Dynamic Economic Dispatch of Thermal-Wind-Storage Systems Based on Reinforcement Learning

1st Yuheng Li
School of Mathematics
Southeast University
Nanjing, China
843634544@qq.com

2nd Chengfang Hu
School of Mathematics
Southeast University
Nanjing 210096, China
hu_chengfang@foxmail.com

3rd Junjie Fu*†
**School of Mathematics*
Southeast University
Nanjing 210096, China
†Purple Mountain Laboratories
Nanjing 211111, China
fujunjie@seu.edu.cn

4th Shuai Wang
Research Insitute for Frontier Science
Beihang University
Beijing 100191, China
wangshuai@buaa.edu.cn

***Abstract*—This paper studies a dynamic economic dispatch (DED) problem which includes thermal and wind-storage hybrid units, aiming at minimizing the total generation cost and penalty costs involving generation regulation, load shedding, and wind curtailment. Each unit is assigned with a fixed, discrete, constrained virtual action set, and its cost function is unknown. Based on the developed model, a reinforcement learning algorithm is applied to solve the DED problem under the wind uncertainty. Simulation results illustrate the effectiveness of the algorithm.**

***Index Terms*—Dynamic economic dispatch (DED), reinforcement learning, wind power, storage**

I. Introduction

Owing to the urgent need to relieve the pressure of environmental pollution and fossil energy shortages, more and more attention has been paid to renewable energy sources, especially wind power on account of its advantages like non-pollution, high feasibility, and inexhaustible feature. However, the increasing wind power generation challenges not only the security and stability of the grid, but also the economic operation and management of power systems due to its uncertainty and intermittency. As one of the most important issues of power systems [1], [2], the economic dispatch problem (ED) [3]–[5] of wind power systems is also threatened by the complicated process of large-scale distribution and the volatility of wind power. Therefore, the ED problem considering the wind power uncertainty requires more in-depth study [6], [7].

To cope with the ED problem under the wind power uncertainty, much research effort has been devoted to developing ED optimization algorithms such as robust optimization [8]–[10] and stochastic optimization [11]–[13]. [8] introduced the concept of dynamic uncertainty sets to model the temporal and spatial correlations of wind power. For a multi-energy cooperation system, a robust operational optimization framework was designed with multi-energy devices and integrated energy networks in [9]. Aimed at eliminating the uncertain impact of wind power, [10] proposed a two-level robust method to address the multi-microgrid ED problem subject to wind power. [11] presented the mean-tracking model to minimize the deviation in generation cost between the trial and the pre-schedule solution. [12] established a multi-stage stochastic optimization model which mitigated the conservatism of previous methods by exploiting the probabilistic distribution of the renewable energy generation. [13] proposed a difference-of-convexity optimization method for the stochastic robust real-time power dispatch problem based on known empirical distributions. In the above research on robust optimization, the methods typically suffered from the conservatism for the sake of immunizing against the worst-case uncertainty realization. And the aforementioned stochastic optimization methods uniformly assumed that the probability distribution of wind uncertainty was known. It's challenging to cope with the DED problem due to fluctuant and non-schedulable qualities of wind sources under the premise of little prior information.

Reinforcement learning (RL) is a method to obtain the optimal policy by interacting with the environment. It requires little or no prior information of the detailed model of the environment and plays an active role in handling the uncertainty of the decision-making problem, such as hybrid electric vehicles [14], zero-sum games [15], nonlinear control systems [16], event-triggered [17] and multi-agent tasks [18]. To address the intrinsic randomness of the DED problem, RL algorithms have aroused wide-spread interest [19]–[22]. [19] applied the reinforcement learning algorithm to achieve

This work was supported by the Ministry of Industry and Information Technology of China through Grant No. JCKY2021602B035.

optimal decision-making considering the effective cooperation of wind power and energy storage. [20] integrated a diffusion strategy with distributed RL algorithms to solve the dynamic ED problem. [21] proposed to obtain the global information utilizing average-consensus algorithms. [22] took advantage of quadratic functions to approximate the state-action value function in the RL method.

