Data-driven Optimal Dynamic Dispatch for Hydro-PV-PHS Integrated Power Systems Using Deep Reinforcement Learning Approach

Jingxian Yang, Jichun Liu, *Senior Member*, *IEEE*, Yue Xiang, *Senior Member*, *IEEE*, Shuai Zhang, and Junyong Liu, *Member*, *IEEE*

***Abstract*—To utilize electricity in a clean and integrated manner, a zero-carbon hydro-photovoltaic (PV)-pumped hydro storage (PHS) integrated power system is studied, considering the uncertainties of PV and load demand. It is a challenge for operators to develop a dynamic dispatch mechanism for such a system, and traditional dispatch methods are difficult to adapt to random changes in the actual environment. Therefore, this study proposes a real-time dynamic dispatch strategy considering economic operation and complementary regulatory ability. First, the dynamic dispatch of a hydro-PV-PHS integrated power system is presented as a multi-objective optimization problem and the weight factor between different goals is effectively calculated using information entropy. Afterwards, the dispatch model is converted into the Markov decision process, where the dynamic dispatch decision is formulated as a reinforcement learning framework. Then, a deep deterministic policy gradient (DDPG) is deployed towards the online decision for dispatch in continuous action spaces. Finally, a case study is applied to evaluate the performance of the proposed method based on a real hydro-PV-PHS integrated power system in China. Simulations show that the system agent reduces the power volatility of supply by 26.7% after hydropower regulating and further relieves power fluctuation at the point of common coupling (PCC) to the upper-level grid by 3.28% after PHS participation. The comparison results verify the effectiveness of the proposed method.**

***Index Terms*—DDPG, dynamic economic dispatch, hydro-PV-PHS integrated power system, information entropy, uncertainties.**

I. Introduction

WITH many serious ecological and environmental problems, such as acid rain and ozone layer destruction resulting from traditional fossil fuel energy, many countries have set a target of transferring to zero carbon to achieve the goals of the Paris Climate Agreement [1]. For instance, China is projected to achieve net-zero emission by 2060 [2]. It is reported that 41% of carbon dioxide emission is from fossil-fuel power plants, so, it is an important measure to utilize clean energy sources instead of fossil energy to realize zero carbon energy production and consumption [3]. With the world-wide concern about zero-carbon, more and more renewable energy resources (RERs) are being incorporated into the power system. The installed capacity of RERs in the world has reached 2.799 TW in 2020 according to statistics [4]. However, its strongly uncertain and random features causes the fluctuations on both supply and demand sides which introduces new challenges in regards to maintaining system reliability [5], [6]. Driven by this problem, the multi-energy complementary development mode, which utilizes other energy sources to compensate for the variability of renewable energy to balance load and generation in real-time, has become a new development direction in the utilization of RERs [7], [8].

Of all renewable sources, solar energy generation has gained momentum among RERs owing to its safety, low cost and maintainability. Today, China is a leading consumer of solar energy, followed by the United States, Japan, and Germany [9]. As the "Renewable Energy Statistic (2020)" stated, by the end of 2020, the total global net solar photovoltaic installed capacity reached 760.4 GW, a 21.4% increase over 2019. However, the PV output which depends on uncertain weather is unstable and represents a rising challenge to schedule the entire energy system [10]. To acquire a reliable output, energy storage devices have been combined to compensate for the fluctuation of PV power. In recent years, PHS which often stores energy by pumping water from a lower reservoir to an upper one during off-peak periods and releases the stored water to generate electricity by turbines in peak periods has attracted attention for applications in hybrid systems [11], [12]. PHS is usually utilized to handle the intermittent characteristics of PV output owing to its long lifetime, start/stop flexibility, low operational and maintenance costs, and no pollution. Therefore, PHS is considered the most promising system for handling such integrated power systems, and many PHS plants were installed by the end of 2020 with capacities of 172.5 GW which accounts for 90.3% of the global energy storage market [13]. However, the suppressing effect of PHS is limited because of its limited installed capacity [14]. Therefore, hydropower which can be adjusted instantly according to the demand and the fluctuation of PV power output has gradually become one of the more preferred forms of RESs used for electricity. The largest hydropower energy producer and

Manuscript received September 25, 2021; revised November 24, 2021, accepted December 30, 2021. Date of online publication August 18, 2022; date of current version August 24, 2022. This work was supported by the National Key R&D Program of China under Grant 2018YFB0905200.

J. X. Yang, J. C. Liu, Y. Xiang (corresponding author, email: xiang@scu.edu.cn), and J. Y. Liu are with the College of Electrical Engineering, Sichuan University, Chengdu 610065, China.

S. Zhang is with the State Grid Sichuan Electric Power Company, Chengdu 610065, China.

DOI: 10.17775/CSEEJPES.2021.07210

2096-0042 © 2021 CSEE. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

consumer country is China, which accounted for 28.6% of the usage in the world, in 2020, and hydropower energy and solar energy provided 16.8% and 11.5% of total electricity output respectively [15]. Actually, with the successful operation of the world's largest complementary hydro-PV complementary power station-Longyangxia hydro-PV integrated power system in China, the hydro/PHS-PV complementary power system has attracted attention in areas with abundant hydroelectric and solar resources, such as western China [16]–[18].

In general, generation scheduling of such an integrated power system is a very complicated optimization task which is difficult to determine the most economical revenue in meeting the load demands in a schedule horizon of one month or one day [19]. Over all the scheduling periods, the power balance constraint, generation capacity limits, and operation limits of all devices in an integrated power system are satisfied. As the increase in the number of integrated power systems with high-share renewables, more and more research has primarily focused on short-term schedules which often minimize the deviation between the generation scheduling plan and actual integrated power in the optimized model by using some methods, such as model predictive control, for considering the fluctuation feature of RESs to improve the dependability of the whole system [20]. Then the security-economic dispatch is taken fully into account to ensure the power flow of this integrated system is employed, which coordinates different behaviors of the different controllable components at different times [21]. Therefore, the dispatch of such a system is considered as a multi-objective optimization problem which considers economic profit, volatility, and other aspects while achieving safe and stable operations and meeting various requirements.

A large number of short-term operation models of multi-energy complementary power systems have been proposed. Wei *et al*. [22] analyzed the spatial-temporal correlation between wind farms and PV plants through generative adversarial networks, and proposed a two-stage approach to minimize the operation costs based on uncertain scenarios in a hybrid hydro-wind-solar system. Ming *et al*. [23] adopted a robust optimization method and investigated a two-layer nested framework to reduce the average water consumption of hydro power stations considering the prediction error of PV power, and performed a simulation in Longyangxia hydro-PV plant using the proposed method. Cerejo *et al*. [24] analyzed the effects of uncertainty from wind forecasting on the operations of hydro producers and formulated an optimization problem of a hydro-wind plant to maximize the revenues of the co-producers. However, the above-discussed studies only considered the economic factors and the uncertainty of RERs but did not explore the complementary features and synergies of energy sources with respect to smoothing the fluctuation of the hybrid power system. Wang *et al*. [25] investigated a model for minimizing the thermal output fluctuation and increasing power generation to realize a complementary coordinated operations in hybrid power systems. However, it only considered the fluctuation of power sources which is unsuitable for applying in the hybrid system due to the lack of the fluctuation of load. In [26], load fluctuation is taking into account for matching the balance between supply and demand, but the index by fluctuation change rate may not adequately measure the fluctuation of hydro-PV power output in the multi-energy system due to not considering the intermittent PV output. Meanwhile, related studies primarily focused on the intermittence or instability of PV power generation caused by unpredictable weather or different seasons, but paid limited attention to the uncertainty of water inflow in abound, normal and dry periods which is critical for the short-term operation of the hydropower plant. The complementary operation of the hydro-PV-PHS integrated power system is also a typical multi-objective optimization problem, and how to trade-off different targets is still a difficult problem which is worthy of studying.

Many methods have been put forward to determine the optimal dispatch strategy of a hydro-PV-PHS integrated power system through forecasting values of PV, water inflow, load, and price. Ref. [27] proposed an optimization model to maximize the peak-load regulation ability and minimize the deviations between supply and demand based on a Non-dominated Sorting Genetic Algorithm (NSGA-II). Makhdoomi *et al*. [28] solved different optimization targets by a modified crow search algorithm in a hybrid power system. Ref. [29] transformed the original problem into a mixed-integer linear programming model by using several linearizations and approximation methods for the short-term joint operation for an integrated power system. However, these methods are deterministic optimization rules which have difficulties to achieve optimal results due to the high uncertainty of RERs. In [30] and [31], a model predictive control framework is presented for calculating the optimal dispatch to improve the robustness of control strategy in a hybrid power system. A chance-constrained programming-based stochastic optimization model is proposed to obtain the optimal operation strategy of a hydro-PV integrated power system [32]. Ref. [33] proposed a stability-constrained two-stage robust optimization method to obtain a robust operation plan under the worst case in an integrated hydrogen hybrid energy system. However, stochastic optimization is difficult to apply in actual engineering applications, because the optimal solution has a certain level of risk owing to taking into account the prediction error through a certain confidence interval, and the robust optimization solution tends to be conservative. The optimization plan needs to be recalculated once the data of the integrated power system changes with these approaches. So, this may result in being very time-consuming. Moreover, the detailed appliances and models are essential in the operation of integrated power systems, and the optimized scheduling plan is highly dependent on the accuracy of the model. However, the accuracy of the fixed model cannot be guaranteed due to the variable environment. Therefore, various uncertainty variables should be considered, although they are not easy to be accurately modeled.

