



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/200675/>

Version: Accepted Version

---

**Article:**

Wang, X., Zhou, J., Qin, B. et al. (2023) Coordinated power smoothing control strategy of multi-wind turbines and energy storage systems in wind farm based on MADRL. IEEE Transactions on Sustainable Energy. ISSN: 1949-3029

<https://doi.org/10.1109/tste.2023.3287871>

---

© 2023 The Authors. Except as otherwise noted, this author-accepted version of a journal article published in IEEE Transactions on Sustainable Energy is made available via the University of Sheffield Research Publications and Copyright Policy under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is properly cited. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: <https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.

# Coordinated Power Smoothing Control Strategy of Multi-Wind Turbines and Energy Storage Systems in Wind Farm based on MADRL

Xin Wang, *Member, IEEE*, Jianshu Zhou, Bin Qin, Lingzhong Guo, *Member, IEEE*

**Abstract--** The randomness and volatility of wind power greatly affect the safety and economy of the power systems, and the wake effect of the wind farm aggravates the wind energy loss and the wind power fluctuation. Taking into consideration the wake effect of the wind farm, a new coordinated wind power smoothing control strategy for multi-wind turbines (M-WT) and energy storage systems (ESS) is proposed. The proposed method is based on a multi-agent deep reinforcement learning (MADRL), in which the relationship between output power and wake effect is firstly analyzed, and a power smoothing control model of the M-WT and ESS is established. MADRL is then introduced to optimize the power control of M-WT and ESS. In order to further increase the learning and training efficiency, an improved MADRL algorithm based on the partitioned experience buffer and prioritized experience replay is proposed, where the experience buffer is divided into positive, negative, and neutral experiences, and the experiences are sampled according to experience priority. The effectiveness of the proposed strategy is verified on the SimWindFarm platform. The results show that the proposed control strategy can maximize the economic benefits while further smoothing wind power fluctuations and increasing power generation.

**Index Terms--** wind farm, energy storage systems, power control, wake effect, multi-agent deep reinforcement learning

## I. INTRODUCTION

Wind power generation is one of the most important methods to solve environmental pollution and energy crises [1]. Smoothing wind power fluctuation can effectively reduce its negative influence on the reliability and stability of power systems [2]. Increasing the active power and generating capacity of a wind farm has become an important way for wind farm industries to reduce investment risks [3].

At present, there are two methods to stabilize wind power fluctuations: installing energy storage systems and power smoothing control of a wind turbine (WT) [4-7]. The former can smooth wind power fluctuations effectively with high operability and little power loss but requires additional equipment costs. In [7], a hybrid energy storage system of supercapacitors and lithium batteries was used to smooth wind power fluctuations, and a stochastic optimization scheduling strategy for wind power smoothing was proposed. Lin et al. [8]

proposed a long-term stable operation control method for wind power smoothing based on a dual-battery energy storage system (DBESS), to ensure the long-term stable operation of DBESS while meeting the wind power demand.

The latter is generally realized through pitch angle control, rotor speed control, DC side capacitor control, as well as grid-connected side converter control [9,10]. Liao et al. [11] proposed a low-pass virtual filter method for power smoothing control of wind power generation systems. A new low-pass virtual filter in the rotor energy control loop of a wind power generation system is introduced so that the system has more power smoothing capability and stability. Xue et al. [12] proposed a power smoothing control strategy based on an adaptive capability of WT. Power smoothing can be achieved through DC voltage control, rotor speed control and pitch angle control. The power control of WT is to smooth the power fluctuations at the cost of losing some power and increasing the fatigue load of the wind turbine. Most of the power smoothing control methods in the existing literatures only consider at the individual wind turbine level, while the actual power smoothing control should be considered at the wind farm level. Moreover, power smoothing at the individual WT level does not necessarily represent the total output power smoothing of the wind farm.

In terms of the power smoothing control of a wind farm, Zhu et al. [13] proposed a power smoothing control strategy for a wind farm based on power distribution. The output power of wind turbines (WTs) is controlled through the machine-side and grid-side converter control, and the output power of each WT is controlled by setting power distribution rules. This method can ensure the maximum output power of the wind farm while smoothing the output power fluctuations. However, the wind farm wake effect is not considered, which may cause the uneven distribution of wind speed, affect the operation status of each WT in the wind farm, further decrease the output power of the WTs and increase its fluctuations. Considering the wake effect of a wind farm, Howlader et al. [14] proposed a smooth wind power coordinated control strategy for multi-wind turbines (M-WT). The influence of the wake effect on the output wind power from the perspective of M-WT in the wind farm is considered. In addition, the wake effect of the wind farm on power smoothing under different tower spacing is also considered.

---

This work was supported in part by the National Natural Science Foundation of China under Grants 62033014 and 61673166, and in part by the Natural Science Foundation of Hunan Province under Grants 2021JJ50006 and 2022JJ50074.

Xin Wang, Jianshu Zhou, and Bin Qin are with the School of Electrical & Information Engineering, Hunan University of Technology, Zhuzhou, 412007, China (e-mail: qinbin99p@163.com).

Lingzhong Guo is with the Department of Automatic Control and Systems Engineering, The University of Sheffield, Sheffield, S1 3JD, UK.

However, the control structure in the above literatures is designed based on the wind farm modeling, ignoring the error and uncertainty of the model.

In the process of wind farm control, accurate wind farm dynamics analysis is required, and the inevitable modeling errors and uncertainties lead to significant degradation of the control performance. In contrast, Deep Reinforcement Learning (DRL) [15] can interact with complex environment with no models or inaccurate models to search optimal control strategies that can achieve long-term rewards and enhance adaptability and robustness. Aiming at the wake effect between WT's and the randomness of the environment, a robust deep reinforcement learning method was proposed to deal with uncertain environmental conditions and strong aerodynamic interaction between WT's to realize wind farm power tracking [16]. Huang et al. [17] proposed a DRL-based control strategy for wind-solar energy storage systems to maximize the long-term benefits. The limitation of the method is that it is difficult to solve the M-WT control problem by a single agent, and the complex control problem needs to be decomposed into a multi-agent cooperative problem. Multi-agent deep reinforcement learning (MADRL) [18] applies DRL's principles and algorithms to multi-agent systems. It can organize multi-agents to conduct self-learning and realize cooperative solutions to complex problems through the interaction between agents. In addition, compared with a single agent, multiple agents can share risks and improve system reliability. Therefore, MADRL has the potential to solve the control problems of complex uncertain, and nonlinear systems such as the wind farm.

In this study, aiming at overcoming the limitations of the current wind power smoothing methods, a MADRL-based coordinated control strategy for M-WT and energy storage systems (ESS) is proposed. The mainstream WT control method is only studied for controlling individual WT. The smoothing power of individual WT does not necessarily represent the smoothing output power of a wind farm. Moreover, such methods do not apply to wind farms consisting of multiple turbines. In this paper, the output power of individual WT in the wind farm is coordinated and controlled so that the sum of the powers of the WT's is smoothed. The proposed method is studied at the wind farm level, which avoids the inapplicability of individual WT control to the wind farm. Due to the high controllability and fast response of ESS, the ESS is used to smooth the high-frequency fluctuations that are difficult to be handled by the internal control of the wind farm. The M-WT coordinated smooth power control can smooth part of the power fluctuations in the wind farm, undertaking the task of power smoothing. The ESS of the proposed method deals with fewer power fluctuations than that of the individual control, and the ESS capacity configuration can be appropriately reduced, which reduces the investment of ESS cost. At present, the wind farm model is difficult to establish accurately, and the inaccurate model will lead to the unsatisfactory control performance. To solve this problem, a power optimization control of the M-WT and ESS based on a multi-agent twin delayed deep deterministic policy gradient (MATD3) algorithm is proposed. The MATD3 algorithm is used to optimize the power control of the M-WT and ESS, and the power is corrected and compensated when the model has errors or the parameters are time-varying, so as to reduce the

negative impact caused by the inaccurate model. In addition, the computational complexity and experiences of the MATD3 algorithm under multi-agent tasks will increase exponentially, the learning ability of the algorithm will decrease and the convergence speed will slow down. In response to the problem, an improved MATD3 algorithm, based on the partitioned buffer and priority experience replay, is proposed to enhance the efficiency of the MATD3 algorithm, where the experience buffer is divided into positive experiences, negative experiences, and neutral experiences, and then the experiences are sampled according to the experience priority.

The main contributions of the paper are as follows:

1) Different from the individual power smoothing control of WT and ESS, the proposed control strategy combines the M-WT and ESS smooth power controls. Some power fluctuations are smoothed through coordinated power control among wind turbines, while the ESS smooth high-frequency fluctuations that are difficult to be handled by internal control of the wind farm. The M-WT and ESS bear wind power fluctuations and relieve the pressure of smooth power to each other.

2) A new power optimization control method based on MATD3 for M-WT and ESS is proposed to reduce the negative effect caused by model uncertainty and to enhance reliability and robustness. Meanwhile, the power smoothness of the wind farm, the power generation of the wind farm, the load of the WT, and the loss of the ESS are taken as the reward functions of MATD3, to reduce the system loss on the premise of ensuring the power smoothness.

3) An improved MATD3 algorithm based on the partitioned experience buffer and priority experience replay (PEPE-MATD3) is proposed to enhance algorithm efficiency, where the experience buffer is divided into positive experiences, negative experiences, and neutral experiences according to the reward values of learning. The experiences are preferentially sampled according to the experience priority to filter out more useful experiences for policy learning, so as to improve the learning ability of the algorithm.

The rest of this article is organized as follows. Section 2 introduces the coordinated control system of an offshore wind farm and ESS. Section 3 introduces a power-optimized compensation based on PEPE-MATD3. Section 4 verifies the effectiveness and feasibility of the proposed method through SimWindFarm simulations, and conclusions are drawn in Section 5.

## II. COORDINATED CONTROL SYSTEM OF WIND FARM AND ESS

### A. System structure

The structure of the combined power generation system of an offshore wind farm and ESS is shown in Fig. 1. The offshore wind farm structure is mainly composed of M-WT, ESS, transformers, and controller. According to the data of the wind farm and the ESS collected by the wind farm monitoring system, the controller controls the output power of the WT's and the ESS of the wind farm.

