energies

MDPI

Article

Tracking Photovoltaic Power Output Schedule of the Energy Storage System Based on Reinforcement Learning

Meijun Guo [1], Mifeng Ren [1,*], Junghui Chen [2,*], Lan Cheng [1] and Zhile Yang [3]

[1] College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan 030024, China; mmmguomeijun@163.com (M.G.); taolan_1983@126.com (L.C.)
[2] Department of Chemical Engineering, Chung-Yuan Christian University, Taoyuan 320314, Taiwan
[3] Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China; zyang07@qub.ac.uk
* Correspondence: renmifeng@126.com (M.R.); jason@wavenet.cycu.edu.tw (J.C.)

Abstract: The inherent randomness, fluctuation, and intermittence of photovoltaic power generation make it difficult to track the scheduling plan. To improve the ability to track the photovoltaic plan to a greater extent, a real-time charge and discharge power control method based on deep reinforcement learning is proposed. Firstly, the photovoltaic and energy storage hybrid system and the mathematical model of the hybrid system are briefly introduced, and the tracking control problem is defined. Then, power generation plans on different days are clustered into four scenarios by the K-means clustering algorithm. The mean, standard deviation, and kurtosis of the power generation plant are used as the features. Based on the clustered results, the state, action, and reward required for reinforcement learning are set. In the constraint conditions of various variables, to increase the accuracy of the hybrid system for tracking the new generation schedule, the proximal policy optimization (PPO) algorithm is used to optimize the charging/discharging power of the energy storage system (ESS). Finally, the proposed control method is applied to a photovoltaic power station. The results of several valid experiments indicate that the average errors of tracking using the Proportion Integral Differential (PID), model predictive control (MPC) method, and the PPO algorithm in the same condition are 0.374 MW, 0.609 MW, and 0.104 MW, respectively, and the computing time is 1.134 s, 2.760 s, and 0.053 s, respectively. The consequence of these indicates that the proposed deep reinforcement learning-based control strategy is more competitive than the traditional methods in terms of generalization and computation time.

Keywords: deep reinforcement learning; energy storage system; photovoltaic power output; schedule tracking control

Citation: Guo, M.; Ren, M.; Chen, J.; Cheng, L.; Yang, Z. Tracking Photovoltaic Power Output Schedule of the Energy Storage System Based on Reinforcement Learning. *Energies* **2023**, *16*, 5840. https://doi.org/10.3390/en16155840

Academic Editor: Peter D. Lund

Received: 15 June 2023
Revised: 23 July 2023
Accepted: 31 July 2023
Published: 7 August 2023

1. Introduction

China has announced to the world its goal of achieving carbon neutrality by 2060, which fully reflects the responsibility of a major country and further emphasizes the important position of ecological civilization construction in the national strategy. The key to achieving China's carbon neutrality goal is building a clean, low-carbon, recycling economic system and a green, carbon-reducing, secure, and highly efficient energy system [1]. To continuously promote the energy revolution and achieve the carbon peaking and carbon neutrality goals, the National Energy Administration plans that the country will install more than twice the current amount of capacity of installed wind and solar power in the next 10 years [2].

Wind and photovoltaic power generation are characterized by randomness, fluctuation, and intermittence. If wind or photovoltaic power with randomness is directly connected to the grid, it would bring great instability to the power grid. To solve this problem, energy storage has been employed in renewable green and clean energy power stations, and it has been proven to be an effective way of fluctuation smoothing [3], peak

cutting, and valley filling [4]. In [5], a pumped storage power station was used to establish an optimization model to overcome the risks associated with excessive power fluctuations in wind power generation. Ref. [6] equips energy storage systems in energy communities to sell surplus energy to energy retailers when consumers have excess energy over demand energy and to buy energy from the storage system when consumers have insufficient energy. It is important to form a combined wind-power storage system that can effectively make up for the shortcomings of renewable energy generation for the stability of the power grid [7].

Currently, some scholars have conducted research on applications of the ESS in tracking new energy power generation schemes. In [8], an integrated control approach to internal energy coordination control and multi-objective optimization control was adopted to realize the power tracking control of photovoltaic power stations. In [9], a charge/discharge controlling strategy for an ESS with five control parameters was established, and a method of real-time optimization of control coefficients by the particle swarm optimization algorithm was proposed to track the power generation schedule. Ref. [10] proposed an optimal control technique for power flow control of hybrid renewable energy systems, combining the whale optimization algorithm and the artificial neural network, the simulation results show that the proposed technique is successfully used to resolve the optimal power flow problem of the hybrid system In [11], a fuzzy model predictive control for the ESS has been proposed, and simulation based on the historical operation data of the photovoltaic plant shows that the proposed control method has flexibility and adaptability. Although those methods can effectively realize the tracking control problem, they can only be used for a fixed power generation schedule, which is difficult to dynamically adapt to the random fluctuations of scenery and achieve online control. For the multi-time scale scheduling problem, the above methods are easy to fall into the local optimum due to the dimensional disaster.

Reinforcement learning is an adaptive model-free machine learning method, which has a good ability to extract historical data features and can avoid the problems of uncertainty modeling and dimensionality disaster [12]. Reinforcement learning has now been applied to the energy scheduling problem [13,14]. In [15], the q-learning algorithm was used for minimizing the photovoltaic power generation cost installed in the microgrid, and its results indicate that q-learning was superior to the rule-based heuristic algorithm. In [16], the deep reinforcement learning theory was applied to integrated energy distribution. It can respond dynamically to uncertainties in the environment and achieve the effect of improving the economy of system operation. In [17], an improved K-means algorithm was used to achieve energy storage grouping. The multi-agent deep deterministic policy gradient (MADDPG) algorithm was then used to tackle the grouped multi-agent system. The experiments showed that the suggested scheduling strategy can suppress the fluctuation impacts in the wind power output and improve the operational efficiency of the hybrid system. The application of reinforcement learning reduces costs and minimizes fluctuations in hybrid storage systems. It provides a good way to track generation plans.

In this paper, a charging and discharging control strategy for an energy storage system based on the PPO algorithm is proposed. The major contributions of the paper are as follows:

(1) A charge and discharge power control method for the energy storage system is proposed based on deep reinforcement learning. It can adapt to different power generation plans.

(2) The K-means algorithm is used to solve the problem that the control parameters will be different in different weather conditions.