A data-driven approach is an attractive paradigm for addressing stochastic optimal control problems to overcome all the mentioned disadvantages. Reinforcement learning (RL), as a significant branch of machine learning, is a better choice for solving the above problems [34], [35]. With the development of the deep neural network (DNN), deep reinforcement learning (DRL) which combined RL with DNN has attracted more

attention because it can effectively deal with high-dimensional and continuous action domains [36]–[38]. In recent years, DRL methods have made huge progress in some applications. In [39]–[41], the authors introduced a DRL framework for energy management without building the dynamical model and requiring the prior knowledge of uncertain RERs. The simulation results validated the effectiveness of the data-driven approach. In this paper, the dynamic dispatch of the integrated power system is a random sequential decision problem which can be converted to a Markov decision process (MDP). Thus, the DRL method can be implemented to solve this problem. To the best of our knowledge, few studies have applied the DRL method to solve the dynamic intelligent dispatch of hydro-PV-PHS integrated power systems.

Inspired by the above studies, this paper introduces a DRL-based framework for the dynamic dispatch of a hydro-PV-PHS integrated power system. The operation principle of this system is built as an optimized target considering the uncertainty of PV power, load, price, and water inflow. The main contributions of this paper are summarized as follows:

1) A dynamic reward function is put forward to represent the optimal dispatch of the hybrid system. Additionally, the information entropy is utilized to quantify the weights of different objectives to establish a trade-off among economy, fluctuation of power sources and systems.

2) The dynamic dispatch problem is converted to a MDP, and the system agent is developed by a deep deterministic policy gradient (DDPG) algorithm to achieve the optimal dispatch strategy of the hybrid system.

3) The proposed method is verified on a real-world data set in simulation experiments which can directly ignore the system modeling error. The simulation results demonstrate the validity and advancement of the proposed method.

The remainder of this paper is organized as follows: the system model is presented in Section II. A DRL framework of dynamic dispatch of a hydro-PV-PHS integrated power system is described in Section III. Section IV introduces the DDPG algorithm in detail. A case study is performed in Section V and conclusions are presented in Section VI.

II. Formulation

A diagram of a hydro-PV-PHS integrated power system is presented in Fig. 1. As shown, it consists of small hydropower plants, pumped hydro storage, photovoltaic generating units, and local load. All utilities are monitored by a control center which balances supply and demand. With the promise of full absorption of PV, the hydropower station adjusts its power generation to reduce the fluctuation of PV power output to obtain a smooth output curve for the source. Additionally, pumped hydro storage is introduced to further stabilize the fluctuation on the PCC and cooperates with other utilities to supply stable power for the upper-level grid. The surplus power will be sold to the main grid to obtain revenues based on the spot prices when the local load demand has been met, otherwise, the hydro-PV-PHS integrated power system will purchase electricity from the upper-level power grid to satisfy the local demand.

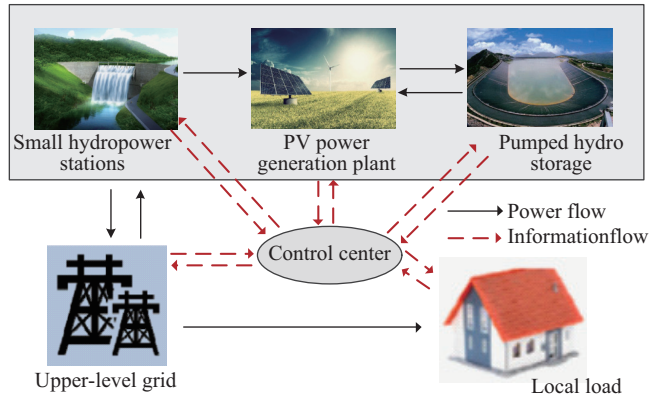

Fig. 1. The architecture of a hydro-PV-PHS integrated power system.

The control center decides the exchanging volume with the main grid via monitoring real-time PV generation, load, and spot price, as well as the state of PHS, then, sends the corresponding signal to different entities.

A. Mathematical Description of Hydropower Model

At each hour, the output of the ith hydropower plant can be expressed as follows:

$$p_{\mathrm{hydro},i,t} = 9.81\eta_i H_i Q_{i,t}/1000 \tag{1}$$

where t indicates the time and η_i stands for the overall efficiency. H_i represents the water head which is equal to the fore-bay water lever minus the tail-race level and $Q_{i,t}$ is the power released from the reservoir. $P_{\mathrm{hydro},i,t}$ is the power output of the ith hydropower plant at the tth time which can be calculated from (1).

$$V_{i,t} = V_{i,t} + 3600(I_{i,t} - Q_{i,t}) \tag{2}$$

$$Q_{i,\min} \le Q_{i,t} \le Q_{i,\max} \tag{3}$$

$$V_{i,\min} \le V_i \le V_{i,\max} \tag{4}$$

Water should satisfy balance constraints as determined in (2), where $I_{i,t}$ and $Q_{i,t}$ represent the in flow and the outflow of the reservoir in time period t respectively. $V_{i,t+1}$ and $V_{i,t}$ are reservoir storages at time t and $t+1$. The limitation of reservoir storage and reservoir release is shown in (3) and (4), where $Q_{i,\min}$ and $Q_{i,\max}$ stand for the minimum and maximum release volume respectively. $V_{i,\min}$ and $V_{i,\max}$ denote the lower and upper storage limits respectively.

B. Mathematical Description of PHS

PHS has two modes: one is pump mode which pumps water to the upper reservoir by the additional power when the price is low. The other one is turbine mode which generates power to satisfy the load demand or sell electricity to the main grid when the price is high. We can obtain different operating conditions through the sign of $P_{\mathrm{PHS},t}$ that represents the power output of the PHS plant as follows:

$$P_{\mathrm{PHS},t} = \begin{cases} P_{\mathrm{turbine},t}, & \text{if } P_{\mathrm{PHS},t} > 0 \\ P_{\mathrm{pump},t}, & \text{if } P_{\mathrm{PHS},t} < 0 \end{cases} \tag{5}$$

where a positive value at the tth time means the discharged power is in turbine mode $P_{\mathrm{turbine},t}$, otherwise, a negative value

means the charged power is shown as $P_{\mathrm{pump},t}$ at pump mode. The above two modes are defined by (6)–(7),

$$P_{\mathrm{turbine},t} = \eta_{\mathrm{turbine}} g H q_{\mathrm{turbine},t}/1000 \tag{6}$$

$$P_{\mathrm{pump},t} = g H q_{\mathrm{pump},t}/1000\eta_{\mathrm{pump}} \tag{7}$$

where η_{turbine} and η_{pump} are the power generation efficiency in the two modes. H is the generation head and g is the acceleration of gravity. $q_{\mathrm{turbine},t}$ and $q_{\mathrm{pump},t}$ denote discharged volume and charged volume of PHS respectively.

The upriver reservoir capacity constraint of PHS is shown by:

$$V_t = V_{t-1} + (\psi_t + q_{\mathrm{pump},t} - q_{\mathrm{turbine},t}) \tag{8}$$

where ψ_t denotes inter-regional flow at time t and V_t represents the upriver reservoir capacity of PHS.

C. Objective Function

The joint operation of the hydro-PV-PHS integrated power system is used to address the intermittent and unstableness of PV output regarding the economy of the hybrid system. The economic benefits of the hybrid power system should be maximized while the fluctuation should be minimized. Three objective functions are described as follows:

Objective 1: maximizing the revenue of the integrated power system

$$ER_t = \lambda_t \left(P_{\mathrm{PV},t} + \sum_{i=1}^{N} P_{\mathrm{hydro},i,t} + P_{\mathrm{PHS},t} - P_{\mathrm{load},t} \right) \tag{9}$$

where ER_t stands for the revenue at time t and $P_{\mathrm{PV},t}$ represents the power output of PV. $P_{\mathrm{load},t}$ stands for the local demand and λ_t denotes the spot price at time t.

Objective 2: minimizing the fluctuation of the power source

$$\Delta P_{\mathrm{source},t} = P_{\mathrm{PV},t} + \sum_{i=1}^{N} P_{\mathrm{hydro},i,t} - \overline{P_{r,t}} \tag{10}$$

The uncertainty of PV output includes the random fluctuation, intermittent fluctuation, and fixed fluctuation, so the compensation to PV also covers three aspects: first, the jagged random fluctuation should be eliminated for smoothing the power curve to close the envelope of a sunny day. Secondly, hydropower should be adjusted to compensate for the inherent fluctuation to obtain a stable power output based on the first compensation. Thirdly, hydropower should be regulated to compensate for the intermittent fluctuation at night to allow for a stable output in one day. Considering the above various fluctuations, we set $\overline{P_{r,t}}$ to be a three-stage line as a reference which can be regarded as the target of the source output. In the simulation, the segmental average values of hydro-PV power output are set according to experience.

Objective 3: minimizing the fluctuation of on-grid power

$$\Delta P_t = P_{\mathrm{grid},t} - P'_{\mathrm{grid},t} \tag{11}$$

$$P'_{\mathrm{grid},t} = P_{\mathrm{PV},t} + \sum_{i=1}^{N} P_{\mathrm{hydro},i,t} - P_{\mathrm{load},t} \tag{12}$$

$$P_{\mathrm{grid},t} = P_{\mathrm{PV},t} + \sum_{i=1}^{N} P_{\mathrm{hydro},i,t} + P_{\mathrm{PHS},t} - P_{\mathrm{load},t} \tag{13}$$

where ΔP_t represents the fluctuation value at the interval of Δt and $P_{\mathrm{grid},t}$ denotes the on-grid power exchanged between the hydro-PV-PHS integrated power system and the main grid at the tth time step where $P'_{\mathrm{grid},t}$ is the previous exchanged power on PCC before PHS participating. If we ignore the internal loss of the system network, (12) can be regarded as the power balance constraints. The upper and lower limit of $P_{\mathrm{grid},t}$ is described as (13). The limit of exchanged power is shown by:

$$P_{\mathrm{grid,min}} \le P_{\mathrm{grid},t} \le P_{\mathrm{grid,max}} \tag{14}$$

where $P_{\mathrm{grid,min}}$ and $P_{\mathrm{grid,max}}$ denote the range of exchanged power on the PCC.

