energies

MDPI

Article

Adaptive Optimization Operation of Electric Vehicle Energy Replenishment Stations Considering the Degradation of Energy Storage Batteries

Yuhang Bai and Yun Zhu *

School of Electrical Engineering, Guangxi University, Nanning 530004, China; 2112392003@st.gxu.edu.cn
* Correspondence: zhuyun@gxu.edu.cn

Abstract: As the construction of supporting infrastructure for electric vehicles (EV) becomes more and more perfect, an energy replenishment station (ERS) involving photovoltaics (PV) that can provide charging and battery swapping services for electric vehicle owners comes into the vision of humanity. The operation optimization of each device in the ERS is conducive to improving the service capacity of the ERS, extending the service life of the energy storage batteries (ESB), and enhancing the economic benefits of the ERS. However, traditional model-based optimization algorithms cannot fully consider the stochastic nature of EV owners' charging and battery swapping demands, the uncertainty of PV output, and the complex operating characteristics of ESB. Therefore, we propose a deep reinforcement learning-based adaptive optimal operation method for ERS considering ESB's losses. Firstly, a mathematical model of each device in the ERS is established, and a refined energy storage model is introduced to describe ESB's capacity degradation and efficiency decay. Secondly, to solve the dimensional disaster problem, the state space and action space selection method, and the charging strategy of batteries in the battery swapping station (BSS) are proposed to apply to the ERS, thus modeling the ERS optimization operation problem as a Markov decision process. Then, the solution is performed using a neural network-based proximal policy optimization (PPO) algorithm, consisting of a recurrent neural network that extracts information about the PV outflow trend and a deep neural network used to generate a control policy. Finally, the effectiveness of the proposed method is verified by simulation calculations, which not only enable adaptive decision-making under different PV output scenarios, but also consider the availability of EV battery swapping services, energy storage losses, and the economic benefits of the ERS.

Keywords: electric vehicles; energy replenishment station; deep reinforcement learning; adaptive optimized operation energy storage; uncertainty

check for updates

Citation: Bai, Y.; Zhu, Y. Adaptive Optimization Operation of Electric Vehicle Energy Replenishment Stations Considering the Degradation of Energy Storage Batteries. *Energies* **2023**, *16*, 4879. https://doi.org/10.3390/en16134879

Academic Editors: Surender Reddy Salkuti and Brian Azzopardi

Received: 30 May 2023
Revised: 16 June 2023
Accepted: 20 June 2023
Published: 22 June 2023

Copyright: © 2023 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 (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, transportation electrification has been considered an effective measure to achieve energy saving and emission reduction, and improve energy utilization efficiency [1]. Thanks to both policy and market stimulation, the number of electric vehicles in China exceeds 10.45 million by the end of 2022, and the resulting diversified demand for energy replenishment poses a new challenge to the construction of electric vehicle replenishment infrastructure in China. ERS is an integrated power facility that provides charging and battery swapping services for electric vehicle owners. It integrates photovoltaic power generation equipment and ESB, battery swapping equipment, and DC charging posts. In addition to purchasing electricity from the higher grid, the ERS can also provide charging and battery swapping services for EV users through clean energy generated by photovoltaic power generation equipment and the electricity stored in ESB [2–4]. Compared with traditional charging stations and battery swapping stations, ERS has many advantages, such as broad service categories, high functional integration, low carbon, and environmental protection. However, in practical application, the fluctuation of PV equipment output and

the uncertainty of EV owners' replenishment demand significantly increase the difficulty of real-time control of energy storage equipment and battery swapping equipment in the ERS [5].

Scholars from different perspectives have studied the above issues. Regarding the optimal operation of battery swapping equipment, the authors in [6] proposed a charging strategy that maximizes PV self-consumption, while ensuring the quality of BSS service based on the predicted values of battery swapping demand. In [7], the differences in battery charging power and capacity within the BSS are further considered, and the optimal scheduling strategy is solved using mixed integer linear programming. In [8], EV owners are considered a link in the battery exchange chain, and a linear programming-based charging strategy solution method is proposed. It should be noted, however, that implementing the above approach is highly dependent on the prediction of uncertainty in EV battery swapping demand, and may yield sub-optimal results in practical application scenarios.

In modeling and capacity estimation of battery energy storage systems in ERS, the literature [9,10] set the energy storage charging/discharging efficiency and energy storage capacity to constant values without considering the capacity degradation of ESB and the decrease of charging/discharging efficiency. In [11], a binomial fit was made to the operating efficiency of an energy storage device consisting of multiple battery cells to obtain a relationship function between charge and discharge efficiency and battery SOC. In [12], a hybrid battery model based on circuit and analytical battery models was proposed to enable accurate tracking and lifetime prediction of the battery state of charge (SOC). Literature [13] proposed a semi-empirical capacity decline model for lithium batteries under irregular operation by considering the lithium battery charge and discharge power, SOC, operating temperature, depth of discharge (DOD), and operating time. Literature [14] applied the above-refined energy storage model to the degradation estimation of short-term ESB and used deep reinforcement learning algorithms to improve the full life-cycle benefits of energy storage stations substantially.

However, the above studies focus on estimating the residual capacity of energy storage devices and optimizing BSS charging strategies. Few scholars have conducted academic research on ERS as a whole. It is worth noting that BSS can not only provide battery swapping service for EVs, but also has strong energy storage characteristics and the potential to participate in PV consumption and power peak-shaving. Suppose we can realize the information interoperability, cooperative control, and unified decision of battery swapping equipment, PV power generation equipment, energy storage equipment, and EV replenishment demand. In that case, it will further enhance the self-consumption rate of PV, improve the economic efficiency of ERS, and complete the construction of EV-related infrastructure.

In this paper, we focus on the real-time operation strategy of energy storage devices and battery swapping equipment in ERS and propose a deep reinforcement learning-based method for the optimal real-time operation of ERS. Firstly, a mathematical model of ERS considering the type of electric vehicle and the degradation of ESB capacity is established; secondly, the state space, action space, and reward function are determined for the ERS real-time scheduling problem; then, a PPO-based method for solving the ERS optimal operation problem is constructed. In addition, solar irradiation intensity time-series data is introduced as input, and a long short-term memory network (LSTM) is used to assist the agent in decision-making. Simulations show that the proposed method in this paper can achieve adaptive decision-making in a more stochastic scenario, and has significant advantages in a scenario where the three factors of battery swapping service availability, ESB degradation, and energy station economic efficiency are considered.

2. Mathematical Model for Electric Vehicles Energy Replenishment Station

2.1. ERS Composition and Application Scenario

ERS is a public service place integrating charging and battery swapping services, which are mainly composed of photovoltaic power generation equipment, DC fast charging piles, battery swapping station, energy storage equipment, and AC/DC conversion components; its structure schematic is shown in Figure 1.

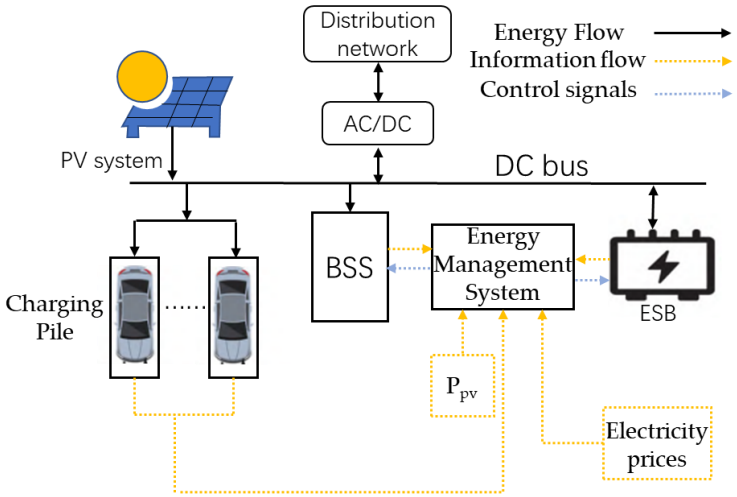


Figure 1. Structure of the ERS.

The energy management system (EMS) collects PV output information, tariff information, charging pile load information, BSS operating conditions, and ESB's status information in real-time and issues dispatch orders to energy storage equipment and BSS to maximize ERS revenue while taking into account EV users' energy replenishment demand and energy storage capacity degradation. Considering that the highway service area is located in an open area and has the basic conditions for installing PV systems and energy storage equipment, this paper uses the ERS in a highway service area as a specific scenario to conduct the research.

2.2. Objective Function

In order that the scheduling strategy of the EMS can maximize the economic benefits of ERS while taking into account the replenishment demand of EV users and the capacity degradation of ESB, this paper introduces the degradation cost C_{BESS} of energy storage batteries and the battery replacement waiting penalty C_{BSS} into the objective function, which constitutes an ERS operation optimization model to maximize the comprehensive benefits.

$$f = \max \sum_{t=1}^{T} \left(B_{FC,t} + B_{BSS,t} - C_{grid} - C_{BESS} - C_{BSS} \right) \tag{1}$$

where, $B_{FC,t}$ and $B_{BSS,t}$ are the revenue obtained by ERS from providing charging and battery replacement services to EV customers, respectively, C_{grid} is the cost of interaction between the ERS and the grid, C_{BESS} is the cost of ESB degradation, and C_{BSS} is the battery exchange waiting penalty. T is the optimization period considered in this paper to ensure the real-time optimization operation and combine the battery replacement time of the current mainstream BSS. The EMS outputs control signals with a time interval of $\Delta t = 6\text{min}$, $T = 240$, i.e., one day is divided into 240 time periods. Many variables are involved in the above cost calculation, so this paper will start the modeling analysis of the above variables one by one in the following subsections.

