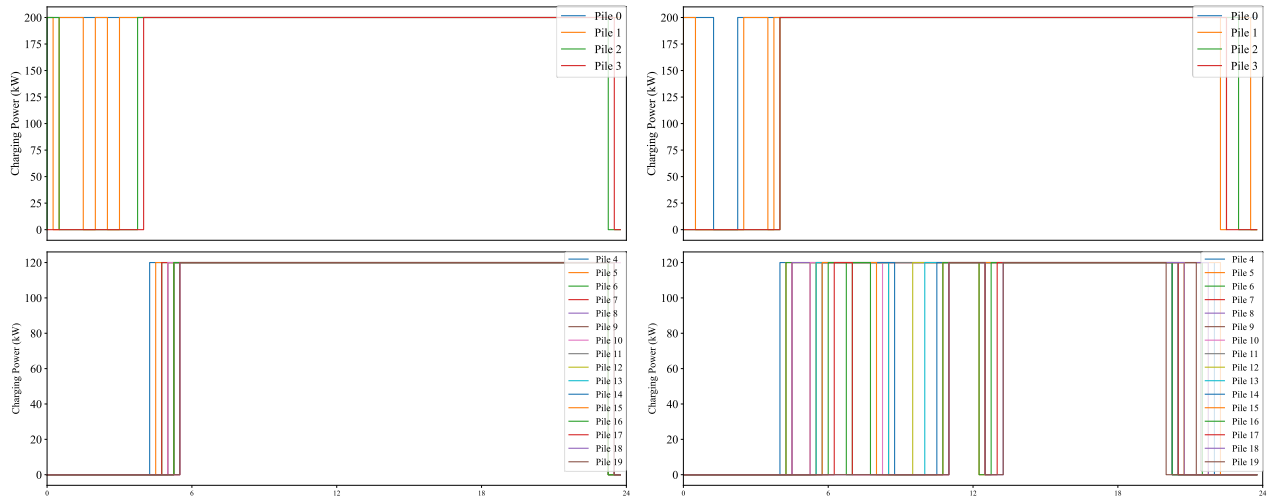
c TD3 algorithm

d The proposed method

Fig. 14. Number of arrivals, departures, queue, spillage and charging piles in use

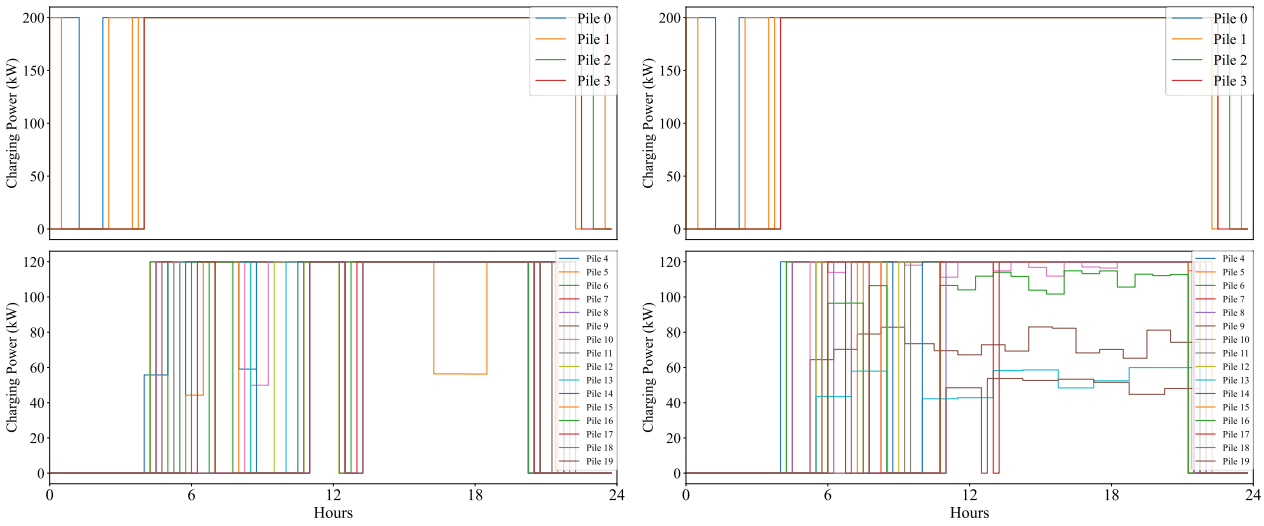
In the intra-day stage, the number of arrivals, departures, queue, spillage and charging piles in use with or without price incentives are shown in Fig. 14. Due to the difference between day-ahead (1h) and intra-day (15 min) time scales, and the day-ahead calculation is based on the predicted data of traffic flow while the intra-day calculation is based on the actual data, the number curves in Fig. 13 and Fig. 14 are different. As can be seen from Fig. 14, there is no full parking of charging station with optimal price incentives of latter three methods.