This paper intends to solve the challenging dynamic economic dispatch problem under the wind uncertainty and realize the demand-supply power balance when finding the optimal power allocation to minimize the total generation cost. We establish a thermal-wind-storage hybrid DED model, where each wind power generator is equipped with a finite energy storage device to reduce the adverse effect of wind curtailment and load shedding. Based on the proposed model, a reinforcement algorithm involving penalties is proposed which resolves the issue of DED in the presence of uncertain wind power generation.

The rest of this paper is structured as follows. Section II formulates the problem and introduces necessary basic assumptions. The reinforcement learning algorithm is proposed in Section III. Simulation results to demonstrate the effectiveness of the algorithm are given in Section IV and the conclusion is presented in Section V.

II. Problem Statement

We consider a smart grid which consists of N thermal and M wind power generators. Each wind power generator is equipped with an energy storage device to balance the uncertainty of wind power. It's required that the total electricity generation equals to the total power demand at each time slot t. The objective of the dynamic economic dispatch problem is to find the optimal power allocation such that the total generation cost of all generators is minimized. The total cost is described by

$$min \qquad \sum_{t=1}^{T}[\sum_{i=1}^{N} C_i(p_{i,t}) + C_t^P], \tag{1}$$

where $C_i(\cdot)$ is the ith thermal generation cost function and $p_{i,t}$ is the ith thermal power output at time slot t. C_t^P is the wind power system stochastic cost consisting of load shedding power $p_{LS,t}$ and wind curtailment power $p_{AS,t}$, i.e.,

$$C_t^P = C_{LS}p_{LS,t} + C_{AS}p_{AS,t}, \tag{2}$$

where C_{LS} and C_{AS} are fixed penalty coefficients. In the problem of DED, we should consider the following constraints:

a) power balance constraint:

$$\begin{aligned} &\sum_{i=1}^{N} p_{i,t} + \sum_{w=1}^{M} \hat{p}_{w,t}^{wind} + \sum_{w=1}^{M} R_{w,t} = D_t, \\ &t = 1, 2, ..., T, w = 1, 2, ..., M, \end{aligned} \tag{3}$$

where D_t is the total power demand at time t, $\hat{p}_{w,t}^{wind}$ is the wth wind power output prediction at time t, and $R_{w,t}$ is the charge-discharge electric power.

b) generation capacity constraints:

$$\begin{aligned} &\underline{p}_i^{ther} \le p_{i,t} \le \overline{p}_i^{ther}, i = 1, 2, ..., N, \\ &\underline{p}_w^{wind} \le \hat{p}_{w,t}^{wind} \le \overline{p}_w^{wind}, w = 1, 2, ..., M, \end{aligned} \tag{4}$$

where $\underline{p}_i^{ther}$ and $\overline{p}_i^{ther}$ are the minimum and maximum power output of the ith thermal generator, respectively. Similarly, $\underline{p}_w^{wind}$ and $\overline{p}_w^{wind}$ are the minimum and maximum power output of the wth wind power generator.

c) storage device constraints:

$$\begin{aligned} &R_w^{min} \le R_{w,t} \le R_w^{max}, \\ &E_w^{min} \le E_{w,t} \le E_w^{max}, \end{aligned} \tag{5}$$

where R_w^{min} and R_w^{max} are the minimum and maximum power output of the wth storage device. $E_{w,t}$ represents the current electricity stored at the wth device which obeys $E_{w,t} = E_{w,t-1} - R_{w,t}\Delta T$ and ΔT is the fixed length of each time slot. E_w^{min} and E_w^{max} are the minimum and maximum electricity storage of the wth device.

Due to the uncertainty of wind power generation, there're often some differences between real wind power $p_{w,t}^{wind}$ and its prediction $\hat{p}_{w,t}^{wind}$. We assume wind power error $\Delta p_{w,t}^{wind}$ obeys normal distribution [23] and suppose

$$\begin{aligned} \Delta P_t &= \sum_{i=1}^{N} p_{i,t} + \sum_{w=1}^{M} p_{w,t}^{wind} - D_t, \\ E_t^{rest} &= \sum_{w=1}^{M} E_{w,t-1} \\ \Delta E_t^{up} &= \sum_{w=1}^{M} (E_w^{max} - E_{w,t-1}), \\ \Delta E_t^{down} &= \sum_{w=1}^{M} (E_{w,t-1} - E_w^{min}), \end{aligned} \tag{6}$$

where ΔP_t represents the error between the total power generation and the total power demand at time slot t, E_t^{rest} is the total energy storage, ΔE_t^{up} and ΔE_t^{down} are the total maximum charging and discharging capacity of storage devices at time slot t, respectively.

