Coordinated Control of Wind Turbine and Hybrid Energy Storage System Based on Multi-Agent Deep Reinforcement Learning for Wind Power Smoothing

Xin Wang [a,b], Jianshu Zhou [a,b], Bin Qin [a,b*], Lingzhong Guo [c]

[a]*School of Electrical & Information Engineering, Hunan University of Technology, Zhuzhou, Hunan, 412007, China*

[b]*Hunan Engineering Research Center of Electrical Drive and Regenerative Energy Storage and Utilization, Zhuzhou, Hunan, 412007, China*

[c]*Department of Automatic Control and Systems Engineering, University of Sheffield, S1 3JD, UK*

Elsevier use only: Received 9 June 2022; Received in revised form 6 November 2022; Accepted 28 November 2022

Abstract

Due to the inherent fluctuation, wind power integration into the large-scale grid brings instability and other safety risks. In this study by using a multi-agent deep reinforcement learning, a new coordinated control strategy of a wind turbine (WT) and a hybrid energy storage system (HESS) is proposed for the purpose of wind power smoothing, where the HESS is combined with the rotor kinetic energy and pitch control of the wind turbine. Firstly, the wind power output is forecasted and decomposed into high, medium, and low-frequency components through an adaptive variational mode decomposition (VMD). The optimal secondary allocation of the reference power of the high-frequency and medium-frequency is then performed through a multi-agent twin-delay deep deterministic policy gradient algorithm (MATD3) to smooth the power output. To improve the exploration ability of the learning, a new type of α-state Lévy noises is injected into the action space of the MATD3 and the noises are dynamically adjusted. Simulation and RT-LAB semi-physical real-time experimental results show that the proposed control strategy can make full use of the smoothing output power of the WT and HESS combined generation system reasonably, extend the life of the energy storage elements and reduce the wear of the WT.

Key words: Wind power smoothing; Hybrid energy storage system (HESS); Pitch control; Rotor kinetic energy control; Coordination control; Multi-agent deep reinforcement learning TD3

1 Introduction

The fluctuation of wind output power is an important factor that affects the stability and the cost of the wind power system operation [1]. The essence of smoothing wind power is to improve the controllability of the wind power, optimize the power quality, and eliminate the adverse effects of the grid connection of the wind power [2].

At present, the methods dealing with the output power fluctuation of wind power systems mainly include the regulation control of a wind turbine (WT) and the indirect power control of energy storage systems (ESS) [3- 6], where the latter is more popular.

Regarding the regulation control of WT, Kim et al. proposed a power smoothing method based on rotor variable speed control [7], which added a frequency regulation control loop to the traditional rotor variable speed control. The advantage of this method is that the influence of output power fluctuation on the system can be reduced, and the narrow-band regulation of the system frequency can be realized, but the fatigue load of wind turbines is not considered. Tang et al. introduced a pitch control on the basis of rotor speed control and proposed an improved wind power smoothing control method by coordinating rotor speed and pitch angle

[8], which further increases power smoothing ability of wind turbines. Moreover, based on the rotor speed change, a control mode switching method was designed to reduce the unnecessary pitch angle action, which greatly reduces the fatigue load of wind turbines. In [9], another wind power smoothing scheme was realized by DC voltage control, rotor speed control, and pitch control according to the adaptive ability of PMSG wind turbines, and based on hierarchical rules, a power allocation strategy was proposed, which can reasonably control the actions of three parts according to the degree of the power fluctuation and the hierarchical rules, greatly improve the power smoothing capability of wind power and reduce control pressure of each part. In general, the advantage of such methods is that they do not need to install additional equipment so as to reduce the cost. However, the power smoothing capability of the regulation and control of wind turbines is generally weak, and the control structure of wind turbines has to be changed, resulting in influence on the maximum power tracking performance of wind energy, Moreover, a part of wind power has to be sacrificed for power smoothing.

Compared with WT regulation control, ESS has a stronger capability to smooth output power and provide rapid response to the fluctuation of output power, and at the same time there is no wind power loss during power smoothing process. ESS control focuses on allocating and setting the reference power of energy storage components. Its control methods can be divided into two categories: actual wind power as control inputs and forecasting wind power as control inputs. Jiang et al. proposed a hybrid energy storage system (HESS) coordinated control method for wind power smoothing via new wavelet analysis[10]. The proposed wavelet analysis consists of primary filtering (PF) and secondary filtering (SF). PF decomposes the actual wind power into low-frequency signals to meet the requirements of wind power grid connection, while SF decomposes the fluctuation signals into the reference power of lithium batteries and supercapacitors on the premise of considering the remaining power. Compared with the traditional filtering allocation methods, this method can ensure the HESS to operate in a safe range. Nguyen et al. proposed an adaptive control strategy for virtual capacity of HESS [11]. The smoothing time constant of the wavelet transform and the variation range of the state of charge (SOC) of the HESS are increased through the adaptive virtual capacity of the HESS so that the HESS has a stronger power smoothing ability. In [12], a long-term stable operation control with a dual-battery energy storage system (DESS) based on real-time operating status and wind power fluctuations was proposed to adaptively fine-tune the low-pass filter time constant, charge the battery throughput power in real-time, and optimize SOC of the two battery packs. In the latest literatures, this type of methods with actual power as control input can ensure ESS to smooth wind power under safe and stable conditions. However, they are difficult to prepare for random fluctuation of power in the future, which will lead to the controller's inability in making the optimal strategies from a long-term perspective, ensuring the maximum use of the smooth ability of the energy storage system, and ensuring the energy storage system in the optimal state during the control period.

Some researchers have proposed control strategies based on wind power forecasting for wind power smoothing. These strategies forecast the future power output and schedule the ESS in advance to optimize the operation of the ESS. In [13], based on wind power forecasting, an SOC optimal control method for battery energy storage systems (BESS) was proposed, in which the future optimal SOC values are calculated using wind power forecasting information and the time constant of low-pass filtering is adjusted by fuzzy control to keep the SOC within the optimal range in the future. Zhou et al. made scenario analysis to SOC optimization control based on wind power forecasting [14]. The possible situation of BESS is forecasted, and the uncertainty of wind power forecasting is considered through scenario analysis, so as to reduce its influence. In [15], a model predictive control based on operational constraints was proposed to smooth the wind power, in which a double closed-loop structure taking into account the operational constraints was designed. The primary power is set according to the predicted ESS average power and SOC values in the outer loop, and the new ESS charge and discharge power are calculated under the premise of considering the operation constraints in the inner loop, so that the actual output power follows the reference values. Wan et al. proposed a stochastic optimization regulation method based on probability forecasting for the HESS to smooth wind power fluctuations [16]. The power is forecasted according to the probability model, and then decomposed by an adaptive VMD to obtain the reference power of the HESS, which is optimized by a stochastic optimization model. This method further increases the power smoothing ability of the HESS and ensures the economic operation of the HESS.

Reinforcement learning is to use feedback information to gradually acquire decision-making ability through the continuous interaction between the learning system and the environment and improve decision-making ability through continuously learning [17]. Given the variability and uncertainty of wind power output, deep learning can naturally cope with this uncertainty and continuously learn to maximize the benefits of the wind energy storage system. Deep learning has been used for time-varying forecasting to establish the correlation between current events and future events and reveal the spatial-temporal correlation characteristics contained in wind farms [18, 20]. Deep reinforcement learning is the combination of reinforcement learning and deep learning and has the advantages of both. Motivated by these studies, deep reinforcement learning will be applied in smoothing power control of a wind energy storage system in this paper.

Considering the existing problems in wind power smoothing research, a new coordinated control strategy for the WT-HESS combined generation system based on multi-agent deep reinforcement learning is proposed to smooth wind power. In this strategy, the HESS is combined with rotor kinetic energy and pitch control to smooth the grid-connected power of the WT and at the same time share the working load of smoothing power fluctuations. Based on long short-term memory (LSTM), a wind power forecasting model is established, and an adaptive variational modal decomposition (VMD) is used to perform primary power allocation to the future wind power to obtain high, medium, and low-frequency power. Then, through the multi-agent deep reinforcement learning (MADRL), the high and medium frequency power are allocated to maximize the working efficiency of the WT-HESS combined generation system under the premise of ensuring power smoothing.

The main contributions of this paper are as follows:

1) Using the energy storage systems to smooth wind power has become an important research topic nowadays. However, the energy storage systems and their operation and maintenance are expensive, and the energy storage equipment with a limited capacity will be inevitably overcharged and over-discharged. Considering the wind turbine itself has great potential in power smoothing, a hybrid energy storage system (HESS) combined with the rotor kinetic energy and pitch control of a wind turbine is proposed in this paper to smooth the grid connected power. The rotor kinetic energy and pitch control are used to reduce some pressure of the HESS in wind power smoothing, which are conducive to the life protection of HESS and prolong the operation life of HESS.

2) Considering that the output power of wind power systems naturally contains uncertainty, which fits the framework of the Markov decision process, the basis of applying deep reinforcement learning. Therefore, in this paper, a new coordinated control strategy for the WT-HESS combined generation system based on multi-agent deep reinforcement learning is proposed. A deep reinforcement learning algorithm termed twin-delay deep deterministic policy gradient (MATD3) is introduced to optimize power allocation. Considering the complexity of the optimal allocation of power and the reliability and scalability of the system, a multi-agent deep reinforcement learning is used to decompose the optimization problem into sub-problems so as to simplify the problem, which is helpful to obtain better results.

3) An improved multi-agent twin-delay deep deterministic policy gradient algorithm based on dynamic adjustment of α-state Lévy noises (SLA-MATD3) is proposed. The α-state Lévy noises are introduced during agents training and they are dynamically adjusted according to the SOC changes of the HESS, so as to guide agents to better explore the environment, which helps to avoid local optimal solutions.

The rest of this paper is organized as follows. Section 2 introduces a coordinated control strategy based on the WT and HESS for power smoothing. Section 3 presents the HESS, pitch angle, and rotor kinetic energy control. Section 4 presents the power primary allocation based on an adaptive VMD. Secondary power allocation based on an improved MADRL is described in detail in Section 5. The proposed method is verified by the MATLAB simulation and RT-LAB semi-physical real-time experiment in Section 6. Finally, the conclusion is drawn in Section 7.