The charging power of charging piles with different methods is shown in Fig. 15.



a Average allocation with the PV-WP-ES

b Optimal method based on the day-ahead price incentives



c TD3 algorithm d The proposed method

Fig. 15. Charging power with different methods

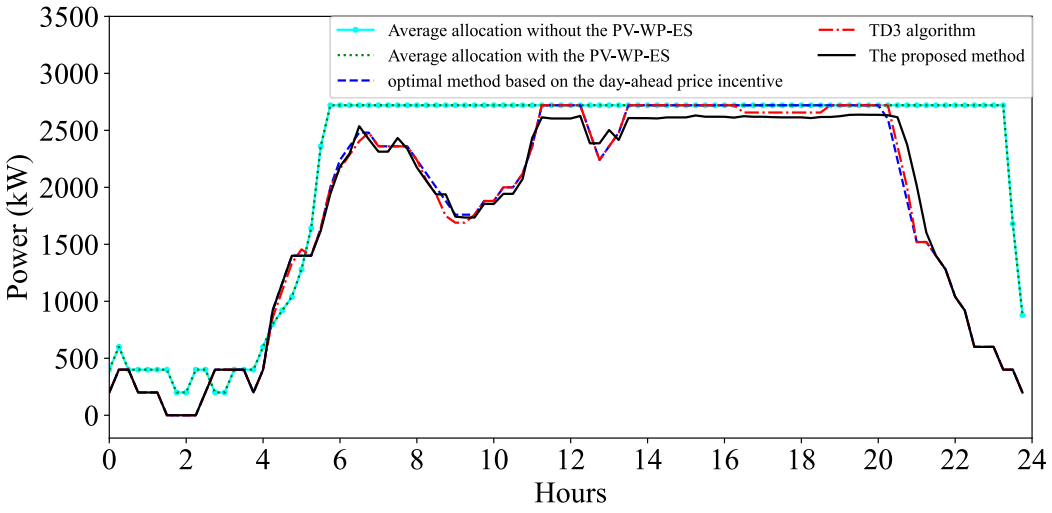


Fig. 16. Total charging power with different methods

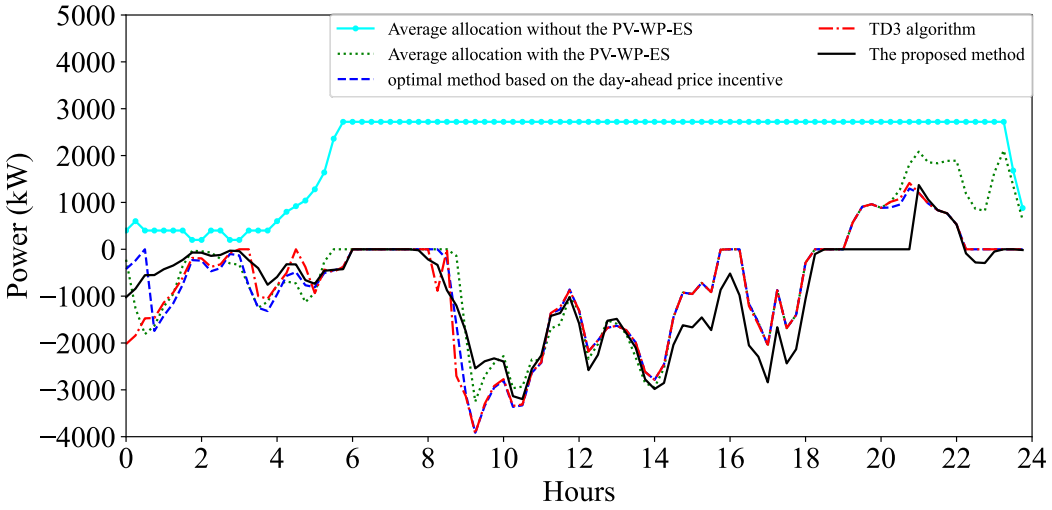


Fig. 17. Grid-connected power with different methods

The total charging power and grid-connected power with different methods are shown in Fig. 16 and Fig. 17. If the power on the grid-connected side is greater than 0, the power is purchased from the grid; if the power is less than 0, the power is transmitted to the grid. As can be seen from Fig. 17, at 19~21h, the grid-connected power of the proposed method is zero, and electricity isn't purchased from the grid. The grid-connected power of the other methods is greater than zero, and buy electricity from the grid. It shows that the proposed method can better use the output of the charging station to fill its demand and reduce the external input. At 14~19h and 22~23h, the grid-connected power curve of other methods is higher than that of the proposed method, which indicates that the proposed method can output more clean energy to the grid.