If $\Delta P_t > 0$ which means the actual total generated power is more than the total power demand, wind curtailment power $p_{AS,t}$ is formulated as follows,

$$p_{AS,t} = \begin{cases} 0, & \Delta P_t \cdot \Delta T \le \Delta E_t^{up} \\ \Delta P_t \cdot \Delta T - \Delta E_t^{up}, & \Delta P_t \cdot \Delta T > \Delta E_t^{up}. \end{cases} \tag{7}$$

If $\Delta P_t < 0$, load shedding power $p_{LS,t}$ is

$$p_{LS,t} = \begin{cases} 0, & -\Delta P_t \cdot \Delta T \le \Delta E_t^{down} \\ -\Delta P_t \cdot \Delta T - \Delta E_t^{down}, & -\Delta P_t \cdot \Delta T > \Delta E_t^{down}. \end{cases} \tag{8}$$

For the thermal generation cost, the most common formulation is

$$C_i(p_{i,t}) = a_i p_{i,t}^2 + b_i p_{i,t} + c_i, \tag{9}$$

where a_i, b_i and c_i are coefficients for the ith thermal generator [24].

Besides, three standard assumptions are made to guarantee the existence of an optimal solution to (1).

Assumption 1: The graph topology of the smart grid is undirected and connected.

Assumption 2: $\mathcal{P}_i, \mathcal{R}_w$ are finite discrete sets of admissible electricity output of the ith thermal unit and wth storage device respectively. For any time slot t, at least there exists a feasible combination $\{p_{i,t}, R_{w,t} | i = 1, 2, ..., N, w = 1, 2, ..., M\}$ such that $\sum_{i=1}^{N} p_{i,t} + \sum_{w=1}^{M} \hat{p}_{w,t}^{wind} + \sum_{w=1}^{M} R_{w,t} = D_t$, $p_{i,t} \in \mathcal{P}_i$, $R_{w,t} \in \mathcal{R}_w$.

Assumption 3: There exists an optimal solution $p_{i,t}^*, R_{w,t}^*$ satisfying $\sum_{i=1}^{N} p_{i,t}^* + \sum_{w=1}^{M} \hat{p}_{w,t}^{wind} + \sum_{w=1}^{M} R_{w,t}^* = D_t$ for all time slot t.

III. Q-learning Reinforcement Learning Algorithm

In order to solve the dynamic economic dispatch problem with uncertain wind power generation and unknown cost functions, we apply a Q-learning reinforcement learning algorithm with a central decision-making agent to find the optimal policy of all units including thermal generators, wind power generators and storage devices.

In general, reinforcement learning in discrete time is usually formulated as a tuple $\{\mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}\}$, where $\mathcal{S}$ is the set of states, $\mathcal{A}$ is the set of actions. $\mathcal{T} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ is a state transition function, and $\mathcal{R} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to \mathbb{R}$ is a reward function where $\mathcal{T}$ and $\mathcal{R}$ are usually static. Besides, the policy is defined as $\pi : \mathcal{A} \to \mathcal{S}$, where $\pi(a|s)$ is the probability of action $a \in \mathcal{A}$ in the state $s \in \mathcal{S}$. In the Q-learning algorithm, $\mathcal{T}$ and $\mathcal{R}$ are unknown. The Q-function $Q_\pi(s, a)$ called state-action-value function for policy π is denoted as the expected discount of the long-term reward to the agent at the state s taking action a.

Suppose that the central agent can discover the total power demand D_t, wind power generation total prediction $\hat{p}_t^{wind} = \sum_{w=1}^{M} \hat{p}_{w,t}^{wind}$, the total energy storage E_t^{rest} at time slot t and obtain the total generation cost $C_t = \sum_{i=1}^{N} C_i(p_{i,t})$. Besides, it can calculate all feasible combinations of units for all time slot t.