2.3. Calculation Model of Charging Revenue $B_{FC,t}$

Considering that the ERS is deployed in a highway service area, the primary objective for the EV user is to quickly recharge the EV and then continue the rest of the trip. Therefore, it is considered that the fast charging power $P_{ev,i}$ remains constant from the moment the EV is connected to the fast charger until the end of charging. Note that the charging power of

the EV depends not only on the maximum charging power available from the fast charger, but also on the maximum charging power acceptable to the vehicle [15].

$$\begin{cases} P_{ev,i} = P_{st}\ , P_{ev,i}^{\max} \geq P_{st} \\ P_{ev,i} = P_{ev,i}^{\max}\ , P_{ev,i}^{\max} < P_{st} \end{cases} \tag{2}$$

where $P_{ev,i}$ is the actual charging power of the ith EV, P_{st} and $P_{ev,i}^{\max}$ are the maximum charging power that can be provided by the charging piles and the maximum charging power that can be accepted by the EV, respectively. If all the fast-charging piles are occupied, the vehicle owner has to wait for other users to finish charging before charging. When the battery SOC approaches 0.9, the fast-charging power will drop rapidly, and to avoid overcharging damage to the battery, the upper limit of battery SOC is limited to 0.9, i.e., after the battery SOC reaches 0.9, the charging pile will no longer provide power to the EV and the EV charging will be finished. EMS calculates the total fast-charging load by adding up the charging power of fast-charging piles at time t.

$$P_{ev,sum}^{t} = \sum_{i=1}^{n} P_{ev,i}^{t} \tag{3}$$

where n is the total number of fast-charging piles at the ERS, and $P_{ev,sum}^{t}$ is the total fast-charging load. Then, the charging service revenue $B_{FC,t}$ at time t can be expressed as:

$$B_{FC,t} = P_{ev.sum}^{t} \cdot E_{ev,t} \cdot \Delta t \tag{4}$$

$E_{ev,t}$ is the electricity purchase price for EV customers at the time t. This price consists of two parts: the energy station purchase price and the service charge, i.e., $E_{ev,t} = E_{c,t} + E_s$ where E_s is the service charge per unit of electricity.

2.4. Calculation Model of Battery Swapping Service Revenue $B_{BSS,t}$

Referring to the current mainstream commercial BSS operation mode, this paper assumes that the BSS in ERS adopts the single-channel battery swapping mode, i.e., the exchange station only provides the battery swapping service to one EV at a time, and the number of chargers in the BSS is the same as the number of batteries. The total number of batteries in the BSS is constant M at any time. The possible states of batteries in the BSS are fully charged batteries (FB), pending batteries (PB), and charging batteries (CB). The battery conversion relationship is shown in Figure 2 and is as follows:

$$\begin{cases} N_{t+1}^{f} = N_{t}^{f} + \Delta N_{t}^{cf} - \Delta N_{t}^{fh} \\ N_{t+1}^{c} = N_{t}^{c} + \Delta N_{t}^{pc} - \Delta N_{t}^{cf} \\ N_{t+1}^{p} = N_{t}^{p} + \Delta N_{t}^{hp} - \Delta N_{t}^{pc} \\ N_{t}^{p} + N_{t}^{c} + N_{t}^{f} = M \end{cases} \tag{5}$$

where N_t^f, N_t^c, N_t^p, and ΔN_t^h are the number of FB, CB, PB, and batteries participating in battery swapping service, respectively, and $\Delta N_t^{kk'}$ denotes the number of batteries that have moved from state k to state k' in the period t. $k, k' \in \{f, c, p, h\}$. where f, c, p and h indicate the four states of the battery in the battery swapping station, i.e., fully charged, charging, waiting to be charged, and being swapped. In order to avoid damage to the battery by overcharging, the upper limit of battery SOC is limited to 0.9., i.e., the battery SOC charged to 0.9 is considered a fully charged battery. The formula for calculating the SOC of the battery in the charging stage is as follows:

$$S_{t+1}^{BSS,i} = S_{t}^{BSS,i} + P_{BSS} \cdot \Delta t \cdot {\eta_c}^{BSS} / C^{BSS} \tag{6}$$

where $S_t^{BSS,i}$ is the SOC of the ith battery in the substation at time t, P_{BSS}, ${\eta_c}^{BSS}$, and C^{BSS} are the charging power of the charger, charging efficiency, and battery capacity, respectively. In this paper, assuming that the chargers in the BSS are in constant power charging mode, the total electrical load at time t of the BSS is as follows:

$$P_{BSS,sum}^t = P_{BSS} \cdot {n_t}^{BSS} \tag{7}$$

where $P_{BSS,sum}^t$ is the total load at time t and n_t^{BSS} is the number of chargers charging at time t. The benefits to BSS from providing battery swapping services $B_{BSS,t}$ are as follows:

$$B_{BSS,t} = (0.9 - \Delta N_t^h) \cdot C^{BSS} \cdot E_{ev,t} \cdot \Delta t \tag{8}$$

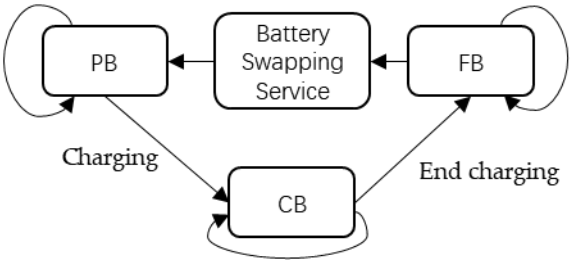


Figure 2. Battery conversion relationship.

2.5. Model for Calculating the Cost of Interaction with the Grid C_{grid}

The ERS operation is subject to the following power balance constraints:

$$P_{pv,t}^* + P_{grid,t} = P_{ev,sum}^t + P_{BSS,sum}^t + P_{BESS}(t) \tag{9}$$

where $P_{pv,t}^*$ is the PV panel power output at the moment t, $P_{ev,sum}^t$ and $P_{BSS,sum}^t$ are calculated by Equations (3) and (7), respectively, $P_{BESS}(t)$ is the power output of the energy storage battery at time t. This value is output by EMS control as part of the action space, so the energy station and grid interaction power $P_{grid,t}$ can be calculated by Equation (9). The ERS interaction cost with the upper grid is as follows:

$$C_{grid} = \begin{cases} E_{c,t} \cdot P_{grid} \cdot \Delta t P_{grid,t} \geq 0 \\ E_{d,t} \cdot P_{grid} \cdot \Delta t P_{grid,t} < 0 \end{cases} \tag{10}$$

If the P_{grid} is positive, it means that ERS purchases electricity from the grid, and the purchase price is $E_{c,t}$. If it is negative, it means that ERS feeds electricity to the grid, and the feed price is $E_{d,t}$.

2.6. Energy Storage Battery Degradation Model

The degradation of ESB can manifest as charge and discharge efficiency degradation and battery capacity decay. This paper introduces a refined energy storage model to model and analyze the cost generated by the two degradation parts separately.

2.6.1. Charge/Discharge Efficiency Model

In this paper, lithium-ion batteries are used as energy storage batteries, and the steady-state circuit diagram of its minimum energy storage unit is shown in Figure 3:

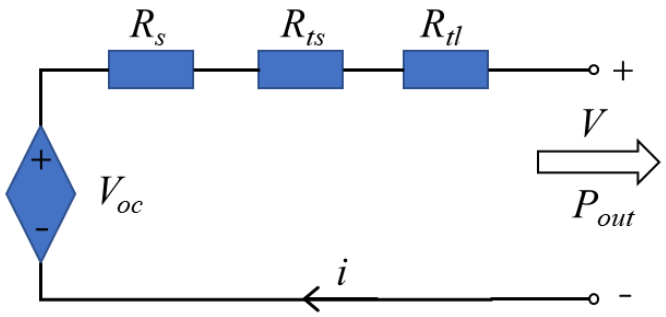


Figure 3. Steady-state equivalent circuit of the energy storage unit.

The circuit consists of an open circuit voltage V_{oc} and three equivalent resistors R_s, R_{ts}, and R_{tl},which represent different chemical processes: ohmic loss, charge transfer, and membrane diffusion. All the above four variables have a nonlinear relationship with the energy storage SOC [11].

$$\begin{cases} V_{oc} = a_0 e^{(-a_1 S_b)} + a_2 + a_3 S_b - a_4 S_b^2 + a_5 {S_b}^3 \\ R_s = b_0 e^{(-b_1 S_b)} + b_2 + b_3 S_b - b_4 {S_b}^2 + b_5 {S_b}^3 \\ R_{ts} = c_0 \cdot e^{-c_1 \cdot S_b} + c_2 \\ R_{tl} = d_0 \cdot e^{-d_1 \cdot S_b} + d_2 \\ R_{tot} = R_s + R_{ts} + R_{tl} \end{cases} \tag{11}$$

where $a_0 \cdots a_5, b_0 \cdots b_5, c_0 \cdots c_2, d_0 \cdots d_2$ are fitting coefficients, S_b is the energy storage battery SOC, and R_{tot} is the sum of equivalent resistance. Assuming that the internal energy storage system has a balanced circuit and the batteries are wired in series and parallel, the total battery output current can be expressed as $i_{out} = i \cdot N_{parai}$ and the total battery output voltage as $V_{out} = V \cdot N_{seriesi}$, where N_{parai} and $N_{seriesi}$ are the number of parallel and series batteries, respectively. Where i and V are the single battery current and output voltage, respectively. From Figure 3, it can be seen that the single-cell current i can be solved by (12) to obtain

$$i^2 - i\frac{V_{oc}(S_b)}{R_{tot}(S_b)} + \frac{1}{N_{seriesi} N_{parai}} \cdot \frac{P_{BESS}}{R_{tot}(S_b)} = 0 \tag{12}$$

$$\begin{cases} \eta^{ch} = \frac{V_{oc}(S_b)}{V_{oc}(S_b) - i \cdot R_{tot}(S_b)} \\ \eta^{dis} = \frac{V_{oc}(S_b) - i \cdot R_{tot}(S_b)}{V_{oc}(S_b)} \end{cases} \tag{13}$$

In simultaneous Equations (12) and (13), the relationship between battery charge/discharge efficiency and output power and SOC can be obtained, as shown in Equation (14) [11].

$$\begin{cases} \eta^{ch} = e_0 + e_1 S_b + e_2 {S_b}^2 + e_3 P_{BESS} + e_4 P_{BESS}^2 + e_5 P_{BESS} S_b \\ 1/\eta^{dis} = h_0 + h_1 S_b + h_2 {S_b}^2 + h_3 P_{BESS} + h_4 P_{BESS}^2 + h_5 P_{BESS} S_b \end{cases} \tag{14}$$

where η^{ch} and $1/\eta^{dis}$ are the charging and discharging efficiency of the ESB, respectively, and $e_0 \cdots e_5, h_0 \cdots h_5$ are the fitting coefficients. S_b is the energy storage battery SOC. For ease of expression, the charging and discharging efficiency η_{BESS} of the ESB is expressed as shown in (15):

$$\eta_{BESS} = \begin{cases} \eta^{ch}, P_{BESS}(t) < 0 \\ \frac{1}{\eta^{dis}}, P_{BESS}(t) \geq 0 \end{cases} \tag{15}$$

where $P_{BESS}(t)$ is the charging and discharging power of the ESB at the moment t, $P_{BESS}(t)$ is positive for discharging and $P_{BESS}(t)$ is negative for charging. In summary, the energy storage battery SOC at time t can be expressed as:

$$S_b(t) = S_b(t-1) - \frac{\eta_{BESS} P_{BESS}(t) \Delta t}{Q(t)} \tag{16}$$

where $S_b(t)$ and $S_b(t-1)$ are the SOC of the ESB at moments t and $t-1$, respectively. $Q(t)$ is the capacity of the ESB at the time t, Δt is the time interval between EMS output control signals, η_{BESS} is the electric energy storage charging and discharging efficiency coefficient. The decrease in ESB charging and discharging efficiency impacts ERS's power purchase cost in the form of a change in the charging and discharging efficiency coefficient η_{BESS}, which is implicitly included in the transaction cost with the upper grid. Therefore, there is no additional penalty item.