D. Model of System

For the hybrid system, the regulation of hydropower generation is primrily used to mitigate the fluctuation of the source, and PHS is added to relieve the power fluctuation on PCC to the upper-level grid. Meanwhile, the economic benefit of the hybrid system should be maximized considering spot price conditions.

The capacity of the existing reservoir of hydropower stations and PHS are similar to the state of charge of the battery, so we can obtain the following equations:

$$SOC_{\mathrm{hydro},i,t} = V_{i,t}/V_{i,\max} \tag{15}$$

$$SOC_{\mathrm{PHS},t} = V_t/V_{\max} \tag{16}$$

$$SOC_{\mathrm{hydro},i,\min} \le SOC_{\mathrm{hydro},i,t} \le SOC_{\mathrm{hydro},i,\max} \tag{17}$$

$$SOC_{\mathrm{PHS,min}} \le SOC_{\mathrm{PHS},t} \le SOC_{\mathrm{PHS,max}} \tag{18}$$

where $SOC_{\mathrm{hydro},i,\max}$, $SOC_{\mathrm{hydro},i,\min}$, $SOC_{\mathrm{PHS,max}}$, $SOC_{\mathrm{PHS,min}}$ are the maximum and minimum of the capacity ratio of the upriver reservoir capacity of the ith hydropower station and PHS. Both the reservoir constraints of hydropower plants and PHS should be in the range of upper and lower limits, and the exceeded parts should be transformed to the penalty terms incorporated into the optimizing target.

For the hydro-PV-PHS integrated power system, the bus voltage and the feeder current should be constrained with the admissible range as flows:

$$V_{\min} \le V_{i,t} \le V_{\max} \tag{19}$$

$$I_{j,\min} \le I_{j,t} \le I_{j,\max} \tag{20}$$

where $V_{i,t}$ and $I_{j,t}$ represent the voltage of the ith nodal and the current of the jth feeder at the tth time step. $V_{\min}$, $V_{\max}$, $I_{j,\min}$ and $I_{j,\max}$ are the range of nodal voltage and feeder current respectively.

It is important to take into account of earnings of the maximum revenue under different runoffs. Additionally, the hybrid energy system outputs a stable power in order to ensure the secure integrated power into the main grid to promote the consumption of PV. The objective function of the hybrid power system includes: maximizing the economic benefits, minimizing the fluctuation of the power source and PCC to the upper-level grid which is shown as follows:

$$R_{\mathrm{total},t} = \max\left(\beta_1 ER_t - \beta_2 \left(\Delta P_{\mathrm{source},t}\right)^2 - \beta_3 \left(\Delta P_t\right)^2\right) \tag{21}$$

where β_1, β_2 and β_3 are the weight factors. To address the trade-off between different targets, we can utilize entropy theory to obtain the weights [42]. Each objective function is calculated in the total T time step to form a matrix R_T as follows:

$$R_T = \begin{bmatrix} ER_1 & \Delta P_{\text{source},1} & \Delta P_1 \\ ER_2 & \Delta P_{\text{source},2} & \Delta P_2 \\ \vdots & \vdots & \vdots \\ ER_{T-1} & \Delta P_{\text{source},T-1} & \Delta P_{T-1} \\ ER_T & \Delta P_{\text{source},T} & \Delta P_T \end{bmatrix} \tag{22}$$

where each column of the matrix R_T is one type of target and each row is one type of time step. We use $r_{i,j}$ to represent the element of the matrix. Then the matrix can be normalized as follows:

$$c_{i,j} = \frac{\max(r_{i,j}) - r_{i,j}}{\max(r_{i,j}) - \min(r_{i,j})} \tag{23}$$

Then we can obtain the information entropy of the jth target H_j as follows:

$$H_j = -\sum_{i=1}^{T} c_{ij} \ln(c_{ij}) \tag{24}$$

At last, the weights of the jth target are then calculated through information entropy.

$$\beta_j = (1 - H_j) / \sum_{j=1}^{3} (1 - H_j) \tag{25}$$

where the weights satisfy $0 \le \beta_j \le 1$ and $\sum_{j=1}^{3} \beta_j = 1$.

III. Reinforcement Framework of Dynamic Dispatch

In this study, the dynamic dispatch of hydro-PV-PHS integrated power system aims to satisfy local demand in real-time by regulating the power output of hydropower plants and PHS plants in the case of uncertain solar energy and load. Reinforcement learning is very suitable for solving the optimal decision problem containing uncertain factors. To solve this problem, the mathematical model of dynamic economic dispatch of this system is transformed into a Markov decision process (MDP) and a DDPG algorithm is introduced to solve the MDP to obtain an optimal real-time dispatching strategy.

A. Environment

The reinforcement learning process has two conditions: agent and environment. The agent is trained to obtain the most cumulative reward according to the interactive environmental information. The environment represents the optimization model of a hydro-PV-PHS integrated power system, which refers to dynamic energy consumption, production, and exchange, as shown in Fig. 1.

B. Action

The agent is equivalent to the system operator. The operation agent determines the power generation of hydropower plants and the PHS plant according to the local environment information and a setpoint value offered by the system. Therefore, the agent action and action space are described as follows:

$$a_t = \{p_{\text{hydro},i,t}, P_{\text{PHS},t}\} \tag{26}$$

$$p_{\text{hydro},i,t} \in [p_{\text{hydro},i,\min}, p_{\text{hydro},i,\max}] \tag{27}$$

$$P_{\text{PHS},t} \in [P_{\text{PHS,min}}, P_{\text{PHS,miax}}] \tag{28}$$

where $p_{\text{hydro},i,\min}$ and $p_{\text{hydro},i,\max}$ are the range of the ith hydropower output. $P_{\text{PHS,min}}$ and $P_{\text{PHS,miax}}$ are the lower and upper limits of PHS output.

C. State

The state is the observed value obtained through interacting with the actual environment. It offers a reference data to help the agent to effectively train according to the perceived state information. The operation agent determines the regulation output though the observed information of the hydro-PV-PHS integrated power system. At the tth time step, the state space is given by:

$$s_t = (t, \lambda_t, P_{\text{PV},t}, P_{\text{load},t}, SOC_{\text{hydro},i,t}, SOC_{\text{PHS},t}) \tag{29}$$

D. Reward

The reward is the most important index to RL. The agent can be guided by a reasonable value function to advance in the "right direction." The target of this hybrid system is to maximize the revenue and minimize different types of fluctuations. The goal of the agent is to maximize accumulated reward in the learning process. However, the optimal policy must satisfy the constraints in the complementary model, at this point, the constraints should be reasonably converted into part of the reward. Here, the reward function is defined as:

$$r_t = 1/50000(R_{\text{total},t} - R_{\text{punish},t}) \tag{30}$$

$$R_{\text{punish},t} = (|c_t - c_{\min}| + |c_t - c_{\max}| - |c_{\max} - c_{\min}|)/2 \tag{31}$$

where r_t represents the reward at the tth time step. $R_{\text{punish},t}$ denotes the penalty term when (17)–(20) are not satisfied, where c_t includes $SOC_{\text{hydro},i,t}$, $SOC_{\text{PHS},t}$, $V_{i,t}$ and $I_{j,t}$.

The agent's objective is to find the optimal dispatch strategy from the largest accumulated reward using the Bellman equation via

$$Q_\pi(s,a) = E_\pi\left(\sum_{k=0}^{T} \gamma^k r_{t+k}(s_{t+k}, a_{t+k}) | s_t = s, a_t = a\right) \tag{32}$$

where E_π is the expectation under the strategy π and γ is a discount factor. The policy which follows the optimal action-value function can be described as:

$$\pi^* = \arg\max Q_\pi(s,a) \tag{33}$$

IV. Dispatch Policy Based on DDPG Algorithm

In this paper, the DDPG algorithm is introduced to solve

the MDP problem and to learn the optimal dynamic dispatch policy. DDPG is a model-free method based on the deterministic policy gradient which has the advantages of operating in continuous state and action space. It can explore the action space more thoroughly for better search results compared with other RL methods, such as DQN. In this way, it can not only estimate the optimal policy function through deep neural networks but also avoid the curse of dimensionality and save the entire action domain [43]. This method, which is based on dynamic interactions and evaluative feedback, does not require a forecasting model to be available. The agent is learning-while-doing though its trial-and-error with environment information which has a significant advantage of little or no domain knowledge requirements. The framework includes an actor-network (the policy network with θ^{π} as the parameter to approximate the policy function) and a critic network (the value network with θ^{Q} as the parameter to approximate the action-value function). Each network has its own target networks $\theta^{\pi'}$ and $\theta^{Q'}$ where π' and Q' represent target strategy and target Q value respectively. In the training process, the actor network and the critic network will fight with each other to maximize respective targets in order to obtain the total optimal regulation policy. The Actor's parameter θ^{π} and the Critic's parameter θ^{Q} are updated by the following equation:

$$L(\theta^{Q}) = \frac{1}{M} E\left(y_t - Q\left(s_t, a_t|\theta^{Q}\right)\right) \tag{34}$$

where (34) optimizes the parameter though minimizing the loss function $L(\theta^{Q})$ and where M is the size of the mini-batch. s_t is the current observation and a_t is the current action chosen by the actor. y_t denotes the target Q value which is computed as follows:

$$y_t = r_t + \gamma Q'\left(s_{t+1}, \pi'\left(s_{t+1}|\theta^{\pi'}\right)|\theta^{Q'}\right) \tag{35}$$

$$\nabla_{\theta^{Q}} L\left(\theta^{Q}\right) = E\left(2\left(y_t - Q\left(s_t, a_t|\theta^{Q}\right)\right)\right) \nabla_{\theta^{Q}} Q\left(s_t, a_t\right) \tag{36}$$

where r_t is the reward obtained at the tth step in an episode and s_{t+1} is the next observation. We can obtain (36) by applying equation (35) to (34) where $y_t - Q(s_t, a_t|\theta^{Q})$ is the timing differential error.