2 A coordinated control strategy based on WT and HESS for power smoothing

2.1 System structure

The structure of the wind power generation system is shown in Fig. 1. It is mainly composed of a wind turbine, a permanent magnet direct drive generator, a back-to-back converter, and HESS. The HESS consists of two DC/DC converters, supercapacitors, and lithium batteries. The supercapacitors and lithium batteries are connected to the DC bus of the wind turbine's back-to-back converter through the DC/DC converters. P_w is the output of wind power, and P_g is the grid-connected power.

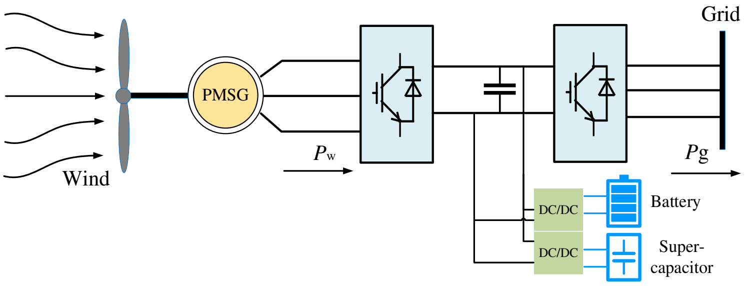


Fig.1. Wind power system structure

2.2 System control strategy

The control strategy of the WT-HESS combined generation system is shown in Fig. 2. The system control strategy is divided into two parts: power allocation and power control. The power allocation is composed of primary power allocation based on an adaptive VMD and secondary power allocation based on the SLA-MATD3. The former is to decompose the predicted power into high, medium, and low-frequency power. The latter is to optimally allocate the high and medium-frequency power, and obtain the power reference values of four parts: the pitch control, rotor kinetic energy control, lithium battery control, and supercapacitor control. The power control is composed of the pitch control, rotor kinetic energy control, lithium battery control, and supercapacitor control, and is responsible for completing the corresponding power smoothing.

The following control steps are considered: firstly, a wind power forecast model based on the LSTM is established to forecast the future output of wind power. Then, the value of the wind power forecast is preliminarily decomposed into three layers using an adaptive VMD: 1) the high-frequency preset reference power $P_{\mathrm{ref\_H}}$, 2) the medium-frequency preset reference power $P_{\mathrm{ref\_M}}$, 3) the low-frequency preset reference power $P_{\mathrm{ref\_L}}$ The low-frequency $P_{\mathrm{ref\_L}}$ is used as the expected smooth output power, and the high-frequency $P_{\mathrm{ref\_H}}$ and the medium frequency $P_{\mathrm{ref\_M}}$ are re-allocated through SLA-MATD3 to obtain the power reference values of $P_{\mathrm{sc\_ref}}$, $P_{\mathrm{b\_ref}}$, P_{pitch} and P_{rotor}, respectively. According to these power reference values, the respective smoothing tasks can be completed through the lithium battery control, supercapacitor control, pitch control, and rotor kinetic energy control.

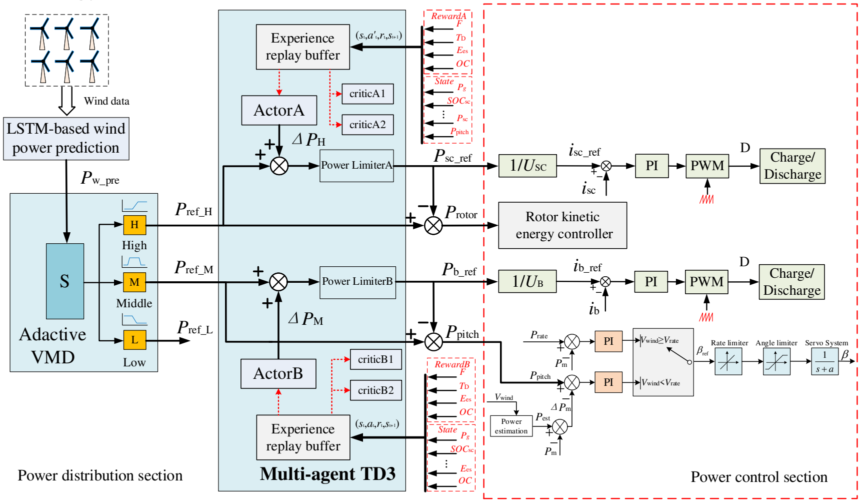


Fig.2. Control strategy for the WT-HESS combined generation system

3 Power smoothing control

3.1 Pitch control

Based on pitch control, the power smoothing is to adjust blade pitch angle β_w to achieve output power smoothing. The smoothing instruction is introduced into the pitch control to guide and determine the pitch angle reference β_{ref} [9,21]. The pitch control is divided into the control above the rated wind speed and below the rated wind speed. When the wind speed is above the rated wind speed, the error value of the rated power P_{rate} of the WT and the actual mechanical power P_{m} is sent to the PI controller. The control below the rated wind speed is the error between the mechanical power P_{est} calculated by the power estimation and the actual mechanical power P_{m} to obtain the mechanical power change value ΔP_{m}. The error value of P_{pitch} and ΔP_{m} obtained by an allocation algorithm is sent to the PI control, so as to control the pitch angle of the WT. The mechanical power estimate $P_{\mathrm{est}}=\tau(V_{\mathrm{wind}}/ V_{\mathrm{rate}})^3$, τ is an empirical parameter, V_{wind} is the actual wind speed and V_{rate} is the rated wind speed. The pitch control for wind power smoothing is shown in Fig. 3.

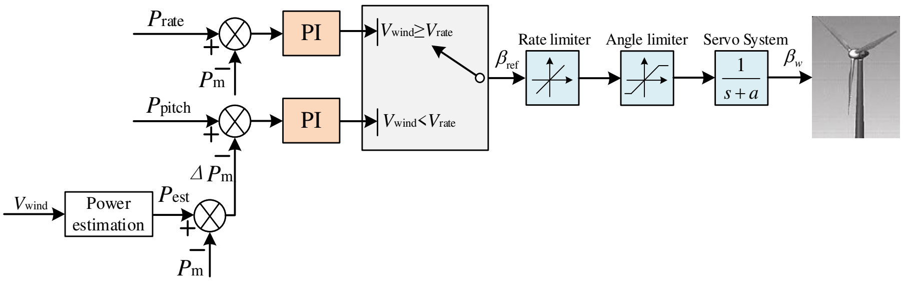


Fig.3. Pitch control

3.2 Rotor kinetic energy control

The principle of using the rotor kinetic energy to power smoothing is that the rotor has inertia during rotation, so corresponding energy is released and stored when the rotor speed changes. The rotor kinetic energy is treated as a virtual storage device for adjustment when the output power fluctuates.

According to the law of conservation of kinetic energy, the equation of rotor kinetic energy E can be obtained as follows,

$$E = \frac{1}{2} J\omega^2 \tag{1}$$

Where J and ω are the moments of inertia and rotor speed, respectively. In order to have wind power smoothing, the power instruction is introduced to obtain the new rotor kinetic energy,

$$E' = \frac{1}{2} J\omega^2 + \int P_{\text{rotor}} dt \tag{2}$$

Where P_{rotor} is the reference value for power variation of rotor kinetic energy control. By transforming the above equation, the new reference value of rotational speed is as follows,

$$\omega'_{\text{ref}} = \sqrt{\frac{2E'}{J}} = \sqrt{\frac{2(1/2\,J\omega^2 + \int_{t_0}^{t_1} P_{\text{rotor}} dt)}{J}} \tag{3}$$

The maximum power P_{mppt} is obtained through the maximum power tracking so as to obtain the new rotor reference power $P_{\text{ref\_rotor}}$ [9].

$$P_{\text{ref\_rotor}} = (P_{\text{mppt}} + P_{\text{rotor}})\frac{\omega}{\omega'_{\text{ref}}} \tag{4}$$

The power smoothing based on the rotor kinetic energy can be realized by the reference speed of the generator as shown in Eq. (4). The new power reference value $P_{\text{ref\_rotor}}$ can be obtained by Eq. (4) and used as the input of the power-current double-loop control of the generator side converter. The rotor kinetic energy control strategy for power smoothing is shown in Fig. 4. i_{d} and i_{q} are the currents of the d and q axes respectively.

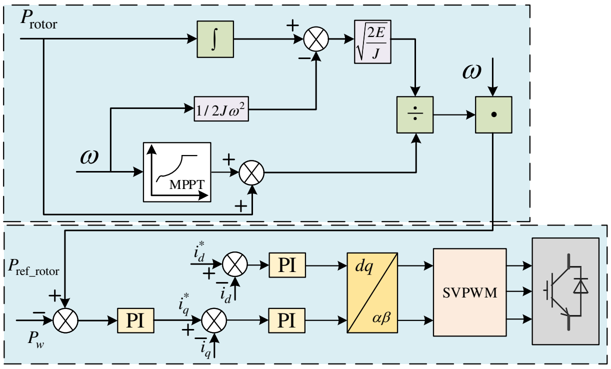


Fig.4. Rotor kinetic energy control

3.3 Control of the HESS

The control structure of the HESS is shown in Fig. 5. The power reference values of supercapacitors and lithium batteries are obtained through the multi-agent deep reinforcement learning. The supercapacitors process the high-frequency power signal $P_{\text{ref\_sc}}$ and the lithium batteries process the medium-frequency power signal $P_{\mathrm{ref\_b}}$. The HESS control adopts the current loop control. The reference power P_{ref} of the energy storage elements and the real-time voltage U are calculated to obtain the current reference value i_{ref}, and then the output power of the energy storage elements is adjusted through the current loop. K_{SC} and K_{b} is the current limiting modules of the supercapacitors and lithium batteries, respectively. By detecting the SOC value of the energy storage elements in real-time, the size of the current reference value is limited so as to realize the real-time protection of overcharge and over-discharge of the energy storage elements [22,23]. K_{SC} can be divided into charge current limiting coefficient $K_{\mathrm{SC\_c}}$ and discharge current limiting coefficient $K_{\mathrm{SC\_dis}}$. K_{b} can be divided into charge current limiting coefficient $K_{\mathrm{b\_c}}$ and discharge current limiting coefficient $K_{\mathrm{b\_dis}}$, as shown in Eq. (5) and Eq. (6).