The power equation of the combined generation system of the offshore wind farm and the ESS is as follows:

$$P_{\text{farm}} = \sum_{i=1}^n P_{WTi} \quad (1)$$

$$P_{\text{grid}} = P_{\text{farm}} + P_{\text{es}} \quad (2)$$

where  $P_{WT_i}$  is the output power of the  $i^{\text{th}}$  wind turbine,  $P_{\text{farm}}$  is the output power of the wind farm,  $P_{\text{es}}$  is the output power of the ESS, and  $P_{\text{grid}}$  is the grid-connected power.

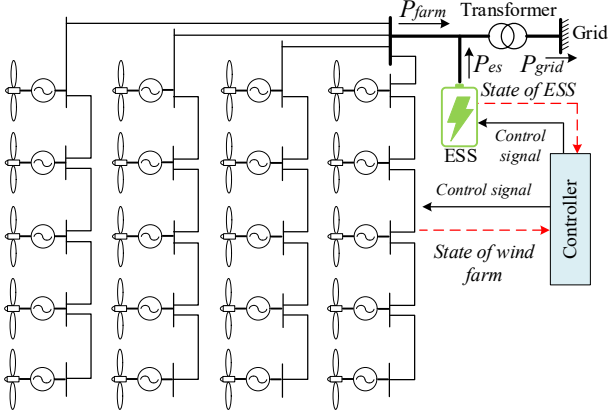


Fig.1 Combined power generation system of an offshore wind farm and ESS

### B. Overall control strategy

As shown in Fig.2, the control structure consists of the power control based on a wake model and the power optimization based on MADRL. The former is composed of a wake model and the reference power setting. In the wake model part, the input wind speed of each wind turbine  $[V_1, V_2, \dots, V_n]$  is calculated according to the input wind speed of the wind farm  $V_{\text{farm}}$  and the thrust coefficients of the WTs  $[C_{T1}, C_{T2}, \dots, C_{Tn}]$ . Meanwhile, the power of WT  $i$   $P_{WT_i, \text{pre}}$  and the power of the wind farm  $P_{\text{farm, pre}}$  are calculated. In the reference power setting part, the low-frequency power of the wind farm  $P_L$  is obtained by filtering  $P_{\text{farm, pre}}$  through Bessel low-pass filtering, and the reference power of each wind turbine  $P_{\text{ref}, i}$  is determined according to the proportion of each wind turbine power  $P_{WT_i, \text{pre}}/P_{\text{farm, pre}}$ . The reference power of the ESS  $P_{\text{ref}, \text{es}}$  is obtained by the difference between  $P_L$  and  $P_{\text{grid}}$ , and the ESS is used to deal with the power fluctuations that are difficult to be handled by WTs.

The power optimization aims to improve wind power generation, further smooth the power fluctuations and reduce the loss of WT and ESS. PEPE-MATD3 is used to optimize the power control of M-WT and the ESS is used to reduce the impact of model errors and uncertainties. The cooperation among agents in PEPE-MATD3 is used to optimize and compensate for the reference power of each wind turbine and the ESS. The input wind speed of each column of the WTs in the wind farm is similar. It is optimized and compensated by an agent, thereby reducing the computational complexity of the PEPE-MATD3. Agent 1 in PEPE-MATD3 optimizes the power with reference adjustment value  $\Delta P_{\text{ref}, 1-3}$ , so as to obtain the new reference power  $P'_{\text{ref}, 1-3}$  of the wind turbines. Other agents act in the same way. The power optimization of the ESS is controlled by a single Agent  $_{\text{es}}$ . The optimization method of the ESS is the same as those of the WTs.

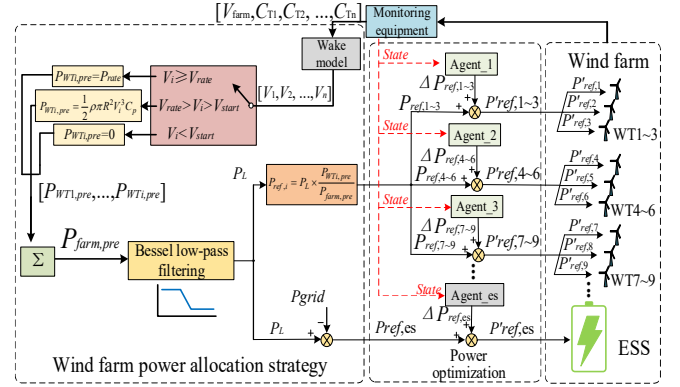


Fig.2 Control structure

### C. Jensen wake model

In practical operation, a simplified wind farm wake model is crucial to reduce the calculation load and improve the real-time performance of active power output dispatching of the wind farm. Therefore, Jensen's wake model can be used to calculate wake flow [20,21]:

$$D_w = D(1 + 2ks) \quad (3)$$

$$V_{1,L} = V_0 \left[ 1 - \frac{1 - \sqrt{1 - C_T}}{1 + 2ks/D} \right]^2 \quad (4)$$

where  $s$  is the distance from a certain position of the downstream wind turbine to the upstream wind turbine,  $V_0$  is the incoming wind speed at infinity, and  $D$  is the diameter of the upstream wind turbine.  $V_{1,L}$  and  $D_w$  are the wind speed and wake section diameter at  $s$  of the upstream turbine in the wake, respectively.  $k$  is the expansion rate.

It is necessary to consider whether the downstream wind turbine is within the wake influence radius of the upstream wind turbine when calculating the wind speed of the downstream wind turbine. If it is within the wake influence radius, the wake influence of the upstream wind turbine should be considered.

The wind wheel wake of a wind turbine affected by a single wake is superimposed. The wind speed at the location of the downstream wind turbine is expressed as:

$$V_{1,L} = V_0 - V_0 \left( 1 - \sqrt{1 - C_T} \right) \left( \frac{D_w}{D} \right)^2 \frac{A_{\text{overlap}}}{A_0} \quad (5)$$

$$D = D_w - 2ks \quad (6)$$

where  $A_{\text{overlap}}$  is the overlap area of the wake of the upstream wind turbine at the downstream wind turbine and wheel.  $A_0$  is the swept area of the downstream wind turbine.

### D. Active power distribution of wind farm

The active power distribution of the wind farm is shown in Fig. 3.  $V_i$  is calculated according to the wake model, and then  $P_{WT_i, \text{pre}}$  is calculated according to Eq. (7),

$$P_{WT_i, \text{pre}} = \begin{cases} P_{\text{rate}} & V_i > V_{\text{rate}} \\ \frac{1}{2} \rho \pi R^2 V_i^3 C_p & V_{\text{rate}} \geq V_i > V_{\text{start}} \\ 0 & V_i \leq V_{\text{start}} \end{cases} \quad (7)$$

where  $P_{\text{rate}}$  is the rated power of the wind turbine,  $\rho$  is the air density,  $R$  is the radius of the wind wheel,  $V_{\text{rate}}$  is the rated wind speed, and  $V_{\text{start}}$  is the starting wind speed.

When  $V_i$  of the  $i^{\text{th}}$  wind turbine is less than  $V_{\text{start}}$ , the wind turbine is in the start mode and  $P_{WTi,pre}=0$ . When  $V_i$  of the  $i^{\text{th}}$  wind turbine is greater than  $V_{\text{start}}$  and less than  $V_{\text{rate}}$ , the wind turbine is in the maximum wind energy tracking mode, and  $P_{WTi,pre}=1/2\rho\pi R^2 V_i^3 C_p$ . When  $V_i$  of the  $i^{\text{th}}$  wind turbine is greater than  $V_{\text{rate}}$ , the wind turbine is in the constant power mode, and  $P_{WTi,pre}=P_{\text{rate}}$ .

$P_{\text{farm,pre}}$  is obtained by summing  $P_{WTi,pre}$  of each wind turbine

$$P_{\text{farm,pre}} = \sum_{i=1}^n P_{WTi,pre} \quad (8)$$

According to  $P_{\text{farm,pre}}$ ,  $P_L$  is obtained by Bessel low-pass filtering. Bessel low-pass filtering extracts  $P_L$  from the original power signal when the power fluctuation is less than 10% of the rated power and within 1 min. The  $P_{\text{ref},i}$  of each wind turbine is calculated as follows,

$$P_{\text{ref},i} = P_L \times \frac{P_{WTi,pre}}{P_{\text{farm,pre}}} \quad (9)$$

It is used as the inputs of the WT internal control, so as to control WT output power. The methods in references [22,23] are adopted for the modeling and control of the WTs. Because the power control of WTs is difficult to fully track the set reference power, ESS is used to deal with the power fluctuations.  $P_{\text{ref,es}}$  is calculated by the difference between  $P_L$  and  $P_{\text{grid}}$ .