(3) The charge/discharge power limit of the energy storage system and the residual capacity limit of the energy storage system are considered.

This paper proposes an effective scheduling scheme based on reinforcement learning. The whole scheme will be detailed in the following sections. First, the mathematical model, constraint conditions, and the objective function of the hybrid system are established in Section 2; then, the state, action, and reward required by the PPO algorithm and the

optimization process are introduced in Section 3; next, the proposed control method is applied to a photovoltaic power station in Section 4; finally, the whole paper is summarized in Section 5.

2. Problem Formulation

The photovoltaic and energy storage hybrid system includes a photovoltaic power generation system, a control center, and an ESS. The structure of the hybrid system is described in Figure 1. The photovoltaic power generation station delivers the day-ahead forecast power to the dispatching center as the power generation plan every day. If the actual power generation is too distinct from the power generation plan, it needs to be adjusted through the ESS.

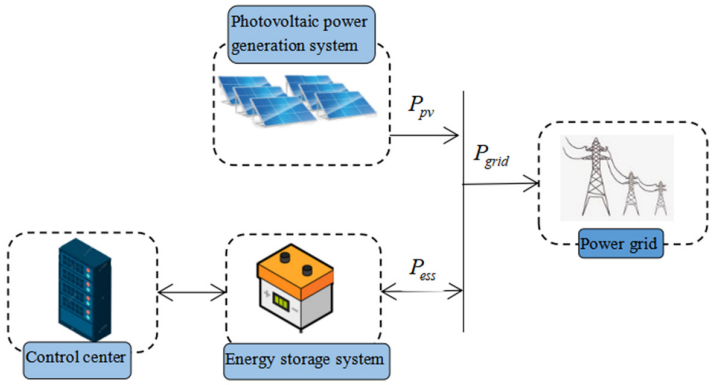

Figure 1. Photovoltaic and energy storage hybrid system.

In an ideal condition, the power generation of a photovoltaic power station should be equal to the power generation plan, but in practice, because photovoltaic power generation is affected by solar radiation and other meteorological factors, so the power generation of the photovoltaic power station and the power generation plan cannot be equal; there will always be deviations, which is not conducive to the stability of the power grid. Thus, it is necessary to connect the energy storage system to regulate the power generation of photovoltaic power stations. This means that to ensure stable operation of the power grid, a simple way is to make the hybrid system (the power generation of photovoltaic and energy storage) and the power generation plan as close as possible. When the power generation of the photovoltaic power station is smaller than the power generation plan, the energy storage system discharges to provide the missing power; when the power generation of the photovoltaic power station is larger than the power generation plan, the energy storage system charges to absorb the excess power.

The mathematical model of the hybrid system can be established as (1).

$$\begin{cases} C_{ess}(t) = (1-\rho)C_{ess}(t-1) - \Delta C_{ess}(t) \\ \Delta C_{ess}(t) = \begin{cases} P_{ess}(t)\eta_c \Delta t, P_{ess}(t) \le 0 \\ P_{ess}(t)\Delta t/\eta_d, P_{ess}(t) > 0 \end{cases} \\ P_{grid} = P_{ess}(t) + P_{pv}(t) \end{cases}, \quad (1)$$

where $C_{ess}(t)$ is the residual capacity of the ESS at the end of time t, MW·H, $P_{ess}(t)$ is the charging and discharging power value of the ESS at time t, the charging power is negative and the discharging power is positive, ρ is the self-discharge rate of the ESS, η_c and η_d are the charging and discharge efficiency of the ESS, Δt is the time interval, and $P_{pv}(t)$ is the photovoltaic power generation at time t.

To make the ESS run healthily for a long time and save costs, the design of the hybrid system should satisfy some constraints.

1. The capacity constraint of the ESS:

$$SOC_{\min} \cdot C_{rated} \leq C_{ess}(t) \leq SOC_{\max} \cdot C_{rated}, \tag{2}$$

where $SOC = \frac{Cess(t)}{C_{rated}}$, $SOC_{\min}$ and $SOC_{\max}$ are the lower and upper limits of the SOC, and C_{rated} is the rated capacity of the ESS.

2. Charge/discharge power constraint of the ESS:

$$-P_{\max} \leq P_{ess}(t) \leq P_{\max}, \tag{3}$$

where $P_{\max}$ is the maximum charge/discharge power value, and $P_{ess}(t)$ can take all real numbers in this range.

To make the residual capacity still satisfy the constraint after charging/discharging, $P_{ess}(t)$ should satisfy the following requirements:

(a) Charging:

$$-\min(P_{\max}, \frac{SOC_{\max} \cdot C_{rated} - (1-\rho)C_{ess}(t)}{\eta_c \cdot \Delta t}) \leq P_{ess}(t) \leq 0. \tag{4}$$

(b) Discharging:

$$0 \leq P_{ess}(t) \leq \min(P_{\max}, \frac{[(1-\rho)C_{ess}(t) - SOC_{\min} \cdot C_{rated}] \cdot \eta_d}{\Delta t}). \tag{5}$$

During the tracking of the generation plan, the target power curve is the planned output curve (day-ahead forecast of photovoltaic power generation) issued by the dispatch center. Take the time interval to be 15 min for example. There are 96 time periods in a day, and each period corresponds to a planned output value. The objective function in this paper is composed of (1) the deviation between the power generation and the power generation plan of the hybrid optical storage system, and (2) the deviation between the residual capacity and the ideal capacity, as shown in (6). The first part of the objective function describes the economics of energy storage, and the second part describes the tracking effect of the hybrid system.

$$J = \alpha\left|C_{ess}(t) - C_{ideal}\right| + \beta\left|P_{pv} + P_{ess} - P_{aim}\right|, \tag{6}$$

where C_{ideal} is the ideal capacity of the ESS, P_{aim} is the power generation plan for each time interval, α and β are the weight coefficients, and $\alpha + \beta = 1$.

This work aims to design a charge/discharge power control method for an ESS to satisfy two requirements simultaneously: (1) The power generation of the hybrid system follows the power generation plan as closely as possible; (2) The residual capacity of the ESS is close to the ideal capacity in the condition of satisfying the constraint conditions.