$$\theta^{Q} \leftarrow \theta^{Q} - \mu_Q \nabla_{\theta^{Q}} L(\theta^{Q}) \tag{37}$$

$$\nabla_{\theta^{\pi}} \pi = \nabla_a Q(s_t, a_t|\theta^{Q})|s = s_t$$

$$a = \pi(s_t) \cdot (\nabla_{\theta^{\pi}} \pi(s_t|\theta^{\pi})|s = s_t) \tag{38}$$

$$\theta^{\pi} \leftarrow \theta^{\pi} - \mu_{\pi} \nabla_{\theta^{\pi}} L(\theta^{Q}) \tag{39}$$

The critic network can be updated through a gradient to form (37) with a small learning rate θ^{Q} and the policy network can be updated by the strategy gradient according to (39) with a small learning rate μ_{π}.

Finally, the agent softly updates the target networks $\theta^{\pi'}$ and $\theta^{Q'}$ with a small update rate τ:

$$\theta^{Q'} \leftarrow \tau\theta^{Q} + (1-\tau)\theta^{Q'}$$

$$\theta^{\pi'} \leftarrow \tau\theta^{\pi} + (1-\tau)\theta^{\pi'} \tag{40}$$

where $0 \leq \tau \leq 1$. To fully explore the action space for interacting with the environment in order to learn a better dispatch strategy and avoid being caught into a local optima, a random noise $\eta(t)$ obeying Gaussian distribution which is added to DDPG and is given as follows:

$$a_t = \pi(s_t|\theta^{\pi}) + \eta(t) \tag{41}$$

To enhance the stability of the DDPG algorithm, the experience replay buffer is introduced. The whole training process is divided into three parts: exploration, learning, and convergence. In the exploratory stage, the agent observes the current information s_t, then chooses an action a_t and reaches a new state s_{t+1}, and at last obtains payoff r_t. The tuple (s_t, a_t, r_t, s_{t+1}) is stored in the experience replay buffer and the old tuples will be replaced by the new ones when their memory is full. In the learning process, mini-batch tuples are sampled to update the parameters of the actor and critic networks. through the game between the two networks, the agent learns an optimal strategy and keeps the cumulative rewards steady in the convergence stage. The computational process of the DDPG algorithm is shown in Fig. 2.

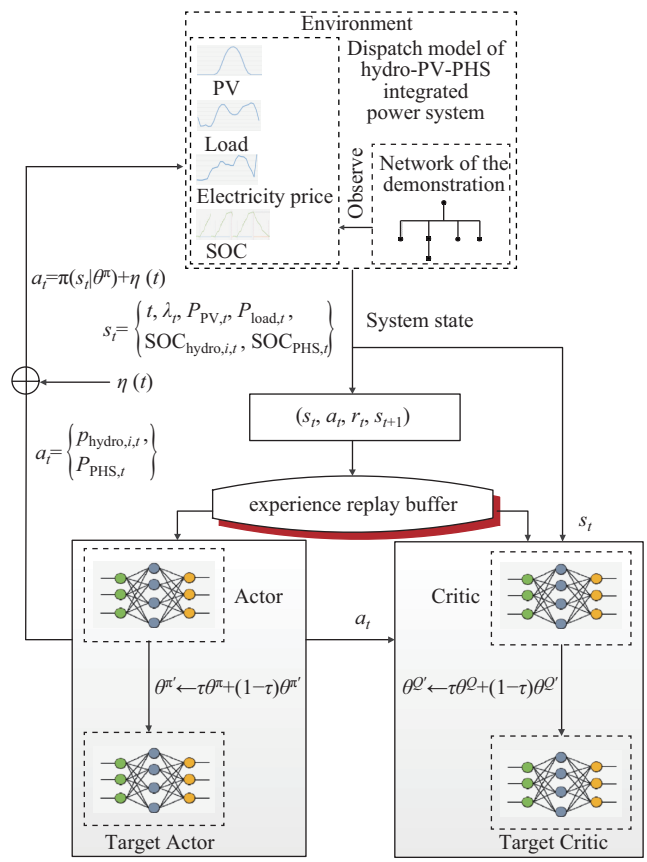

Fig. 2. DDPG-based training framework of hydro-PV-PHS integrated power system.

V. Case Study

The tested hydro-PV-PHS integrated power system which is shown in Fig. 3 consists of hydropower plants, PV power generation plants, and pumped- storage hydroelectric stations, which is a typical clean energy system. The part circled by the red dotted line is the simulating demonstration which is located in a county in the southwest of China. It includes two

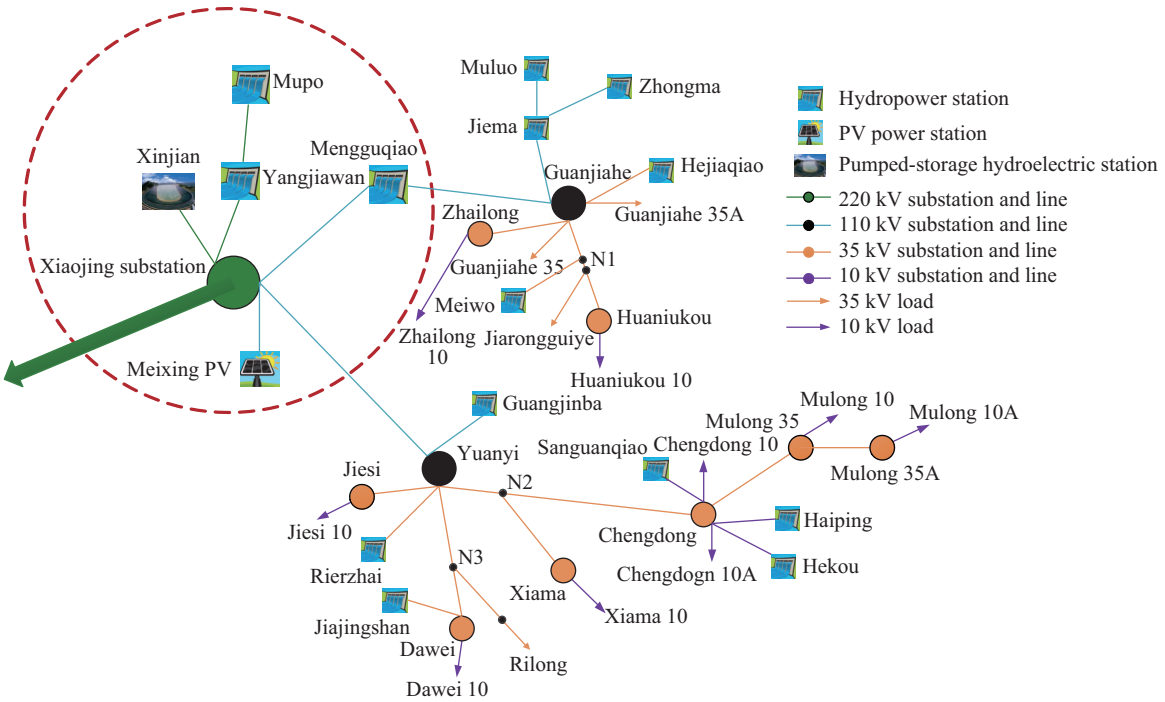

Fig. 3. Network of the tested hydro-PV-PHS integrated power system.

small hydropower plants with daily regulation named Mupo and Yangjiawan hydropower plants, a run-off hydropower plant named Mengguqiao hydropower plant, a 20-MW PHS power station named xinjian PHS power station, and a 100-MW PV power plant named Meixing PV power plant. The installed capacity of Mupo and Yangjiawan daily regulation hydropower plants are 45 MW and 60 MW respectively and the run-off Mengguqiao hydropower plant is 36 MW. For the sake of representation, we label Mupo, Yangjiawan, and Mengguqiao hydropower plants as hydropower plant #1, #2, and #3 respectively in this paper.

The environment simulation for the learning process was based on the data set on an hourly basis from 1 January 2017 to 7 November 2018 which consists of 3,370 (674×5) daily profiles. The daily profiles include the collected data of the PV power generation, load, price, and the average inflow of hydropower plants #1 and #2. The daily input data are shown in Fig. 4, as noted, we can see that PV output and price follow a strong stochastic process and the inflow of hydropower plants are variable in different periods which illustrates the uncertainty of the hydro-PV-PHS integrated power system. The weight factors of the objective function are set 0.5126, 0.0906, and 0.3968 respectively through the computation of entropy information. 95% of the dataset are trained by the agent to obtain an optimal dynamic dispatch strategy and the other 5% are allocated to form the test data to validate the effectiveness of the trained strategy. Different hyperparameters will have different performances of computation, so it is important to properly select the parameters for improving the performance of the algorithm. In this study, both actor-network and critic-network have two hidden layers with 128, 64 neurons respectively. The learning rate for the actor and critic are set to 1e-3 and 2e-3 respectively. The noise additional in DDPG is centered around 0 with a variance 0.45 and the soft replacement coefficient is set to 1e-2. The mini-batch size is set to 32 and the size of replay buffer is set to 8e4 in the experiment.