$$K_{SC}\begin{cases} K_{SC\_c}=\begin{cases} 1 & SOC_{sc}<0.8 \\ 10\times(0.9\text{-}SOC_{sc}) & 0.8\le SOC_{sc}<0.9 \\ 0 & SOC_{sc}\ge 0.9 \end{cases} \\ K_{SC\_dis}=\begin{cases} 0 & SOC_{sc}<0.1 \\ 10\times(SOC_{sc}-0.1) & 0.1\le SOC_{sc}<0.2 \\ 1 & SOC_{sc}\ge 0.2 \end{cases} \end{cases} \tag{5}$$

$$K_{b}\begin{cases} K_{b\_c}=\begin{cases} 1 & SOC_{b}<0.8 \\ 10\times(0.9\text{-}SOC_{b}) & 0.8\le SOC_{b}<0.9 \\ 0 & SOC_{b}\ge 0.9 \end{cases} \\ K_{b\_dis}=\begin{cases} 0 & SOC_{b}<0.1 \\ 10\times(SOC_{b}-0.1) & 0.1\le SOC_{b}<0.2 \\ 1 & SOC_{b}\ge 0.2 \end{cases} \end{cases} \tag{6}$$

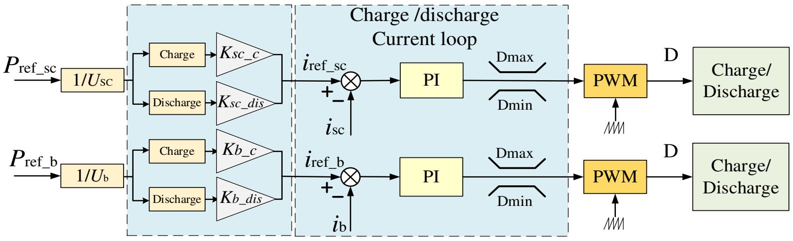


Fig.5.The HESS control

4 Primary power allocation based on adaptive VMD

4.1 Wind power forecast based on LSTM

The LSTM network is an improved network of the Recurrent Neural Network (RNN). Under the condition that the time feedback mechanism is basically unchanged, the mechanism of the Memory Cell and Gate (forgetting gate, input gate, and output gate) is introduced to realize the storage and control information [24].

Wind power forecast based on LSTM is to use LSTM network to carry out the dynamic time modeling of multivariate time series, so as to realize the wind power forecast. The prediction model based on LSTM network is shown in Fig. 6. Its forecasting process is as follows: Firstly, the network structure and parameters of the LSTM are initialized, and the wind speed and wind power data are normalized. Then, the wind speed and power data at time t-1 are used as the input of the LSTM network. The LSTM network is updated and the wind power value at time t is predicted. Finally, the wind power value at time t is obtained by renormalization. The above steps are repeated until the end of the process.

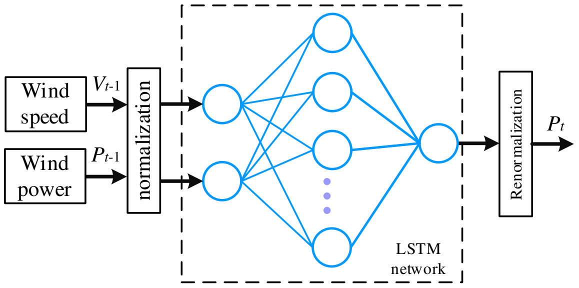


Fig. 6 The prediction model based on LSTM network

4.2 Adaptive VMD

VMD is a new signal analysis method with completely non-recursive and quasi-orthogonal [25]. It can decompose the original signal $f(t)$ into the different intrinsic modal functions $u_k(t)$ of K central frequencies ω_k by the preset decomposition scale K. The specific equation is as follows,

$$f(t)=\sum_{k=1}^{K}u_k(t) \tag{7}$$

$$u_k(t)=A_k(t)\cos(\phi_k(t)) \tag{8}$$

The realization of the VMD algorithm needs to pre-set the decomposition scale. The pre-set decomposition scale is highly subjective, and an unreasonable decomposition scale will cause over-decomposition or under-decomposition, while the adaptive VMD can independently determine the optimal decomposition scale to ensure a better decomposition effect[16]. An adaptive VMD primary power allocation method is used in this paper, and its flowchart is shown in Fig. 7. Firstly, the grid-connected wind power is extracted from the original power signal, which satisfies that the power fluctuations is lower than 5% of the rated power according to the requirements of grid-connected wind power in Denmark [26]. Secondly, according to the characteristics of the energy storage elements and the smooth control of the WT, the appropriate high-frequency and medium-frequency power signals are decomposed. The rules for the adaptive VMD: (1) The power fluctuations of the low-frequency P_L are less than 5% of the rated power, (2) the high-frequency P_H meets the characteristics of the supercapacitor control and rotor kinetic energy control, (3) the medium-frequency P_M meets the characteristics of the lithium battery control and pitch control. The expressions of high, medium, and low-frequency wind power signals extracted by the adaptive VMD method are as follows,

$$\begin{cases} P_L=u_1(t) \\ P_M=\sum_{i=2}^{m}u_i(t) \\ P_H=\sum_{i=\mathrm{m}}^{k}u_i(t) \end{cases} \tag{9}$$

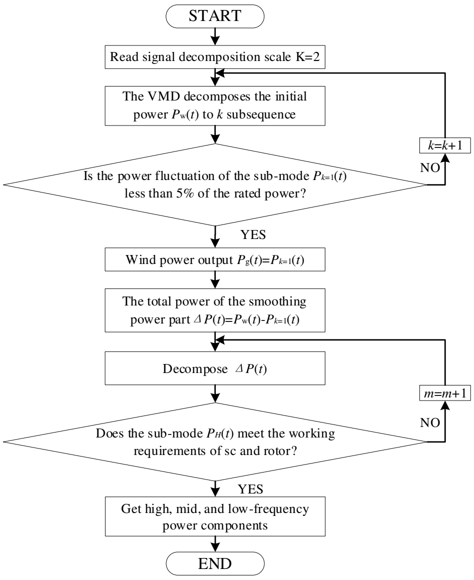

START
Read signal decomposition scale K=2
The VMD decomposes the initial power $P_w(t)$ to k subsequence
Is the power fluctuation of the sub-mode $P_{k=1}(t)$ less than 5% of the rated power?
NO
k=k+1
YES
Wind power output $P_g(t)$=$P_{k=1}(t)$
The total power of the smoothing power part $\Delta P(t)$=$P_w(t)$-$P_{k=1}(t)$
Decompose $\Delta P(t)$
Does the sub-mode $P_H(t)$ meet the working requirements of sc and rotor?
NO
m=m+1
YES
Get high, mid, and low-frequency power components
END

Fig. 7. Adaptive VMD primary allocation flow chart

5 Secondary power allocation based on an improved MADRL

5.1 Multi-agent twin-delay deterministic strategy gradient algorithm

The twin-delayed deep deterministic policy gradient algorithm (TD3) is a new reinforcement learning algorithm improved from the deep deterministic policy gradient (DDPG) algorithm [27,28]. The multi-agent TD3(MATD3) extends TD3 to multi-agent domains [29]. The MATD3 uses a centralized training and decentralized execution setting. The critic has access to all agents' past actions, observations, and rewards, as well as policies. Two centralized critics $Q^{\pi}{}_{i,\theta 1,2}(s,a_1,\ldots,a_n)$ are learned by exploiting this information. Each agent has a decentralized actor.

To solve the problem of overestimation in the Actor-Critic structure, the MATD3 network adopts a structure similar to Double-DQN, namely, two Q networks, one for selecting actions (the strategy of the current state) and one for evaluating the value of the current state. The deviation problem is solved by setting two critic values in the MATD3 and setting the target value as the minimum of the two Q values. In the learning process, the target network of two critic networks is used to obtain the smaller value of the next estimated value to calculate the update target, that is, the update target value is as follows,

$$y_i = r_i + \gamma \min\nolimits_{n=1,2} Q^{\pi}_{i,\theta'_n}(s', \pi_{\mu}(s')) \tag{10}$$

In Bellman's update of actor-network, the regularization technique of the smoothing target policy is used to reduce the high variance target value generated by the deterministic strategy method when updating actor. The actor-network update gradient,

$$\nabla_{\theta_{\mu,i}} J(\phi) \approx N^{-1} \sum \nabla_{\theta} \pi_{\phi} \nabla_{\phi}(s) Q_{\theta_i}(s, a_1, \ldots a_N)\,|_{a_i=\pi_{\phi}(s_i)} \tag{11}$$

Finally, two target network parameters are updated by the soft update strategy,

$$\begin{cases} \theta'_{i,n} \leftarrow \tau\theta_{i,n} + (1-\tau)\theta'_{i,n}, n = 1,2 \\ \phi' \leftarrow \tau\phi + (1-\tau)\phi' \end{cases} \tag{12}$$

5.2 Dynamic adjustment of SLA-MATD3 strategy based on *α*-state Levy noises

Exploration is an important part of the continuous action space learning. In the power allocation training, the agents need to explore new strategies to achieve optimal power allocation. Therefore, *α*-state Lévy noises are introduced to optimize the deterministic strategy generated by the MATD3 and adjust the noise dynamically to guide the agents to explore the strategy.

According to the Lévy-Itô decomposition theorem [30], the Lévy process can be regarded as the sum of a Gauss type continuous term and a Poisson type jump term. It indicates that the Lévy process has the characteristics of both the Gaussian White noise and Poisson noise.