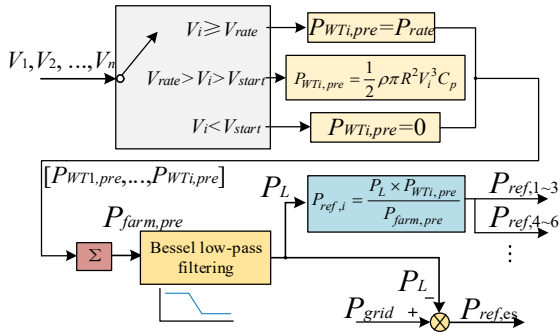


Fig.3 Active power control for wind farm

### E. Active power distribution of wind farm

The control structure of the ESS is shown in Fig. 4.  $P_{\text{ref,es}}$  is optimized by agent\_es of MADRL to obtain the new value  $P'_{\text{ref,es}}$ . Overcharge and discharge protection is incorporated into the ESS control [24,25]. When the SOC of ESS  $> 0.8$ , the overcharge protection acts, and the charging power is multiplied by a charge protection parameter  $K_{\text{es,c}}$ . When the SOC of ESS is  $< 0.2$ , the over-discharge protection acts, and the discharge power is multiplied by a discharge protection parameter  $K_{\text{es,dis}}$ . The calculation of  $K_{\text{es,c}}$  and  $K_{\text{es,dis}}$  are shown in Eq. (10) and Eq. (11), respectively. The error is obtained by the difference between the power reference value protected by SOC  $P^*_{\text{ref,es}}$  and the actual power  $P_{\text{es}}$ , and it is sent to PI controller to control the charge and discharge of the ESS.

$$K_{\text{es,c}} = \begin{cases} 1 & SOC < 0.8 \\ 10 \times (0.9 - SOC) & 0.8 \leq SOC < 0.9 \\ 0 & SOC \geq 0.9 \end{cases} \quad (10)$$

$$K_{\text{es,dis}} = \begin{cases} 0 & SOC < 0.1 \\ 10 \times (SOC - 0.1) & 0.1 \leq SOC < 0.2 \\ 1 & SOC \geq 0.2 \end{cases} \quad (11)$$

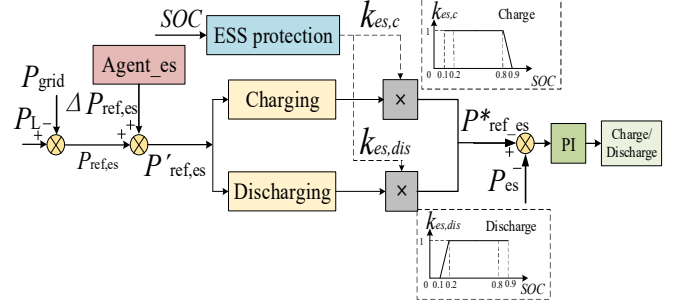


Fig. 4 ESS control structure

## III. POWER-OPTIMIZED COMPENSATION BASED ON PEPE-MATD3

### A. MATD3

As a new MADRL algorithm for solving continuous problems, MATD3 extends TD3 to the multi-agent domain [19,26,27]. MATD3 is similar to the multi-agent deterministic policy gradient (MADDPG) [28], which also uses a centralized training and decentralized execution framework. Each agent has two centralized critics and one actor (policy). This algorithm does not need to establish real communication rules, and it is easy to converge to the global optimum. Two critics  $Q_{i,\theta,2}^\pi$  are introduced into MATD3 and the minimum of the two is taken when calculating the target value to reduce the impact caused by network overestimation. The TD target value  $y_i$  of the  $i^{\text{th}}$  agent,

$$y_i = r_i + \gamma \min_{n=1,2} Q_{i,\theta_n}^\pi(S', \pi_{\phi,1}(S'_1) + \varepsilon, \dots, \pi_{\phi,N}(S'_N) + \varepsilon) \quad (12)$$

where  $r_i$  is the reward value of the  $i^{\text{th}}$  agent,  $\gamma$  is the conversion factor,  $\pi_\phi$  is the strategy of the algorithm, and  $\varepsilon$  is Gaussian noise. During training, the two critics of each agent can access the actions, states, rewards, and strategies of all agents from the experience buffer [19], thus realizing the interaction among agents. These two critics complete a centralized training task, namely, they evaluate the values of their actions not only according to their own state, but also considering the behavior states of other agents. On the other hand, the actor does a decentralized task according to a policy, namely, it only needs to take its own state into account and act accordingly.

### B. Partitioned experience buffer and priority experience replay

The structure of the partitioned experience buffer and priority experience replay (PEPE) is shown in Fig. 5. The proposed experience replay method first stratifies the experience buffer according to the impact of rewards on agents' learning and then sets priority sampling.

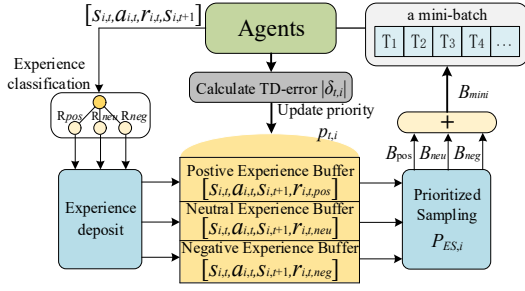


Fig. 5 PEPE structure

Different experiences play different roles in the agents' learning. Positive experiences can accelerate the agents' learning, negative experiences can improve the agents' generalization ability and anti-risk ability [29]. Agents generate many negative experiences in the early stage of training, which affects agents' learning rate. In the middle stage of training, the experience buffer stores more neutral experience, which affects the learning rate and anti-risk ability of the agents. In the late stage of training, the experience buffer stores a lot of positive experiences, which affects agents' anti-risk ability. To solve this problem, the experience buffer  $R(s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t})$  is divided into three areas according to the range of reward values, and a dynamic sampling ratio is set for each buffer. In the early stage of training, the number of positive experiences is increased to improve the learning rate of agents. In the middle stage of training, the number of positive and negative experiences is increased to improve the anti-risk ability and the learning rate of agents. In the late stage of training, the number of negative experiences is increased to improve the anti-risk ability of agents.

$$\begin{cases} R_{\text{pos}} \subset (s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{pos}}), r_{i,t,\text{pos}} \in [+o_p, +\infty) \\ R_{\text{neg}} \subset (s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{neg}}), r_{i,t,\text{neg}} \in [-\infty, -o_n) \\ R_{\text{neu}} \subset (s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{neu}}), r_{i,t,\text{neu}} \in (-o_n, +o_p) \end{cases} \quad (13)$$

where  $R_{\text{pos}}(s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{pos}})$  is the positive experience area,  $R_{\text{neg}}(s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{neg}})$  is the negative experience area, and  $R_{\text{neu}}(s_{i,t}, a_{i,t}, s_{i,t+1}, r_{i,t,\text{neu}})$  is the neutral experience area.  $O_p$  and  $O_n$  are two boundary coefficients, respectively.

The sampling number in each area is determined as follows:

$$\begin{cases} B_{\text{pos}} = (N_{\text{neg}} + N_{\text{neu}} / N_{\text{sum}}) * B_{\text{size}} \\ B_{\text{neg}} = (N_{\text{pos}} + N_{\text{neu}} / N_{\text{sum}}) * B_{\text{size}} \\ B_{\text{neu}} = (N_{\text{pos}} + N_{\text{neg}} / N_{\text{sum}}) * B_{\text{size}} \end{cases} \quad (14)$$

where  $B_{\text{pos}}$ ,  $B_{\text{neg}}$ , and  $B_{\text{neu}}$  are the number of experiences sampled from the positive experience area, negative experience area, and neutral experience area, respectively.  $N_{\text{pos}}$ ,  $N_{\text{neg}}$ , and  $N_{\text{neu}}$  are the number of experiences in the positive experience area, negative experience area, and neutral experience area, respectively.  $N_{\text{sum}}$  is the total number of experiences in the experience buffer area, and  $B_{\text{size}}$  is the number of batches.  $B_{\text{pos}}$ ,  $B_{\text{neg}}$ , and  $B_{\text{neu}}$  are determined according to the sampling probability  $P_{ES,i}$  from three experience areas respectively, and then aggregated into a minibatch to train the agents.

Sampling priority in experience replay is set, and the most useful experience is preferentially sampled to update the agents

and improve the agents' learning efficiency. The experience priority  $p_{t,i}$  is determined based on TD-error  $|\delta_{t,i}|$ ,

$$p_{t,i} = |\delta_{t,i}| + \delta \quad (15)$$

$$\delta_{t,i} = y_i - Q_{i,\theta_n}^\pi(S, a_1, a_2, \dots, a_N) \quad (16)$$

where  $\epsilon$  is an infinitesimal positive number (To prevent  $p_{t,i}$  from zero). The larger the TD error, the greater the role of this experience, and the higher priority of this experience. The smaller the TD error, the lower priority of this experience. The  $P_{ES,i}$  of experience being selected is as follows:

$$P_{ES,i} = \frac{p_i^\mu}{\sum p_i^\mu} \quad (17)$$

where  $\mu$  is the weight factor of sampling. It represents the influence degree of priority on sampling probability. The algorithm pseudocode is shown in Table I.

TABLE I  
MATD3 with PEPE

MATD3 with PEPE
Initialize the two critic networks $Q_{i,\theta_1}^\pi, Q_{i,\theta_2}^\pi$ and the network parameters $\theta_{i,1}, \theta_{i,2}, \phi_i$ of the actor network for each agent $i$ ;
Assign network parameters to corresponding target network parameters: $\theta'_{i,1} \leftarrow \theta_{i,1}, \theta'_{i,2} \leftarrow \theta_{i,2}, \phi'_i \leftarrow \phi_i$ and initialize the experience buffer $R$ .
<b>for</b> $t=1,2,\dots,T$ <b>do</b>
For each agent $i$ , select random action $a_i \sim \pi_i(s_i) + \epsilon$ , and the noise is dynamically adjusted to explore the current deterministic strategy;
Execute $a_{1,t}, \dots, a_{N,t}$ and observe reward $r_{i,t}$ and new state $s_{i,t+1}$ .
<b>if</b> $r_{i,t} > +o_p$ <b>then</b>
Put the experience transition $(s_{1,t}, \dots, s_{N,t}, a_{1,t}, \dots, a_{N,t}, r_{1,t}, \dots, r_{N,t}, s_{1,t+1}, \dots, s_{N,t+1})$ into the positive experience space $R_{\text{pos}}$ .
<b>else if</b> $r_{i,t} < -o_n$
Put the experience transition into the negative experience space $R_{\text{neg}}$ .
<b>else</b>
Put the experience transition into the neutral experience space $R_{\text{neu}}$ .
<b>end if</b>
<b>end if</b>
<b>for</b> agent $i=1$ to $N$ <b>do</b>
Calculate the sampling priority $P_{ES,i} = p_i^\mu / \sum p_i^\mu$ ;
According to the sampling probability $P_{ES,i}$ , the number of experiences of $B_{\text{pos}}$ , $B_{\text{neg}}$ , and $B_{\text{neu}}$ is sampled from the positive, negative, and neutral experience areas, respectively. Then, these three experiences are summarized into a minibatch $(s_{1,t}, \dots, s_{N,t}, a_{1,t}, \dots, a_{N,t}, r_{1,t}, \dots, r_{N,t}, s_{1,t+1}, \dots, s_{N,t+1})$ as the training data of the policy network and value network;
Update target value
$y_i = r_i + \gamma \min_{n=1,2} Q_{i,\theta'_n}^\pi(S', \pi_{\phi'_1}(S') + \epsilon, \dots, \pi_{\phi'_N}(S') + \epsilon)$
Update the priority of each experience transition
$p_{t,i} =  y_i - Q_{i,\theta_n}^\pi(S, a_1, a_2, \dots, a_N)  + \delta$
Update critic network parameter.
<b>if</b> $t \bmod d$ <b>then</b>
Update action parameters by policy gradient
$\nabla_{\theta_{i,d}} J(\phi) \approx N^{-1} \sum \nabla_{\theta} \pi_{\phi} \nabla_{\phi} (s) Q_{\theta} (s, a_1, \dots, a_N)  _{a_i = \pi_{\phi}(s_i)}$
Update the target value network and policy network parameters $\theta_{i,1}, \theta_{i,2}, \phi_i$ ,
$\begin{cases} \theta'_{i,n} \leftarrow \tau \theta_{i,n} + (1-\tau) \theta'_{i,n}, n=1,2 \\ \phi' \leftarrow \tau \phi + (1-\tau) \phi' \end{cases}$
<b>end if</b>
<b>end for</b>
<b>end</b>