The parameters of the test hydro-PV-PHS integrated power system are shown in Table I.

TABLE I
PARAMETERS OF HYDRO-PV-PHS INTEGRATED MODEL

Parameter	Value	Parameter	Value	Parameter	Value
η_1	0.898	$P_{\text{hydro},2,\max}$	60 MW	$SOC_{\text{hydro},1,\min}$	0.2
η_2	0.833	$P_{\text{hydro},3,\max}$	36 MW	$SOC_{\text{hydro},1,\max}$	1
η_3	0.776	$P_{\text{hydro},1,\min}$	4.5 MW	$SOC_{\text{hydro},2,\min}$	0.2
η_{turbine}	0.898	$P_{\text{hydro},2,\min}$	7.8 MW	$SOC_{\text{hydro},2,\max}$	1
η_{pump}	0.89	$P_{\text{hydro},3,\min}$	6 MW	$SOC_{\text{PHS},\min}$	0.2
$P_{\text{hydro},1,\max}$	45 MW	P_c	20 MW	$SOC_{\text{PHS},\max}$	1

VI. PARAMETERS OF HYDRO-PV-PHS INTEGRATED MODEL

A. Training Process

In the experiment, the DDPG agent is carried for 6,000 times of trial-and-error training to learn the dispatch strategy for the hydro-PV-PHS integrated power system. Each episode includes 24-time steps representing a whole day and the reward discount is chosen to 0.9. Therefore, the agent samples a day state from the training data to form a state set concluding 24 points $S = [s_1, s_2, \cdots, s_{24}]$ in each episode, then the agent predicts an action set $A = [a_1, a_2, \cdots, a_{24}]$ according to the input state and obtains a corresponding cumulative reward by computing the reward set $R = [r_1, r_2, \cdots, r_{24}]$. In the training process, the agent tries to maximize R by interacting with the random environment to achieve a better dispatch ability. The details are shown in Fig. 5.

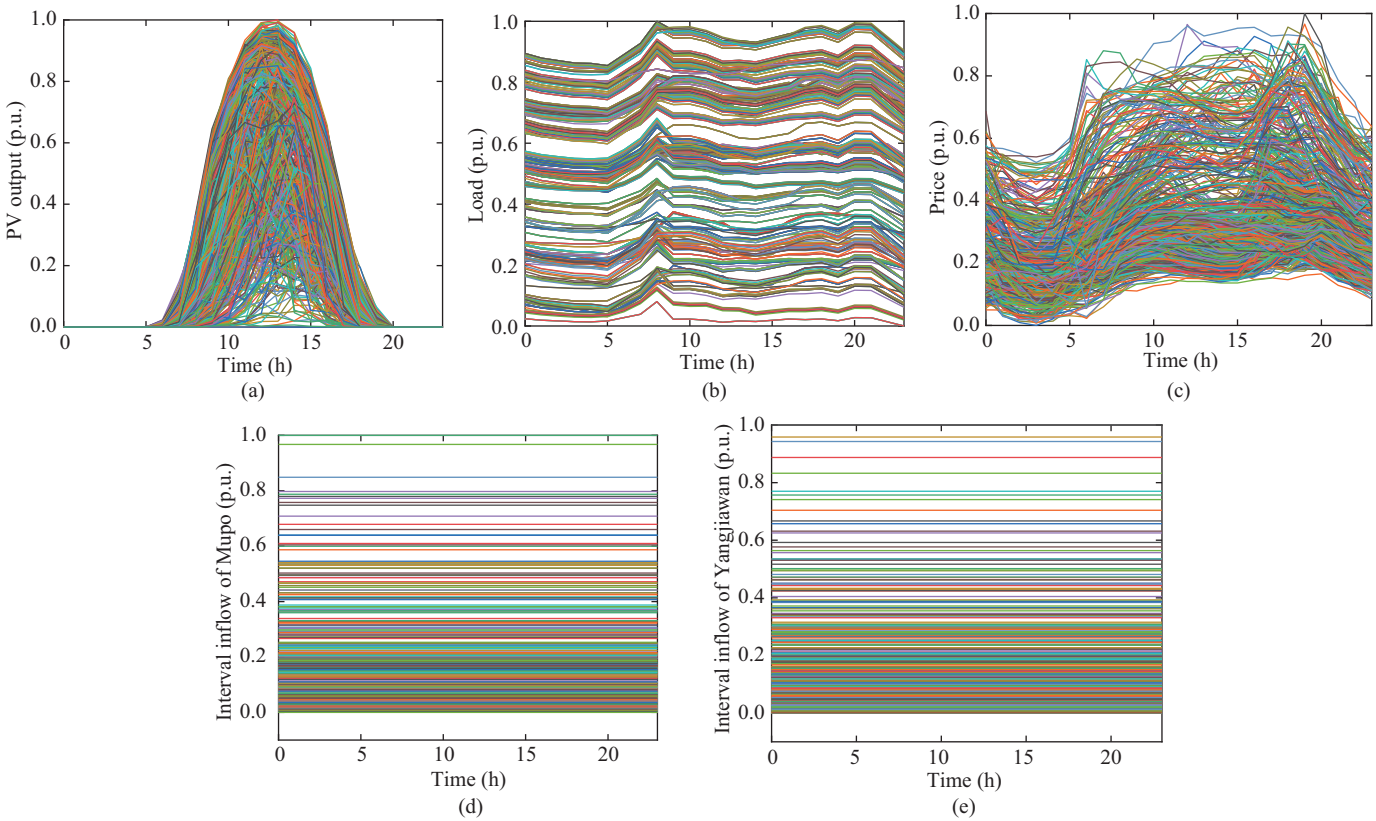

Fig. 4. Historical sample data of hydro-PV-PHS integrated power system. (a) PV output. (b) Load demand. (c) Price. (d) Interval inflow of hydropower plant #1. (e) Interval inflow of hydropower plant #2.

The cumulative reward is shown in Fig. 6. It can be seen that the cumulative reward is low and negative at the initial stage, this is primarily because the system agent is unfamiliar with the uncertain environment and tries to explore the optimal dispatch strategy to deal with random circumstances. The reward keeps increasing throughout the training and converges to a stable value after approximately 2,500 rounds of episodes which means the system agent has learned the optimal dispatch strategy to regulate the output of hydropower stations and PHS plants at arbitrarily random scenarios. It illustrates that the agent can dispatch the output of different unities based on the real-time status (PV, load, price and interval inflow) after training.

B. Dispatching Results

To verify the dispatch ability of the trained system agent in a random environment, the performance is validated in a test database. The power generation output of this system during anarbitrary 6 days selected in the test set is shown as follows. The dispatch results of hydropower plants are shown in Fig. 7. It can be observed that, the total power output is different at different stages in one day after the hydropower station is added to the demonstration base, but the output of every stage is relatively stable. The output curve, which can be nearly seen as a 3 segments' curve, is much smoother. It also shows that the complementary performance of the hydropower station depends on the average inflow, reservoir regulation capacity, and hydropower installed capacity. The daily average runoff for the six days is different apparently because of different weather conditions. The output power of hydropower plants #1 and #2 in the first two days are more than 25 MW and 32 MW separately and the interval inflow accounts for about 80%–90% of the maximum inflow. The total hydropower output is about 110 MW which shows day 1 and day 2 are obviously in the rainy period. The interval inflow in the following two days is higher than 40% and lower than 60% of the maximum inflow, and the total hydropower output is approximately equal to 70 MW, which indicates the two days are in the normal season. In the last two days, the interval inflow is only 10–20% of the maximum inflow. It can be seen that the output of hydropower plant #1 is under 15 MW and the output of #2 is within 17 MW. So, the fifth and sixth day are distinctly in the dry period.