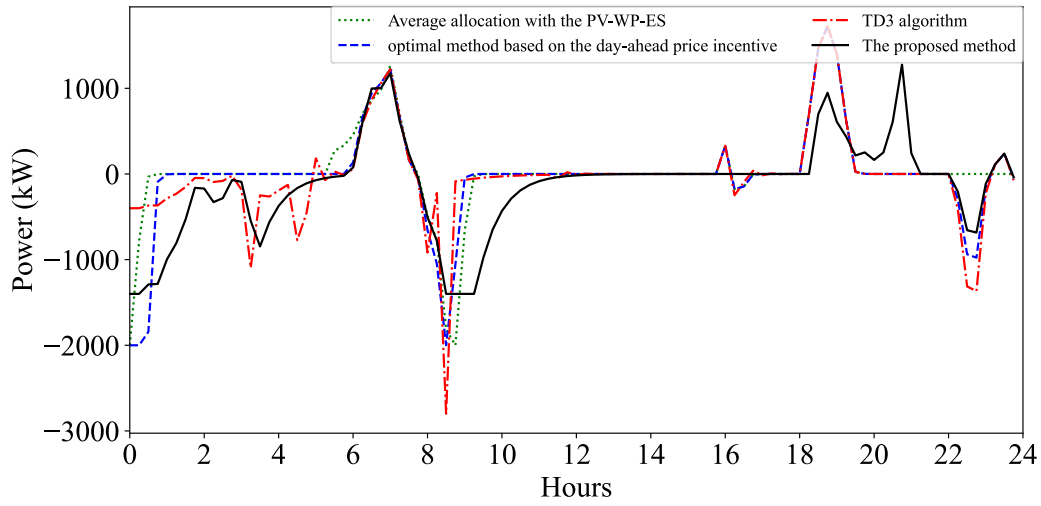


Fig. 18. Charging power of the ESS with different methods

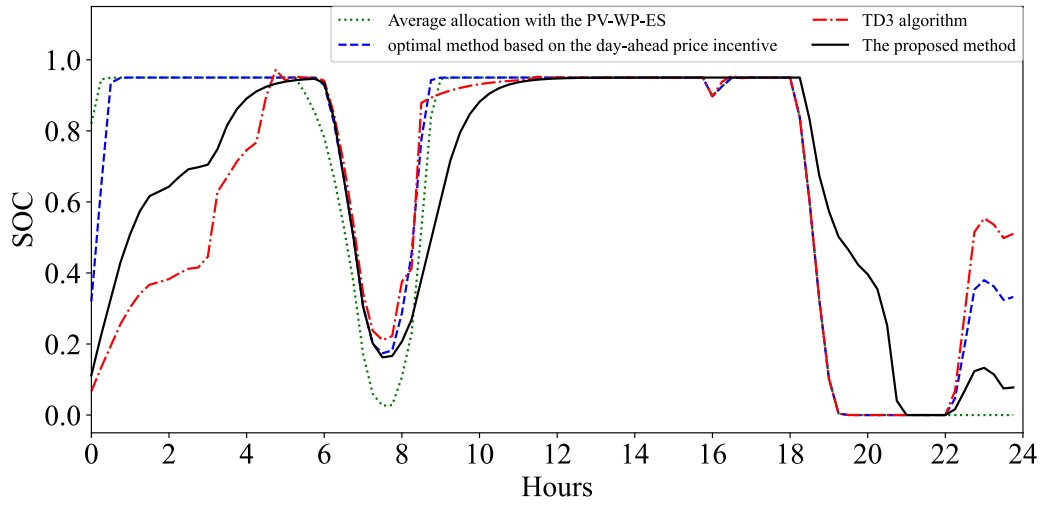


Fig. 19. SOC of the ESS with different methods

Table.5 Overall evaluation indicators

methods	K_{sum}/CNY	$C_{loss,s}$/CNY	Total revenue/CNY	S_{pw}	E_c/kg
Average allocation without the PV-WP-ES	9386	0	9686	0	36840
Average allocation with the PV-WP-ES	37767	10691	27076	87.72%	-10462
Optimal method based on the day-ahead price incentives	64366	10897	53469	93.58%	-14190
TD3 algorithm	64571	2123	62448	93.55%	-14821
The proposed method	58323	2109	56214	96.71%	-16245

The power and SOC of the ESS under different methods are shown in Fig. 18 and 19. The overall evaluation indicators are shown in Table 5. Combined with Table 5 and Fig. 18, it can be seen that compared with other methods, the total revenue of Average allocation without the PV-WP-ES is only 9686 CNY, and the income of traditional highway charging stations is only the difference between charging fee and purchase electricity fee, with a low return rate. At the same time, it can be seen from Fig. 18 that the operation of traditional highway charging stations without the PV-WP-ES relies on the power grid. In the weak highway power grid, this method has weak anti-risk ability. From several other approaches, highway charging stations with the PV-WP-ES make up for this deficiency. The total revenue of the proposed method based on the MATD3 =56214 CNY, is significantly greater than that of other methods, and the revenue of charging stations is greater. At the same time, the self-consumption rate of S_{pw}=96.71% of the proposed method is higher than other methods, and the self-production and self-marketing ability is stronger, and the dependence on the power grid is weaker. Therefore, the proposed method based on optimal decision of the MATD3 can better utilize the scheduling EV charging and the charging and discharging of the ESS to achieve self-consumption.

It can be seen from Fig. 19 and Table 5 that the SOC curve of the ESS of the proposed method based on the MATD3 is at the boundary for a shorter time than that of other methods at 9~12h and 21~22h. At the same time, among other methods, the battery loss cost of the proposed method C_{loss}=2109 CNY is the smallest, which indicates that the proposed method tries to control the ESS to avoid over-discharge and reduce the loss of the battery. As can be seen from Table 5, among other methods, the proposed method based on the MATD3 has the smallest carbon emission E_c=-16245kg, which indicates that