The *α*-stable distribution is a family of four-parameter distribution functions, which is generally denoted as $S(x;\ \text{α}, \beta, \chi, \delta)$). It is also a limit distribution that can maintain the generation mechanism and propagation conditions of the natural noise process, and also a generalized Gaussian distribution. The density function of the *α*-stable distribution is defined as [31],

$$f(x;\alpha,\beta,\chi,\delta) = \int_{-\infty}^{+\infty} \exp\{i\delta x - \chi|t|^{\alpha}[1 + i\beta \,\mathrm{sgn}(t)\omega(t,\alpha)]\} e^{-itx} dt \tag{13}$$

$$\omega(t,\alpha) = \begin{cases} \tan\dfrac{\alpha\pi}{2}, & \text{if } \alpha \neq 1 \\ \dfrac{2}{\pi}\ln|t|, & \text{if } \alpha{=}1 \end{cases} \tag{14}$$

$$\mathrm{sgn}(t) = \begin{cases} 1, & \text{if } t > 0 \\ 0, & \text{if } t = 0 \\ -1, & \text{if } t < 1 \end{cases} \tag{15}$$

The *α*-stable Lévy process is a special case in the *α*-stable [32]. For $0<\alpha<2$, the sample orbit of the *α*-stable Lévy process is shown as follows,

$$L\left(nh\right)=\sum_{i=0}^{n}h^{1/\alpha}Y_i \tag{16}$$

$$Y_i=\begin{cases}\chi X_i+\delta, & if\,\alpha\neq 1\\ \chi X_i+\dfrac{2}{\pi}\beta\chi\ln\chi+\delta, & if\,\alpha=1\end{cases} \tag{17}$$

where h is the time step. α is a stability index of the Lévy process used to describe the decay rate of the Lévy process. δ is used to measure the deviation degree of the Lévy process from the mean. β is used to measure the skew of the Lévy process. χ is used to describe the position of the Lévy process.

During the learning process, parameters in the α-state Lévy noises are dynamically adjusted according to the state of the WT and HESS, to guide the agents to explore more effectively. The dynamic adjustment rules of noise parameters are as follows ($\eta 1$, $\eta 2$, and $\eta 3$ are noise adjustment parameters),

$$L'\left(nh\right)=\begin{cases}\sum_{i=0}^{n}\eta_1 h^{1/\alpha}Y_i, SOC\in(0.1,0.2)\cup(0.8,0.9)\\ \sum_{i=0}^{n}\eta_2 h^{1/\alpha}Y, SOC\in(0.2,0.3)\cup(0.7,0.8)\\ \sum_{i=0}^{n}\eta_3 h^{1/\alpha}Y_i, SOC\in(0.3,0.4)\cup(0.6,0.7)\end{cases} \tag{18}$$

When the SOC is far from 0.5, the energy storage elements may be over-charged and over-discharged, which is not desired. In this case, it is hoped that the noise will be greater so that the policy can be explored more to find a better policy. When the SOC is close to 0.5, the energy storage elements will not be over-charged or over-discharged, which is the desired result. In this case, if the noise will be smaller, then the policy will no longer be over-explored, and it will gradually be stabilized. The variation of noise is determined through SOC range division and noise parameters. Eight split points are divided for SOC according to the distance from 0.5: 0.1, 0.2, 0.3, 0.4, 0.6, 0.7, 0.8 and 0.9. The first ranges are 0.1~0.2 and 0.8~0.9, respectively, which is farthest from 0.5, so η_1 is set to 1.8; the second ranges are 0.2~0.3 and 0.7~0.8, respectively, η_2 is set to 0.5; the third ranges are 0.3~0.4 and 0.6~0.7, respectively, η_3 is set to 0.2. When SOC is in the range of 0.4~0.6, the noise on the action is removed.

The SLA-MATD3 algorithm is shown in Table 1.

Table 1

SLA-MATD3 algorithm

```
SLA-MATD3 algorithm
Initialize the two critic networks $Q^{\pi}_{i,\theta 1}$, $Q^{\pi}_{i,\theta 2}$ 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: $\theta'_{i,1}\leftarrow\theta_{i,1}$, $\theta'_{i,2}\leftarrow\theta_{i,2}$, $\phi'_i\leftarrow\phi_i$ initialize the experience pool $R$;
    for $t$=1,2…$T$ do
        The $\alpha$-state Lévy noise is introduced. For each agent $i$, select random action $a_i\sim\pi_i(s_i)+\varepsilon_i$, and the noise is dynamically adjusted to explore the current deterministic strategy;
        Execute $a_{1,t}$, …, $a_{N,t}$ and observe reward $r_{i,t}$ and new state $r_{i,t+1}$;
        Policy networks store state-transition tuples ($s_t$, $a_{1,t}$, …, $a_{N,t}$, $r_{1,t}$, …, $r_{N,t}$, $s_t$) in the experience pool $R$;
        Randomly sampling $N$ tuples from the experience pool $R$ as a mini-batch of training data for online policy network and value network;
        for agent $i$=1 to $N$ do
            Update target value $y_i=r_i+\gamma\min_{n=1,2}Q^{\pi}_{i,\theta'_n}(s',\pi_\mu(s'))$
            Update critic network parameter
            if $t$ mod $d$ then
                Update action parameters by policy gradient $\nabla_{\theta_{\mu,i}}J(\phi)\approx N^{-1}\sum\nabla_\theta\pi_\phi\nabla_\phi(s)Q_{\theta_i}(s,a_1,\ldots a_N)|_{a_i=\pi_\phi(s_i)}$
                Update the target value network and policy network parameters $\theta_{i,1}$, $\theta_{i,2}$, $\phi_i$,
                    $\begin{cases}\theta'_{i,n}\leftarrow\tau\theta_{i,n}+(1-\tau)\theta'_{i,n}, n=1,2\\ \phi'\leftarrow\tau\phi+(1-\tau)\phi'\end{cases}$
            end if
        end for
    end for
```

end

5.3 Secondary allocation based on SLA-MATD3

The secondary power allocation based on the SLA-MATD3 is shown in Fig. 8. It has set priority. Power fluctuations are preferentially handled by batteries and supercapacitors. When the *SOC* of the energy storage system is out of range ([SOC<0.2&P_{b_ref}>0]， [SOC>0.8&P_{b_ref}<0]) or the power exceeds the rated power, the pitch angle and rotor kinetic energy will smooth the power. This part is mainly realized by the "power limiter" in Fig. 8. The agents of SLA-MATD3 are optimized for secondary power allocation. Taking medium-frequency power allocation as an example, since $P_{pitch}= P_{ref\_M} - P_{b\_ref}$ and the battery reference power P_{b_ref} is determined according to the action value $\varDelta P_M$ of agent B and the "power limiter", the secondary power allocation can be optimized by agent B.

The agents of the SLA-MATD3 are divided into agents A and B. The agents A and B allocate the high-frequency power signal P_{ref_H} and the medium-frequency power signal P_{ref_M} respectively to obtain the optimal reference power controlled by the HESS and WT. Each agent has a decentralized actor, which only accesses its local observations. At the same time, each agent has two centralized critics, which can access the observations and action of all agents. The two agents are both cooperative and competitive. The common goal is to smooth wind power, while the two agents have different goals, which are determined through the reward mechanism.

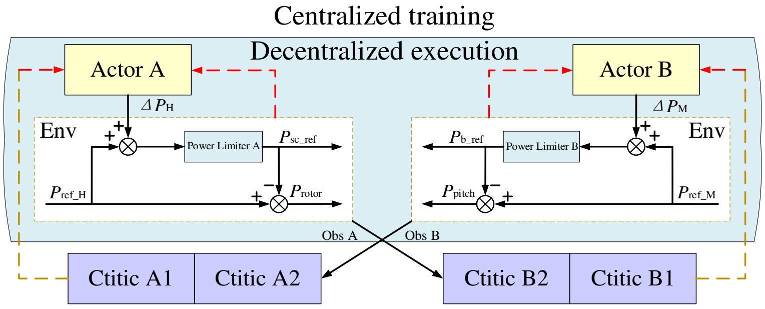


Fig. 8. Secondary allocation structure based on SLA-MATD3

For the WT-HESS model, the environment provides information to the agents A and B: the wind speed V_{wind}, the output P_w of the wind power, the grid-connected power P_g and fluctuation P_{flu_g}, the output P_B and the SOC_b of the lithium batteries, the output P_{sc} and the SOC_{sc} of the supercapacitors, the pitch angle β_w, the rotor speed ω. The state-space of the wind-storage combined power generation model is defined as,

$$S = [V_{\text{wind}}, P_w, P_g, P_{flu\_g}, P_B, P_{SC}, SOC_B, SOC_{SC}, \beta_w, \omega] \tag{19}$$

After observing the state information of the environment, the agents select an action in the action space according to its strategy π. The action spaces A_1 and A_2 are composed of the high-frequency and medium-frequency power compensation values respectively, and their expressions are as follows,

$$\begin{cases} A_1 = \varDelta P_H \\ A_2 = \varDelta P_M \end{cases} \tag{20}$$

In the learning process, the tasks each agent needed are set by the reward function, which determines whether they are cooperative or competitive. Their common reward is the output power smoothness F, and each of them has different rewards. For the agent A, the supercapacitor state of charge reward $y_{sc,i}$, total supercapacitor charge and discharge energy E_{sc}, and supercapacitor output coefficient OC_{sc} are used as reward r_1. For agent B, the lithium battery state of charge reward $y_{b,i}$, total lithium battery charge and discharge energy E_b, lithium battery output coefficient OC_b, and pitch angle standard deviation β_{std} are used as reward r_2,

$$\begin{cases} r_1(t) = \sum_{i=0}^{3} \xi y_{sc,i} - \left(\rho_{sc} E_{sc} + \lambda F + \psi_{sc} OC_{sc}\right) \\ r_2(t) = \sum_{i=0}^{3} \xi y_{b,i} - \left(\rho_b E_b + \lambda F + \psi_b OC_b + \zeta \beta_{std}\right) \end{cases} \tag{21}$$

$$\begin{cases} y_{sc,0} = -4, SOC(t_k) < 0.1 \ or \ 0.9 < SOC(t_k) \\ y_{sc,1} = 0.5, 0.2 \le SOC(t_k) \le 0.8 \\ y_{sc,2} = 1, 0.3 \le SOC(t_k) \le 0.7 \\ y_{sc,3} = 2, 0.4 \le SOC(t_k) \le 0.6 \end{cases} \tag{22}$$

where ξ, ρ_{sc}, ρ_b, λ, Ψ_{sc}, Ψ_b and ζ are the weight coefficients. The calculation methods of $y_{b,i}$ and $y_{sc,i}$ are the same, except that the SOC in Eq. (24) is the SOC of the lithium batteries.