Define $P_t = (p_{1,t}, ..., p_{N,t}, R_{1,t}, ..., R_{M,t})^T \in A_t$, where A_t is the set of action at time slot t, i.e., $A_t = \{P_t | P_t \in \prod_{i=1}^{N} \mathcal{P}_i \times \prod_{w=1}^{M} \mathcal{R}_w\}$.

A Q-learning reinforcement learning algorithm is proposed which repeatedly executing the following four steps.

1) *Select the action by ϵ-greedy policy*: Based on the current Q-function and state information D_t, $\hat{p}_t^{wind}$ and E_t^{rest}, we choose the following action by using the ϵ-greedy policy to protect the agent from being trapped in the past experience, i.e., selecting the current optimal action with probability $1-\epsilon$, and other actions with probability ϵ:

$$\pi(a|s_t) = \begin{cases} 1-\epsilon, & a = \arg\min_{P_t} Q(s_t, P_t) \\ \epsilon, & \text{otherwise,} \end{cases} \tag{10}$$

where $\pi(a|s_t)$ is the probability of agents taking action a at time slot t.

Remark 1: When choosing other actions instead of the current optimal action, here we stochasticly select $p_{i,t} \in \mathcal{P}_i, i = 1, 2, ...N$, and calculate the total error

$$\zeta_t = D_t - \sum_{i=1}^{N} p_{i,t} - \hat{p}_t^{wind} \tag{11}$$

at time slot t. Then choose $R_{w,t} \in \mathcal{R}_w$, s.t.

$$\sum_{w=1}^{M} R_{w,t} = \zeta_t. \tag{12}$$

If there doesn't exist a suitable $R_{w,t}, w = 1, 2, ..., M$ meeting the condition, choose $p_{i,t} \in \mathcal{P}_i$ repeatedly until satisfying (11) and (12).

2) *Measure the total generation cost at time slot t*: According to the real wind power generation $\{p_{1,t}^{wind}, ..., p_{M,t}^{wind}\}$ and the current storage electricity $\{E_{1,t}, ..., E_{M,t}\}$ that we have obtained, the central agent can get the total generation cost C_t and the total penalty C_t^P via (2), (6), (7), (8) and (9).

3) *Update the state-action-value function*: At each trial k, the Q-function $Q(s_t, P_t)$ at time slot t can be updated as follows:

$$Q(s_t, P_t) \leftarrow C_t + C_t^P + \gamma \min_{P_{t+1}} Q(s_{t+1}, P_{t+1}), \tag{13}$$

where $S_t = \{s_t | s_t \in \mathcal{D} \times \hat{\mathcal{P}}^{wind} \times \mathcal{E}^{rest}\}$ is the set of states at time slot t. Suppose $\mathcal{D}, \hat{\mathcal{P}}^{wind}$ and $\mathcal{E}^{rest}$ are admissible discrete sets of the total power demand, total wind power generation prediction, and total energy storage respectively, i.e., $D_t \in \mathcal{D}, \hat{p}_t^{wind} \in \hat{\mathcal{P}}^{wind}$ and $E_t^{rest} \in \mathcal{E}^{rest}$ for all time slot t.

Remark 2: The updating iteration utilizes the historical information $\min_{P_{t+1}} Q(s_{t+1}, P_{t+1})$ and the new experience information C_t, C_t^P ensuring the convergence at each trial. After updating more and more trials, the information learned by the central decision-making agent will become closer and closer to the true Q-value.

4) *Renew operation policy*: The optimal operation policy $\pi(s_t)$ can be renewed like

$$\pi(s_t) \leftarrow \arg\min_{P_t'} Q(s_t, P_t') \tag{14}$$

after (13).

Remark 3: If there are multiple combinations P_t' minimizing $Q(s_t, P_t')$, we can choose one arbitrarily.

The Q-learning reinforcement learning algorithm for dynamic economic dispatch problem with wind power generation is summarized in Algorithm 1.