3. Power Generation Control Strategy Based on the PPO Algorithm

The flow chart is shown in Figure 2. First, establish the mathematical model of the hybrid system, the constraint conditions of each variable, and the optimization objective function, all of which are defined in Section 2. Use the K-means algorithm to divide different days into k classes based on mean, standard deviation, and kurtosis; then, set the state, action, and reward required by the PPO algorithm. Next, the selected action is constantly optimized according to the PPO algorithm in different scenarios. Finally, the output power of the hybrid system can follow the power generation plan under the optimal regulation of the ESS.

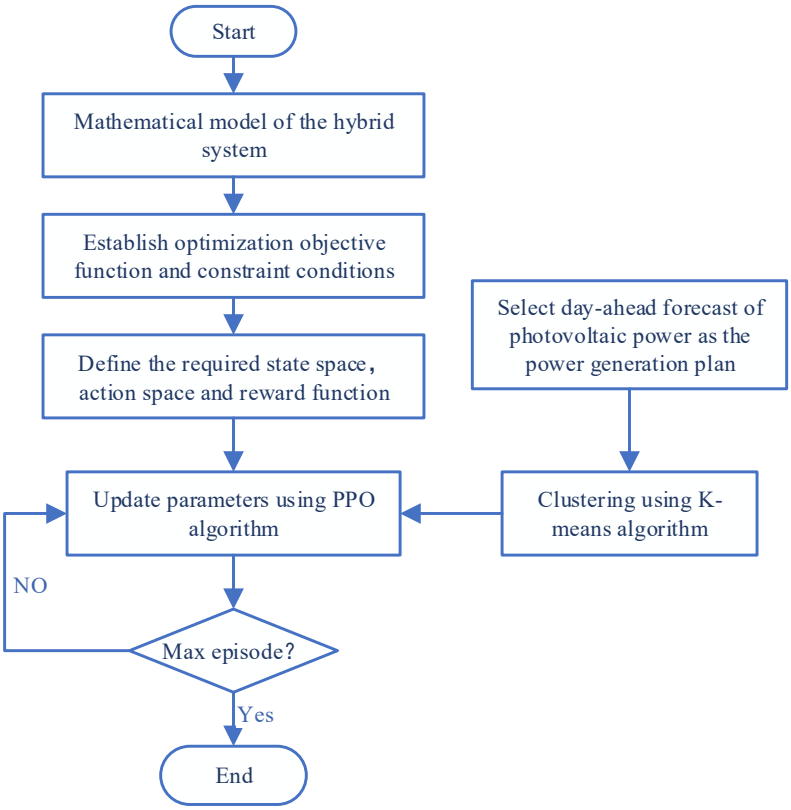

Figure 2. Solution flow chart.

3.1. Scenario Clustering Based on the K-Means Algorithm

As weather conditions vary, the generation power of the photovoltaic power station is a lot different from day to day, which brings difficulties to the control. Before designing the control algorithm, it is essential to cluster the scenarios on different days.

Because there is no definite classification for power generation plans of different days, the clustering algorithm is chosen to cluster different power generation plans, and among the original many power generation plans similar ones are clustered into one class. The K-means algorithm is widely used by researchers because of its simple principle and fast convergence, so it is chosen as the clustering algorithm for power generation plans in this study.

In this paper, the mean, standard deviation, and kurtosis of the power generation plan are taken as characteristics to divide the different power generation plans into different scenarios. All three metrics are used to measure the characteristics of the generation schedule curve. The mean represents the average of the generation schedule at 96 time points per day. (The photovoltaic power output is measured every 15 min, so there are 96 photovoltaic power generation data in a day). The standard deviation reflects the degree of dispersion of the generation schedule. The mean and the standard deviation are the two most important measures to describe the trend of data concentration and the degree of dispersion. Kurtosis is used to measure the steepness of the probability distribution of the generation schedule. The K-means clustering algorithm is a cluster analysis algorithm with an iterative solution [18]. By calculating the distance between different generation plan characteristics, similar generation plans are clustered into one scenario. The clustering process is as follows:

(i) Determine the clustering features: The mean, standard deviation, and kurtosis of the power generation plan are taken as clustering features.
(ii) Select cluster center: Select k objects from the data as the initial clustering center.

(iii) Calculate the distances from each set of features to all cluster centers. Among the calculated distances, each feature that has the minimum distance with the cluster center would be classified into that cluster.
(iv) The center of each cluster is the average of all the features in the corresponding cluster.
(v) Calculate the clustering cost function.
(vi) Stop the calculation if the cost function is below a certain threshold value or the improvement over the previous iteration is below a certain tolerance.

After the above steps, different daily power generation plans in the historical data are divided into k scenarios. Because different scenarios have different control parameters, a reinforcement learning algorithm is adopted to choose a more appropriate charging/discharging strategy in each scenario.

3.2. Tracking Control Based on Reinforcement Learning

Reinforcement learning is a self-learning mechanism that establishes the mapping relationship between environmental states and actions by training agents to constantly interact with the environment [19], as shown in Figure 3.

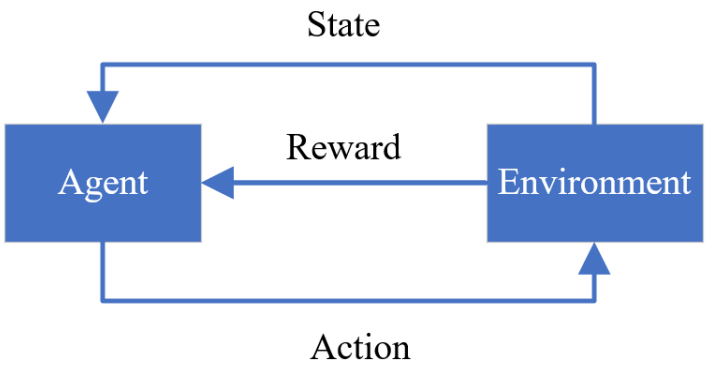

Figure 3. Reinforcement learning.

Reinforcement learning regards learning as a process of exploration and exploitation. The agent chooses an option for action based on the environmental information. After the action is received by the environment, the state changes accordingly and generates reward or punishment feedback for the agent. The agent chooses action according to the reinforcement signal (reward) and the observed state in the current environment and repeats this process until the last state or until the end condition is reached.

During constant interactions with the environment, the agent constantly learns the optimal control strategy that maximizes the total reward value in the whole process. In reinforcement learning, policy-based approaches are more applicable for continuous state and action space problems than value-based approaches.