The output power smoothness of the WT [14,16],

$$F = \sum_{k=1}^{T/\Delta t} \left(\frac{\Delta P_g(k)}{P_{rate}} \right)^2 \tag{23}$$

$$\Delta P_g(k) = \left| P_g(t_0 + k\Delta t) - P_g(t_0 + (k-1)\Delta t) \right| \tag{24}$$

where the power smoothness F represents the smoothing effect of the output power of the wind turbine. The smaller the F, the better the smoothing effect and the smaller the impact on the power grid. ΔP_g is the absolute value of the WT output power fluctuation, P_{rate} is the rated power of the WT, Δt is a time interval, and T is the total time.

Total charge and discharge energy of the ESS,

$$E_{es} = \Delta t \sum_{t=0}^{T} \left| P_{es}(t) \right| \tag{25}$$

where E_{es} represents the total energy absorbed and released by ESS when smoothing wind power. The smaller the E_{es} value is, the smaller the working pressure of ESS is.

The output coefficient of energy storage system is given by,

$$OC = \sqrt{\frac{1}{T-1} \sum_{t=0}^{T} (SOC(t) - 0.5)^2} \tag{26}$$

where OC is the output capacity of the energy storage system. The closer the SOC value is to 0.5, the smaller OC is, and the stronger the ESS's ability to cope with future power fluctuations.

6 Results and analysis of simulation and experiment

6.1 WT-HESS configuration and power forecast

To verify the effectiveness of the proposed method, a 10MW WT-HESS combined generation system was established on MATLAB. The specific parameters are shown in Table 2.

Table 2

Parameters of 10MW wind energy storage system

Item	Value
Rated power /MW	10
Wheel hub height /m	90
Tower height /m	87.6
Rated wind speed /(m/s)	12
Rated speed of generator /(rad/s)	94
Stator phase rated resistance /(Ω)	0.05
Galvanic inductance /(H)	0.0002
Generator inertia /(kg·m^2)	543
Rated dc voltage /(V)	1150
Capacity of the DC side /(F)	0.075
Lithium battery rated power /MW	0.3
Lithium battery capacity /(MW·h)	0.28
Initial capacity of the Lithium battery	50%
Supercapacitor rated power /MW	0.1

Supercapacitor battery capacity /(MW·h)	0.031
Initial capacity of the supercapacitor	50%

By simulating the dispatch operation, the charging/discharging power and stored energy of the HESS can be obtained at any time during the whole operation period. Therefore, the capacity of the HESS can be calculated according to these results. The minimum capacity of the HESS that meets SOC operation requirements is calculated as follows [33],

$$E_{\text{rate}} = \frac{\nu \cdot (\max[E_{\text{flu}}(t)] - \min[E_{\text{flu}}(t)])}{\eta_e \cdot (SOC_{\max} - SOC_{\min})} \tag{27}$$

$$E_{\text{flu}}(t) = \Delta t \sum_{t=0}^{T} P_{es}(t) \tag{28}$$

where $E_{\text{flu}}(t)$ is the energy fluctuation of the HESS at different times relative to the initial state, ν configures the margin for the HESS capacity, and η_e is the charging and discharging efficiency of energy storage.

The Turbsim software developed by National Renewable Energy Laboratory (NREL) generated wind with an average wind speed of 12m/s based on the Kaimal turbulence model. The wind speed is shown in Fig. 9.

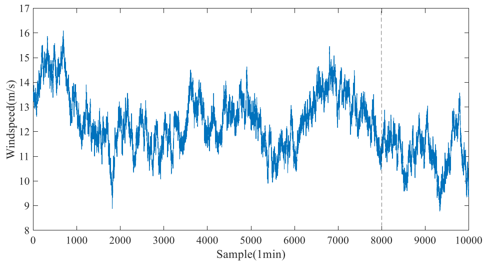
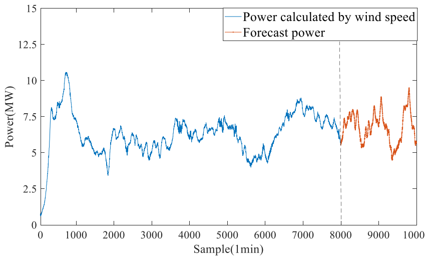


Fig. 9. Turbulent wind speed

Input 10,000 wind speed data with sampling interval of 1min into the wind power model to generate 10,000min wind power output power, which was used as the data of the wind power forecasting model. In the wind power forecasting part, the LSTM time series forecasting toolbox in MATLAB was used to establish the wind power forecasting model and using wind speed data to obtain future wind power. The first 8000min were used as training data and the second 2000min as test data. The LSTM network makes single-step forecast for the wind power time series and updates the network state at each forecast. The comparison between the forecast power value and the power calculated by wind speed is shown in Fig. 10, and the forecast power error is shown in Fig.11.

In the simulation, the model parameters were determined after multiple tests and tuning, in which the number of hidden layers was set to 100, the number of network training was 250, the initial learning rate was 0.005, and the mini-batch was 128 with the Adam optimizer. The gradient threshold was set to 1 to prevent gradient explosion.



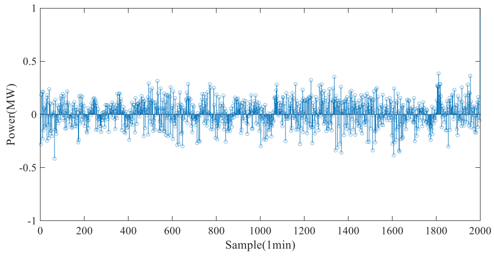
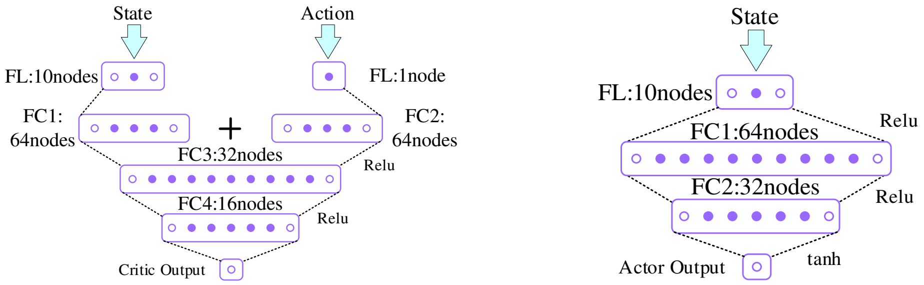


Fig. 10. Comparison of forecast power and power calculated by wind speed

Fig. 11. Forecasting power error

6.2 Analysis of algorithm training results

On the premise that other parameters remain unchanged, the learning ability and speed of the agent A and agent B of different multi-agent deep reinforcement learning algorithms were compared. The structures of the critic-network and the actor-network of the algorithm are shown in Fig. 12. (Note that agent A and agent B have the same network structure).



a Critic-network structure b Actor-network structure

Fig. 12. Network structure of SLA-MATD3 algorithm

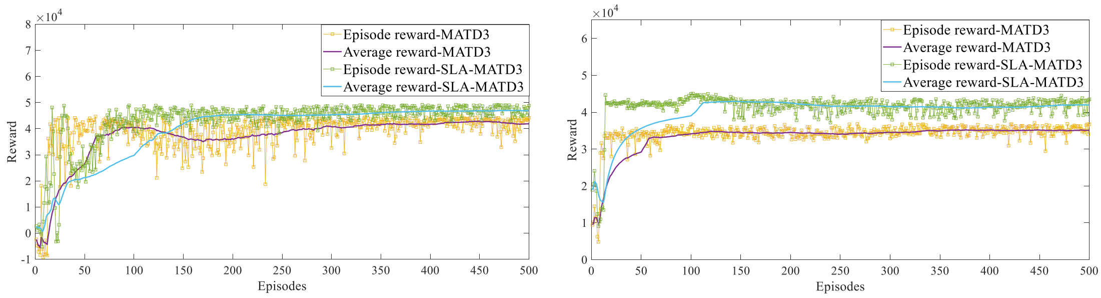
The hyperparameters of the algorithm and the parameters of the reward function were determined through multiple tests and tuning. The specific parameters are shown in Table 3.

Table 3

Hyperparameters and other parameters for algorithm training

Hyperparameters	Value	Other parameters	Value
Number of trainings	500	ξ	50
Mini-batch	512	ρ_{sc}	1×10^{-4}
Experience buffer capacity	1×10^{6}	ρ_b	1×10^{-4}
discount factor	0.995	λ	10
learning rate	0.01	Ψ_{sc}	20
optimizer	Adam	Ψ_b	20
		ζ	40
		$\eta_1/\eta_2/\eta_3$	1.8/0.5/0.2

The training process of the SLA-MATD3 algorithm was carried out with CPU R7 3800x@3.9GHz RTX2060TI hardware. In one iteration of the SLA-MATD3, the training function calls are 2000 times in total, and the operational time of each call is 7.7858×10^{-3}s. Reward values of different algorithms are shown in Fig. 13. To ensure the training effect, the agents conducted 500 trial and error training and took every 50 rewards to calculate the average reward. For agent A, the reward value of the SLA-MATD3 begins to converge at epoch 100, and the final convergence value remains at 46000. The reward value of the MATD3 begins to converge at epoch 75, but the final convergence value is maintained at 41000, lower than that of the SLA-MATD3. For agent B, MATD3 and SLA-MATD3 begin to converge around epoch 25, but the final convergence value of the SLA-MATD3 is 42000, which is obviously higher than that of the MATD3. Through the results and analysis, it can be concluded that the proposed multi-agent deep reinforcement learning can obtain the highest reward value.