Algorithm 1 Q-learning Reinforcement Learning Algorithm

Initialization: $t = 0$, $k = 0$, $\epsilon \in (0, 1)$, initial storing power E_0^w, $Q(s, a) = +\infty$, $\forall a \in \prod_{i=1}^{N} \mathcal{P}_i \times \prod_{w=1}^{M} \mathcal{R}_w$, $\forall s \in \mathcal{D} \times \hat{\mathcal{P}}^{wind} \times \mathcal{E}^{rest}$
repeat
  $k \leftarrow k + 1$
  repeat
    $t \leftarrow t + 1$
    Obtain the current state $s_t = (D_t, \hat{p}_t^{wind}, E_t^{rest})$ at time slot t
    $r = rand(1)$
    if $r < \epsilon$ **then**
      Find a feasible combination $P_t = (p_{1,t}, p_{2,t}, ..., p_{N,t}, R_{1,t}, ..., R_{M,t})^T$ as the following action a_t via (11) and (12)
    else
      Choose the current optimal action as the action a_t
    end if
    Operate the chosen action a_t and obtain the total generation cost C_t and penalties C_t^P via (2), (6), (7), (8) and (9).
    Update $Q(s_t, a_t)$ via (13)
    Renew operation policy $\pi(s_t)$ based on (14)
  until $t = T$
  Reset $t = 0$
until $k = K$, K is the maximum number of trials

TABLE I
PARAMETERS OF THERMAL GENERATORS

Number	$\underline{p}_i^{ther}$	$\overline{p}_i^{ther}$	a_i	b_i	c_i
1	25	100	0.0216	6.6	25
2	25	50	0.0084	7.63	20

IV. SIMULATION

Considering the dynamic economic dispatch problem over multiple time slots, we set three generators including two thermal and one wind power connected via the undirected graph shown in Fig. 1, where $T1, T2$ represent two thermal generators, $W1$ is the wind power generator and $S1$ is the storage device.

The cost function for two thermal generators is taken as $C_i(p_{i,t}) = a_i p_{i,t}^2 + b_i p_{i,t} + c_i$, where a_i, b_i and c_i are generation cost coefficients of the ith thermal unit, $i = 1, 2$, which are shown in Table I. Parameters of the wind power generator and its energy storage device are shown in Table II including the minimum and maximum power output of the wind power generator $\underline{p}_i^{wind}, \overline{p}_i^{wind}$, the maximum and minimum storage electricity E_w^{min}, E_w^{max} and the charge-discharge range of the storage device R_w^{min}, R_w^{max}, $w = 1$. The admissible power outputs of each thermal generator are set as follows: $\mathcal{P}_1 = [0, 25, 26, 27, ..., 100], \mathcal{P}_2 = [0, 25, 26, 27, ..., 50](MW)$. The total power demand D_t of all loads are $900, 950, 1000, 1050, 1100(MW)$ for different time periods $[0, 1), [1, 2), [2, 3), [3, 4), [4, 5)(h)$, respectively. The initial power of the storage device is set as $200(MW \cdot h)$. The set of possible wind power generation prediction $\hat{P}_t^{wind}$ is $[900, 910, ..., 1000](MW)$ and the error between real wind power generation and its prediction obeys normal distribution with the zero mean and $0.01\hat{p}_{w,t}^{wind}$ standard deviation at time slot t.

TABLE II
PARAMETERS OF THE WIND POWER GENERATOR AND ITS STORAGE

Number	$\underline{p}_i^{wind}$	$\overline{p}_i^{wind}$	E_w^{min}	E_w^{max}	R_w^{min}	R_w^{max}
1	900	1000	100	600	-200	200