The PPO algorithm was proposed by Schulman et al. [20]. It is a reinforcement learning algorithm proposed by OpenAI. It can quickly learn the correct strategy in complex scenarios and solve the problem of continuous action and continuous state. Among many reinforcement learning algorithms, the PPO algorithm has the advantages of strong adaptability and stable training. Since the actual residual capacity of the energy storage system and the generation schedule are continuous variables, the PPO algorithm applies to the study of this paper. The PPO algorithm is used to optimize the charging and discharging power decisions of the energy storage system, so that the power generated by the photovoltaic power system can follow the power generation schedule as closely as possible in the regulation of the energy storage system.

As the PPO algorithm is adopted to make sequence decisions, the corresponding state space, action space, and reward should be set according to the problem to be solved. The mathematical function of the hybrid system is analyzed in Section 2. The corresponding state space, action space, and reward function are set as follows.

(1) State space

The definition of the state space is shown in (7). The residual capacity of the ESS $C_{ess}(t-1)$, the ultra-short-term forecast of the photovoltaic power generation $P_{pv-pre}(t)$ and the power generation plan $P_{aim}(t)$ (day-ahead forecast of photovoltaic power generation) is selected as the state space.

$$s(t) = \left\{C_{ess}(t-1), P_{pv-pre}(t), P_{aim}(t)\right\}. \tag{7}$$

(2) Action space

The charging/discharging power of the ESS $P_{ess}(t)$ is selected as the action space, as shown in (8).

$$a(t) = \{P_{ess}(t)\}. \tag{8}$$

(3) Reward function

The reward function in this paper is composed of the objective function. Since reinforcement learning aims to maximize the reward, the negative value of the objective function is taken as the reward function as shown in (9).

$$r(t) = -\alpha\left|C_{ess}(t) - C_{ideal}\right| - \beta\left|P_{pv}(t) + P_{ess}(t) - P_{aim}(t)\right|. \tag{9}$$

Algorithm 1 shows the process of the PPO algorithm. The output power determination procedure based on the PPO algorithm is shown in Figure 4. In the offline process of training, the neural network is trained, and after the training is completed, the state is directly inputted into the trained neural network to complete the online application.

Algorithm 1. PPO algorithm.

Train:

Initialize: policy parameter θ, replay buffer $\mathcal{B}$, and the number of iterations N
for i = 1 to N **do**
Initialize the environmental information s_1.
  for t = 1 to T **do**
    Sample action a_t according to π_θ
    Calculate the reward r_t and observe the next state s_{t+1};
    Store transitions (s_t, a_t, r_t, s_{t+1}) in $\mathcal{B}$
    Compute advantages $\hat{A}_t$ with Eq.(10)
  end for
  for *epoch* = 1 to K **do**
    Sample mini-batches from $\mathcal{B}$
      Update θ by the gradient method with Equations (11) and (12)
end for
  Clear $\mathcal{B}$
end for
Save the policy parameter θ

Test:

Initialize s_t
for t = 1 to T **do**
  Sample a_t according to π_θ
  Execute a_t, and update the environment state to s_{t+1}

end for

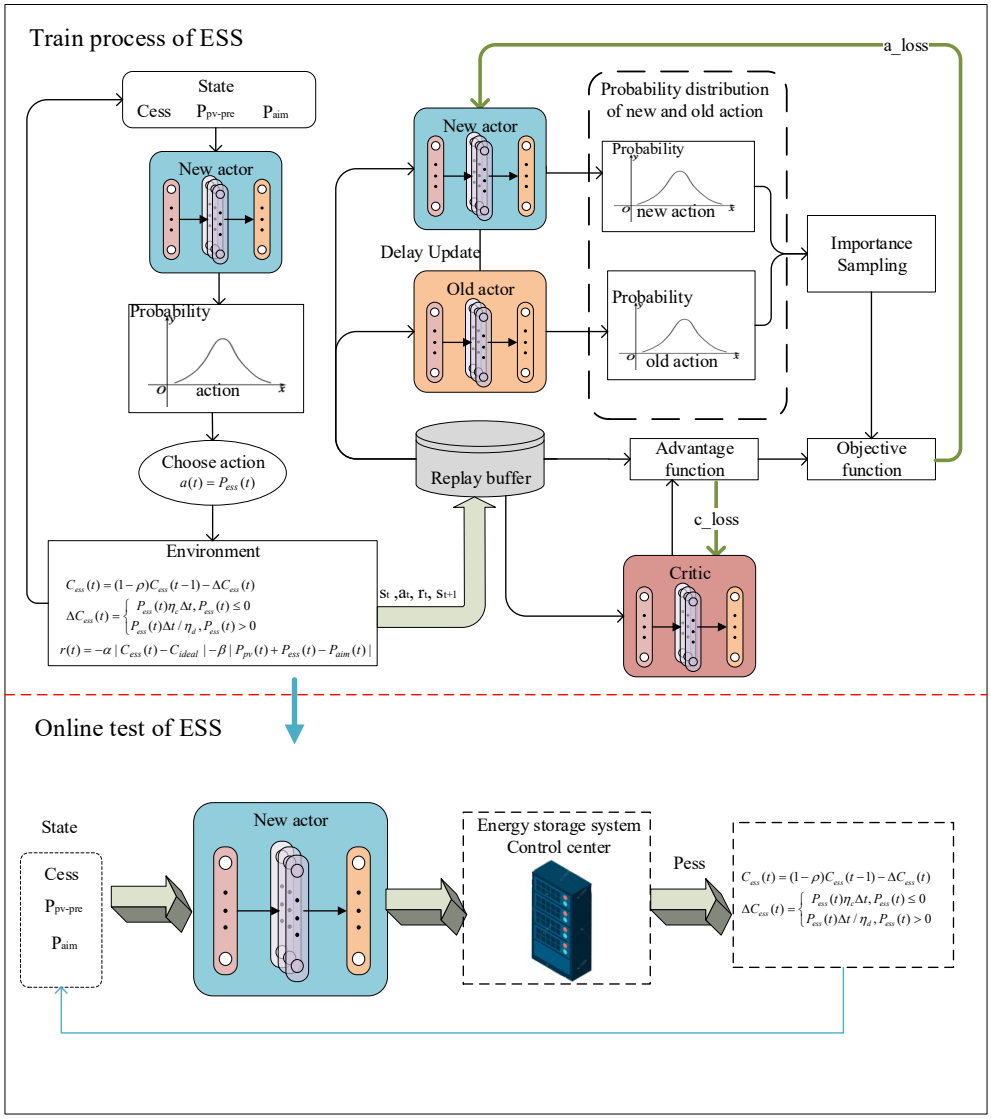

Figure 4. Control strategy based on the PPO algorithm.