a Agent A reward b Agent B reward

Fig. 13. Reward values of different algorithms

To verify the advantages of multi-agent DRL over single-agent DRL in solving complex problems, the global rewards of multi-agent DRL and single-agent DRL were compared. The results are shown in Fig. 14. As can be seen from Fig. 14, the R_{global} of SLA-MATD3 starts to converge at the 100th iteration, and its final average convergence value is maintained at 89354.7. The R_{global} of MATD3 starts to converge at the 75th iteration, and its final average convergence value is maintained at 79464.5. The R_{global} of Single-TD3 starts to converge at the 150th iteration, and its final convergence value is maintained at 72122.1. The final convergence values of R_{global} of SLA-MATD3 and MATD3 are all higher than that of Single-TD3, indicating that the effect of multi-agent deep reinforcement learning is better than that of single agent DRL. After taking into consideration these results, the multi-agent deep reinforcement learning algorithm was chosen to solve the quadratic assignment problem in this study.

$$(R_{\text{global}} = \sum\nolimits_{i=0}^{3} \xi y_{sc,i} + \sum\nolimits_{i=0}^{3} \xi y_{b,i} - (\lambda F + \rho_{sc} E_{sc} + \psi_{sc} OC_{sc} + \rho_b E_b + \psi_b OC_b + \zeta \beta_{std}))$$

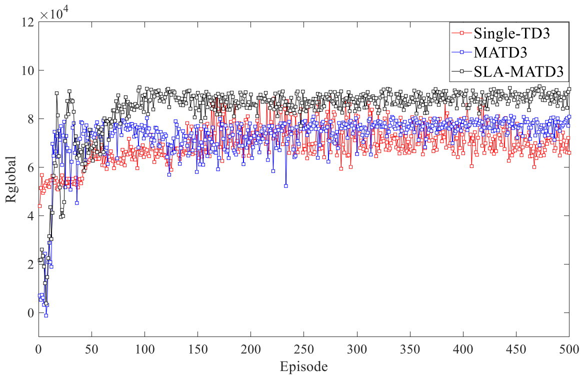
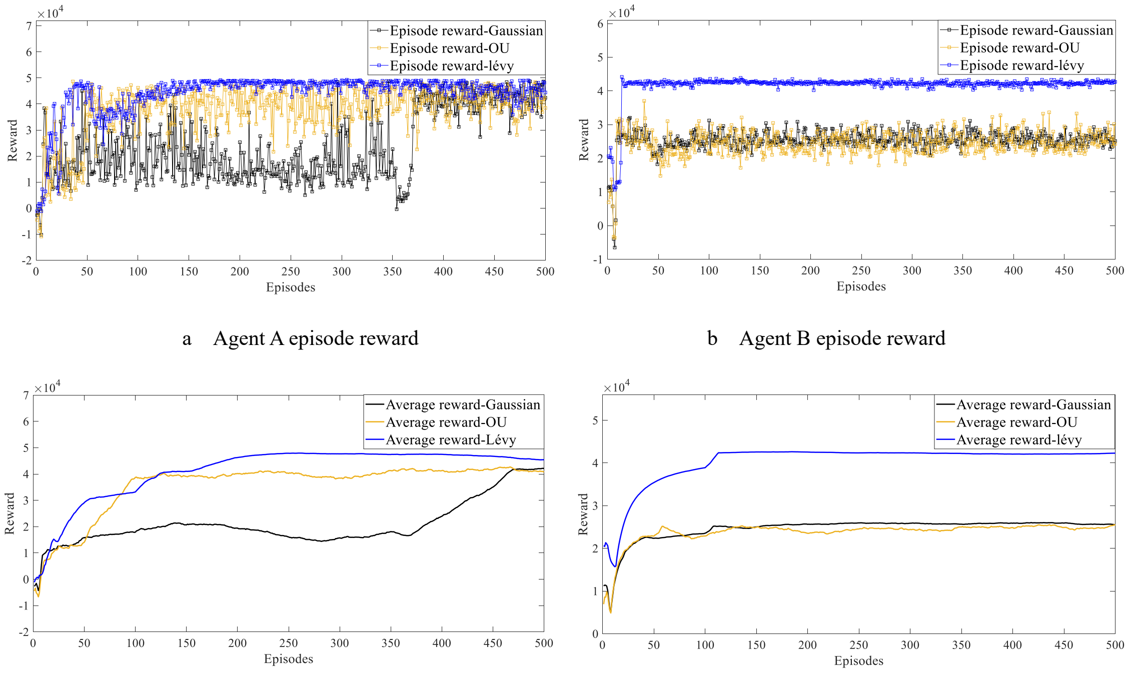


Fig.14 Global index Comparison

Three types of the Gaussian Noise, the OU Process, and the α-state Levy Noise were added to the MATD3 to test their learning efficiency. The reward values of different noises are shown in Fig. 15. By comparing the results, the convergent reward values of agent A and agent B of MATD3 with the introduction of α-state Levy noise are both greater than the reward values of the other two types of noise, indicating that the agent of this method has a stronger exploration ability.



a Agent A episode reward

b Agent B episode reward

c Agent A average reward

d Agent B average reward

Fig. 15. Reward values of different noises

6.3 Analysis of simulation results

Simulations were carried out on the model of the WT-HESS combined generation system. In what follows, the simulation results from four methods will be compared and analyzed: the WT control, the predictive control with HESS, the predictive control with WT and HESS, and the proposed control.

WT control is a method combining pitch control with rotor kinetic energy control for wind power smoothing. The method is to divide the smooth commands into high and low-frequency, where the pitch control processes low-frequency signals, while the rotor kinetic energy control processes high-frequency signals. The prediction-based optimal control of HESS is to determine the smoothing commands of the HESS by predicting power. They are then divided into high and low frequency commands, which are substituted into a dynamic optimization model to determine the new power commands, so that according to which the supercapacitors and lithium batteries can act. The prediction-based optimal control of the HESS and WT is to directly divide the predicted power into four parts of smoothing commands, and they are then substituted into the dynamic optimization model to determine the new power commands, so that the lithium batteries, supercapacitors, pitch angle, and rotor kinetic energy control can act according to the new power commands.

The output power of the WT using the proposed control method and other control methods is shown in Fig. 16. The power evaluation indexes are shown in Table 4.

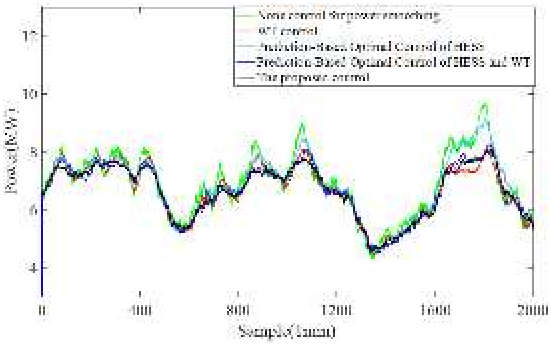


Fig. 16. Output power with different control methods

Table 4

Power evaluation indexes

Control methods	$\Delta P_{g,sum}$(10min) MW	$\Delta P_{g,sum}$(1min) MW
None control for power smoothing	315.40	87.27
WT control	218.48	57.87
Prediction-Based Optimal Control of HESS	187.85	37.39
Prediction-Based Optimal control of HESS and WT	173.46	36.55
The proposed control	147.17	34.33

According to the power and absolute value of power fluctuation results, compared with other control methods, the output power curve of the proposed control method is smoother, especially in the position with a large fluctuation degree. The sum of absolute values of wind power fluctuations of 10min and 1min is 147.17MW and 34.33MW, respectively, which are lower than those of other control methods. The results show the superiority of the proposed control method in smoothing power fluctuation.

The pitch angle β_w with different control methods is shown in Fig. 17. Rotor speed ω with different control methods is shown in Fig. 18.

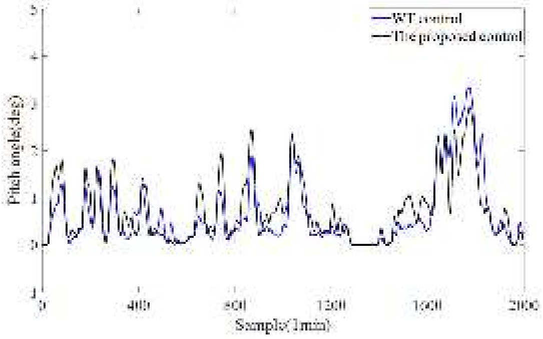


Fig. 17. Pitch angle with different control methods

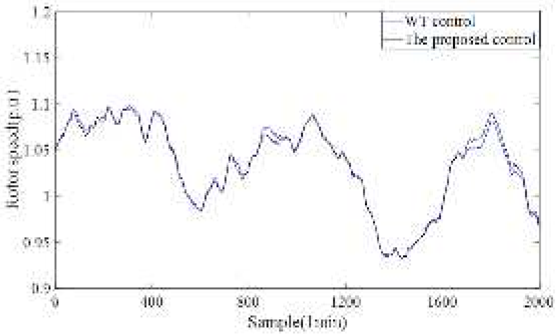
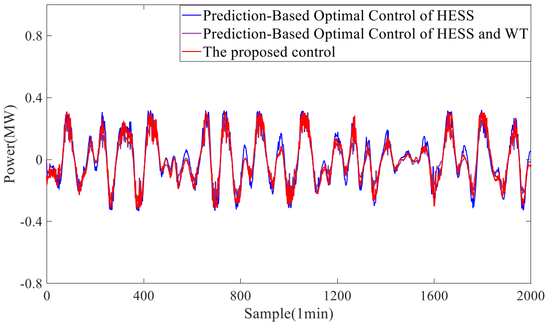


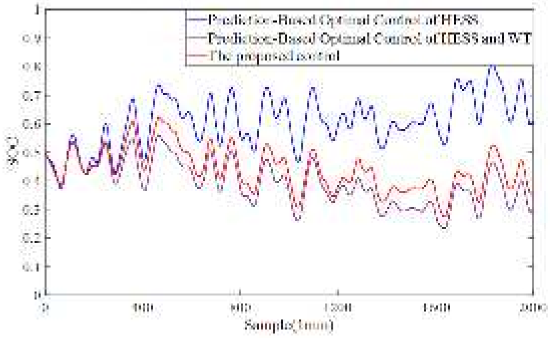
Fig. 18. Rotor speed with different control methods

It can be observed from Fig. 17 that the standard deviation β_{std} of the pitch angle of WT control, and the proposed method are 0.7272, and 0.6705, respectively. Compared with WT control, the pitch angle fluctuation degree of the proposed method is lower, and the fatigue load of blades is reduced to a certain extent. As can be seen from Fig. 18, according to smoothing instructions, the rotor accelerates and decelerates to suppress power fluctuations.