### C. Power optimization strategy based on PEPE-MATD3

In the original MATD3 algorithm, the experience replay is utilized to store the experience data in the experience buffer, randomly sample from the experience buffer, and use the experiences to update the target strategy. The experience utilization of such experience replay algorithm is not high, which affects the learning efficiency of agents. Therefore, the PEPE algorithm is introduced into the MATD3 algorithm. Firstly, the improved PEPE-MATD3 algorithm divides the experience pool into three layers: positive experiences, negative experiences, and neutral experiences according to the different impacts of reward values on learning. In the early, middle, and late stages of training, the number of samples in each layer is determined to improve the utilization efficiency of experiences and the learning efficiency of agents. The experiences are then preferentially sampled according to the experience priority, and the experiences that are more useful for policy learning are screened out based on stratification, so as to improve the learning ability of the algorithm. According to the state and the strategy, the agents make actions (power optimization compensation amount  $\Delta P_{\text{ref},1-10}$ ) to adjust the reference power of the wind turbines and receive the reward of feedback. The agents explore and learn during the trial-and-error training of adjusting reference power until they learn the optimal strategy. The power optimization framework is shown in Fig. 6.

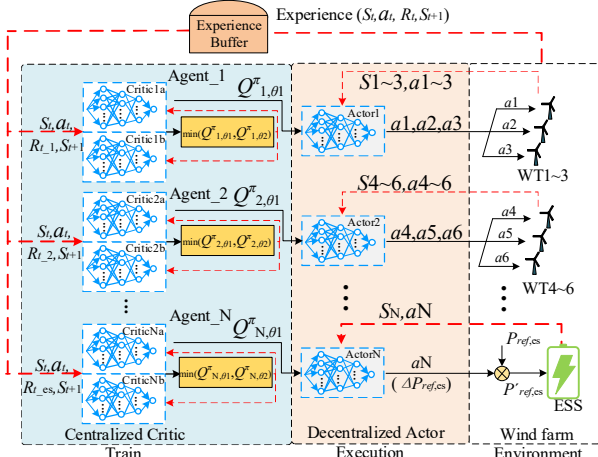


Fig. 6 Power optimization framework based on PEPE-MATD3

A centralized training and decentralized execution architecture is adopted for PEPE-MATD3. The centralized training is to guide actor training through the critics that can observe globally in each agent. The critics of each agent has access to the information of all other agents, which can realize the communication among multiple agents. Decentralized execution means that each agent's actor acts independently according to the state of the environment. Each column of wind turbines is controlled by an agent to reduce the computational load. Agent\_1 optimizes the power through the output action values  $a_1, a_2, a_3$  to obtain the new wind turbine reference power  $\Delta P'_{\text{ref},1-3}$ . Other agents act in the same way. Agent\_N of the ESS optimizes the power through the output action values  $a_N(\Delta P'_{\text{ref},\text{es}})$  to obtain the reference power  $\Delta P'_{\text{ref},\text{es}}$  of the new ESS.

The environment provides  $V_{\text{farm}}, P_{\text{farm}}, P_{\text{grid}}, P_{\text{es}}, SOC, V_i, P_{\text{WT}_i}$ , and  $\beta_{\text{WT}_i}$  information to each agent. The state space of the

combined power generation system of the wind farm and ESS is defined as:

$$S = [V_{\text{wind}}(t), P_{\text{farm}}(t), P_{\text{grid}}(t), P_{\text{es}}(t), SOC(t), V_i(t), P_{\text{WT}_i}(t), \beta_{\text{WT}_i}(t)] \quad (18)$$

After observing the state information of the environment, the agent chooses an action in the action space according to its policy  $\pi$ . The action space  $a_i$  and  $a_N$  are the reference adjustment values  $\Delta P_{\text{ref},i}$  and  $\Delta P_{\text{ref},\text{es}}$  for each wind turbine and ESS, respectively, and their expressions are as follows:

$$\begin{cases} a_i = \Delta P_{\text{ref},i} \\ a_N = \Delta P_{\text{ref},\text{es}} \end{cases} \quad (19)$$

In the learning process, setting the reward function determines the tasks that each agent needs to complete as well as whether they cooperate or compete. In order to solve the problems of power fluctuation, power loss, excessive load of WTs, ESS loss, the wind farm's power generation, grid-connected power smoothness, and pitch angle standard deviation are taken as the reward values for the WTs. For the ESS, it can be protected according to SOC. Therefore, the level of SOC and the grid-connected power smoothness are taken as the reward values,

$$\begin{cases} r_i(t) = \rho E_{\text{farm}} - (\lambda_{\text{WT}} F_g + \zeta \beta_{\text{std},i}) \\ r_{\text{es}}(t) = \sum_{i=0}^3 \xi y_{\text{es}} - \lambda_{\text{es}} F_g \end{cases} \quad (20)$$

where  $r_i(t)$  and  $r_{\text{es}}(t)$  are the reward functions of the agents of the WTs and ESS, respectively.  $\rho, \xi, \lambda_{\text{WT}}, \lambda_{\text{es}}, \sigma$  and  $\zeta$  are the weight coefficients.

$y_{\text{es},j}(j=0,1,2,3)$  is defined as:

$$\begin{cases} y_{\text{es},0} = -2, & SOC(t) < 0.2 \text{ or } 0.8 < SOC(t) \\ y_{\text{es},1} = 0.2, & 0.2 \leq SOC(t) \leq 0.8 \\ y_{\text{es},2} = 0.5, & 0.3 \leq SOC(t) \leq 0.7 \\ y_{\text{es},3} = 1.0, & 0.4 \leq SOC(t) \leq 0.6 \end{cases} \quad (21)$$

$E_{\text{farm}}$  is defined as:

$$E_{\text{farm}} = \sum_{k=0}^{T/\Delta t} |P_{\text{farm}}(k)| \quad (22)$$

where  $E_{\text{farm}}$  is the power generation of the wind farm.  $\Delta t$  is the time interval and  $T$  is the total time.

$F_g$  is defined as [30]:

$$F_g = \sum_{k=1}^{T/\Delta t} \left( \frac{\Delta P_{\text{grid}}(k)}{P_{\text{farm,rate}}} \right)^2 \quad (23)$$

$$\Delta P_{\text{grid}}(k) = |P_{\text{grid}}(t_0 + k\Delta t) - P_{\text{grid}}(t_0 + (k-1)\Delta t)| \quad (24)$$

where  $F_g$  represents the grid-connected power smoothness. The smaller  $F_g$  is, the better the smoothing effect is, and the smaller the impact on the power grid is.  $\Delta P_{\text{grid}}$  is the absolute value of the grid-connected power fluctuation.  $P_{\text{farm,rate}}$  is the rated power of the wind farm.

## IV. RESULTS AND DISCUSSION

### A. System parameter configuration

To verify the effectiveness of the proposed strategy, a simulation model was established on SimWindFarm [22]. The simplified 5MW FAST wind turbine model developed by

NREL was used, which is composed of the pneumatic model, transmission chain, generator, blade and tower model, and the wind turbine control strategies [22,23]. The layout of 10 NREL 5MW wind turbines in the simulation model is shown in Fig. 7.

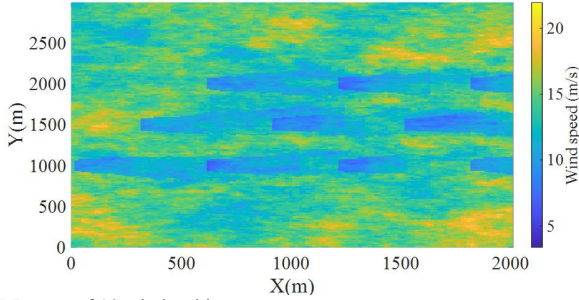


Fig. 7 Layout of 10 wind turbines

The charge/discharge power and the stored energy of the ESS can be obtained at any time during the whole operation period by simulating the scheduling operation. Therefore, the ESS capacity can be calculated and determined according to these results. The rated capacity of the ESS that meets the operational requirements of SOC is calculated as follows [31]:

$$E_{\text{rate}} = \frac{\nu \cdot (\max[E_{\text{flu}}(t)] - \min[E_{\text{flu}}(t)])}{\eta_e \cdot (SOC_{\text{max}} - SOC_{\text{min}})} \quad (25)$$

$$E_{\text{flu}}(t) = T \sum_{k=0}^{T/\Delta t} P_{\text{es}}(k) \quad (26)$$

where  $E_{\text{flu}}(t)$  is the energy fluctuation of the ESS to the initial state at different times,  $P_{\text{es}}(t)$  is the output power of the ESS,  $\nu$  is the capacity configuration margin of the ESS, and  $\eta_e$  is the charging and discharging efficiency of the ESS. The system parameters are shown in Table II.