The advantage function ($\hat{A}_t$) is used to evaluate the advantage value of taking the current action in the current state versus taking the average action as shown in Equation (10).

$$\hat{A}_t = r_t + \gamma r_{t+1} + \gamma^{T-t+1} r_{T-1} + \gamma^{T-t} V(s_T) - V(s_t), \tag{10}$$

where T is the maximum length of a trajectory, γ is a discounted factor, and $V(s_t)$ denotes the value expectation of state s_t.

The parameter update formula of the PPO algorithm is as follows:

$$\theta_{k+1} = \underset{\theta}{\operatorname{argmax}} \underset{s,a \sim \pi\theta_k}{E} [L_t(\theta)], \tag{11}$$

where $L^{\mathrm{clip}(\theta)}$ is the objective function.

$$L_t(\theta) = \sum_{(s_t,a_t)} \min\{r_t(\theta)\hat{A}_t, clip[r_t(\theta), 1-\varepsilon, 1+\varepsilon]\hat{A}_t\}, \tag{12}$$

where ε is the maximum difference between the old and new probability ratios. $\hat{A}_t$ is the advantage function; when $\hat{A}_t > 0$, if $r_t(\theta) > 1+\varepsilon$, the upper limit value is $(1+\varepsilon)\hat{A}_t$; when $\hat{A}_t < 0$, if $r_t(\theta) < 1-\varepsilon$, the lower limit value is $(1-\varepsilon)\hat{A}_t$.

4. Simulation Analysis

4.1. Description of Photovoltaic and Energy Storage Hybrid System

All work mentioned in this paper is based on actual historical power data, day-ahead forecast of the photovoltaic power generation data, and ultra-short-term forecast of photovoltaic power generation data of a 40 MW photovoltaic energy storage power station in Belgium from March to May 2019. The daily power forecast data will be used to provide a strong reference for generation planning for the photovoltaic power station. Considering the actual needs, the allowable error between the output power of the hybrid system and the power generation plan is taken as $\varepsilon = 3\%$.

The maximum storage capacity of the ESS is 20 MW·H, the maximum charging/discharging power value is 10 MW, the charging/discharging efficiency is 1; self-discharge rate $\rho = 0$, $SOC_{\min} = 0.1$, $SOC_{\max} = 0.9$, $\alpha = 0.1$, $\beta = 0.9$, and $\Delta t = 15$ min. The hyper-parameters used in the PPO algorithm are shown in Table 1.

Table 1. The hyperparameters of PPO algorithm.

Parameter	Value
Minibatch size	32
Actor neural network size	(3,32,64,128,1)
Critic neural network size	(3,16,64,1)
Actor learning rate	2×10^{-5}
Critic learning rate	4×10^{-5}
Clipping parameter	0.2
Discount factor	0.9
Maximum episode	1500

4.2. Results and Discussions

According to the mean, standard deviation, and kurtosis of the power generation plan, days from March to May 2019 in Belgium are clustered, and the different days are divided into four scenarios. The four scenarios from March to May are 26 days, 13 days, 28 days, and 25 days, respectively. As shown in Figure 5. The four colors represent four types of scenarios. The generation schedule for one day for each type of scenario is selected and plotted in Figure 6, and it can be seen that the generation schedules for the four scenarios are quite different.

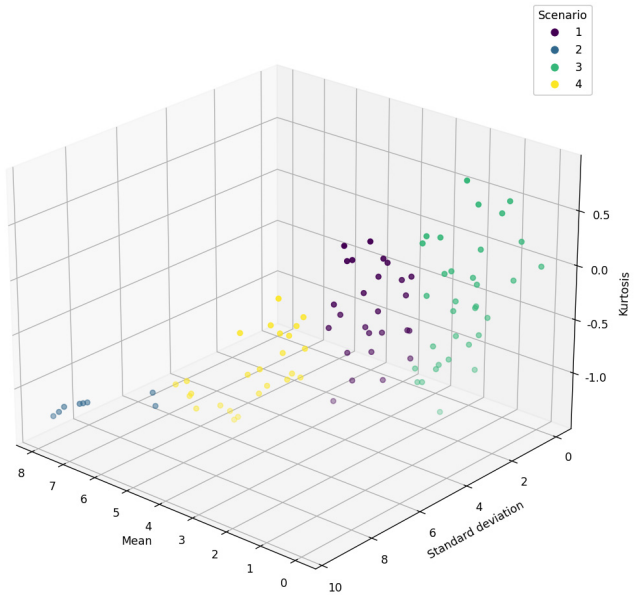

Figure 5. Clustering results.

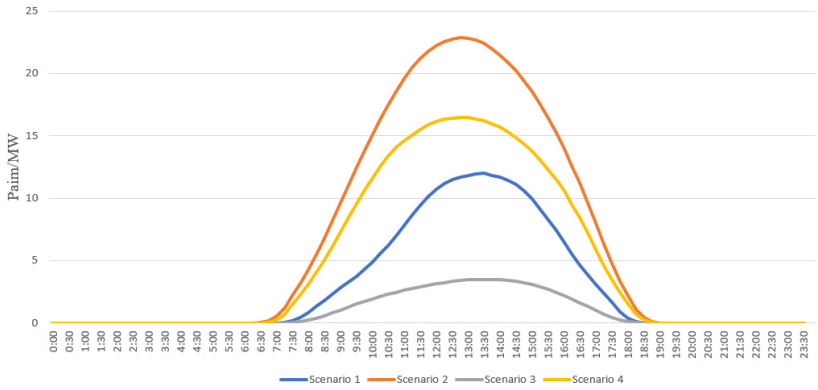

Figure 6. Power generation plans in different 4 scenarios.