We can see that the daily power supply is divided into three stages by the segmented fluctuation-control reward, whether it is in wet period or dry period which have different streamflow levels. The total output of the three stages after the regulated hydropower is relatively stable although it is large in the middle stage and low in the rest of the two stages because of shortage of solar energy. In rainy seasons, the power supply stabilizes at about 150 MW, while the other two stages are fixed at more or less 120 MW and 100 MW respectively. In normal periods, the total supply decreases by about 40 MW due to reduced average runoff inflow and the total output of dry seasons is smaller due to the same reasons. A smooth power output curve can save maintenance costs by reducing the start-

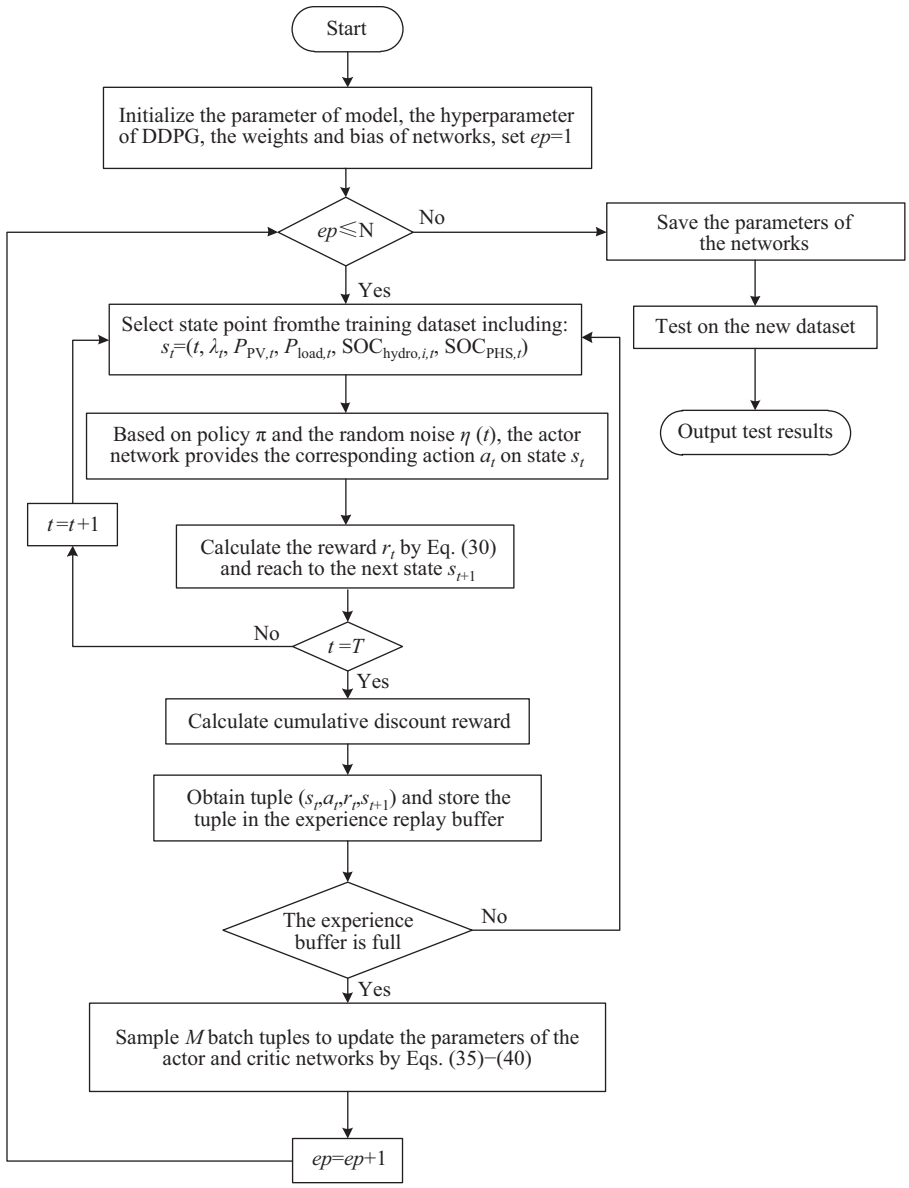

Fig. 5. Flowchart of the dynamic dispatch based DDPG algorithm.

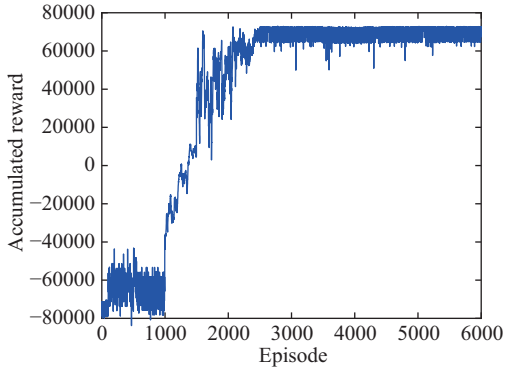

Fig. 6. The training process of the system agent.

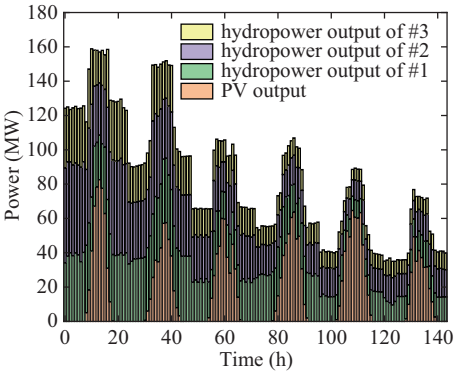

Fig. 7. Dispatch results of hydropower plants based on DDPG.

stop times of hydropower stations. The detailed comparison results of volatility on the daily supply point in the test data are shown in Fig. 8. As shown, the fluctuation of supply has been greatly alleviated after hydropower regulating although the volatility changes at different times. Furthermore, the average volatility of supply in the six days is clearly reduced, as

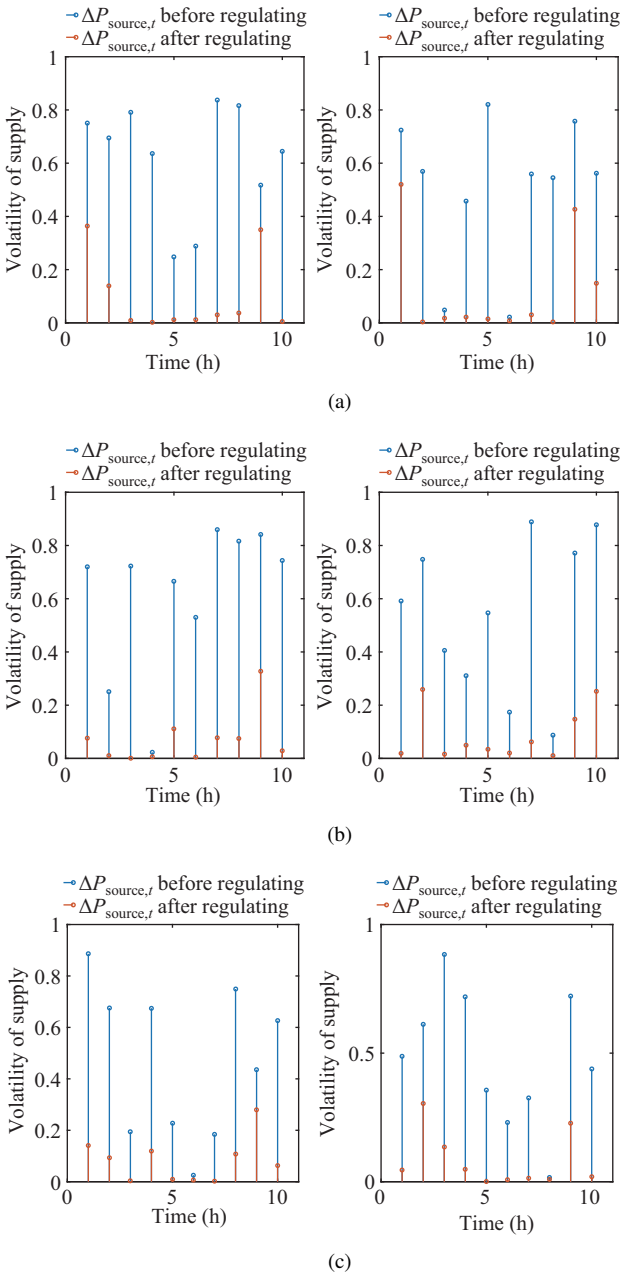

Fig. 8. Volatility of supply in different periods. (a) Wet periods. (b) Normal periods. (c) Dry periods.

shown in Fig. 9. The absolute mean value of volatility has reduced from nearly 32.7% to 6.0% after hydropower stations participating.

There are two sources to decide the PHS's condition of an agent: 1) The prior power fluctuation before PHS participating; 2) Sport price at current moment. Considering the above information, the dispatch results of PHS are shown in Fig. 9, where ΔP_t represents the previous power fluctuation before PHS regulating. It can be seen that the agent can very well determine the generation-pumping condition of the PHS according to price. The PHS usually generates power and sells to the grid to obtain revenue at peak price, but it still generates power to increase the power supply although price is off-peak, such as 129–131 points, because of the local load demand cannot be satisfied. Meanwhile, PHS operates in the pumping mode to reduce the volatility on PCC, such as 66, 108, and 109 points, although the prices are high. The PHS generally purchases power from the grid when the price is low, but the agent controls the PHS to sell power to the grid to relieve fluctuations to satisfy the delivery demand of energy supply, such as the case at the ninth point. Therefore, the agent controls the PHS to operate in the opposite direction of ΔP_t violating the trend that purchasing at a low price and generating electricity at a high price for reducing volatility on PCC.

The volatility on PCC to the upper-level grid after PHS participating is shown in Fig. 10. The mean volatility during the six days is nearly reduced by 3.28%, which indicates the power fluctuation on PCC has been improved after the

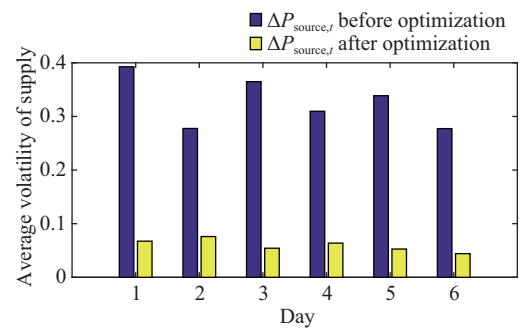

Fig. 9. Average volatility of supply after hydropower participating.

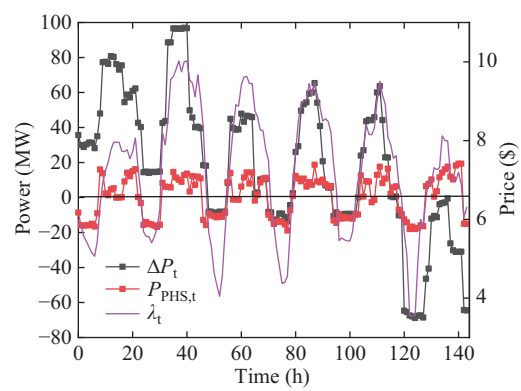

Fig. 10. Dispatch results of PHS based on power fluctuation and price.