2.6.2. Energy Storage Battery Capacity Degradation Model

The degradation process of energy storage capacity is a nonlinear process related to the number of battery cycles and usage time. In order to accurately estimate the capacity degradation of energy storage batteries under non-uniform charge and discharge environments, a semi-empirical capacity degradation model is used in this paper [13].

$$\begin{cases} L = 1 - \alpha_{sei} e^{-\beta_{sei} f_d N} - (1 - \alpha_{sei}) e^{-f_d N}, L = 0 \\ L = 1 - (1 - L') e^{-N f_d}, L \neq 0 \end{cases} \tag{17}$$

In (17), L indicates the percentage of capacity lost in the current cycle to the rated maximum capacity of the battery, e.g., when $L = 0$ it means a brand new battery with no degradation at all and $L = 0.1$ means the remaining capacity of the battery is only 90% of the rated maximum capacity. In this paper, $L = 0.2$ is considered as the end of life of the energy storage device. L' is the ratio of the lost capacity of the battery to its rated maximum capacity before the cycle. α_{sei}, β_{sei} are the constant coefficients related to the formation of the solid electrolyte mask of the new battery, and N is the number of cycles. f_d is the decay function of the battery capacity in a single cycle, and its value is related to the temperature T_C, time t, charge state S_b, and depth of discharge δ, and its expression is shown in (18):

$$f_d = [f_\delta(\delta) + f_t(t_c)] f_{soc}(S_b) f_T(T_C) \tag{18}$$

$$\begin{cases} f_\delta(\delta) = \left(k_{\delta 1} \delta^{k_{\delta 2}} + k_{\delta 3}\right)^{-1} \\ f_t(t) = k_t t \\ f_T(T_C) = e^{k_T (T_C - T_{ref}) \cdot \frac{T_{ref}}{T_C}} \\ f_{soc}(S_b) = e^{k_{soc}(S_b - S_b^{ref})} \end{cases} \tag{19}$$

where $f_\delta(\delta)$, $f_t(t)$, $f_T(T_C)$, and $f_{soc}(S_b)$ are the influence functions of depth of discharge δ, time t, temperature T_c, and SOC on the life of the ESB, respectively, where $k_{\sigma 1} \cdots k_{\sigma 3}, k_t, k_T, k_{soc}$ are the correlation coefficients of the influence functions, and T_{ref}, S_b^{ref} are the reference temperature and reference SOC, respectively.

In this paper, we use the rain flow counting method [13] commonly used in fatigue analysis in conjunction with the semi-empirical model mentioned above to calculate the capacity degradation L of energy storage batteries. We assume that the solid electrolyte film formation period of the new battery ends at $L' = 0.05$. The evaluation process in Figure 4.

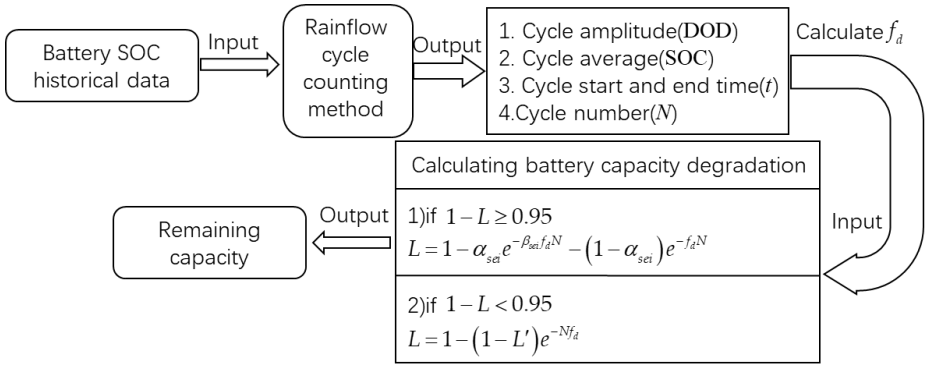


Figure 4. The framework of the battery degradation assessment.

It is worth noting that the refined energy storage loss model used in this paper can only quantify the energy storage loss over a longer period. For the intra-day operation optimization problem with a short time span, the additional cost brought by the energy storage loss is difficult to guide the algorithm to form an optimal strategy, i.e., the agent has little ability to identify which type of action (BSS charging power and ESB charging/discharging power) actually corresponds to a higher reward. This is because the reward is delayed and cumulative. Hence, this paper uses a segmented update method for the cost brought by the energy storage loss, i.e., the energy storage capacity degradation penalty factor σ_k is calculated once for each specific counting period T_1 based on (20) [14].

$$\sigma_k = \frac{Q^{start}(1-L)}{\sum_{t=1}^{T_1}|P_{BESS}(t)|}C_B \tag{20}$$

where Q^{start} is the remaining capacity of the energy storage battery before the start of the counting cycle, L is the percentage of capacity degraded during the cycle, C_B is the energy storage battery kilowatt hour cost coefficient, and $\sum_{t=1}^{T_1}|P_{BESS}(t)|$ denotes the sum of the absolute values of the power of the ESB in the time t to T_1. Therefore, the cost of energy storage capacity degradation C_{BESS} during each decision time interval Δt can be expressed as:

$$C_{BESS} = \sigma_k \cdot |P_{BESS}| \cdot \Delta t \tag{21}$$

2.7. Computational Model of Battery Exchange Service Waiting Penalty C_{BSS}

The battery swapping time is similar to the decision interval Δt of the control center in this paper. Therefore, when EV users generate the demand for power exchange during the period t, only one fully charged battery is needed in the BSS to meet the demand, and the waiting penalty can be expressed as:

$$C_{BSS} = 0.9 - S^{BSS}_{soc,i,\max} \tag{22}$$

where $S^{BSS}_{soc,i,\max}$ is the SOC of the battery with the highest charge among all the batteries in the BSS.

2.8. Constraints on ERS Operation

The constraints include power balance constraints (9), equipment operation constraints (6), (7), (15), and (16), while each device in the ERS has an upper and lower operating limit range, and the operating constraints for BSS and ESB are as follows:

$$\begin{cases} 0 \leq P^t_{BSS,sum} \leq P_{BSS} \cdot M \\ P^{\min}_{BESS} \leq P_{BESS}(t) \leq P^{\max}_{BESS} \end{cases} \tag{23}$$

$$\begin{cases} S_t^{BSS} \leq 0.95 \\ S_b^{\min} \leq S_b(t) \leq S_b^{\max} \end{cases} \tag{24}$$

where P_{BSS} is the charging power of individual chargers in the station, M is the total number of chargers in the station, $P_{BESS}^{\min}$ and $P_{BESS}^{\max}$ are the maximum charging power and maximum discharging power of the storage battery, respectively. $S_b^{\min}$ and $S_b^{\max}$ are the maximum and minimum SOC of the storage battery during the whole period. In order to ensure the normal operation of the storage battery on the next day, the SOC of the storage at the end of the period is required to be a given value:

$$S_b^{t=1} = S_b^{t=T} \tag{25}$$

3. Deep Reinforcement Learning-Based Real-Time Optimization of ERS Operation Model

The arrival time of the EV, the type of EV replenishment, and the PV power are uncertainties in the environment that are difficult to be effectively addressed by traditional model-based optimization algorithms. Essentially, the optimal operation of an energy station is a stochastic sequential decision problem. Reinforcement learning (RL) is a class of efficient algorithms for solving sequential decision problems based on Markov decision processes [16,17]. A Markov decision process can be represented by a five-tuple {S, A, P, R, γ} with elements of the state, action, state transfer probability of the environment, reward, and discount factor. The agent learns adaptively by "trial and error". It gradually adjusts its behavioral strategy by continuously interacting with the environment to obtain the maximum cumulative reward. In this section, the ERS operation model established by (1)–(25) is transformed into a reinforcement learning model quintet according to the Markov decision process, and by improving the training mechanism of the PPO algorithm and introducing LSTM networks to assist in decision-making to help the agent in policy optimization.