In the four scenarios, the process of optimizing the charging/discharging power is the same. The data collected on multiple days in each scenario is used for training, and the remaining data in each scenario is used for testing. The testing results in Scenario 1 are shown in Figure 7. The actual photovoltaic power generation fluctuates greatly, and several periods are outside the range of the power generation plan, deviating from the power generation plan. After the PPO algorithm optimizes the charge/discharge power, the power generation of the hybrid system satisfies the power generation requirements.

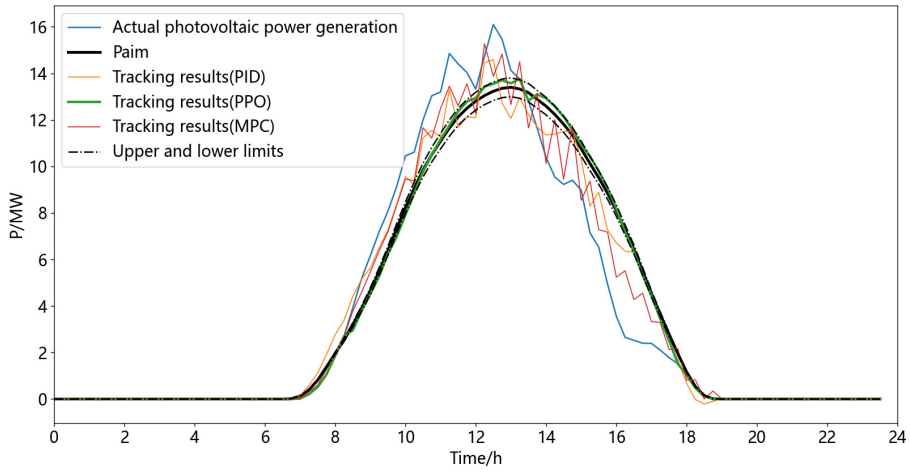

Figure 7. Tracking results of the power generation plan.

It can be seen from Figures 8 and 9 that the proposed method can satisfy all the constraint conditions of the hybrid system. Figure 10 shows the mean deviation under the condition of no ESS, the PID algorithm with the ESS, the MPC algorithm with the ESS and the PPO algorithm with the ESS. Adding the ESS would reduce the deviation, and the deviation from using the PPO algorithm is smaller than that of using PID and MPC. The maximum deviation, average deviation, probability of deviation less than 3%, and execution time in three conditions are shown in Table 2. "Max deviation" and "Mean deviation" represent the maximum deviation and mean deviation between the hybrid generation system and the generation plan after the energy storage system has completed charge/discharge in a day, respectively. "Probability of deviation of less than 3%" represents the probability that the generation plan deviates from the hybrid system generation between 97% and 103% of the generation plan for 96 time points in a day. "Time" represents the average control time. The smaller "Max deviation", "Mean deviation", and "Time", the better; the larger "Probability of deviation of less than 3%", the better. It shows that the maximum deviation, average deviation, and computing time using the PPO algorithm

are smaller than those using PID or MPC, and the Probability of deviation of less than 3% using the PPO algorithm is larger than that of using the PID and MPC method. It is shown that the PPO method can successfully make the power generation of the hybrid system track power generation plans.

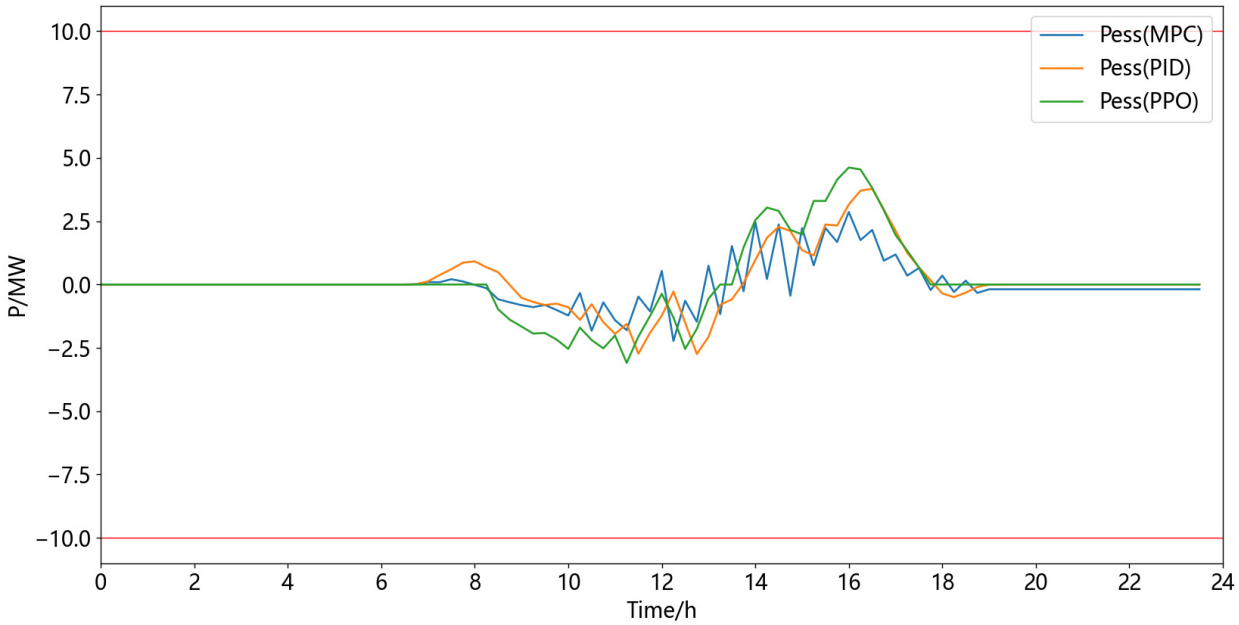

Figure 8. Output power of the ESS.