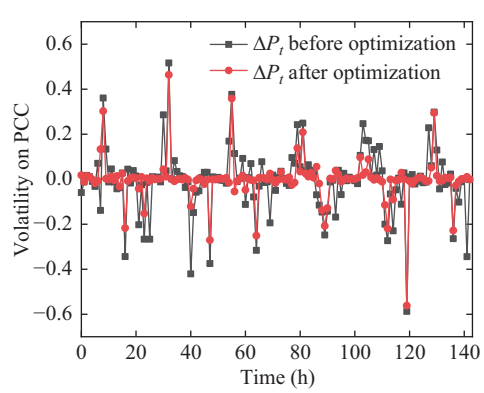

Fig. 11. Volatility on PCC after PHS participating.

dynamic dispatch of the system agent. That is, the agent can determine different operating conditions based on information, including current power output of the source, load, and price, to cope with the stress of fluctuation to the upper-level grid. Moreover, it can be seen that the PHS is frequently generated and pumped in DDPG-based dispatching schemes from Figs. 10 and 11.

Again, it can be observed that the agent maintains the change of SOC as shown in Fig. 12. All SOCs are between 0.2 and 1 during these tested days which means the agent can cause the system to maximize the profit under the safe boundary in a random environment. For mitigating the fluctuation of the power source, the agent regulates the power generation of Mupo-Yangjiawan hydropower stations through the experience learned from training datasets. It can be seen that the reservoir water level frequently fluctuates on account of the high volatility of PV output as shown in Fig. 13.

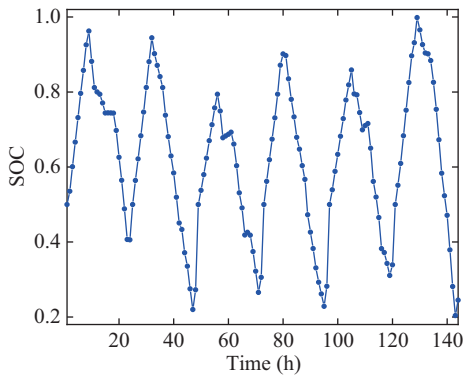

Fig. 12. SOC variation of PHS.

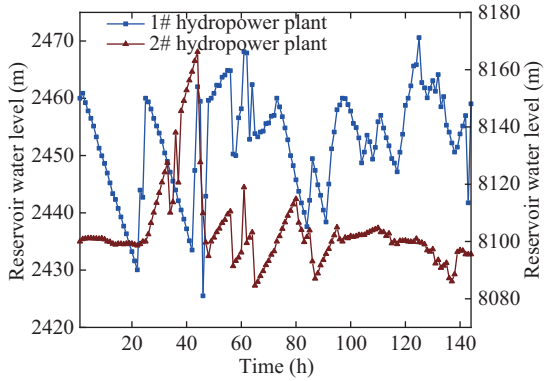

Fig. 13. Reservoir water level of the hydropower station.

C. Comparison Results

Moreover, the other two methods are used as benchmarks for comparison to analyze the performance of DDPG for dynamic dispatch of the hydro-PV-PHS integrated power system. One is stochastic programming (SP) and the other is particle swarm optimization (PSO). Due to the dynamic dispatch being tested in real-time, it needs to give the value of PV output, water inflow, and spot price through prediction for the two benchmarks. In this experiment, the forecasted state error is assumed to be 5%. For the uncertainty modeling, it is assumed that the variation of the uncertain state variables obeys a normal distribution. The mean value of the distribution is the test data and 200 random scenarios can be generated from the confidence interval to represent the stochasticity of the system. Then the number of scenarios is reduced to 50 using scenario reduction technology. At every SP scenario, the optimization problem is solved with the Gurobi solver. The optimal solution of the PSO-based method is given by averaging the optimal results of different samples. The comparison results are shown in Table II. It can be observed that PSO has the lowest revenue and highest volatility in comparison with the other two methods. Compared with the SP-based method, the DDPG algorithm increases the revenue by 3.9%, meanwhile, it reduces the volatility of supply and PCC by 1.46% and 1.33% respectively. In comparison with DDPG, the PSO method decreases the revenue by 7.3% and increases the volatility by about 2.43% and 2.6% respectively. The comparison results confirm that DDPG can achieve better performance which outperforms PSO and SP because of avoiding the prediction of power, load, price, and water inflow. Therefore, the agent can adapt the random dynamic environment to dynamically dispatch the power generation of the hydropower plants and the PHS. The accuracy of the optimization of SP and PSO-based methods depends on the prediction precision, so the dispatch results may be affected by forecasting error of PV, load, price, and so on. In summary, the dynamic dispatch strategy of the DDPG-based method achieves the best performance.

TABLE II
ESTIMATION RESULTS OF DIFFERENT ALGORITHMS

Methods	Revenue ($)	Volatility of power source	Volatility on PCC
SP	24886.9356	0.0744	0.0752
PSO	24030.8045	0.0841	0.0879
DDPG	25923.8877	0.0598	0.0619

The experiment is programmed in Python 3.7 on a personal computer with 3.4 GHz×8 Intel i7-4700 and 16 GB main memory DDR3. The results of testing time are listed in Table III. It can be seen that the running time of SP and PSO are 16.223 s and 90.879 s respectively, however, the computation time of DDPG is only 0.136 s although the offline training is time-consuming, this is primarily because the agent can directly learn the mapping relationships of state-action from offline samples and then make the online decision in the second level. The solution time of SP is 100 times more than DDPG, this may be because SP has to solve the dispatch problem at every time slot with many scenarios which will affect computation speed. PSO needs more time, primarily because it is an iterative-oriented algorithm which does not have the ability to learn and cannot quickly respond to the integrated system. Through the training of the nearly two-year data set, the system agent can immediately respond according

TABLE III
COMPARATIVE RESULTS OF RUNNING TIME

Methods	SP	PSO	DDPG
Running time	16.223 s	90.879 s	0.136 s

the interacted environment information and can apply in a online dispatch decision of the hydro-PV-PHS integrated power system.

VII. Conclusion

The paper has proposed a DDPG-based method for dynamic dispatch of hydro-PV-PHS integrated power systems. Considering economic benefits, power fluctuation of grid-connection points and supply, a DRL-based system agent framework has been proposed. The presented framework can effectively handle complicated online dispatch problems of hydro-PV-PHS integrated power systems. In comparison with other methods, several findings are summarized as follows:

1) The uncertainty of PV, load, electricity price, and water inflow are fully taken into account, a DDPG agent is utilized to make dynamic dispatch decisions online, the presented method does not need to forecast the value of uncertain power, and it can provide a good strategy to improve the economy and relive random fluctuations of supply and PCC.

2) In addition, the dynamic reward function has developed in the intelligent dispatch system, information entropy is introduced to trade-off economy and volatility, which can further improve the performance of the agent.

3) The comparison results show significant improvements over other existing model-based methods. DDPG can obtain an optimal dispatch strategy, which can increase the economic revenue by up to 7.3% over PSO algorithms, while reducing the mean volatility by about 2.5%.

4) The average computation time of the proposed method is 0.136 s, which has a significant advantage in decision speed. Comparison results show that our proposed method can achieve fast decision making which is more efficient than conventional model-based methods. In this way, the computational cost is greatly reduced.

Future studies will focus on using other DRL-based algorithms to obtain a more precise dispatch decision in typical hydro-PV-load scenarios and exploring the DRL-based method which can estimate the boundary to deal with the difficulty to design the reward when containing many constraints.

References

[1] T. M. Alabi, L. Lu, and Z. Y. Yang, "Stochastic optimal planning scheme of a zero-carbon multi-energy system (ZC-MES) considering the uncertainties of individual energy demand and renewable resources: an integrated chance-constrained and decomposition algorithm (CC-DA) approach," *Energy*, vol. 232, pp. 121000, May 2021.

[2] T. M. Alabi, L. Lu, and Z. Y. Yang, "Improved hybrid inexact optimal scheduling of virtual powerplant (VPP) for zero-carbon multi-energy system (ZCMES) incorporating Electric Vehicle (EV) multi-flexible approach," *Journal of Cleaner Production*, vol. 326, pp. 129294, Oct. 2021.

[3] D. Wu, M. Javadi, and J. N. Jiang, "A preliminary study of impact of reduced system inertia in a low-carbon power system," *Journal of Modern Power Systems and Clean Energy*, vol. 3, no. 1, pp. 82–92, Mar. 2015.

[4] L. X. Tian, Y. S. Huang, S. Liu, S. Z. Sun, J. J. Deng, and H. F. Zhao, "Application of photovoltaic power generation in rail transit power supply system under the background of energy low carbon transformation," *Alexandria Engineering Journal*, vol. 60, no. 6, pp. 5167–5174, Jun. 2021.

[5] N. Li and K. W. Hedman, "Economic assessment of energy storage in systems with high levels of renewable resources," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 3, pp. 1103–1111, Jul. 2015.

[6] N. Li, C. Uçkun, E. M. Constantinescu, J. R. Birge, K. W. Hedman, and A. Botterud, "Flexible operation of batteries in power system scheduling with renewable energy," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 2, pp. 685–696, Apr. 2016.