3.1. Selection of State Space

The selection of state space should give priority to the factors that have an impact on the decision. For the ERS operation optimization problem in this paper, the state space should at least reflect the current moment of the ERS-grid interaction tariff, the total electricity demand of the charging piles, the ability of the BSS to provide battery swap service, the storage battery SOC, and the PV output level at the current moment, while the PV output trend of the day will also influence the agent's strategy, which will be verified in the simulation in the next section. At the same time, considering that the application scenario of ERS in this paper is a highway service area, the regulation of the charging power of the user charging post is unacceptable, so the charging power demand of the charging post is the same as the PV output, which is an uncontrollable factor and is expressed in the form of subtraction of the two in the state space. The final state space is expressed as follows:

$$\begin{aligned} S_t = [&t, E_{c,t}, E_{d,t}, S_b^t, N_t^f, P_{BSS,\max}^t, P_{pv-ev}^t, \\ &P_{pv}^{t=60}, P_{pv}^{t=70} \ldots P_{pv}^{t=180}] \end{aligned} \tag{26}$$

The parameters include time t, purchased power price to the grid $E_{c,t}$, feeder price $E_{d,t}$ to the grid, storage battery charge state S_b^t, number of fully charged batteries in the BSS N_t^f, charging power limit at the BSS $P_{BSS,\max}^t$ as in (27), difference between PV output and EV user fast charging demand P_{pv-ev}^t as in (28), and forecast of PV output from 6:00 to 18:00 of the day $P_{pv}^{t=60}, P_{pv}^{t=70} \ldots P_{pv}^{t=180}$.

$$P_{BSS,\max}^t = P_{BSS} \cdot (M - N_t^f) \tag{27}$$

$$P_{pv-ev}^{t} = P_{pv}^{*}(t) - P_{ev,sum}^{t} \tag{28}$$

3.2. Selection of Action Space

The action space generally selects the decision variables in the model, and the controllable variables in the ERS model in this paper include the energy storage battery output, the BSS charging power, and the interaction power between the ERS and the grid. Only two need to be determined, and the rest can be uniquely determined by (9). Therefore, the action space in this paper is expressed as:

$$a^{t} = \left[a_1^{t}, a_2^{t}\right] \tag{29}$$

where a_1^t is the power output of the energy storage device, and a_2^t is the BSS charging action command. The power output command of the energy storage device is the output power of the energy storage in the current period, and its range is shown in (30).

$$a_1^{t} = \begin{cases} P_{BESS}^{\min} & P_{BESS}(t) \le P_{BESS}^{\min} \\ P_{BESS}(t) & P_{BESS}^{\min} \le P_{BESS}(t) \le P_{BESS}^{\max} \\ P_{BESS}^{\max} & P_{BESS}(t) \ge P_{BESS}^{\max} \end{cases} \tag{30}$$

The number of batteries in the BSS is large, and the specific battery number as one of the actions will make the action space too large for the model to converge, so the BSS charging action command a_2^t is, therefore, designed as the ratio of the number of batteries involved in charging to the total number of batteries in the BSS at time t, as shown in (31). The action range of a_2^t is shown in (32), where $M\left(1 - N_t^f\right)$ is the total number of batteries that can be involved in charging in the BSS.

$$0 \le a_2^{t} = \frac{N_t^c}{M} \le 1 \tag{31}$$

$$a_2^{t} = \begin{cases} \frac{N_t^c}{M} & 0 \le \frac{N_t^c}{M} \le M\left(1 - N_t^f\right) \\ M\left(1 - N_t^f\right) & \frac{N_t^c}{M} \ge M\left(1 - N_t^f\right) \end{cases} \tag{32}$$

The a_2^t only gives the number of batteries to be charged, without specifying which batteries will be charged. In order to keep the number of fully charged batteries in the BSS as much as possible, the batteries to be charged with high remaining power in the BSS are selected for charging first, i.e., after the BSS receives instruction a_2^t, and selects the $a_2^t \cdot M$ batteries with the highest SOC among the PB for charging. The charging power of the BSS at the moment t can be calculated by (7), and then the ERS-grid interaction power can be calculated by (9).

3.3. Design of the Reward Function

The objective of the deep reinforcement learning algorithm is to maximize the cumulative gain, consistent with the form of the objective function in this paper. Therefore the value of the objective function at the current decision period is fed back to the agent as the real-time reward of the algorithm.

$$r_1^{t} = B_{FC,t} + B_{BSS,t} - C_{grid} - C_{BESS} - C_{BSS} \tag{33}$$

At the same time, the energy storage action a_1^t directly affects the storage cell SOC for the next period, so a penalty needs to be imposed for actions that cause the storage cell SOC to exceed the upper and lower bounds:

$$r_2^t = \begin{cases} \left| a_1^t - \frac{S_b^t - S_b^{\max}}{\eta_t^{ch} \Delta t} \right| & S_b^{t+1} > S_b^{\max} \\ 0 & S_b^{\min} < S_b^{t+1} < S_b^{\max} \\ \left| a_1^t - \eta_t^{dis} \frac{S_b^t - S_b^{\max}}{\Delta t} \right| & SoC_{t+1} < SoC^{\min} \end{cases} \tag{34}$$

To ensure that the ESB can operate normally on the following day, a penalty item is added to the energy storage end-time SOC according to constraint (25) as follows:

$$r_3^t = \begin{cases} \left| S_b^{t=T} - S_b^{t=1} \right| & t = T \\ 0 & t < T \end{cases} \tag{35}$$

In summary, the reward function can be expressed as follows:

$$r^t = r_1^t \sigma_1 - r_2^t \sigma_2 - r_3^t \sigma_3 \tag{36}$$

where σ_1, σ_2 and σ_3 are the weighting coefficients, and are positive numbers. The cumulative rewards for the entire scheduling cycle T are as follows:

$$R_t = \sum_{t'=t}^{T} \gamma^{t'-t} r^{t'} \tag{37}$$

where R_t is the cumulative rewards earned by the agent within $[t, T]$ and γ is the discount factor, which indicates the importance of future rewards relative to the present.

3.4. Improvement of the PPO Algorithm Mechanism

3.4.1. Principle of the PPO Algorithm

The PPO algorithm in this paper adopts the Actor-Critic framework to enhance the convergence speed and stability of the model through empirical replay buffer and importance sampling techniques, and the process of updating the parameters of the Actor and Critic networks is shown in Figure 5.

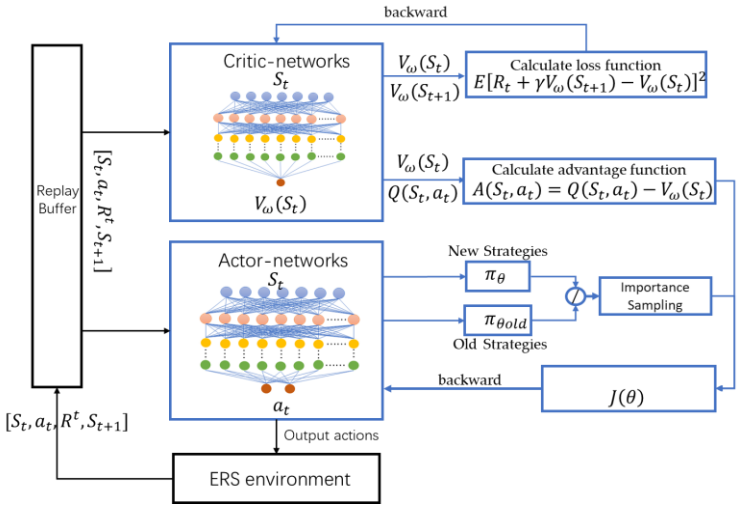


Figure 5. Principle of the PPO algorithm.

The Actor-network is used to generate the operational policy with the network parameters θ; the Critic network is used to evaluate the goodness of the current policy with

the network parameters ω. Firstly, the Actor-network interacts with the ERS environment based on the initial policy $\pi_{\theta \text{old}}$, generating a series of experience data $\{S_t, a_t, r_t, S_{t+1}\}$ and storing it in the replay buffer. Secondly, the Critic network extracts the empirical data $\{S_t, a_t, r_t, S_{t+1}\}$ from the replay buffer and feeds the state S_t into the network to generate the value function $V_\omega(S_t)$. Meanwhile, the Critic-network constructs the loss function $L(\omega)$ in a temporal-difference method as shown in (38), and optimizes the Critic network parameters by gradient descent.

$$L(\omega) = E[R_t + \gamma V_\omega(S_{t+1}) - V_\omega(S_t)]^2 \tag{38}$$

where $E(\cdot)$ is the expectation function, $V_\omega(S_t)$ is the output value of the Critic-network at moment t, and $V_\omega(S_{t+1})$ is the estimated value of the network at moment $t+1$. R_t is the cumulative reward value at time t, γ is the reward discount factor, and the gradient descent updates the Critic-network as shown in (39):

$$\omega \leftarrow \omega - \beta_{critic} \nabla_\omega L(\omega) \tag{39}$$

where β_{critic} is the Critic network learning rate, and $\nabla_\omega L(\omega)$ is the gradient of the loss function $L(\omega)$ with respect to parameter ω. An advantage function is introduced to further assess the current strategy's superiority compared to the average strategy.

$$A(S_t, a_t) = R_t + \gamma V_\omega(S_{t+1}) - V_\omega(S_t) \tag{40}$$

where $A(S_t, a_t)$ is the advantage function, which represents how good or bad the choice of action a_t in state S_t is. R_t is the actual reward obtained at time t, $V_\omega(S_{t+1})$ is an estimate of the output of the Critic-network at time $t+1$, and $V_\omega(S_t)$ is an estimate of the output of the Critic network at time t.

Then, the network weights are updated by optimizing the Actor-network loss function $J(\theta)$ to help the agent learn a better strategy. The loss function $J(\theta)$ is shown in (41), and the Actor-network is updated as shown in (42) [18]:

$$\begin{aligned} J(\theta) = E_{(S_t, a_t) \sim \pi(\cdot;\theta)} [\min(& \tfrac{\pi(a_t|S_t;\theta)}{\pi(a_t|S_t;\theta_{old})} A(S_t, a_t), \\ & \text{clip}(\tfrac{\pi(a_t|S_t;\theta)}{\pi(a_t|S_t;\theta_{old})}, 1-\varepsilon, 1+\varepsilon) A(S_t, a_t))] \end{aligned} \tag{41}$$

where $E(\cdot)$ is the expectation function, $\pi(a_t|S_t;\theta_{old})$, $\pi(a_t|S_t;\theta)$ are the policy output values based on the old and new policy parameters θ_{old}, θ, respectively. ε is the hyperparameter that controls the clipping interval. The clipping function $clip(\cdot)$ then ensures that the sampling probability ratio of the old and new strategies is always within the interval $[1-\varepsilon, 1+\varepsilon]$, preventing the difference between the old and new strategies from crashing the training.

$$\theta \leftarrow \theta + \beta_{Actor} \nabla_\theta J(\theta) \tag{42}$$

where β_{Actor} is the Actor-network learning rate, and $\nabla_\theta J(\theta)$ is the gradient of the loss function $J(\theta)$ for the parameter θ.