TABLE II  
SYSTEM PARAMETERS

Parameter	Value
Total wind farm capacity (MW)	50
Rated power of a wind turbine (MW)	5
Impeller diameter (m)	61.5m
Air density(kg/m <sup>3</sup> )	1.2231
Length of the wind field (m)	2000
Width of wind field (m)	3000
Rated wind speed of a wind turbine (m/s)	11.4
Start wind speed of a wind turbine (m/s)	3
Intensity of turbulence	0.1
ESS capacity (MW·h)	10
Rated power of the ESS (MW)	10
Initial capacity	50%
Efficiency of charging and discharging ( $\eta_e$ )	96%
Margin of capacity configuration ( $\nu$ )	1.2

### B. Analysis of training results

MATD3 was used to optimize M-WT power and ESS to reduce the impact of model errors and uncertainties. The input wind speed of each WT column in the wind farm is similar, and each WT column was optimized and compensated by one agent, thereby reducing the computational complexity. According to the wind farm layout in Fig. 7, 10 WTs are controlled by four agents, and ESS is controlled by another agent. Therefore, the total number of agents in this paper is 5. To verify the training efficiency, an experiment was conducted with the MATD3 algorithm [19], improved PEPE-MATD3 algorithm, and multi-

agent deep deterministic policy gradient (MADDPG) algorithm [28]. The training hyperparameters (which are shared by the three algorithms), shown in Table III were determined by following multiple tests. The structural design of the critic network and actor network of the algorithm is shown in Fig. 8 (the network structure of each agent is the same).

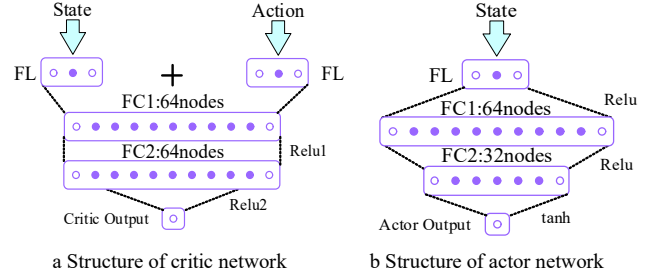


Fig. 8. Structure of the algorithm network

TABLE III  
TRAINING HYPERPARAMETERS

Hyperparameters	Value	weights	Value
Training Number	200	$\rho$	0.01
Batch number	128	$\xi$	0.05
Capacity of the experience buffer	$1 \times 10^6$	$\lambda_{\text{wt}}$	4
Discount factor	0.99	$\lambda_{\text{es}}$	1
learning rate	0.01	$\zeta$	1
optimizer	Adam		

To ensure the training effect, 200 trial and error trainings were conducted by the agents. The global reward index with three algorithms is shown in Fig. 9. As can be observed from Fig. 9, after 200 trial-and-error trainings by agents, the PEPE-MATD3 algorithm can obtain higher reward values than the other two algorithms at the early stage of training, indicating the advantage of improving the positive experiences at the early stage. Although the reward values of the PEPE-MATD3 are lower than that of MATD3 in the 25-60 iterations, they are higher than those of the other two algorithms after 60 iterations, indicating that PEPE-MATD3 has stronger learning ability. In addition, the reward values of the PEPE-MATD3 are maintained at around -870, and they are larger than those of the other two algorithms.

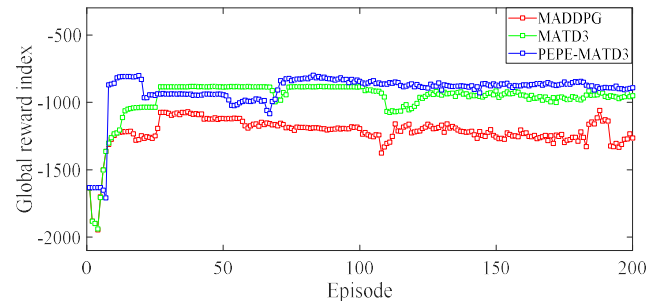


Fig. 9. Global reward index

### C. Simulation experiment analysis

To verify the effectiveness of the proposed control method, it was compared with the traditional control of wind farm (WF), the SOC feedback control of ESS[32], dynamic allocation (DA)-based coordinated control of M-

WT[13], Rule-based coordinated control of WF and ESS, MATD3-based coordinated control of WF and ESS(MATD3). When the average wind speed of the wind farm is 12m/s and the wind direction is  $0^\circ$ , the grid-connected power of the wind farm with six control methods are shown in Fig. 10.  $P_{es}$  and SOC of ESS with different control methods are shown in Fig. 11 and Fig. 12, respectively.

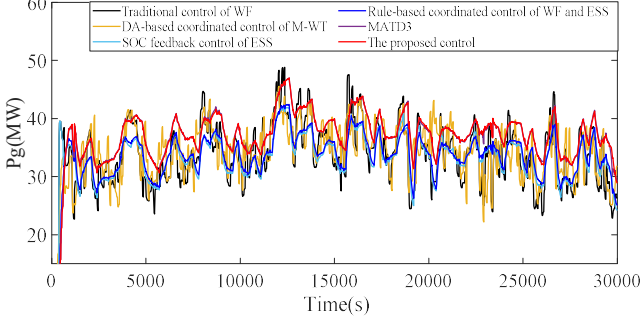


Fig. 10 Grid-connected power of the wind farm

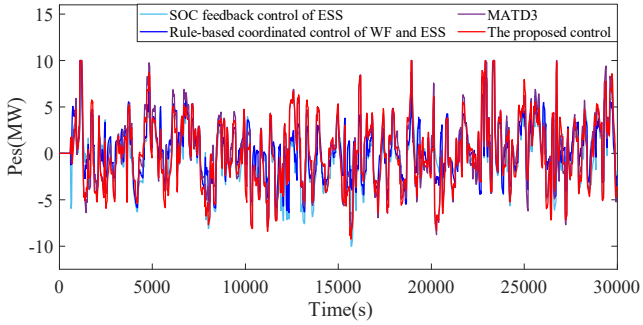


Fig. 11. Output power of the ESS

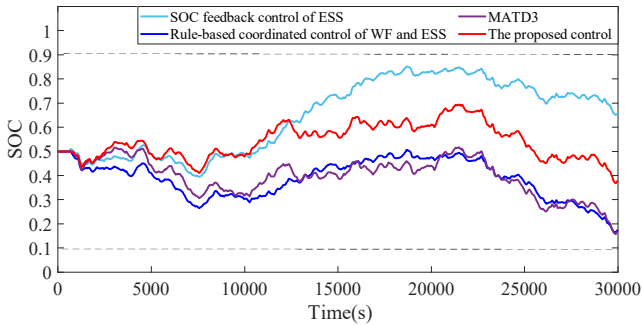


Fig. 12. SOC of the ESS

The output coefficient  $OC$  of the ESS is as follows:

$$OC = \sqrt{\frac{1}{T} \sum_{t=0}^T (SOC(t) - 0.5)^2} \quad (27)$$

The closer the SOC value is to 0.5, the smaller  $OC$ , and the stronger the ESS's ability to cope with future power fluctuations.

The energy evaluation indexes are shown in Table IV, where  $F_f$  is the smoothness of the output power of the wind farm, and

$E_{farm}$  is the total power generation of the wind farm excluding the energy storage part.

TABLE IV  
ENERGY EVALUATION INDEXES

Control methods	$F_f$	$F_g$	$E_{farm}$ ( MW·h)	$OC$
Tradition control of WF	1.0510	1.0510	277.0	—
SOC feedback control of ESS	1.0510	0.5010	277.0	0.2093
DA-based coordinated control of M-WT	0.8051	0.8051	274.1	—
Rule-based coordinated control of WF and ESS	0.8312	0.4234	275.3	0.1376
MATD3	0.7907	0.3902	303.6	0.1278
The proposed control	0.7358	0.3950	305.8	0.0824

It can be observed from Fig. 10 that the grid-connected power curve of the SOC feedback control of the ESS is smoother compared with the DA-based coordinated control of M-WT, especially for high-frequency power fluctuations. At the same time, it can be observed from Fig. 11 that the charging and discharging frequency of the output power of the ESS is high, being able to quickly handle power fluctuations. Therefore, the characteristics of ESS exactly compensate for the lack of M-WT smoothing power. As can be observed from Table IV,  $F_g = 0.4234$  of the rule-based coordinated control of WF and ESS is significantly lower than  $F_g = 0.5010$  of the SOC feedback control of ESS and  $F_g = 0.8051$  of the DA-based coordinated control of M-WT. Moreover, it can be observed from Fig. 10 that the grid-connected power fluctuations of the rule-based coordinated control of WF and ESS are lower than those of the two methods, indicating that it has a better effect on smoothing power fluctuations than the individual control.  $OC = 0.1376$  of the rule-based coordinated control of WF and ESS is lower than  $OC = 0.2093$  of the SOC feedback control of ESS, indicating that it has a stronger ability to smooth future power fluctuations. The output power  $P_{es}$  and SOC of ESS are shown in Fig. 11 and Fig. 12, respectively. It can be observed from Fig. 12 that the SOC of the rule-based coordinated control of WF and ESS is closer to the optimal value 0.5, indicating that the SOC of ESS is kept in a safe range and it is less likely to overcharge and over-discharge, as well as being able to better coping with future power fluctuations. The reason is that it is the ESS control enhanced by the WF smoothing power control, utilizing the power smoothing capability of the WF itself to reduce the workload of ESS so as to improve the system smoothing ability. However, its  $E_{farm}$  is 275.3 MW·h, which is smaller than that of the traditional wind farm, indicating that the power fluctuations are reduced at the cost of energy loss.