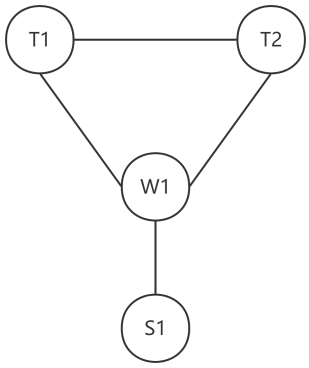
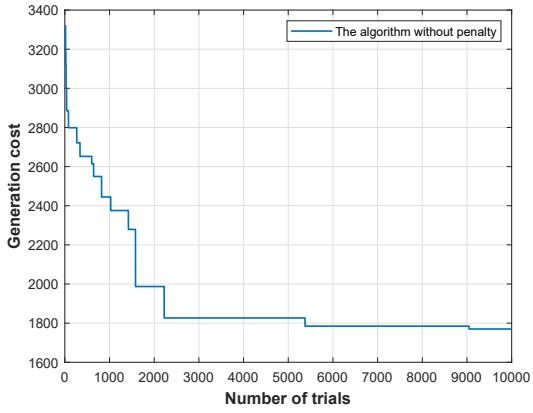


Fig. 1. Communication topology.

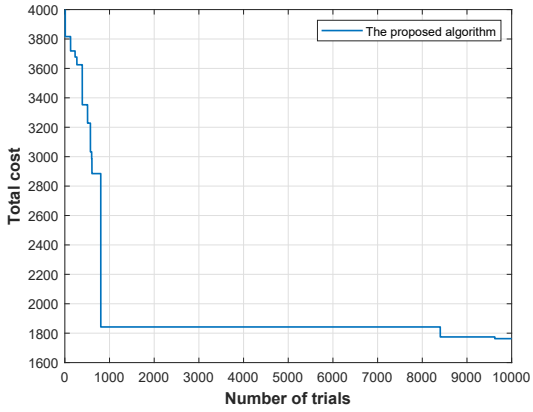
In order to avoid the influence of varying exploration rate ϵ, the exploration rate of ϵ-greedy policy is usually set as a constant. Here we set $\epsilon = 0.2$ and learning rate $\gamma = 0.9$. About penalty parameters C_{LS} and C_{AS}, in Fig. 2 (b) we assume $C_{LS} = 200 \geq C_{AS} = 100$ for the purpose of avoiding load shedding as far as possible.

As shown in Fig. 2 (b), the total cost of the updated policy is getting better and better during the training process through the Q-learning algorithm with penalties. In addition, we also consider the objective with penalty parameters $C_{LS} = C_{AS} = +\infty$. In this case, wind curtailment and load shedding conditions are regarded as non-feasible solutions and abandoned during the updating process. The action set is smaller than the Q-learning algorithm with penalties. Fig. 2 (a) demonstrates the changing curve of the minimum total generation cost in this case. It can be seen that the algorithm can still find high-performance strategies with the smaller action set.

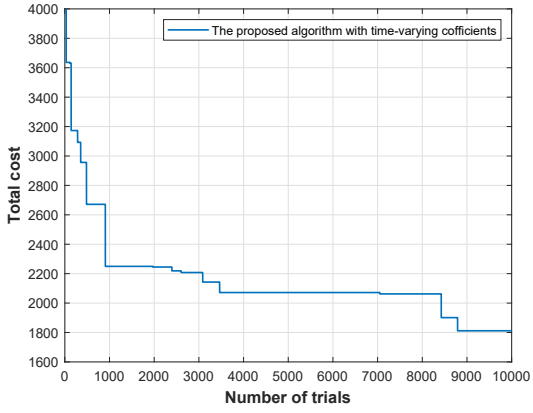
In Fig. 3 (a), the comparison of the maximum power imbalance with the learned strategy and the traditional strategy during each trial is presented. For the traditional DED, the maximum power imbalance during the kth trial can be regarded as the maximum prediction error of the wind power, i.e., $\max_t \sum_{w=1}^{M} \Delta p_{w,t}^{wind}$ since the demand-supply equation is



(a)

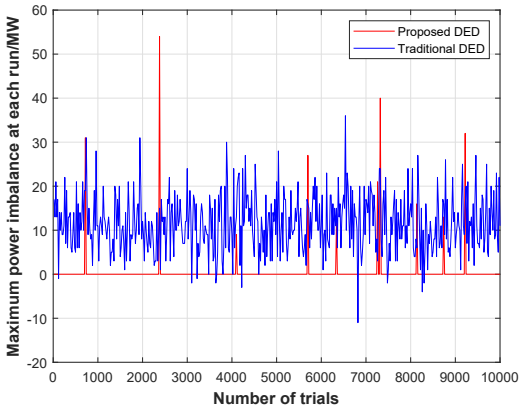


(b)