3.4.2. Learning Rate Decay

The PPO algorithm enhances the policy through the empirical data generated from interacting with the environment. A slightly larger learning rate at the beginning of training can help the agent approach the optimal policy quickly, and a lower learning rate at the later stages of training can avoid falling into an optimal local solution and enhance the smoothness at the later stages of training. In this paper, we use a linear decaying learning rate, i.e., the learning rate of both Actor and Critic networks decays linearly from the initial value to 0 with the increase of training times.

3.4.3. LSTM-Based Pre-Processing of Irradiation Data

PV output varies significantly by season, weather, and time of day. The day's PV output greatly influences whether the agent needs to use the low tariff period to give charging orders during the early morning hours. To assist the agent's decision, LSTM is introduced in this paper to pre-process the solar irradiation data of the day, and its schematic structure is shown in Figure 6. The solar irradiation data of the day obtained from the meteorological center is used as input to the LSTM network, and the output values processed by the network, together with other parameters of the ERS form the state space for interaction with the agent.

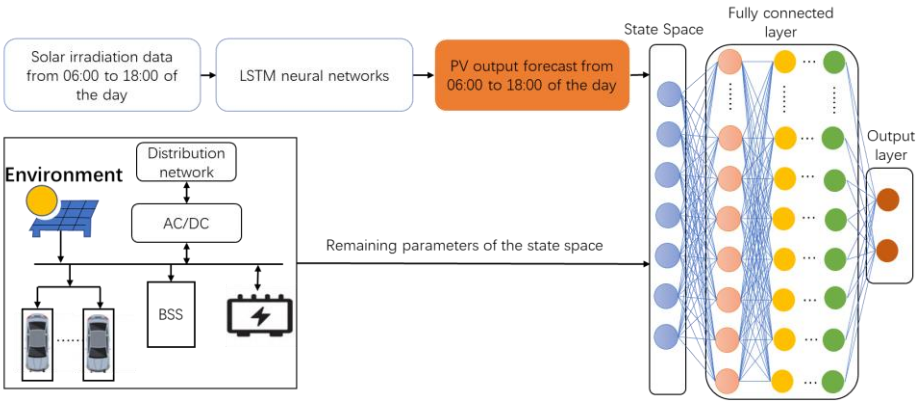


Figure 6. Neural network structure after introducing LSTM.

4. PPO-Based ERS Optimal Operation Problem-Solving Process

4.1. Off-Line Training Process

The operation flow of ERS optimization based on the PPO algorithm is shown in Figure 7. It consists of three parts: ERS main system module, the PPO algorithm module, and the ESB degradation calculation module. Among them, the main system module of the ERS is deployed on each hardware device in the energy station to collect the status information of each device in real-time and form the status space. The ESB degradation calculation module collects the SOC of batteries and the output power in that period in each decision time interval. It updates the remaining capacity of the energy storage system, the charging/discharging efficiency, and the capacity degradation penalty factor σ_k when the amount of SOC data of the energy storage battery in the database is the same as the set counting period T_1. The ERS main system module and the energy storage degradation calculation module constitute the external environment for interaction with the PPO algorithm.

The PPO algorithm is deployed inside the EMS. In the training phase, the algorithm first reads the number of minibatch size N, the number of maximum epoch T, and the maximum episodes $T_{\max}$, and initializes the neural network and the experience buffer. Secondly, the state S_t is observed by interacting with the external environment and generating action a_t based on the existing strategy and obtaining immediate reward r_t, while storing values $\{S_t, a_t, r_t, S_{t+1}\}$ in the experience replay buffer, and so on, to collect experience data.

If the number of experiences in the replay buffer achieves the number of minibatch size N, the experience data are taken out to update the neural network parameters, and in this process, if the maximum number of training steps T is met, the ERS state S_t is initialized and a new training episode is started until the maximum number of training rounds $T_{\max}$ is reached.

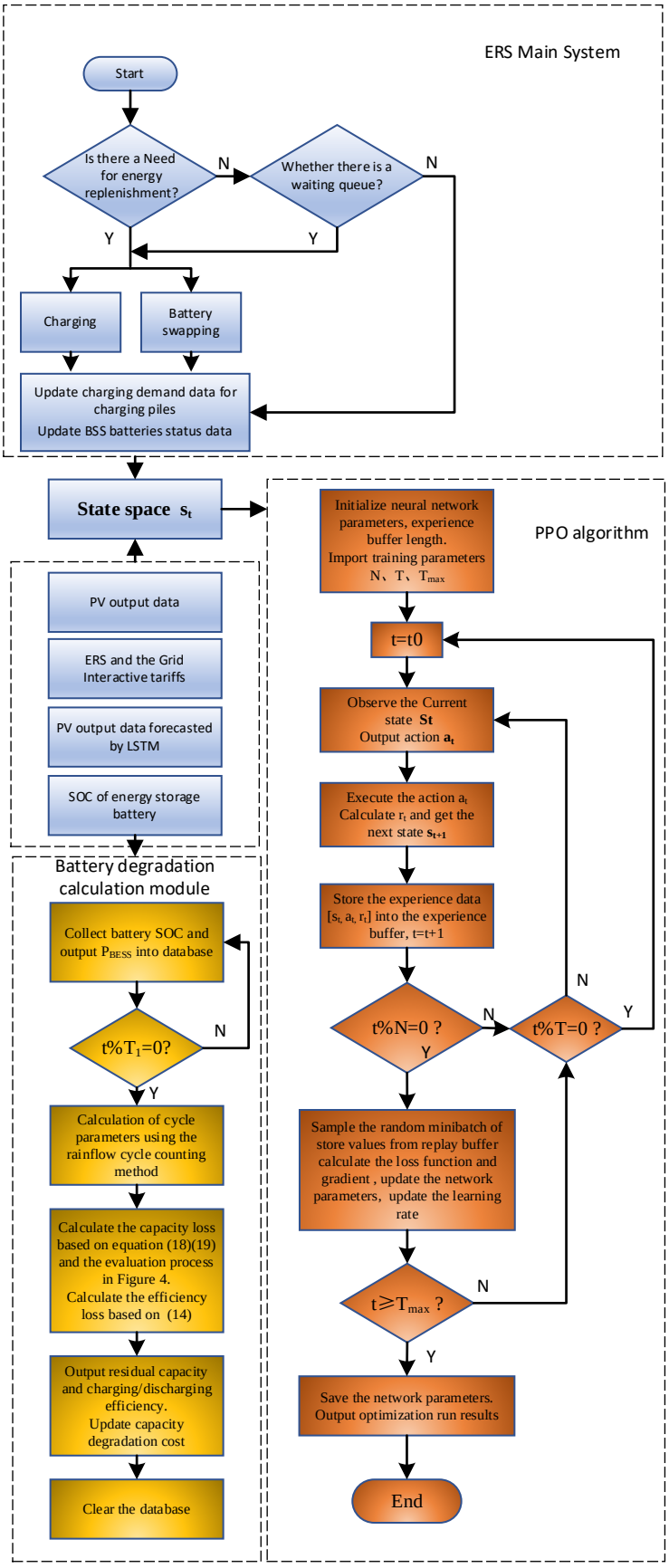


Figure 7. Optimization operation process of the ERS based on the PPO algorithm.

4.2. Online Application Process

After the training is completed, the neural network parameters will be fixed and the Critic network is no longer needed to assist in updating the policy. Therefore, the online application phase only requires the participation of the Actor-network. When the running optimization task starts, the Actor-network receives the environmental information S_t from the ERS and outputs the action a_t until it completes the decision process for all periods T. The online application flow is shown in Figure 8.

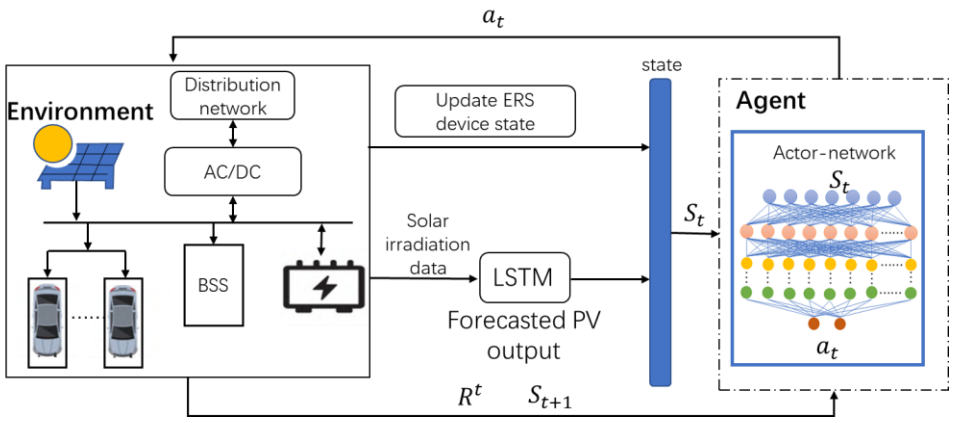


Figure 8. Online optimization operation process based on PPO algorithm.

5. Case Simulation and Performance Analysis

This section uses the ERS shown in Figure 1 as an example for the simulation study, which is located in a high-speed service area in the south of China, with four 120 kW DC fast charging piles and one single small channel BSS. The total number M of batteries in the BSS is 13, and the batteries in the BSS are all charged in constant temperature and power charging mode with charging efficiency $\eta_c^{BSS} = 0.95$ and battery capacity $C^{BSS} = 50$ kW·h. Meanwhile, the station has a 700 kW distributed PV power generation system and energy storage equipment with a rated maximum capacity of 800 kW·h. The Xihe energy big data platform generates the irradiation intensity and PV output data [19], and the calculation parameters related to energy storage battery operating efficiency and capacity degradation are shown in Tables A1 and A2 in Appendix A. The power purchase tariff refers to the peak-valley time-of-use electricity prices of Beijing Electric Power Company, and a service charge E_s = 0.8 is added per kW·h for EV users [20]. The upper and lower limits of energy storage equipment output and SOC constraints are shown in Table 1, and the time-of-use electricity prices are shown in Table 2. The number of EVs arriving at the ERS in each period is a random variable with Poisson distribution $P(\lambda)$, where λ is taken from the average of the number of EVs arriving at the ERS in each period of seven days, see Table A3 in Appendix A. The specific types of EVs, the proportion of the number, and the maximum acceptable charging power are shown in Table 3.