As can be observed from Table IV, the proposed control method has better performance in many aspects compared with other control methods mentioned above. The grid-connected power smoothness  $F_g=0.3950$  of the proposed control method is significantly smaller than that of the rule-based WF and ESS coordinated control  $F_g=0.4234$ . At the same time, it can be observed from Fig. 10 that the grid-connected power fluctuations of the proposed control method are lower than that of the rule-based coordinated control of WF and ESS, demonstrating its better smoothing power fluctuation effect than that of the rule-based coordinated control of WF and ESS without multi-agent deep reinforcement learning. Compared with the rule-based coordinated control of WF and ESS,  $E_{farm} = 305.8$  MW·h of the proposed control method is increased by

10.40%, indicating that the power generation of the wind farm on the premise of ensuring smooth power can be increased by the proposed control method. The reason is that by sacrificing some power generation of upstream WTs, the influence of wake effect on downstream WTs is reduced, and the wind energy obtained by downstream WTs is increased, so as to maximize the power generation of the whole wind farm. The specific process of the strategy is to use the power generation of the wind farm as the reward function of the agents in the PEPE-MATD3 algorithm. Through continuous trial-and-error trainings, agents will find a strategy to maximize the rewards according to the reward function. It can be observed from Fig.

12 that the SOC of the proposed control method is closer to the optimal value 0.5, indicating that the SOC of ESS is kept within a safe range, and it is less likely to overcharge and over-discharge, as well as being able to better coping with future power fluctuations. Although  $F_g=0.3950$  of the proposed control method is slightly larger than  $F_g=0.3902$  of the MATD3-based coordinated control method, the  $E_{farm}$  and  $OC$  are significantly better than that of the latter.

The comparison of the tower root moment of the No. 1 WT with different methods is shown in Fig. 13. The smoothness of the tower root moment of each WT  $F_{M,i}$  is shown in Table V.

TABLE V  
EVALUATION INDEXES OF THE WTS

Control methods	$F_{M,1}$ (Nm)	$F_{M,2}$ (Nm)	$F_{M,3}$ (Nm)	$F_{M,4}$ (Nm)	$F_{M,5}$ (Nm)	$F_{M,6}$ (Nm)	$F_{M,7}$ (Nm)	$F_{M,8}$ (Nm)	$F_{M,9}$ (Nm)	$F_{M,10}$ (Nm)
The tradition control of WF	1.550	1.660	2.445	1.907	1.886	1.853	1.858	1.514	2.094	1.782
DA-based coordinated control of M-WT	1.156	1.797	1.958	1.525	0.6441	1.241	1.695	1.413	1.324	1.447
Rule-based coordinated control of WF and ESS	1.447	1.609	2.308	1.818	1.835	1.754	1.851	1.422	1.997	1.777
MATD3	1.016	0.989	1.683	1.067	1.871	1.064	1.383	0.696	1.672	1.235
the proposed control	0.887	0.980	1.180	0.924	1.339	1.144	1.340	0.697	1.442	1.215

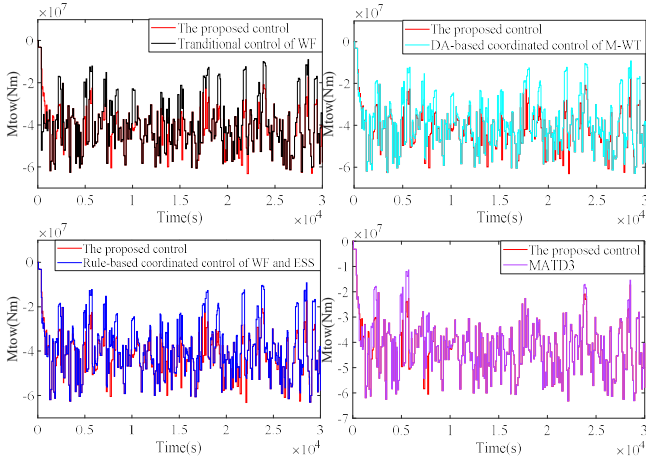


Fig. 13. Tower root bending moment of No. 1 WT with different methods

As can be observed from Fig. 13, the fluctuations of the tower root bending moment with the proposed method are significantly lower than those of the traditional control of WF, the DA-based coordinated control of M-WT, and the rule-based coordinated control of WF and ESS. They are significantly lower than that of MATD3 during 5000s~7000s, although the fluctuation difference is not obvious as a whole. At the same time, it can be found from Table V that the smoothness of the root bending moment of each WT tower with the proposed method is smaller than those of other methods. The results also show that more fatigue load can be reduced with the proposed method.

#### D. Simulation experiments under different environmental conditions

##### 1) Changes in energy storage capacity

On the premise that other parameters remain unchanged, the energy storage capacity configuration was changed from 4MW·h to 15MW·h in the experiment.  $F_g$  of the wind farm and  $OC$  of the ESS with different rated capacities of ESS were analyzed. The results are shown in Fig. 14 and Fig. 15.

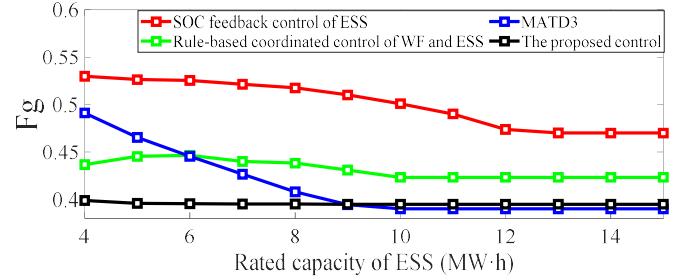


Fig. 14.  $F_g$  with different rated capacities of ESS

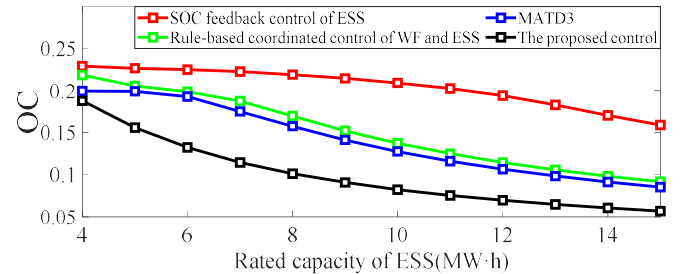


Fig. 15.  $OC$  with different rated capacities of ESS

As can be observed from Fig. 14, the grid-connected power smoothness  $F_g$  of these four methods shows a downward trend when the rated capacity of the ESS increases, indicating that the increased rated capacity can improve the ability of the ESS to smooth wind power. Although the  $F_g$  of the proposed method is slightly lower than that of the MATD3 method when the energy storage capacity ranges from 10MW·h to 15MW·h, it is still under the condition of low rated capacity as a whole and still maintains roughly 0.4, indicating more stable performance in terms of smoothness. As can be observed from Fig. 15, the  $OC$  of the ESS shows a downward trend using these four methods when the rated capacity of the ESS increases, and the  $OC$  of the proposed method is at a lower position than those of the other four methods. This indicates that a better capability of smoothing future power fluctuations is still achieved with the proposed method in the capacity changing environment. Overall, more stable performance and robustness can be

obtained with the proposed method in the capacity changing environment. In addition, the results in different rated capacities verify that the ESS using the proposed method can have many configuration choices according to the investment budget.

2) Average wind speed variation in the wind farm

The average wind speed of the wind farm was varied from 10m/s to 14m/s. The smoothness of grid-connected power, power generation and output coefficient of ESS of the wind farm with different average wind speeds are shown in Fig. 16, Fig. 17 and Fig. 18, respectively.

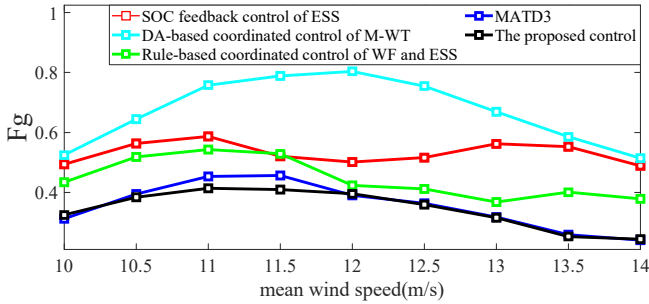


Fig. 16.  $F_g$  with different average wind speeds

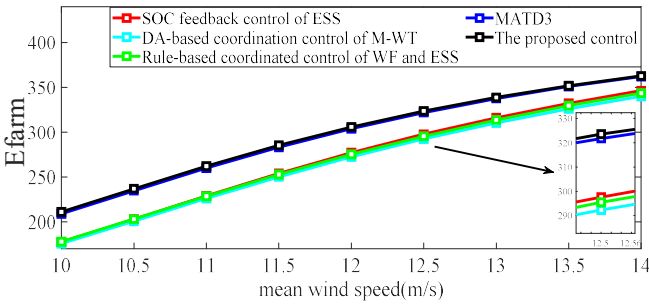


Fig. 17.  $E_{farm}$  with different average wind speeds

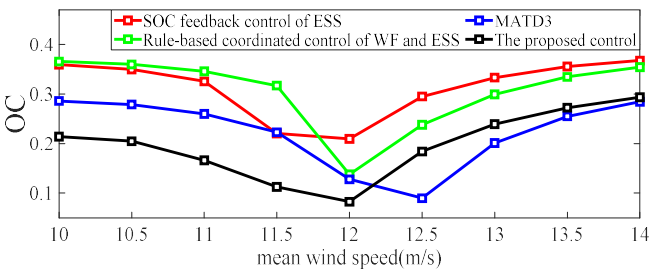


Fig. 18. OC of ESS with different average wind speeds

It can be observed from Fig. 16 that the  $F_g$  of the proposed method is lower than those of other methods when the average wind speed of the wind farm changes, indicating that a better smoothing power performance can still be guaranteed with the proposed method in the average wind speed changing environment. It can be observed from Fig. 17 that the power generation of the proposed method is higher than those of the other four methods when the average wind speed changes from 10m/s to 14m/s, indicating that more energy can be generated in the wind farm with the proposed method at the full wind speed. As can be observed from Fig. 18, although the OC of MATD3 is lower than that of the proposed control method when the average wind speed is from 12.5m/s to 14m/s, the OC of the proposed control method is lower than those of other

algorithms as a whole. In general, a good performance is obtained in smoothness, power generation, and ESS output coefficient with the proposed method under different wind speeds.