The output power and SOC of the lithium batteries with different control methods are shown in Fig. 19. The output power and SOC of the supercapacitors with different control methods are shown in Fig. 20. The overall evaluation indexes are shown in Table 5.

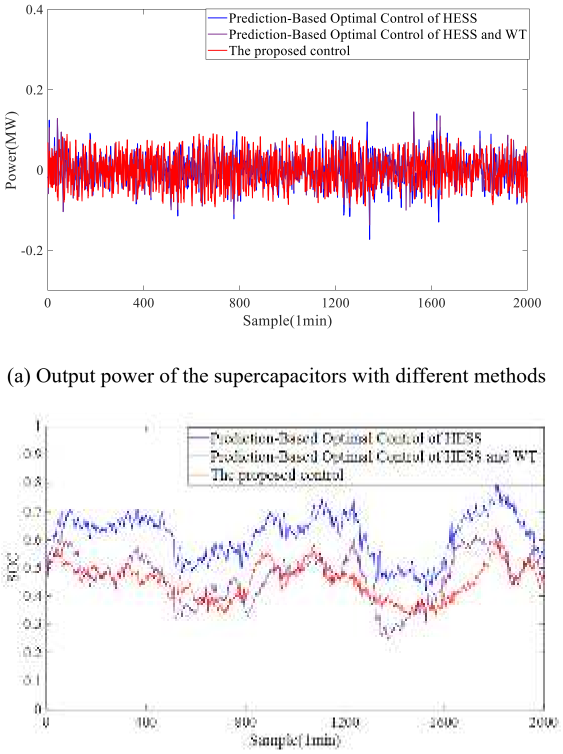


(a) Output power of the lithium batteries with different methods



(b) SOC of the lithium batteries with different methods

Fig. 19. Output power and SOC of the lithium batteries with different control methods



(a) Output power of the supercapacitors with different methods

(b) SOC of the supercapacitors with different methods

Fig. 20. Output power and SOC of the supercapacitors with different control methods

Table 5

Overall evaluation indexes

Control methods	F	E_{B}/MW·h	E_{SC}/MW·h	OC_B	OC_{SC}
None control for power smoothing	1.29	—	—	—	—
WT control	0.72	—	—	—	—
Prediction-Based Optimal Control of HESS	0.51	2.05	0.51	0.1387	0.1390
Prediction-Based Optimal control of HESS and WT	0.41	1.70	0.44	0.1314	0.0924
The proposed control	0.38	1.80	0.45	0.0942	0.0770

It can be observed from the evaluation indexes in Table 5 and Fig. 16 that the wind power smoothness F combined HESS with WT control is lower than that of HESS and WT control alone, indicating that the addition of the WT control on the basis of the HESS control can make the system have better power smoothing ability. The total charging and discharging energy E_{B} and E_{SC} of the lithium batteries and supercapacitors with the prediction-based optimal control of HESS are 2.05MW·h and 0.51MW·h, respectively, which are greater than the prediction-based optimal control of the HESS and WT and the proposed control. The results show that the pitch control and the rotor kinetic energy control can take part in the power smoothing task, further reduce the wind power fluctuation and the output of the lithium batteries and supercapacitors, as well as the working burden of the HESS.

It can be seen from Table 5 that the output coefficients OC of the lithium batteries and supercapacitors of the proposed control method are 0.0942 and 0.0770, respectively, which are significantly smaller than other control methods. It can also be seen from Fig.19b and Fig. 20b that compared with other methods, the SOC curves of the lithium batteries and supercapacitors of the proposed control method are closer to 0.5. At the same time, the wind power smoothness F is 0.38 with the proposed control method, which is smaller than that of other control methods. This result shows that the agents of the SLA-MATD3 in the proposed method can intelligently allocate power, ensure the smoothness of wind power, and make the SOC value of HESS closer to 0.5, so that the HESS has a stronger ability to cope with future power fluctuations. Although the total charge and discharge energy E_{B} and E_{SC} of the proposed control method are a little bit higher than those of the prediction-based optimal control method of the HESS and WT, the wind power smoothness F and the output coefficient OC of the lithium batteries and supercapacitors with the proposed control method are smaller than those of the two methods. The wind power smoothness F reflects the smoothing power effect and the output coefficient OC reflects the ability to cope with future power fluctuations. These two parameters are the main factors for wind power smoothing, which are more important than E_B and E_{SC}. Taking all these factors into account, the control strategy proposed in this paper is better.

6.4 RT-LAB semi-physical real-time experiment

To verify the effectiveness of the proposed control strategy, a model of a wind energy storage system was established on the RT-LAB (OP5600), and a semi-physical real-time simulation experiment was conducted. The experiment platform of the RT-LAB is shown in Fig. 21. The RT-LAB is used to construct wind power permanent magnet synchronous generator set and hybrid energy storage system. The computer processes the deep reinforcement learning part, and the DSP controller (TMS320F2812) is responsible for the realization of the HESS, converter control and pitch control.

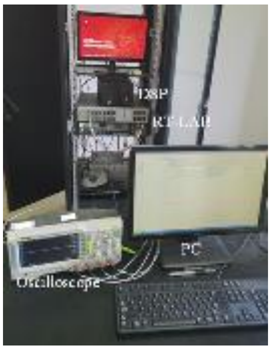


Fig. 21. RT-LAB simulation experiment platform

The experimental conditions were consistent with the simulation situation, and the 2000min simulation was transformed into a real-time simulation experiment of 600s. The actor (policy) of SLA-MATD3 is only executed on the lower-level coordinated control (RTLAB) system according to the states and there is no learning here. In this process, the LSTM function calls are 2000 times in total, and each operational time is 3.6916×10^{-4}s. The test function calls of the deep reinforcement learning are 2000 times in total, and each operational time is 1.2916×10^{-3}s. The RTLAB real-time simulation with 2000 sampling times and a total simulation time of 600s can realize real-time control of this system. The output power of the wind power using the proposed control method is shown in Fig. 22. The output power and SOC of lithium battery and supercapacitor are shown in Fig. 23 and 24, respectively.

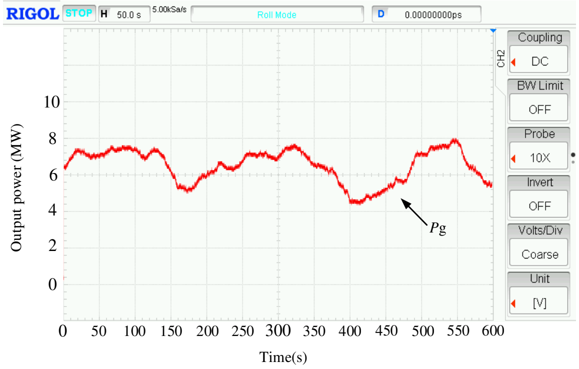


Fig. 22. Output power of the wind power

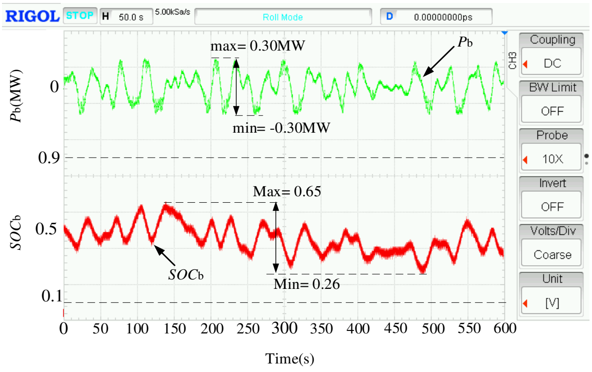


Fig. 23. Lithium battery power output and SOC_b

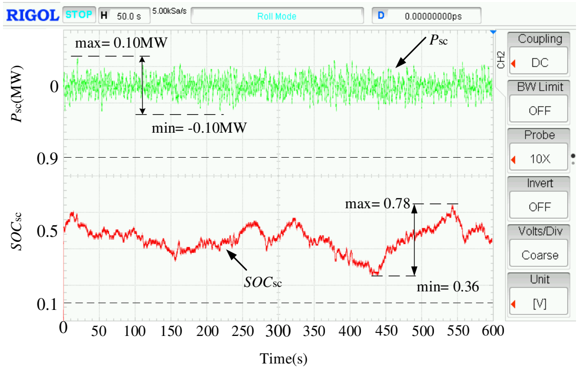


Fig. 24. Supercapacitor output power and SOC_{sc}

It can be seen from Fig. 22 that the proposed control method has a good smoothing effect on the output power of the grid-connected side. As can be seen from Fig. 23 and Fig. 24, the maximum and minimum SOC values of lithium batteries and supercapacitors are SOC_{b_max}=0.65, SOC_{b_min}=0.26, SOC_{sc_max}=0.78, SOC_{sc_min}=0.36, respectively, and the SOC values are all within the range of 0.1~0.9. Neither the lithium batteries nor supercapacitors have been overcharged. The SOC has been kept far away from the boundary value, which indicates that lithium batteries and supercapacitors have a good ability to cope with power fluctuations.