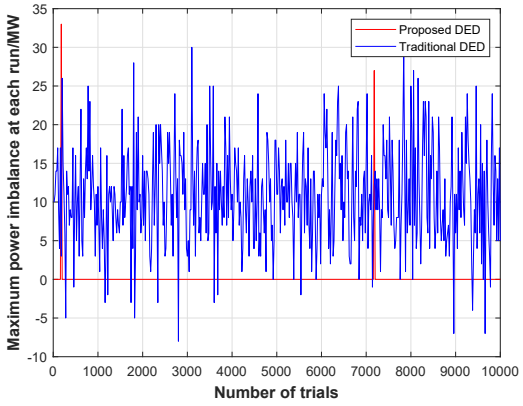


(c)

Fig. 2. Contrasts between algorithms with and without penalty



(a)



(b)

Fig. 3. Power fluctuation curves.

considered with the predicted wind power. With the proposed DED, the power imbalance can be seen as the load shedding or wind curtailment power due to the existence of storage devices, i.e., $\max_t \max\{p_{LS,t}, p_{AS,t}\}$. It can be seen from the figure that the volatility of the hybrid system has been significantly reduced most of the time. However, due to finite options of the thermal unit generation within a certain range and the finite charge-discharge power of storage devices, power imbalance still occasionally occurs.

To reduce the sudden changes of power imbalance, we change the constant penalty parameters into exponential functions which change with the wind curtailment power $p_{AS,t}$ and load shedding power $p_{LS,t}$. We set $C_{AS} = e^{p_{AS,t}}$, $C_{LS} = 2e^{p_{LS,t}}$. In Fig. 2 (c), it's obvious that in this case the total cost is declining during the training process and comparable performance of the learning strategy is achieved. Moreover, as shown in Fig. 3 (b), the effect of reducing the volatility of power imbalance is better and the sudden variation is much less than the constant penalty parameters.

V. Conclusion

Considering the thermal-wind-storage hybrid power system with physical constraints, it brings new challenges to identify feasible actions and achieve cooperation between thermal units, wind power units, and storage devices under wind uncertainty when using reinforcement learning methods. We have proposed a thermal-wind-storage DED model with penalties and a Q-learning based effective reinforcement learning algorithm is proposed to achieve cooperative DED of the hybrid power system. Simulation results show that the algorithm both

enables lower total costs and reduces the power fluctuation caused by the volatility of wind power.

REFERENCES

[1] A. J. Conejo and L. Baringo, Power system operations. Cham, Switzerland: Springer, 2018.

[2] R. Sylwester, "Sources of uncertainty in power system analysis," Przegl. Elektrotech., vol. 84, no. 1, pp. 54-57, January 2008.

[3] X. Xia and A. M. Elaiw, "Optimal dynamic economic dispatch of generation: A review," Electr. Power Syst. Res., vol. 80, no. 8, pp. 975-986, August 2010.

[4] G. Wen, X. Yu, and Z. Liu, "Recent progress on the study of distributed economic dispatch in smart grid: an overview," Front. Inform. Tech. EL., vol. 22, no. 1, pp. 25-39, January 2021.

[5] H. Pourbabak, J. Luo, T. Chen and W. Su, "A novel consensus-based distributed algorithm for economic dispatch based on local estimation of power mismatch," IEEE Trans. Smart Grid., vol. 9, no. 6, pp. 5930-5942, November 2018.

[6] J. Hetzer, D. C. Yu, and K. Bhattarai, "An economic dispatch model incorporating wind power," IEEE Trans. Energy Convers., vol. 23, no. 2, pp. 603-611, June 2008.

[7] W. Zhou, H. Sun, H. Gu, and X. Chen, "A review on economic dispatch of power system including wind farms," Power Syst. Prot. Control, vol. 39, no. 24, pp. 148-154, March 2012.