the proposed method based on the MATD3 realizes the carbon emission in the station is 0, and provides clean energy for the power grid, reducing the external carbon emission by 16245kg. Although the total revenue of other methods is higher than that of the proposed method, the higher part is from the charging income, but the difference is small. For example, the total revenue of TD3 and the proposed method is only 9% difference. Meanwhile, the two indicators of carbon emission and self-consumption rate are better than other methods. The reason is that the proposed method based on the MATD3 sacrifices a small portion of charging revenue through multi-agent games, in exchange for the improvement in carbon emissions and self-consumption rate, reducing dependence and pressure on the power grid and achieving global optimality.

In summary, the proposed method based on the MATD3 algorithm transforms the original complex multi-objective nonlinear non-convex problem into a multi-agent MGM model, greatly reducing the complexity of the optimization model and helping the solving algorithm to find the optimal strategy. At the same time, the proposed method based on the MATD3 algorithm finds the globally optimal solution that balances self-consumption, low carbon, economy, and damage reduction through the game between different charging piles and ESS.

To verify the advantages of the proposed method in terms of computational efficiency, it is compared with the WOA, centralized MATD3, and agent modeling based on the number of objects. The decision time with different methods is shown in Table 6.

Table.6 Decision time with different methods

Methods	WOA	Centralized MATD3	Agent modeling based on the number of objects	The proposed method
Decision time/s	1026.437	2.023	10.142	1.781

As can be seen from Table 6, the decision time =1026.437s of the WOA population-based optimization algorithm is significantly longer than that of these MADRL algorithms. In the intra-day stage, the population-based optimization algorithms must continuously iterate to find the optimal value. The MADRL algorithm makes decisions directly from the learned policy, without taking the learning process into account. Therefore, reinforcement learning has higher computational efficiency. The decision time of the proposed method =1.781s is shorter than that of other methods. This shows that the proposed method improves the computational efficiency of the algorithm by reducing the number of agents based on the type of control object and a centralized training and

decentralized execution mode.