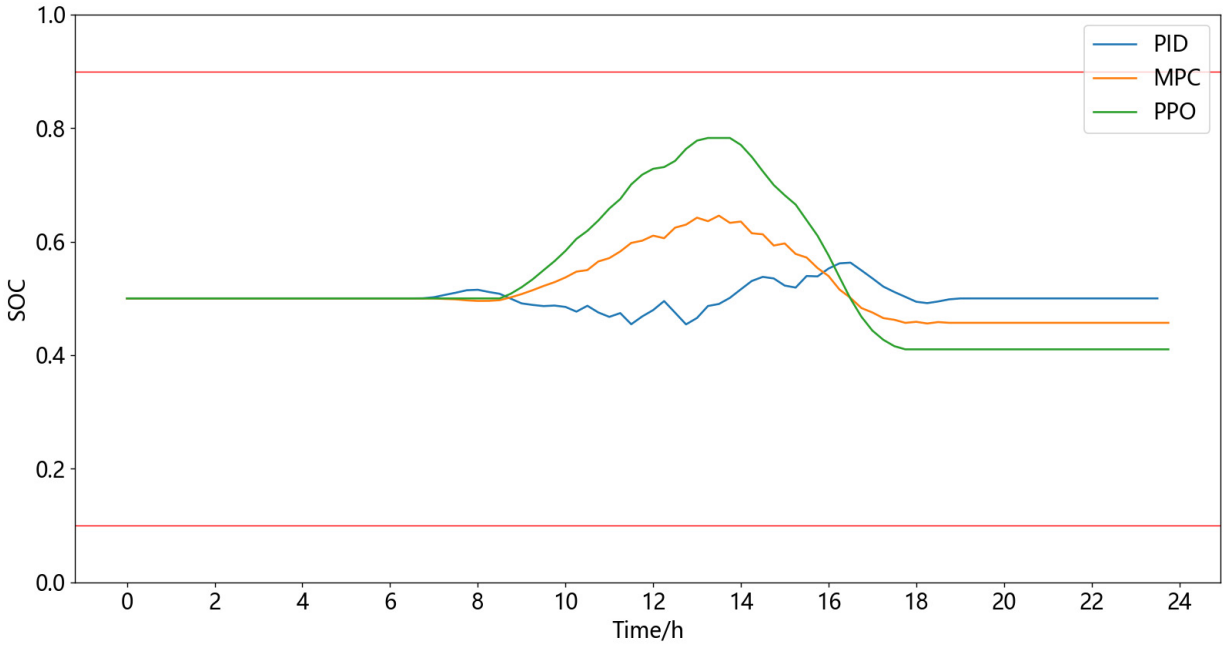

Figure 9. SOC of the ESS.

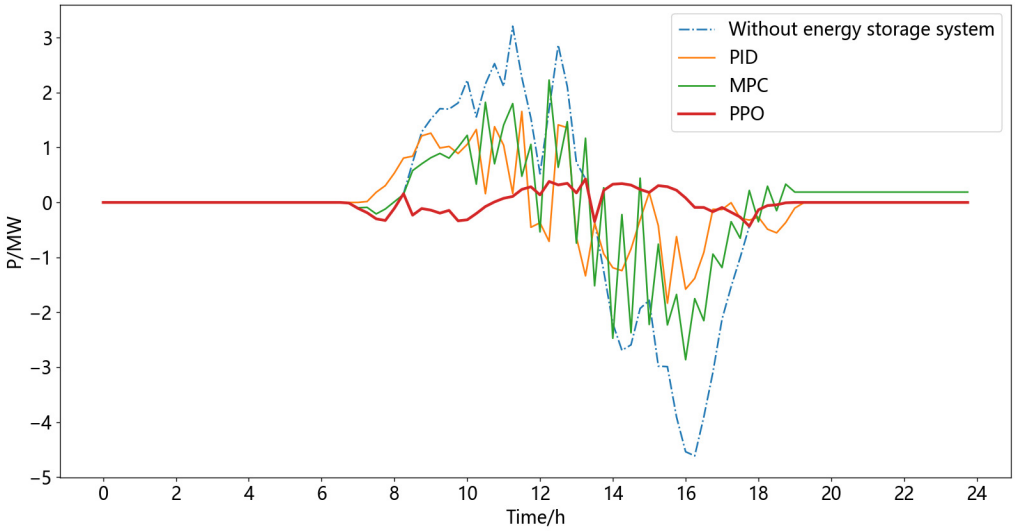

Figure 10. The average errors of tracking for no storage system, and two storage system with PID, MPC, and PPO.

Table 2. Deviation of day 1 (Scenario 1).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability of Deviation of Less than 3%**	**Time/s**
Without an ESS	4.615	0.843	51.04%	
PID	1.835	0.374	64.58%	1.134
MPC	4.395	0.609	54.17%	2.760
The proposed PPO-based control method	**0.377**	**0.104**	**85.40%**	**0.053**

To check the universality of the proposed method, this paper used data from another two days in Scenario 1 for comparison experiments (Tables 3 and 4). In addition, the tracking results under different scenarios are shown in Tables 5–7. The experimental results show that the proposed method is better than PID and MPC on different days in terms of generalization and time consumption.

Table 3. Deviation of day 2 (Scenario 1).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability of Deviation of Less than 3%**	**Time/s**
Without an ESS	3.040	0.808	47.92%	
PID	1.098	0.280	54.16%	1.069
MPC	1.996	0.468	53.13%	1.273
The proposed PPO-based control method	**0.598**	**0.133**	**82.29%**	**0.049**

Table 4. Deviation of day 3 (Scenario 1).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability of Deviation of Less than 3%**	**Time/s**
Without an ESS	3.514	0.716	45.83%	
PID	1.528	0.314	56.25%	0.341
MPC	1.912	0.419	42.71%	1.357
The proposed PPO-based control method	**0.448**	**0.106**	**78.13%**	**0.055**

Table 5. Deviation (Scenario 2).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability Of Deviation of Less than 3%**	**Time/s**
Without an ESS	3.005	0.690	50%	
PID	1.666	0.498	58.33%	0.331
MPC	2.831	0.591	55.83%	1.389
The proposed PPO-based control method	**0.481**	**0.131**	**83.33%**	**0.047**

Table 6. Deviation (Scenario 3).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability of Deviation of Less than 3%**	**Time/s**
Without an ESS	5.102	0.571	65.63%	
PID	3.237	0.432	66.67%	0.334
MPC	4.112	0.816	54.17%	2.150
The proposed PPO-based control method	**1.109**	**0.137**	**84.37%**	**0.051**

Table 7. Deviation (Scenario 4).