Table 1. Parameters of energy storage equipment.

Parameters	Value
$P_{BESS}^{\min}$(kW)	−300
$P_{BESS}^{\max}$(kW)	+300
$SoC_t^{\min}$	0.2
$SoC_t^{\max}$	0.9
Q^{ini}(kW·h)	800

Table 2. Time-of-use electricity price.

Time Period	Electricity Purchase Tariff (CNY/kW·h)	Feed-in Tariff (CNY/kW·h)
00:00–07:00, 23:00–24:00	0.3946	0
7:00–10:00, 15:00–18:00, 21:00–23:00	0.6950	0.5
10:00–15:00, 18:00–21:00	1.0044	0.8

Table 3. EV quantity proportion and parameter information.

Type of Vehicle	Percentage (%)	Acceptable Fast-Charging Power (kW)	Battery Capacity (kW·h)
Large EV	20	120	100
Medium EV	30	120	75
Mini EV	25	55	35
Battery Swapping EV	25	0	50

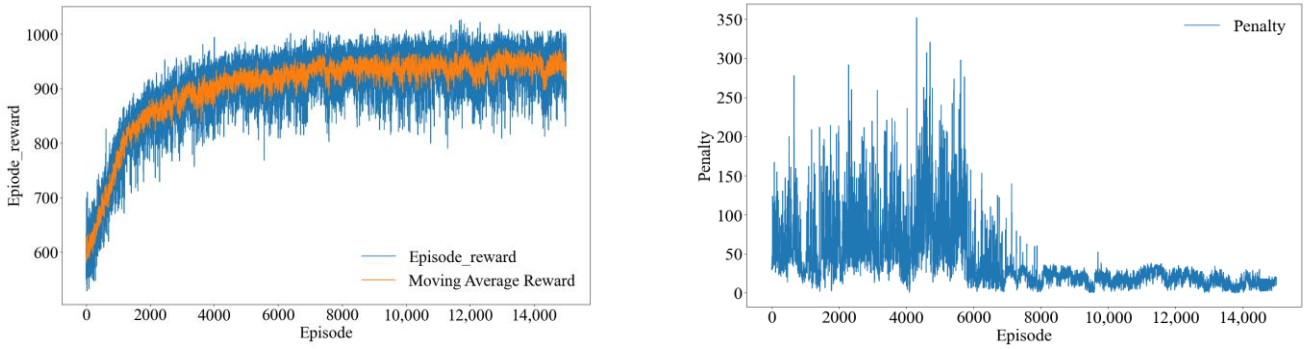
5.1. Training Process

In this paper, the LSTM module uses the Keras framework, and the input layer is a 1×12 time series of solar irradiation data from 06:00 to 18:00 on the same day, and the LSTM output layer has a dimension of 12. The loss function uses mean square error (MSE) to connect another fully connected layer of 12 neurons as the output, representing the PV output prediction from 06:00 to 18:00 on the same day. The deep reinforcement learning method is based on PyTorch 1.5.1 programming, which sets the decision interval $\Delta t = 6\text{min}$ $T = 240$, i.e., one day divided into 240 periods. The Actor and Critic networks contain three hidden layers with 256, 128, and 64 neurons per level, respectively, and the activation function is the Tanh with a reward discount factor $\gamma = 0.99$. The learning rate of both Actor and Critic networks is 0.0003 and linearly decays to 0. The network weights are updated using Adam optimizer, the experience pool size is 120,000, the batch size is 240, the maximum training episode is 15,000, and other hyperparameters are shown in Table 4. The PV output data are selected from the data set for two years, from January 2021 to January 2023, and discrete processing is done by 240 time slots per day, with one year of data selected as the training set and the rest of the data as the test set.

Table 4. Other hyperparameters.

Hyperparameters	Value	Hyperparameters	Value
ε	0.2	σ_2	1
σ_1	0.2	σ_3	100

Figure 9a,b shows the reward and penalty values of the model in each training cycle, respectively. At the early stage of training, the agent lacks knowledge of the environment and has difficulty forming an effective scheduling strategy. As the training proceeds, the average reward shows a stable upward trend with increased training times and finally converges around 950. The penalty value tends to level off and finally converges around 35 after experiencing large fluctuations. At the same time, it is noted that the actual reward value curve has large fluctuations, mainly considering the following three reasons: First, the action of the PPO algorithm is obtained by random sampling on the probability distribution of the output of the policy function, so even under the same state input, there may be differences in the output action. Secondly, the EV models and battery capacities arriving at the station for replenishment at each moment are generated based on the probabilities shown in Table 3, which makes the actual profitability of the replenishment station vary for each training cycle, even if it is run with the optimal strategy. Finally, the uncertainty of the PV output is also an influencing factor for the fluctuation of the reward value.



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

Figure 9. Reward and penalty curve of the training process (**a**) Reward Curve (**b**) Penalty Curve.

5.2. Training Process

To verify the effectiveness of the ERS operation strategy proposed in this paper, Figure 10 shows the dynamic scheduling results of a randomly selected day from the test set, and the BSS battery charging scheduling visualization is shown in Figure 11. Combining Figures 10 and 11 shows that the electrical power of the energy storage system (ESS) and BSS are correlated with fluctuations in PV output, EV charging load, and electricity prices. ESS and BSS's charging behavior appears to occur during periods of high PV power and the valley hours of electricity prices. During peak and normal tariff periods and times of low PV output, ESS devices tend to release stored power, and BSS reduces the number of charging cells to lower charging costs.

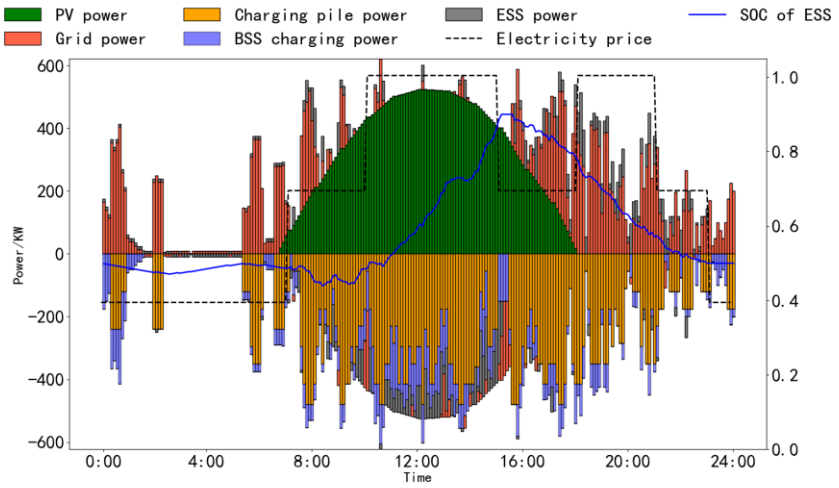


Figure 10. Dispatch results of electric power.

During (00:00–07:00, 23:00–24:00), when electricity prices are in the valley hours, and PV output is 0, the BSS charges a high number of batteries and quickly boosts the number of fully charged batteries from 6 to 13 to reserve available batteries for the upcoming peak battery swapping. The SOC of the ESS changes slightly during this time because it has a certain amount of initial energy storage, and the agent predicts by judging that there is sufficient light on that day and waiting for the PV output to rise before charging can yield more benefits, while avoiding the storage capacity degradation penalty from multiple charging and discharging.

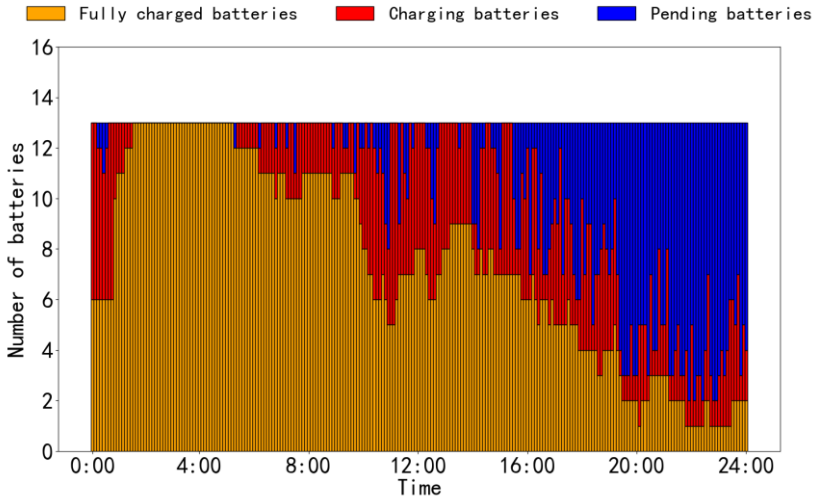


Figure 11. Number of batteries in three states of the ERS.

(07:00–15:00) The tariff rises from normal to peak hours, and the PV output gradually rises to the maximum and then slowly decreases. EVs arriving at the station to swap batteries are at their peak during this time, and the BSS charges at a higher power to meet the demand for EVs, while the ESS makes real-time power adjustments based on PV output, fast charging load, and BSS power to achieve full PV consumption.

(15:00–24:00) The tariff goes through peak hours, normal hours, and eventually drops to valley hours at 23:00, with PV output gradually dropping to 0. The high number of pending batteries during this time in Figure 10 is a result of the BSS actively reducing charging batteries number during this time to expand its profit margin to reduce operating costs. The ESS continues to discharge, and its SOC gradually drops from 0.9 to 0.5 and remains constant.

In summary, the agent is able to make reasonable decisions in different external environments. As seen in Figure 11, the BSS can maintain more than one usable battery throughout the day, indicating that Actor's strategy can meet the user's battery swapping demand. The SOC of ESS at the end of the day is the same as that at the beginning, which shows that the strategy learned by the Actor-network has the ability for sustainable operation. To further validate the proposed method's adaptive decision-making capability when the PV output is insufficient, the dynamic dispatching results of a day in winter in the test set are shown in Figures 12 and 13.

As seen from Figures 12 and 13, the agent can use the day-ahead forecasting link to get an idea of the PV output trend on the test day to aid decision-making. The ESS and BSS adaptively regulate their electrical power when the PV output is low. During the valley hours, the ESS and BSS are charged to obtain sufficient power reserves and fully charged battery reserves. During normal and peak hours, the ESS outputs the stored power to generate revenue, while the BSS reduces the number of batteries being charged to lower charging costs.