[8] L. Alvaro and X. A. Sun, "Adaptive robust optimization with dynamic uncertainty sets for multi-period economic dispatch under significant wind," IEEE Trans. Power Syst., vol. 30, no. 4, pp. 1702-1713, July 2015.

[9] E. A. Martínez Ceseña and P. Mancarella, "Energy systems integration in smart districts: Robust optimisation of multi-energy flows in integrated electricity, heat and gas networks," IEEE Trans. Smart Grid., vol. 10, no. 1, pp. 1122-1131, January 2019.

[10] H. Qiu, B. Zhao, W. Gu and R. Bo, "Bi-level two-stage robust optimal scheduling for AC/DC hybrid multi-microgrids," IEEE Trans. Smart Grid., vol. 9, no. 5, pp. 5455-5466, September 2018.

[11] Z. Lin et al., "Mean-tracking model based stochastic economic dispatch for power systems with high penetration of wind power," Energy, vol. 193, pp. 1151-1163, February 2020.

[12] R. Lu et al., "Multi-stage stochastic programming to joint economic dispatch for energy and reserve with uncertain renewable energy," IEEE Trans. Sustain. Energy, vol. 11, no. 3, pp. 1140-1151, July 2020.

[13] K. Qu, X. Zheng, X. Li, C. Lv and T. Yu, "Stochastic robust real-time power dispatch with wind uncertainty using difference-of-convexity optimization," IEEE Trans. Power Syst., in press.

[14] T. Liu, Y. Zou, D. Liu, and F. Sun, "Reinforcement learning of adaptive energy management with transition probability for a hybrid electric tracked vehicle," IEEE Trans. Ind. Electron., vol. 62, no. 12, pp. 7837C7846, December 2015.

[15] S. Mukhopadhyay, O. Tilak and S. Chakrabarti, "Reinforcement learning algorithms for uncertain, dynamic, zero-sum games," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), vol. 5, no.3, pp. 48-54, 2018.

[16] R. C. B. Rego and F. M. U. d. Arajo, "Nonlinear control system with reinforcement learning and neural networks based Lyapunov functions," IEEE Latin Am. Trans., vol. 19, no. 8, pp. 1253-1260, August 2021.

[17] Y. Wan, C. Long, R. Deng, G. Wen, X. Yu, "Adaptive event-triggered strategy for economic dispatch in uncertain communication networks," IEEE Trans. Control Netw. Syst., vol. 8, no. 4, pp. 1881-1891, December 2021.

[18] G. Wen, J. Fu, P. Dai, and J. Zhou, "DTDE: A new cooperative multi-agent reinforcement learning framework," The Innov., vol. 2, no. 4, pp. 100162, November 2021.

[19] G. Liu , X. Han, S. Wang, M. Yang, and M. Wang, "Optimal decision-making in the cooperation of wind power and energy storage based on reinforcement learning algorithm," Power Syst. Technol., vol. 40, no. 9, pp. 2729-2736, September 2016.

[20] W. Liu, P. Zhuang, H. Liang, J. Peng, and Z. Huang, "Distributed economic dispatch in microgrids based on cooperative reinforcement learning," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 6, pp. 2192-2203, June 2018.

[21] F. Li, J. Qin, and W. X. Zheng, "Distributed q-learning-based online optimization algorithm for unit commitment and dispatch in smart grid," IEEE Trans. Cybern., vol. 50, no. 9, pp. 4146-4156, September 2020.

[22] P. Dai, W. Yu, G. Wen, and S. Baldi, "Distributed reinforcement learning algorithm for dynamic economic dispatch with unknown generation cost functions," IEEE Trans. Ind. Inform., vol. 16, no. 4, pp. 2258-2267, April 2020.

[23] Y. V. Makarov, P. V. Etingov, J. Ma, Z. Huang and K. Subbarao, "Incorporating uncertainty of wind power generation forecast into power system operation, dispatch, and unit commitment procedures," IEEE Trans. Sustain. Energy, vol. 2, no. 4, pp. 433-442, October 2011.

[24] S. Yang, S. Tan, and J. Xu, "Consensus based approach for economic dispatch problem in a smart grid," IEEE Trans. Power Syst., vol. 28, no. 4, pp. 4416-4426, November 2013.