5 Conclusion

Aiming at the existing problems of charging station scheduling, an optimal self-consumption scheduling of the renewable energy charging station on highways based on the MATD3 is proposed. The PV, WP, and ESS are combined with the charging station to reduce the dependence and pressure on the power grid. The power of the charging piles and ESS are flexibly scheduled by MATD3 to realize self-sustainability and self-consumption of the renewable energy charging station on highways. By MATD3's ability to deal with the randomness of the environment and multi-agent cooperation, the complex multi-agent and multi-objective scheduling non-convex problems in the charging station are solved, and the generalization and solving ability of the system is improved. Through simulation experiments, compared with other methods, the self-consumption rate of the proposed method can reach 96.71%, and the carbon emission can reach -16245 kg. The proposed method can efficiently utilize the resources in the station to achieve economy and low-carbon operation and resist risk ability. Future research will focus on how to ensure system stability and security in the trial-and-error training process of deep reinforcement learning, and how to maximize the training effect of agents at acceptable costs of trial-and-error.

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.

Acknowledgement

This work was supported in part by the Sichuan Science and Technology Program (2024YFHZ0312).

Reference

[1] Hyeon Woo, Yongju Son, Jintae Cho, Sung-Yul Kim, Sungyun Choi, Optimal expansion planning of electric vehicle fast charging stations, Applied Energy, 2023, 342: 121116,

[2] H. Jinglin et al. Planning of Electric Vehicle Charging Station on Highway Considering Existing Service Areas and Dynamic Traffic Simulations, 2018 China International Conference on Electricity Distribution (CICED), Tianjin, China, 2018, pp. 2645-2649.

[3] H. Yao, Y. Xiang, C. Gu and J. Liu, Peer-to-Peer Coupled Trading of Energy and Carbon Emission Allowance: A Stochastic Game-Theoretic Approach, IEEE Internet of Things Journal, 2023(Early access).

[4] Y. Xiang et al., Distributionally robust expansion planning of electric vehicle charging system and distribution networks, CSEE Journal of Power and Energy Systems.

[5] Yan, Dongxiang, and C. Ma. Stochastic planning of electric vehicle charging station integrated with photovoltaic and battery systems, IET Generation Transmission & Distribution, 2020, 14(2): .

[6] K. Chaudhari, A. Ukil, K. N. Kumar, U. Manandhar and S. K. Kollimalla, Hybrid Optimization for Economic Deployment of ESS in PV-Integrated EV Charging Stations, IEEE Transactions on Industrial Informatics, 2018, 14(1): 106-116.

[7] M. Fang et al., Data-Driven Load Pattern Identification Based on R-Vine Copula and Random Forest Method, IEEE Transactions on Industry Applications, 2022, 58(6): 7919-7929.

[8] Q. Yan, B. Zhang and M. Kezunovic, Optimized Operational Cost Reduction for an EV Charging Station Integrated With Battery Energy Storage and PV Generation, IEEE Transactions on Smart Grid, 2019, 10(2): 2096-2106.

[9] N. A. El-Taweel, H. Farag, M. F. Shaaban and M. E. AlSharidah, Optimization Model for EV Charging Stations