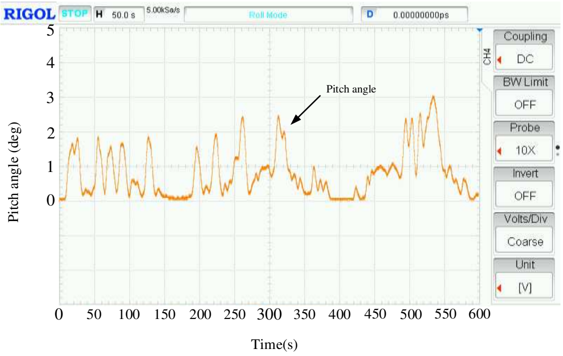


Fig. 25. Pitch angle

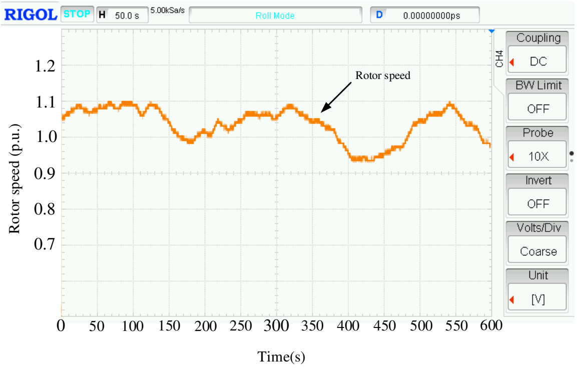


Fig. 26. Rotor speed

The pitch angle and rotor speed under the proposed control method are shown in Fig. 25 and Fig. 26, respectively. As can be seen from the resulting figure, the pitch angle changes according to the smoothing instruction on the premise of satisfying the rotor speed ring control, which meets the expected requirements. The unit value of rotor speed is maintained in the range of 0.7~1.2 to complete the corresponding smoothing task.

Comprehensive experimental results have shown that the proposed control method has good performance in output power fluctuation smoothing, SOC of lithium battery and supercapacitor, pitch angle, and rotor speed, which shows the feasibility and effectiveness of the proposed control method.

7 Conclusion

A coordinated control strategy for a wind turbine and hybrid energy storage system based on multi-agent deep reinforcement learning for wind power smoothing is proposed in this paper. The strategy uses the rotor kinetic energy and pitch control of WT and HESS to smooth the wind power output so that HESS and WT share the working load of smoothing power with each other. Based on the predicted wind power, the adaptive VMD can pre-allocate power for future wind power changes, and then intelligently allocate power through the improved deep reinforcement learning SLA-MATD3 algorithm. On the premise of smoothing the wind power output, HESS always has sufficient charging and discharging capacity to maximize the benefit of the energy storage system, and at the same time, it reduces the pressure of wind power control to suppress the wind power fluctuation. In the improved deep reinforcement learning algorithm, α-state Levy noises are introduced and dynamically adjusted to guide agents to better explore and help avoid local optimal solutions. The simulation and RT-LAB semi-physical real-time experiments show that the proposed control strategy is superior to other methods in the same case in many aspects.

Acknowledgment

Financial support was provided in part by the National Natural Science Foundation of China (62033014, 61673166), Natural Science Foundation of Hunan Province (2021JJ50006, 2022JJ50074). Hunan Engineering Research Center of Electric Drive and Regenerative Energy Storage and Utilization.

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, 2021; 137: 1364-0321.https://doi.org/10. 1016/j.rser.2020.110455.

[2] Tong Zheming, Cheng Zhewu, Tong Shuiguang. A review on the development of compressed air energy storage in China: Technical and eco-nomic challenges to commercialization. *Renewable and Sustainable Energy Reviews*, 135, 2021. https://doi.org/10.1016/j.rser.2020. 110178.

[3] Belaid S , Rekioua D , Oubelaid A , et al. A power management control and optimization of a wind turbine with battery storage system. *Journal of Energy Storage*, 2022, 45: 103613. https://doi.org/10.1016/j.est.2021.103613.

[4] Sun Y , Pei W , Jia D , et al. Application of integrated energy storage system in wind power fluctuation mitigation. *Journal of Energy Storage*, 2020, 32(2):101835. https://doi.org/10.1016/j.est.2020.101835

[5] Karaipoom, T. and Ngamroo, I. Optimal superconducting coil integrated into DFIG wind turbine for fault ride through capability enhancement and output power fluctuation suppression. *IEEE Transactions on Sustainable Energy*, 2015, 6(1): 28-42. https://doi.org/10.1109/TSTE.2014.2355423.

[6] Zhai, Y. et al. Research on the application of superconducting magnetic energy storage in the wind power generation system for smoothing wind power fluctuations. *IEEE Transactions on Applied Superconductivity*, 2021, 31(5): 1-5. https://doi.org/10.1109/TASC. 2021.3064520.

[7] 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 Transactions on Sustainable Energy*, 2017; 8(3): 990-999. https://doi.org/ 10.1109/TSTE.2016. 2637907.

[8] X. Tang, M. Yin, C. Shen, et al. Active Power Control of Wind Turbine Generators via Coordinated Rotor Speed and Pitch Angle Regulation. *IEEE Transactions on Sustainable Energy*. 2019; 10(2): 822-832. https://doi.org/ 10.1109/TSTE.2018.2848923.

[9] Lyu, X., Zhao, J., Jia, Y., Xu, Z., and Wong, K. Po. Coordinated control strategies of PMSG-based wind turbine for smoothing power fluctuations. *IEEE Transactions on Power Systems*, 2019; 34(1): 391-401. https://doi.org/ 10.1109/TP WRS.2018.2866629.

[10] Jiang, Q. and Hong, H. Wavelet-based capacity configuration and coordinated control of hybrid energy storage system for smoothing out wind power fluctuations. *IEEE Transactions on Power Systems*, 2013, 28(2): 1363-1372. https://doi.org/10.1109/TPWRS.2012.221 2252.

[11] Nguyen, V. T. and Shim, J. W. Virtual capacity of hybrid energy storage systems using adaptive state of charge range control for smoothing renewable intermittency. *IEEE Access*, 2020, 8: 126951-126964. https://doi.org/ 10.1109/ACCESS.2020.30085 18.

[12] 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*, 2021, 129(9):106878. https://doi.org/10.1016/j.ijepes.2021.106878.

[13] Zou, J., Peng, C., Shi, J., et al. State-of-charge optimising control approach of battery energy storage system for wind farm. *IET Renew. Power Gener*, 2015, 9 (6): 647–652. https://doi.org/10.1049 /iet-rpg.2014.0202.

[14] Zhou, Y., Yan, Z., and Li, N. A novel state of charge feedback strategy in wind power smoothing based on short-term forecast and scenario analysis. *IEEE Transactions on Sustainable Energy*, 2017, 8(2): 870-879. https://doi.org/10.1109/TSTE.2016.2625305.

[15] Cao M , Xu Q , Qin X , et al. Battery energy storage sizing based on a model predictive control strategy with operational constraints to smooth the wind power. *International Journal of Electrical Power & Energy Systems*, 2020, 115:105471.1-105471.10. https://doi.org/10.1016/j.ijepes.2019.105471.

[16] C. Wan, W. Qian, C. Zhao, et al. Probabilistic forecasting based sizing and control of hybrid energy storage for wind power smoothing. *IEEE Transactions on Sustainable Energy*, 2021; 12(4): 1841-1852. https://doi.org /10.1109/TSTE.2021.3068043.

[17] Sutton, R.S., Barto, A.G. Reinforcement Learning: An Introduction; MIT Press Ltd.: Cambridge, MA, USA, 2018.

[18] Ghoushchi S J , Manjili S , Mardani A , et al. An extended new approach for forecasting short-term wind power using modified Fuzzy Wavelet neural network: A case study in wind power plant. *Energy*, 2021; 223. https://doi.org/ 10.1016/j.energy.2021.120052.

[19] Chen Xiaojiao, Zhang Xiuqing, Dong Mi, et al. Deep learning-based prediction of wind power for multi-turbines in a wind farm. *Frontiers in Energy Research*, 2021; 9. https://doi.org/10.3389/ fenrg.2021.723775

[20] Liu H., Yu C.Q., Wu H.P., et al. A new hybrid ensemble deep reinforcement learning model for wind speed short term forecasting. *Energy*, 2020; 202. https://doi.org/ 10.1016/j.energy.2020.117794.

[21] X. Wang, J. Zhou, B. Qin, et al. Individual pitch control of wind turbines based on SVM load estimation and LIDAR measurement. *IEEE Access*, 2021; 9: 143913-143921. https://doi.org/10.1109/ ACCESS.2021.31205 43.

[22] 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*, 2022; 245. https://doi.org/10.1016/j. energy.2022.123263.

[23] 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*. 2021:1-13. https:// doi.org/10.1049/rpg2. 12199.

[24] Farah Shahid, Aneela Zameer, Muhammad Muneeb. A novel genetic LSTM model for wind power forecast. *Energy*, 2021; 223. https://doi.org/10.1016/j.energy.2021. 120069.

[25] K. Dragomiretskiy, D. Zosso. Variational mode decomposition. *IEEE Trans. Signal Processing*, 2013; 62(3): 531-544. https:// doi.org/10.1109/TSP.2013.2288675.

[26] Denmark Energinet. "Wind turbines connected to grids with voltages above 100 kV," Technical Regulation for the Properties and the Regulation of Wind Turbines. 2004.

[27] Lillicrap TP, Hunt JJ, Pritzel A, et al. Continuouscontrol with deep reinforcement learning. *arXiv preprint*, 2015. https://doi.org/10.48550/arXiv.1509.02971.

[28] Fujimoto S, Van Hoof H, Meger D. Addressing function approximation error in actor-critic methods. *arXiv preprint*, 2015. https://doi.org/10.48550/arXiv.1802.09477.

[29] Ackermann J , V Gabler, Osa T, et al. Reducing overestimation bias in multi-agent domains using double centralized critics. *arXiv preprint*, 2019. https://doi.org/10.48550/arXiv.1910.01465.

[30] Applebaum, D. *Levy Processes and Stochastic Calculus* (Second Edition), Cambridge University Press, 2008.

[31] Janicki A, Weron A. Simulation and chaotic behavior of alpha-stable stochastic processes. *Hsc Books*, 1994; 457(6): 663-664. https://doi.org/10.2307/2983310.

[32] Yu Z G., Anh V., Wang Y., et al. Modeling and simulation of the horizontal component of the geomagnetic field by fractional stochastic differential equations in conjunction with empirical mode decomposition. *Journal of Geophysical Research Space Physics*, 2010, 115(A10). https://doi.org/10.1029/2009JA015206.

[33] Nazaripouya, Hamidreza, Chu, et al. Engineering energy storage sizing method considering the energy conversion loss on facilitating wind power integration. *IET Generation, Transmission Distribution*, 2019, 13(9): 1751-8687. https://doi.org/10.1049/iet-gtd.2018.6358.