It is also noted that the size of the PV output will indirectly affect the ERS operating strategy. The operation strategy of ESS and BSS is essentially a disguised peak-valley arbitrage when the PV output exceeds the load demand, the excess is stored by the energy storage battery and BSS and sold to EV owners to earn revenue when the electricity price is higher; when the PV output is insufficient, the strategy changes to traditional peak-valley arbitrage, i.e., using the peak-valley electricity price to "buy low and sell high". The agent has mastered the strategy during training and can therefore show strong adaptive decision-making ability in different environments, and also demonstrate the robustness and generality of the proposed method.

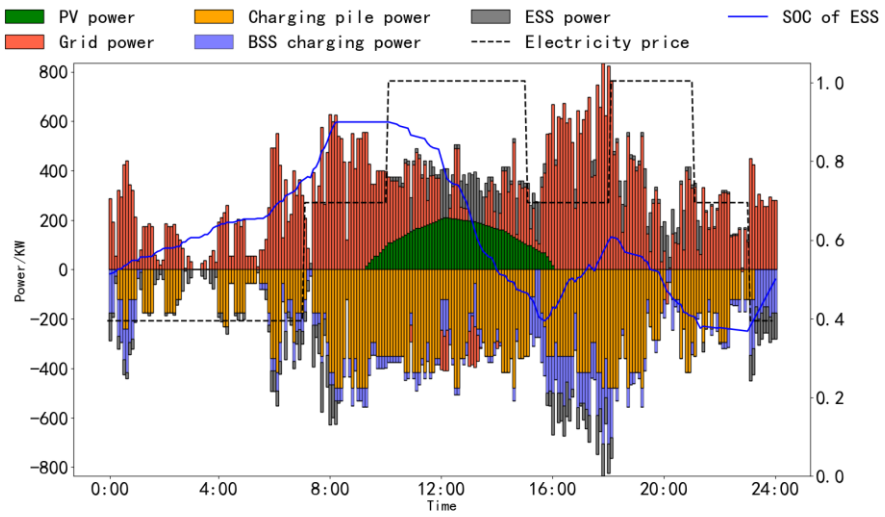


Figure 12. Results of dynamic scheduling of electric power under low PV output.

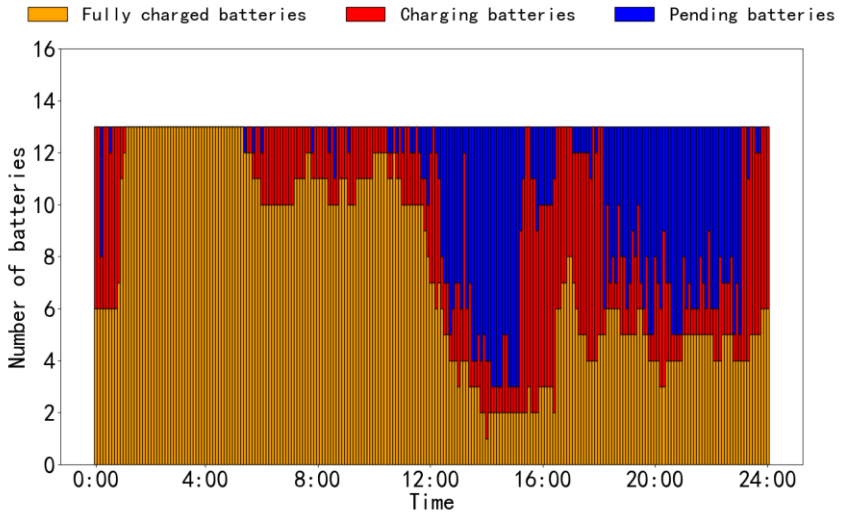


Figure 13. Number of batteries in three states of the ERS under low PV output.

5.3. Comparison of Algorithm Performance

In order to verify the performance of the proposed method in this paper, four methods were selected for comparative analysis. Method 1, the method in this paper, considers the ESB capacity degradation penalty term and incorporates an LSTM network to pre-process the same-day irradiance data. Method 2 is the Deep Deterministic Policy Gradient (DDPG) algorithm. This RL algorithm is commonly used to solve continuous action space problems. It has the same number of layers and training rounds as Method 1, but does not use the LSTM network to generate PV output predictions to assist in decision-making. In addition, to test the ability of the proposed method to cope with uncertainty, Method 3 uses a model-based optimization algorithm as a comparison, which assumes that all uncertainties can be accurately predicted, including EV replenishment type, arrival time, battery capacity, remaining power, maximum acceptable charging power, and PV output. Under these assumptions, the ERS operational strategy optimization problem can be modeled as a deterministic optimization problem, and the Gurobi solver is invoked to solve it as the theoretically optimal solution. Method 4 is Model Predictive Control (MPC), which predicts at each time step the type of EV replenishment, battery capacity, and PV output for a future period, and then solves the revenue maximization problem at that time step, and so on, until the end of the sequence. In this paper, the length of the control sequence is set to 10, i.e. 1 h,

and the length of the prediction sequence to 30, i.e., 3 h. At the same time, considering that the degradation of the ESB capacity and the decrease of the charging/discharging efficiency will take a longer cycle time to manifest, we keep cycling the test set until the storage battery is disposed of to calculate its full life-cycle benefit.

The SOC of the above four methods for the ESS on a given day in the test set is shown in Figure 14. The ESB whole life-cycle capacity degradation curve is shown in Figure 15, and the performance comparison on the test set is shown in Table 5, where the battery swapping service availability is expressed as:

$$\frac{\sum_{d=1}^{365}\sum_{t=0}^{T} h}{365 \times T} \times 100\% \tag{43}$$

where $\sum_{d=1}^{365}\sum_{t=0}^{T} h$ denotes the sum of the periods within the test set when the BSS is unable to provide battery swap services due to insufficient fully charged batteries, where h is a 0–1 variable that counts as 0 if the EV user has a demand for the swap and the BSS is unable to provide the swap, otherwise it counts as 1.

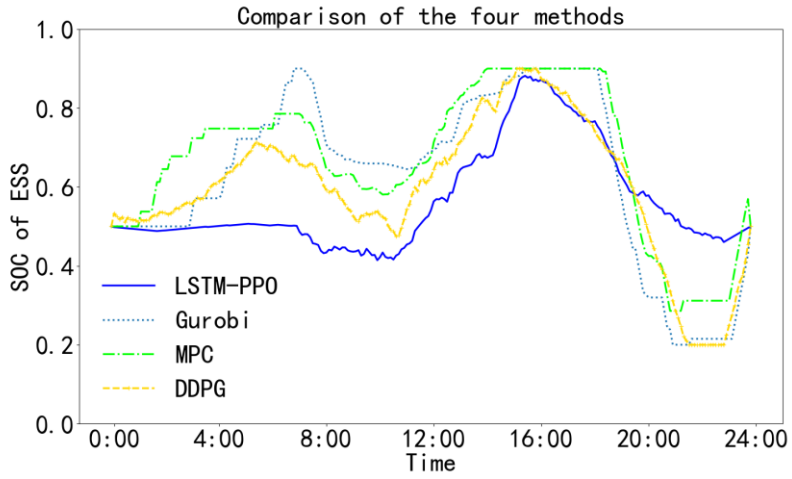


Figure 14. Comparison of the four methods.

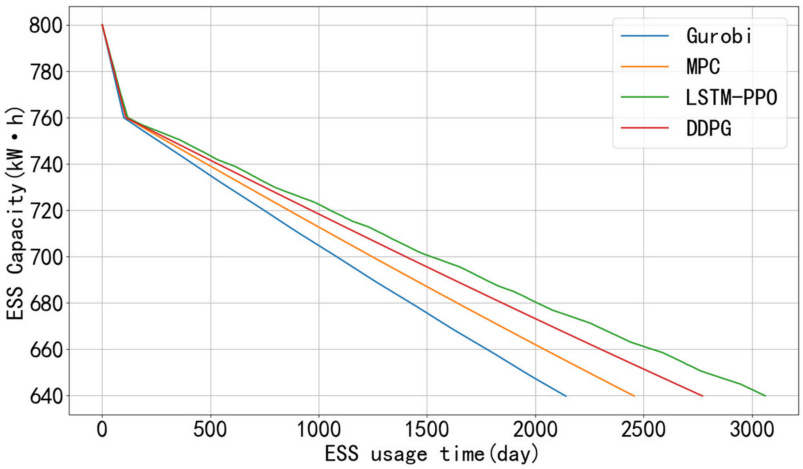


Figure 15. Curves of ESS capacity degradation.

Table 5. Comparison of the four methods.

Methods	Test Set Annual Revenue (CNY/Year)	BSS Service Availability (%)	ESS Lifetime (Day)	ESS Life-Cycle Benefits (CNY)
LSTM-PPO	2,783,018.64	98.58	3065	23,881,738.9
DDPG	2,744,767.22	97.20	2771	20,857,962.3
Gurobi	3,090,575.72	100	2141	18,214,762.3
MPC	2,605,694.14	90.33	2466	17,725,860.7

Combining Figure 14 and Table 5 shows that the theoretical optimal solution based on the Gurobi solver performs the best in the test set, has the highest annual return, and can fully satisfy the EV customer switching demand. However, the ESB lifetime of this method is only 2141 days, the shortest among the four methods, and its whole life-cycle benefit is also the lowest, at 17,725,860.7 CNY. This is because the algorithm only considers the optimal solution within the decision period, and the ESS charging/discharging power and DOD are larger. At the same time, the actual scenario has a large randomness, so the theoretical optimal solution is difficult to reproduce in engineering practice. The LSTM-PPO method in this paper has the second highest test set annual yield and battery swapping service availability, differing by only 11.05% and 1.42%, respectively, compared to the theoretically optimal solution. However, the method has the highest energy storage lifetime and total life cycle yield, which are 43.1% and 34.72% higher than the Gurobi method, as can be seen in Figures 14 and 15.