With PV Farm Transactive Energy, IEEE Transactions on Industrial Informatics, 2022, 18(9): 4608-4621.
[10] Jiawei Feng, Shengya Hou, Lijun Yu, Optimization of photovoltaic battery swapping station based on weather/traffic forecasts and speed variable charging, Applied Energy, 2020, 264: 114708.
[11] J. Zhang et al., A Hierarchical Distributed Energy Management for Multiple PV-Based EV Charging Stations, IECON 2018 - 44th Annual Conference of the IEEE Industrial Electronics Society, Washington, DC, USA, 2018, pp. 1603-1608.
[12] S. Li, H. Wu, X. Bai and S. Yang, Optimal Dispatch for PV-assisted Charging Station of Electric Vehicles, 2019 IEEE PES GTD Grand International Conference and Exposition Asia (GTD Asia), Bangkok, Thailand, 2019, pp. 854-859.
[13] G. Yu, X. Zhang, H. Wang, Y. Zheng, E. Chung and R. Zhu. Low Carbon Planning of PV-Charging Stations for Self-Sustained Highway Transportation Energy System Considering the Retirement of Gas Stations. IEEE Transactions on Industry Applications, vol. 60, no. 1, pp. 1208-1218, Jan.-Feb. 2024.
[14] Ahmed H. Hammam, Mohamed A. Nayel, Mansour A. Mohamed, Optimal design of sizing and allocations for highway electric vehicle charging stations based on a PV system, Applied Energy, Volume 376, Part B, 2024, 124284.
[15] K. He, H. Jia, Y. Mu, X. Yu, X. Dong and Y. Deng. Coordinated Planning of Fixed and Mobile Charging Facilities for Electric Vehicles on Highways. IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 9, pp. 10087-10098, Sept. 2023.
[16] T. -Y. Zhang, Y. Yang, Y. -T. Zhu, E. -J. Yao and K. -Q. Wu. Deploying Public Charging Stations for Battery Electric Vehicles on the Expressway Network Based on Dynamic Charging Demand. IEEE Transactions on Transportation Electrification, vol. 8, no. 2, pp. 2531-2548, June 2022.
[17] Y. Zhang, Z. Yin, H. Xiao and F. Luo. Coordinated Planning of EV Charging Stations and Mobile Energy Storage Vehicles in Highways With Traffic Flow Modeling. IEEE Transactions on Intelligent Transportation Systems.
[18] A. Gusrialdi, Z. Qu and M. A. Simaan. Distributed Scheduling and Cooperative Control for Charging of Electric Vehicles at Highway Service Stations. IEEE Transactions on Intelligent Transportation Systems, vol. 18, no. 10, pp. 2713-2727, Oct. 2017
[19] Giuseppe Napoli, Antonio Polimeni, Salvatore Micari, Laura Andaloro, Vincenzo Antonucci, Optimal allocation of electric vehicle charging stations in a highway network: Part 1. Methodology and test application, Journal of Energy Storage, Volume 27, 2020, 101102.
[20] X. Wang, J. Zhou, B. Qin and L. Guo, Coordinated Power Smoothing Control Strategy of Multi-Wind Turbines and Energy Storage Systems in Wind Farm based on MADRL, IEEE Transactions on Sustainable Energy, 2023(Early access).
[21] Yue Xiang, Yu Lu, Junyong Liu, Deep reinforcement learning based topology-aware voltage regulation of distribution networks with distributed energy storage, Applied Energy, 2023, 332: 120510.
[22] Ruiyang Jin, Yuke Zhou, Chao Lu, Jie Song, Deep reinforcement learning-based strategy for charging station participating in demand response, Applied Energy, 2022, 328: 1201400.
[23] Kang Wang, Haixin Wang, Zihao Yang, Jiawei Feng, et al. A transfer learning method for electric vehicles charging strategy based on deep reinforcement learning, Applied Energy, 2023, 343: 121186.
[24] M. Wang, J. Cheng and H. Zhai. Life Prediction for Machinery Components Based on CNN-BiLSTM Network and Attention Model. 2020 IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, 2020, pp. 851-855.
[25] G. Wang, X. Ji, B. Zhou, H. Li and H. Wang. A radial basis function based approach for electric vehicle charging load forecasting. The 11th IET International Conference on Advances in Power System Control, Operation and Management (APSCOM 2018), Hong Kong, China, 2018, pp. 1-5.

[26] Khojasteh Eghbali, S., Mousavi, S. M., & Salimian, S.. Designing blood supply chain networks with disruption considerations by a new interval-valued fuzzy mathematical model: M/M/C queueing approach. Computers and Industrial Engineering, 2023, 182: 109260.
[27] Liu D, Zhang T, Wang W, Peng X, Liu M,Jia H, Su S. Two-Stage Physical Economic Adjustable Capacity Evaluation Model of Electric Vehicles for Peak Shaving and Valley Filling Auxiliary Services, Sustainability, 2021, 13(15):8153.
[28] Sierzchula, W.; Bakker, S., et al.The influence of financial incentives and othersocio-economic factors on electric vehicle adoption, Energy Policy, 2014, 68: 183-194.
[29] Fujimoto S, Hoof H V, Meger D. Addressing Function Approximation Error in Actor-Critic Methods. arXiv preprint, 2018.
[30] Ackermann J , V Gabler, Osa T, et al. Reducing overestimation bias in multi-agent domains using double centralized critics. arXiv preprint, 2019.