To verify the adaptability and robustness of the proposed control method, the average wind speed of the wind farm, the rated capacity of the ESS, and the turbulence intensity were changed simultaneously, and 200 sets of Monte Carlo simulations were conducted. The simulation results are shown in Fig. 19.

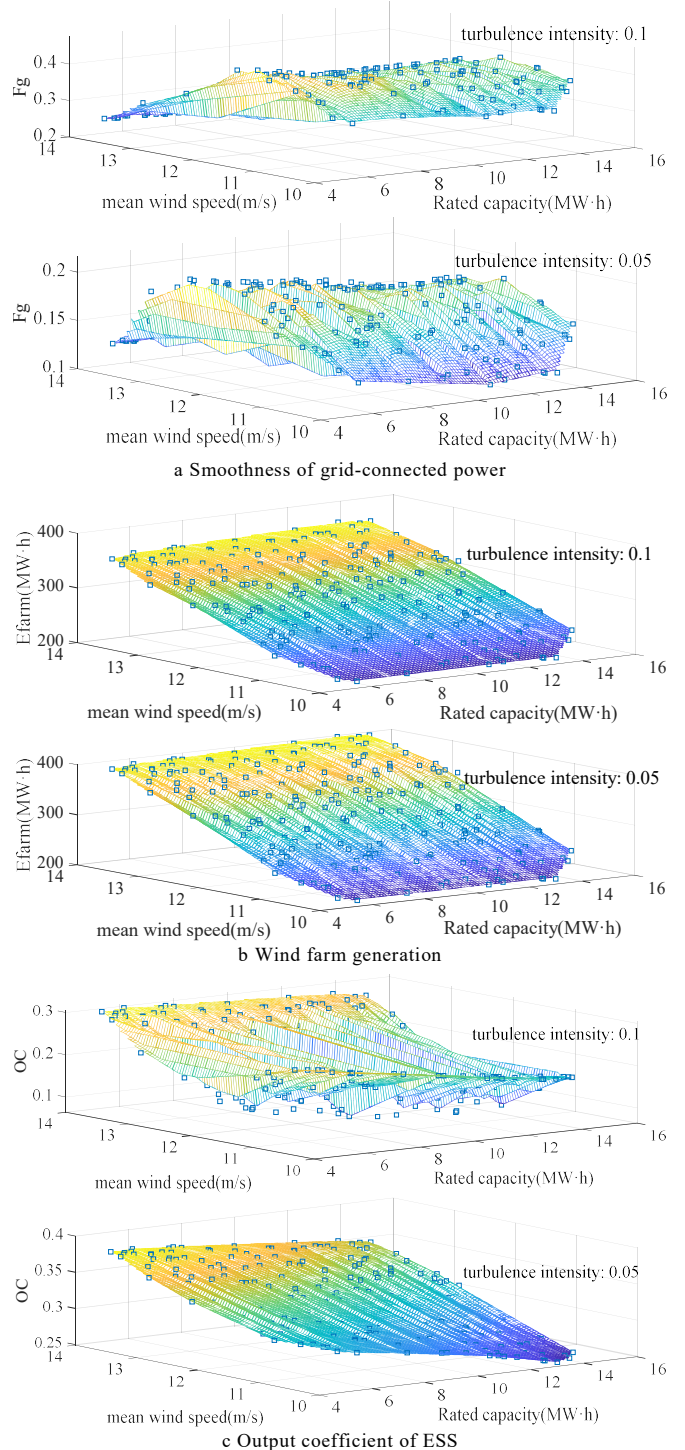


Fig. 19. Monte Carlo simulation results under various uncertain conditions

It can be observed from Fig. 19 that good performance in grid-connected power smoothness, power generation, and ESS output coefficient is obtained with the proposed method when different parameters vary at the same time, which also shows that a variety of environments with different parameters can be better adapted by the proposed method. However, it can be observed from Fig. 19 (a) that the  $F_g$  of the proposed method becomes larger when the turbulence coefficient is 0.05 and the rated capacity is as low as  $4\text{MW}\cdot\text{h}$ , indicating that a good smoothing power performance is difficult to be guaranteed with the proposed method in this case. It can be observed from Fig. 19 (c) that the OC of the proposed method also becomes larger when the turbulence coefficient is 0.05, the rated capacity is as low as  $4\text{MW}\cdot\text{h}$  and the average wind speed is  $14\text{m/s}$ . It is weaker for the proposed method in dealing with future power fluctuations and easier to make SOC overcharge and over-discharge. It is also difficult for the proposed method to ensure good control performance when the average wind speed of the wind farm is too large or too small or the rated capacity is too low. Overall, good adaptability and robustness are obtained with the proposed method in a random uncertain environment.

### 3) Results with model errors

To simulate the errors between the wake model established and the actual wake environment, interference signals (Band-Limited White Noise) were added to the wake model on the SimWindFarm. The simulations on control effects with/without model errors were conducted by the rule-based coordinated control of WF and ESS and the proposed control. The results are shown in Fig. 20, Fig. 21 and Fig. 22, and the corresponding evaluation indexes are shown in Table VI.

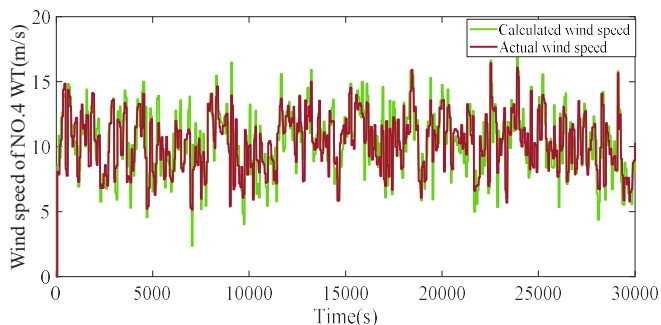


Fig. 20. Wind speed of NO.4 WT

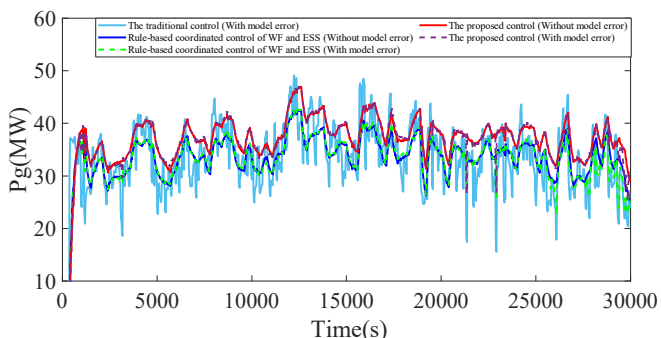


Fig. 21.  $P_g$  with/without model errors

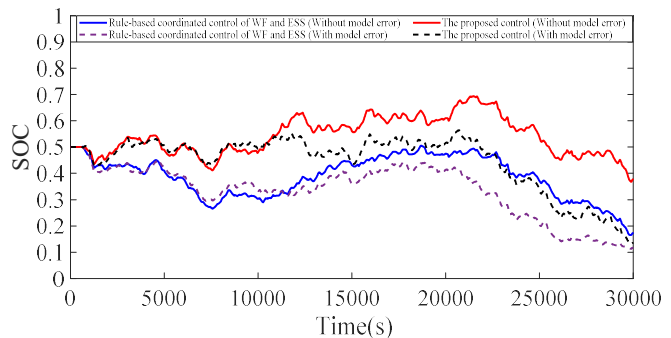


Fig. 22. SOC of the ESS with/without model errors

Table. VI Evaluation indexes

Control methods	Rule-based coordinated control of WF and ESS		The proposed control	
	Without model error	With model error	Without model error	With model error
$F_g$	0.3902	0.4965	0.3950	0.4168
$E_{\text{farm}}/\text{MW}\cdot\text{h}$	303.6	273.5	305.8	303.0
OC	0.1278	0.1924	0.0824	0.1169

As can be observed from Fig. 20, the calculated wind speed of the NO.4 WT is obviously different from the real wind speed if there exist errors in the model. It can be observed from Fig. 21 and Fig. 22 that the fluctuations of the grid-connected power are larger with the rule-based coordinated control of WF and ESS and the proposed control method when the model has errors. Moreover, the SOC of the two methods changes obviously, which are further deviated from 0.5, and in an unsafe range during 25000~30000s, indicating the control effect is affected by the model errors. It can be observed from Fig. 21, Fig. 22 and Table VI that the  $F_g$  of the rule-based coordinated control method of WF and ESS is increased by 0.1063, the OC is increased by 0.0646, and the  $E_{\text{farm}}$  is decreased by  $30.1\text{MW}\cdot\text{h}$  when the model errors occur. On the other hand, the  $F_g$  of the proposed method is increased by 0.0218, the OC is increased by 0.0345, and the  $E_{\text{farm}}$  is reduced by  $2.8\text{MW}\cdot\text{h}$ . It can also be observed from the above data that compared with the rule-based coordinated control method of WF and ESS, the smoothness, output coefficient and energy generation of the proposed control method have smaller changes with the model errors, and a better control effect can be maintained with the proposed control method, indicating that the impact of model inaccuracy is reduced with the optimization compensation of the PEPE-MATD3 in the proposed control method, and the advantage of insensitivity to the model is also reflected by using PEPE-MATD3 with deep reinforcement learning in the proposed control.

## V. CONCLUSION

Aiming at the problems of wind power smoothing, a coordinated power smoothing control strategy for M-WT and ESS based on PEPE-MATD3 algorithm is proposed in this paper. Firstly, a coordinated power control system of the M-WT and ESS based on the wake model is established. The coordinated power control between the M-WT is used to smooth power fluctuations, while the ESS is used to smooth high-frequency fluctuations that are difficult to be handled by internal control of the wind farm. The smooth wind power

capabilities of the M-WT and ESS are combined to compensate the shortcomings of individual WT control. Then, the PEPE-MATD3 algorithm is used to optimize the coordinated power control of the M-WT and ESS, and the negative impact caused by the inaccurate model is reduced by the model-free feature of the PEPE-MATD3 algorithm, so as to reduce the loss of the system on the premise of ensuring smooth power. Experimental results show that the proposed method is superior to M-WT and ESS alone in terms of power smoothness, generation, and output capacity of the ESS, and the proposed method is further improved by optimizing the PEPE-MATD3 algorithm. The simulation results under different scenarios and model errors show that the proposed method can reduce the influence of model uncertainty, decompose the complex multi-objective optimization problem into a multi-agent cooperative problem, simplify the complex problem, and improve the robustness and stability of the system. Future work will focus on practical wind farm applications, and industrial experiments will be conducted when the experimental conditions are sufficient. In addition, how to ensure the system stability and safety in the process of trial-and-error training of deep reinforcement learning and how to maximize the training effect of agents at acceptable trial and error costs will also be investigated.