As can be seen in Figure 14, compared to the other three algorithms, the method proposed in this paper judges that there is sufficient light on that day and therefore forgoes charging the ESB from 0:00 to 7:00 using the valley hours of electricity prices, and also avoids battery degradation due to charging/discharging. As seen in Figure 15, the method in this paper has the most extended battery life, which is further evidence of the effectiveness of the agent strategy. At the same time, note that the ESS capacity degradation curve of this paper's method has small fluctuations compared to the other algorithms in the figure. This is because in the test set, the PV output is strongly correlated with the weather and season, and the agent adaptively adjusts the ESS charging and discharging strategy in both well and poorly-illuminated environments, resulting in differences in battery losses per unit period. The comparison method, on the other hand, lacks the perception of irradiance intensity, and the charging and discharging strategies converge in training, so the loss curve is approximately a sloping straight line.

In this paper, DDPG is used as the baseline algorithm for deep reinforcement learning, and its test set annual revenue, battery replacement service availability, ESB lifetime, and full life-cycle revenue are all lower than the LSTM-PPO method. It is worth noting that the MPC algorithm is also unsatisfactory in all metrics. This is because MPC only considers the current control period's optimality, ignoring the current decision's impact on subsequent periods. MPC will reduce the BSS charging power as much as possible while controlling the deep discharge of the ESB to reduce the power purchase cost in the current period, which, coupled with the intense uncertainty of the replenishment behavior of EV users in the scenario, makes the method ineffective. It is also further demonstrated that the proposed method is well adapted to the uncertainties in engineering applications.

In summary, compared to the other three methods, the proposed method in this paper has the most advantages in the scenarios where the availability of battery swapping services, degradation loss of ESB capacity, and economic benefits of ERS are taken into account. It shows good robustness to uncertainties in engineering applications. The algorithms in this paper were all done on an Intel (R) Core (TM) i7-10700F CPU @ 2.90 GHz computer, with a training time of about 6.5 h. After the training, the Actor-network takes only 8 ms for one forward propagation, which is suitable for optimizing the operation strategy in the real-time phase.

6. Conclusions

This paper takes ERS with PV participation as the research object and proposes a deep reinforcement learning-based method for optimizing the operation strategies of ESS and BSS in ERS accounting for ESB degradation. The method transforms the power control problem of ESS and BSS into a reinforcement learning problem with continuous state space and action space, avoiding the difficulty of solving the problem due to dimensional catastrophe and model uncertainty, and aids the agent in making reasonable decisions by introducing an LSTM network to sense the PV power output trend. Simulation results show that the proposed method can cope with the uncertainty of EV vehicle type, replenishment demand, and environmental PV output. Compared with two other model-based optimization algorithms and a model-free optimization algorithm, the proposed method performs well in scenarios that consider three factors: availability of battery swapping services, degradation of ESS, and economic efficiency of ERS.

In addition, this paper uses ERS as the research object and advanced artificial intelligence algorithms as the solution method, intending to provide a different perspective on the operational optimization of EV-related infrastructure and promote the application of artificial intelligence technology on the ground. Due to the introduction of neural networks in the method of this paper, there are problems of poor interpretation and difficulty in tuning the parameters. In the future, in-depth research will be carried out on the above issues as the ERS involved in PV has the potential to operate in isolation and participate in grid peaking and frequency regulation.

Author Contributions: Conceptualization, Y.Z. and Y.B.; methodology, Y.B.; software, Y.B.; validation, Y.Z. and Y.B.; formal analysis, Y.B.; investigation, Y.Z.; resources, Y.Z.; data curation, Y.Z.; writing—original draft preparation, Y.Z. and Y.B.; writing—review and editing, Y.Z. and Y.B.; visualization, Y.B.; supervision, Y.B.; project administration, Y.Z.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Guangxi Special Fund for Innovation-Driven Development (AA19254034).

Data Availability Statement: Publicly available datasets were analyzed in this study. These data can be found in the References section.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the study's design; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Energy storage operation efficiency parameters.

Parameters	Value	Parameters	Value
e_0	1.00	h_0	1.00
e_1	4.00×10^{-3}	h_1	-4.60×10^{-3}
e_2	-3.11×10^{-3}	h_2	4.13×10^{-3}
e_3	-4.77×10^{-3}	h_3	5.00×10^{-7}
e_4	3.06×10^{-3}	h_4	4.23×10^{-13}
e_5	9.66×10^{-8}	h_5	-1.36×10^{-7}

Table A2. Energy storage operation efficiency parameters.

Parameters	Value	Parameters	Value
α_{sei}	0.0575	k_t	4.14×10^{-10}
β_{sei}	121	k_T	0.0693
$k_{\delta 1}$	140,000	k_{soc}	1.04
$k_{\delta 2}$	−0.501	$k_{ref}/^{\circ}\mathrm{C}$	25
$k_{\delta 3}$	−123,000	S_b^{ref}	0.5

Table A3. Probability distribution of EV arrival at ERS.

Time Period	λ	Time Period	λ	Time Period	λ	Time Period	λ
00:00–01:00	1	06:00–07:00	6	12:00–13:00	30	18:00–19:00	31
01:00–02:00	0	07:00–08:00	15	13:00–14:00	24	19:00–20:00	17
02:00–03:00	0	08:00–09:00	10	14:00–15:00	18	20:00–21:00	15
03:00–04:00	0	09:00–10:00	11	15:00–16:00	8	21:00–22:00	4
04:00–05:00	0	10:00–11:00	16	16:00–17:00	16	22:00–23:00	3
05:00–06:00	1	11:00–12:00	26	17:00–18:00	25	23:00–24:00	1

References

1. Leone, C.; Longo, M.; Fernández-Ramírez, L.M.; García-Triviño, P. Multi-Objective Optimization of PV and Energy Storage Systems for Ultra-Fast Charging Stations. *IEEE Access* **2022**, *10*, 14208–14224. [CrossRef]
2. El-Taweel, N.A.; Farag, H.; Shaaban, M.F.; AlSharidah, M.E. Optimization Model for EV Charging Stations With PV Farm Transactive Energy. *IEEE Trans. Ind. Inform.* **2022**, *18*, 4608–4621. [CrossRef]
3. Chaudhari, K.; Ukil, A.; Kandasamy, N.K.; Manandhar, U.; Kollimalla, S.K. Hybrid Optimization for Economic Deployment of ESS in PV-Integrated EV Charging Stations. *IEEE Trans. Ind. Inform.* **2018**, *14*, 106–116. [CrossRef]
4. Liao, Y.T.; Lu, C.N. Dispatch of EV Charging Station Energy Resources for Sustainable Mobility. *IEEE Trans. Transp. Electrif.* **2017**, *1*, 86–93. [CrossRef]
5. Sadeghianpourhamami, N.; Deleu, J.; Develder, C. Definition and Evaluation of Model-Free Coordination of Electrical Vehicle Charging With Reinforcement Learning. *IEEE Trans. Smart Grid* **2019**, *11*, 203–214. [CrossRef]
6. Liu, N.; Chen, Q.; Lu, X.; Liu, J.; Zhang, J. A Charging Strategy for PV-Based Battery Switch Stations Considering Service Availability and Self-Consumption of PV Energy. *IEEE Trans. Ind. Electron.* **2015**, *62*, 4878–4889. [CrossRef]
7. Shalaby, A.A.; Shaaban, M.F.; Mokhtar, M.; Zeineldin, H.H.; El-Saadany, E.F. A Dynamic Optimal Battery Swapping Mechanism for Electric Vehicles Using an LSTM-Based Rolling Horizon Approach. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 15218–15232. [CrossRef]
8. Ko, H.; Pack, S.; Leung, V. An Optimal Battery Charging Algorithm in Electric Vehicle-Assisted Battery Swapping Environments. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 3985–3994. [CrossRef]
9. Yan, D.; Yin, H.; Li, T.; Ma, C. A Two-Stage Scheme for Both Power Allocation and EV Charging Coordination in A Grid Tied PV-Battery Charging Station. *IEEE Trans. Ind. Inform.* **2021**, *17*, 6994–7004. [CrossRef]
10. Bhatti, A.R.; Salam, Z. A Rule-Based Energy Management Scheme for Uninterrupted Electric Vehicles Charging at Constant Price Using Photovoltaic-Grid System. *Renew. Energy* **2018**, *125*, 384–400. [CrossRef]
11. Morstyn, T.; Hredzak, B.; Aguilera, R.P.; Agelidis, V.G. Model Predictive Control for Distributed Microgrid Battery Energy Storage Systems. *IEEE Trans. Control Syst. Technol.* **2017**, *26*, 1107–1114. [CrossRef]
12. Kim, T.; Qiao, W. A Hybrid Battery Model Capable of Capturing Dynamic Circuit Characteristics and Nonlinear Capacity Effects. *IEEE Trans. Energy Convers.* **2011**, *26*, 1172–1180. [CrossRef]
13. Xu, B.; Oudalov, A.; Ulbig, A.; Andersson, G.; Kirschen, D.S. Modeling of Lithium-Ion Battery Degradation for Cell Life Assessment. *IEEE Trans. Smart Grid* **2016**, *9*, 1131–1140. [CrossRef]
14. Cao, J.; Dan, H.; Fan, Z.; Morstyn, T.; Li, K. Deep Reinforcement Learning-Based Energy Storage Arbitrage With Accurate Lithium-Ion Battery Degradation Model. *IEEE Trans. Smart Grid* **2020**, *11*, 4513–4521. [CrossRef]
15. Wang, L.; Qin, Z.; Slangen, T.; Bauer, P.; Wijk, T.V. Grid Impact of Electric Vehicle Fast Charging Stations: Trends, Standards, Issues and Mitigation Measures—An Overview. *IEEE Open J. Power Electron.* **2021**, *2*, 56–74. [CrossRef]
16. Sutton, R.S.; Barto, A.G. Reinforcement Learning. *A Bradf. Book* **1998**, *15*, 665–685.
17. Franois-Lavet, V.; Henderson, P.; Islam, R.; Bellemare, M.G.; Pineau, J. An Introduction to Deep Reinforcement Learning. *Mach. Learn.* **2018**, *11*, 219–354.
18. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:abs/1707.06347.
19. Xihe Energy Big Data Platform. Available online: https://xihe-energy.com/#climate (accessed on 22 April 2023).
20. State Grid Beijing Electric Power Company's Electric Vehicle Public Implementation of Peak and Valley Tariffs for Public Charging Facilities. Available online: http://www.bj.sgcc.com.cn/html/main/col1/column_1_1.html (accessed on 1 May 2022).

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.