	The Evaluation Index			
	Max Deviation/ MW	**Mean Deviation/ MW**	**Probability of Deviation of Less than 3%**	**Time/s**
Without an ESS	2.873	0.293	60.41%	
PID	0.827	0.135	58.33%	0.355
MPC	1.540	0.162	61.46%	2.321
The proposed PPO-based control method	**0.460**	**0.063**	**69.79%**	**0.049**

5. Conclusions

This paper presents a scheduling strategy for photovoltaic and energy storage hybrid systems based on the PPO algorithm. The proposed method can adapt to the uncertainty of photovoltaic power generation by learning the historical output data of photovoltaic power

generation and has good generalization. The experimental results show the feasibility and effectiveness of the control strategy. However, in this paper, only the deviation between the residual capacity of the energy storage system and the ideal capacity is used to measure the economics and lifetime of energy storage. The smaller deviation between the residual capacity of the energy storage system and the ideal capacity means that the charging and discharging consumption of the energy storage system is smaller and the economy is better. However, in the actual process, the focus of the economy on the consumption of the energy storage system including charging and discharging, the operation of the energy storage system, the construction of the energy storage system, and post-maintenance would be considered. This aspect will be studied in the future.

Author Contributions: Methodology, M.G. and J.C.; Software, M.G.; Validation, M.R. and Z.Y.; Formal analysis, J.C.; Investigation, M.R. and L.C.; Writing—original draft, M.G. and M.R.; Writing—review & editing, J.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (61973226 and 62073232) and the Shanxi Provincial Natural Science Foundation, China (20210302123189).

Data Availability Statement: The data used in this article comes from Solar-PV power generation data (elia.be).

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

References

1. Dai, H.; Su, Y.; Kuang, L.; Liu, J.; Gu, D.; Zou, C. Contemplation on China's Energy-Development Strategies and Initiatives in the Context of Its Carbon Neutrality Goal. *Engineering* **2021**, *7*, 1684–1687. [CrossRef]
2. Hong, F.; Song, J.; Meng, H.; Wang, R.; Fang, F.; Zhang, G. A novel framework on intelligent detection for module defects of PV plant combining the visible and infrared images. *Solar Energy* **2022**, *236*, 406–416. [CrossRef]
3. Qin, L.; Sun, N.; Dong, H.Y. Adaptive Double Kalman Filter Method for Smoothing Wind Power in Multi-Type Energy Storage System. *Energies* **2023**, *16*, 1856. [CrossRef]
4. Li, S.; Xu, Q.; Huang, J. Research on the integrated application of battery energy storage systems in grid peak and frequency regulation. *J. Energy Storage* **2023**, *59*, 106459. [CrossRef]
5. Ding, H.; Hu, Z.; Song, Y. Stochastic optimization of the daily operation of wind farm and pumped-hydro-storage plant. *Renew. Energy* **2012**, *48*, 571–578. [CrossRef]
6. Mignoni, N.; Scarabaggio, P.; Carli, R.; Dotoli, M. Control frameworks for transactive energy storage services in energy communities. *Control Eng. Pract.* **2023**, *130*, 105364.
7. Guo, X.; Xu, M.; Wu, L.; Liu, H.; Sheng, S. Review on Target Tracking of Wind Power and Energy Storage Combined Generation System. In Proceedings of the 2018 2nd International Conference on Power and Energy Engineering, Xiamen, China, 3–5 September 2018.
8. Chunguang, T.; Li, T.; Dexin, L.; Xiangyu, L.; Xuefei, C. Control Strategy for Tracking the Output Power of Photovoltaic Power Generation Based on Hybrid Energy Storage System. *Trans. China Electrotech. Soc.* **2016**, *31*, 75–83.
9. Yan, H.; Li, X.J.; Ma, X.F.; Hui, D. Wind power output schedule tracking control method of energy storage system based on ultra-short term wind power prediction. *Power System Technol.* **2015**, *39*, 432–439.
10. Venkatesan, K.; Govindarajan, U. Optimal power flow control of hybrid renewable energy system with energy storage: A WOANN strategy. *J. Renew. Sustain. Energy* **2019**, *11*, 015501.
11. Qi, X.; He, S.; Wang, W.; Jiang, C.; Hao, L.; Zhu, W. FMPC based control strategy for tracking PV power schedule output of energy storage system. *Renew. Energy Resour.* **2019**, *37*, 354–360.
12. Zhang, S.; Ma, C.; Yang, Z.; Yang, Z.; Wang, Y.; Wu, H.; Ren, Z. Joint dispatch of wind-photovoltaic-storage hybrid system based on deep deterministic policy gradient algorithm. *Electric Power* **2023**, *56*, 68–76.
13. Yang, D.; Wang, L.; Yu, K.; Liang, J. A reinforcement learning-based energy management strategy for fuel cell hybrid vehicle considering real-time velocity prediction. *Energy Convers. Manag.* **2022**, *274*, 116453. [CrossRef]
14. Zhou, K.; Zhou, K.; Yang, S. Reinforcement learning-based scheduling strategy for energy storage in microgrid. *J. Energy Storage* **2022**, *51*, 104379. [CrossRef]
15. Kolodziejczyk, W.; Zoltowska, I.; Cichosz, P. Real-time energy purchase optimization for a storage-integrated photovoltaic system by deep reinforcement learning. *Control Eng. Pract.* **2021**, *106*, 104598. [CrossRef]
16. Yang, T.; Zhao, L.; Li, W.; Zomaya, A.Y. Dynamic energy dispatch strategy for integrated energy system based on improved deep reinforcement learning. *Energy* **2021**, *235*, 121377. [CrossRef]

17. Xinlei, C.; Cui, Y.; Kai, D.; Zijie, M.; Yuan, P.; Zhenfan, Y.; Jixing, W.A.; Xiangzhan, M.; Yang, Y. Day-ahead optimal scheduling approach of wind-storage joint system based on improved K-means and MADDPG algorithm. *Energy Stor. Sci. Technol.* **2021**, *10*, 2200–2208.
18. Wang, Q.; Wang, C.; Feng, Z.; Ye, J.F. Review of K-means clustering algorithm. *Electr. Design Eng.* **2012**, *20*, 21–24.
19. Feinberg, V.; Wan, A.; Stoica, I.; Jordan, M.I.; Gonzalez, J.E.; Levine, S. Model-Based Value Estimation for Efficient Model-Free Reinforcement Learning. *arXiv* **2018**, arXiv:1803.00101.
20. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal Policy Optimization Algorithms. *arXiv* **2017**, arXiv:1707.06347v2.