## VI. REFERENCES

- [1] P.H.A. Barra, W.C. de Carvalho, T.S. Menezes, et al. "A review on wind power smoothing using high-power energy storage systems," *Renewable and Sustainable Energy Reviews*, vol. 137, pp. 110455, Mar. 2021.
- [2] Tong Zheming, Cheng Zhewu, Tong Shuiguang. "A review on the development of compressed air energy storage in China: Technical and economic challenges to commercialization," *Renewable and Sustainable Energy Reviews*, vol. 135, pp. 110178, Jan. 2021.
- [3] Tani, A., Camara, M. B., et al. "Energy management in the decentralized generation systems based on renewable energy—ultracapacitors and battery to compensate the wind/load power fluctuations," *IEEE Trans. on Ind. Appl.*, vol. 51 no. 2, pp. 1817-1827, Mar. 2015.
- [4] Qais M H, Hasanien H M, Alghuwainem S. "Output power smoothing of wind power plants using self-tuned controlled SMES units," *Electric Power Systems Research*, vol. 178, pp. 106056, Jan. 2020.
- [5] Hemmati R, Ghiassi S, Entezariharsini A. "Power fluctuation smoothing and loss reduction in grid integrated with thermal-wind-solar-storage units," *Energy*, vol. 152, pp.759-769, Jun. 2018.
- [6] Siqueira L, Peng W. "Control strategy to smooth wind power output using battery energy storage system: A review," *Journal of Energy Storage*, vol. 35, no. 45, pp.102252, Mar. 2021.
- [7] C. Wan, W. Qian, C. Zhao, et al. "Probabilistic Forecasting Based Sizing and Control of Hybrid Energy Storage for Wind Power Smoothing," *IEEE Trans. Sustain. Energy*, vol. 12, no. 4, pp. 1841-1852, Oct. 2021.
- [8] Lin L, Jia Y, Ma M, et al. "Long-term stable operation control method of dual-battery energy storage system for smoothing wind power fluctuations," *International Journal of Electrical Power & Energy Systems*, vol. 129, no. 9, pp. 106878, Jul. 2021.
- [9] J. I. Yoo, Y. C. Kang, D. Yang, et al. "Power Smoothing of a Variable-Speed Wind Turbine Generator Based on a Two-Valued Control Gain," *IEEE Trans. Sustain. Energy*, vol. 11, no. 4, pp. 2765-2774, Oct. 2020.
- [10] Y. Kim, M. Kang, E. Muljadi, et al. "Power Smoothing of a Variable-Speed Wind Turbine Generator in Association With the Rotor-Speed-Dependent Gain," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 990-999, Jul. 2017.
- [11] K. Liao, D. Lu, M. Wang, et al. "A Low-Pass Virtual Filter for Output Power Smoothing of Wind Energy Conversion Systems," *IEEE Trans. Ind. Electron.*, vol. 69, no. 12, pp. 12874-12885, Dec. 2022.
- [12] X. Lyu, J. Zhao, Y. Jia, et al. "Coordinated Control Strategies of PMSG-Based Wind Turbine for Smoothing Power Fluctuations," *IEEE Trans. Power Syst.*, vol. 34, no. 1, pp. 391-401, Jan. 2019.
- [13] Y. Zhu, R. Zhao and J. Zhao, "Output power smoothing control for the PMSG based wind farm by using the allocation of the wind turbines," In 2017 20th International Conference on Electrical Machines and Systems (ICEMS), 2017, pp. 1-6.
- [14] A. M. Howlader, T. Senjyu and A. Y. Saber, "An Integrated Power Smoothing Control for a Grid-Interactive Wind Farm Considering Wake Effects," *IEEE Syst. J.*, vol. 9, no. 3, pp. 954-965, Sep. 2015.
- [15] K. Arulkumaran, M. P. Deisenroth, M. Brundage and A. A. Bharath, "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26-38, Nov. 2017.
- [16] H. Dong and X. Zhao, "Wind-Farm Power Tracking Via Preview-Based Robust Reinforcement Learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 3, pp. 1706-1715, March. 2022.
- [17] S. Huang, P. Li, M. Yang, Y. Gao, J. Yun and C. Zhang, "A Control Strategy Based on Deep Reinforcement Learning Under the Combined Wind-Solar Storage System," *IEEE Trans. Ind. Appl.*, vol. 57, no. 6, pp. 6547-6558, Nov.-Dec. 2021.
- [18] Nguyen, Thanh Thi, Ngoc Duy Nguyen, et al. "Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications." *IEEE trans. Cybernet.*, vol. 50, no. 9, pp. 3826-3839, Sep. 2020.
- [19] Ackermann J, V Gabler, Osa T, et al. "Reducing overestimation bias in multi-agent domains using double centralized critics," arXiv preprint, 2019.
- [20] Katic I, J Højstrup, Jensen N O. "A Simple Model for Cluster Efficiency," 1987.EWEA Conference&Exhibition, 1986: 407-410.
- [21] Gil M, Gomis-Bellmunt O, Sumper A, et al. "Power generation efficiency analysis of offshore wind farms connected to a SLPC (single large power converter) operated with variable frequencies considering wake effects," *Energy*, vol. 37, no. 1, pp. 455-468, Jan. 2012.
- [22] Grunnet, J. D., Soltani, M., Knudsen, T., et al. "Aeolus toolbox for dynamics wind farm model, simulation and control". In European Wind Energy Conference and Exhibition 2010, vol. 4, pp. 3119–3129, 2010.
- [23] Jonkman J M, Butterfield S, Musial W, et al. "Definition of a 5MW Reference Wind Turbine for Offshore System Development," office of scientific & technical information technical reports, 2009.
- [24] X. Wang, Y. Luo, B. Qin, et al. Power dynamic allocation strategy for urban rail hybrid energy storage system based on iterative learning control. *Energy*, vol. 245, pp. 123263, Jan. 2022.
- [25] X. Wang, Y. Luo, B. Qin, et al. "Hybrid energy management strategy based on dynamic setting and coordinated control for urban rail train with PMSM," *IET Renew Power Gener.*, vol. 15, no. 12, pp. 2740, May. 2021.
- [26] K. Arulkumaran, M. P. Deisenroth, M. Brundage et al. "Deep Reinforcement Learning: A Brief Survey," *IEEE Signal Process. Mag.*, vol. 34, no. 6, pp. 26-38, Nov. 2017.
- [27] Fujimoto S, Hoof H V, Meger D. "Addressing Function Approximation Error in Actor-Critic Methods," arXiv, preprint, 2018.
- [28] Lowe R, Wu Y, Tamar A, et al. "Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments," arXiv preprint, 2017.
- [29] J. Zhang, L. Li and C. Hu, "Learning how to drive using DDPG Algorithm with Double Experience Buffer Priority Sampling," 2020 Chinese Automation Congress (CAC), pp. 2598-2603, 2020.
- [30] Y. Zhou, Z. Yan and N. Li. "A Novel State of Charge Feedback Strategy in Wind Power Smoothing Based on Short-Term Forecast and Scenario Analysis," *IEEE Trans. Sustain. Energy*, vol. 8, no. 2, pp. 870-879, Apr. 2017.
- [31] Nazari-pouya, Hamidreza, Chu, et al. Engineering energy storage sizing method considering the energy conversion loss on facilitating wind power integration. *IET Generation, Transmission & Distribution*, vol. 13, no. 9, pp. 1751-8687, Apr. 2019.
- [32] Zhang F, Meng K, Xu Z, et al. "Battery ESS Planning for Wind Smoothing via Variable-interval Reference Modulation and Self-adaptive SOC Control Strategy," *IEEE Transactions on Sustainable Energy*, vol. 8, no. 2, pp. 695-707, 2017.



**Xin Wang** (Member, IEEE) received her M.S. degree in Computer and its Application from Central South University of Forestry & Technology, Changsha, China in 2004 and Ph.D. degree in Control Science and Engineering from Central South University, Changsha, China in 2010, respectively. She is currently the professor at the School of Electrical and Information Engineering in the Hunan

University of Technology. Her current research interests relate to the intelligent methods for complex systems modeling, control and optimization, and the application of these to resolving various engineering problems.



**Jianshu Zhou** was born in huaihua, Hunan province in 1998. He is currently pursuing a master's degree in electrical engineering at Hunan University of Technology. His research interest covers intelligent modeling, control and optimization methods for wind power systems.



**Bin Qin** received his M.S. degree in automatic control in 1988 and Ph.D. degree in Control theory and Control Engineering in 2006 from Central South University, Changsha, China, respectively. From 2007 to 2008, he was a Visiting Professor with the University of Sheffield, UK. Since 2002, he has been a Professor at the School of Electrical and Information Engineering,

Hunan University of Technology. His current research interests include complex industrial process modeling and intelligent control, Wind power generation and machine learning, production system optimization, and scheduling.



**Lingzhong Guo** received the B.S. degree in Mathematics and the M.Sc. degree in Applied Mathematics in China, and the Ph.D. degree in Control Systems Engineering from Bristol Robotic Laboratory, University of the West of England, UK in 2003. He is currently a lecturer at the Department of Automatic Control and Systems Engineering,

University of Sheffield, UK. His research interests include theory and methods for the identification, analysis, and control of nonlinear dynamic systems as well as applications to structure vibration control and fault detection, urban traffic systems, underground hydrocarbon reservoirs dynamics, energy storage systems, and biomedical engineering systems etc.