[7] M. L. Rahman, S. Oka, and Y. Shirai, "Hybrid power generation system using offshore-wind turbine and tidal turbine for power fluctuation compensation (HOT-PC)," *IEEE Transactions on Sustainable Energy*, vol. 1, no. 2, pp. 92–98, Jul. 2010.

[8] Z. Z. Zhang, Y. Zhang, Q. Huang, and W. J. Lee, "Market-oriented optimal dispatching strategy for a wind farm with a multiple stage hybrid energy storage system," *CSEE Journal of Power and Energy Systems*, vol. 4, no. 4, pp. 417–424, Dec. 2018.

[9] J. N. Yu, Y. M. Tang, K. Y. Chau, R. Nazar, S. Ali, and W. Iqbal, "Role of solar-based renewable energy in mitigating CO_2 emissions: evidence from quantile-on-quantile estimation," *Renewable Energy*, vol. 182, pp. 216–226, Jan. 2022.

[10] J. X. Yang, S. Zhang, Y. Xiang, J. C. Liu, J. Y. Liu, X. Y. Han, and F. Teng, "LSTM auto-encoder based representative scenario generation method for hybrid hydro-PV power system," *IET Generation, Transmission & Distribution*, vol. 14, no. 24, pp. 5935–5943, Dec. 2020.

[11] S. Makhdoomi and A. Askarzadeh, "Daily performance optimization of a grid-connected hybrid system composed of photovoltaic and pumped hydro storage (PV/PHS)," *Renewable Energy*, vol. 159, pp. 272–285, Oct. 2020.

[12] T. Ma, H. X. Yang, L. Lu, and J. Q. Peng, "Pumped storage-based standalone photovoltaic power generation system: modeling and techno-economic optimization," *Applied Energy*, vol. 137, pp. 649–659, Jan. 2015.

[13] S. Ali, R. A. Stewart, and O. Sahin, "Drivers and barriers to the deployment of pumped hydro energy storage applications: systematic literature review," *Cleaner Engineering and Technology*, vol. 5, pp. 100281, Dec. 2021.

[14] C. Li, D. Q. Zhou, H. Wang, H. B. Cheng, and D. D. Li, "Feasibility assessment of a hybrid PV/diesel/battery power system for a housing estate in the severe cold zone-A case study of Harbin, China," *Energy*, vol. 185, pp. 671–681, Oct. 2019.

[15] F. Bilgili, D. B. Lorente, S. Kuşkaya, F. Ünlü, P. Gençoğlu, and P. Rosha, "The role of hydropower energy in the level of CO_2 emissions: an application of continuous wavelet transform," *Renewable Energy*, vol. 178, pp. 283–294, Nov. 2021.

[16] Y. S. Zhang, C. Ma, Y. Yang, X. L. Pang, L. Liu, and J. J. Lian, "Study on short-term optimal operation of cascade hydro-photovoltaic hybrid systems," *Applied Energy*, vol. 291, pp. 116828, Jun. 2021.

[17] F. L. Zhu, P. A. Zhou, Y. M. Sun, B. Xu, Y. F. Ma, W. F. Liu, D. C. Zhang, and J. M. Dawa, "A coordinated optimization framework for long-term complementary operation of a large-scale hydro-photovoltaic hybrid system: nonlinear modeling, multi-objective optimization and robust decision-making," *Energy Conversion and Management*, vol. 226, pp. 113543, Dec. 2020.

[18] W. Fang, Q. Huang, S. Z. Huang, J. Yang, E. H. Meng, and Y. Y. Li, "Optimal sizing of utility-scale photovoltaic power generation complementarily operating with hydropower: a case study of the world's largest hydro-photovoltaic plant," *Energy Conversion and Management*, vol. 136, pp. 161–172, Mar. 2017.

[19] K. P. Wong and Y. W. Wong, "Hybrid genetic/simulated annealing approach to short-term multiple-fuel-constrained generation scheduling," *IEEE Transactions on Power Systems*, vol. 12, no. 2, pp. 776–784, May 1997.

[20] M. Petrollese, L. Valverde, D. Cocco, G. Cau, and J. Guerra, "Real-time integration of optimal generation scheduling with MPC for the energy management of a renewable hydrogen-based microgrid," *Applied Energy*, vol. 166, pp. 96–106, Mar. 2016.

[21] Z. W. Zhang, C. F. Wang, H. C. Lv, F. Q. Liu, H. Z. Sheng, and M. Yang, "Day-ahead optimal dispatch for integrated energy system considering power-to-gas and dynamic pipeline networks," *IEEE Transactions on Industry Applications*, vol. 57, no. 4, pp. 3317–3328, Jul./Aug. 2021.

[22] W. Hu, H. X. Zhang, Y. Dong, Y. T. Wang, L. Dong, and M. Xiao, "Short-term optimal operation of hydro-wind-solar hybrid system with improved generative adversarial networks," *Applied Energy*, vol. 250, pp. 389–403, Sep. 2019.

[23] B. Ming, P. Liu, S. L. Guo, L. Cheng, Y. L. Zhou, S. D. Gao, and H. Li, "Robust hydroelectric unit commitment considering integration of large-scale photovoltaic power: a case study in China," *Applied Energy*, vol. 228, pp. 1341–1352, Oct. 2018.

[24] A. Cerejo, S. J. P. S. Mariano, P. M. S. Carvalho, and M. R. A. Calado, "Hydro-wind optimal operation for joint bidding in day-ahead market: storage efficiency and impact of wind forecasting uncertainty," *Journal*

of Modern Power Systems and Clean Energy, vol. 8, no. 1, pp. 142–149, Jan. 2020.

[25] X. B. Wang, J. X. Chang, X. J. Meng, and Y. M. Wang, "Short-term hydro-thermal-wind-photovoltaic complementary operation of interconnected power systems," *Applied Energy*, vol. 229, pp. 945–962, Nov. 2018.

[26] B. Xiao, J. Zheng, G. G. Yan, Y. G. Shi, M. X. Jiao, Y. Wang, L. Dong, M. C. Wang, and H. Z. Yang, "Short-term optimized operation of multi-energy power system based on complementary characteristics of power sources," in *Proceedings of 2019 IEEE Innovative Smart Grid Technologies - Asia*, 2019, pp. 3770–3775.

[27] J. X. Yang, J. C. Liu, and S. Zhang, "Optimization for short-term operation of hybrid hydro-PV power system based on NSGA-II," in *Proceedings of 2020 IEEE 4th Conference on Energy Internet and Energy System Integration*, 2020, pp. 2267–2271.

[28] S. Makhdoomi and A. Askarzadeh, "Optimizing operation of a photovoltaic/diesel generator hybrid energy system with pumped hydro storage by a modified crow search algorithm," *Journal of Energy Storage*, vol. 27, pp. 101040, Feb. 2020.

[29] L. Lu, W. L. Yuan, C. G. Su, P. L. Wang, C. T. Cheng, D. H. Yan, and Z. N. Wu, "Optimization model for the short-term joint operation of a grid-connected wind-photovoltaic-hydro hybrid energy system with cascade hydropower plants," *Energy Conversion and Management*, vol. 236, pp. 114055, May 2021.

[30] C. A. Hans, P. Sopasakis, J. Raisch, C. Reincke-Collon, and P. Patrinos, "Risk-averse model predictive operation control of islanded microgrids," *IEEE Transactions on Control Systems Technology*, vol. 28, no. 6, pp. 2136–2151, Nov. 2020.

[31] J. Sachs and O. Sawodny, "A two-stage model predictive control strategy for economic diesel-PV-battery island microgrid operation in rural areas," *IEEE Transactions on Sustainable Energy*, vol. 7, no. 3, pp. 903–913, Jul. 2016.

[32] W. L. Yuan, X. Q. Wang, C. G. Su, C. T. Cheng, Z. Liu, and Z. N. Wu, "Stochastic optimization model for the short-term joint operation of photovoltaic power and hydropower plants based on chance-constrained programming," *Energy*, vol. 222, pp. 119996, May 2021.

[33] Q. Li, Y. B. Qiu, H. Q. Yang, Y. Xu, W. R. Chen, and P. Wang, "Stability-constrained two-stage robust optimization for integrated hydrogen hybrid energy system," *CSEE Journal of Power and Energy Systems*, vol. 7, no. 1, pp. 162–171, Jan. 2021.

[34] L. Zhou, A. Swain, and A. Ukil, "Reinforcement learning controllers for enhancement of low voltage ride through capability in hybrid power systems," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 8, pp. 5023–5031, Aug. 2020.

[35] Z. M. Yan and Y. Xu, "Real-time optimal power flow: a *Lagrangian* based deep reinforcement learning approach," *IEEE Transactions on Power Systems*, vol. 35, no. 4, pp. 3270–3273, Jul. 2020.

[36] Y. C. Liang, C. L. Guo, Z. H. Ding, and H. C. Hua, "Agent-based modeling in electricity market using deep deterministic policy gradient algorithm," *IEEE Transactions on Power Systems*, vol. 35, no. 6, pp. 4180–4192, Nov. 2020.

[37] L. Lin, X. Guan, Y. Peng, N. Wang, S. Maharjan, and T. Ohtsuki, "Deep reinforcement learning for economic dispatch of virtual power plant in internet of energy," *IEEE Internet of Things Journal*, vol. 7, no. 7, pp. 6288–6301, Jul. 2020.

[38] X. Zhang, Y. B. Liu, J. J. Duan, G. Qiu, T. J. Liu, and J. Y. Liu, "DDPG-based multi-agent framework for SVC tuning in urban power grid with renewable energy resources," *IEEE Transactions on Power Systems*, vol. 36, no. 6, pp. 5465–5475, Nov. 2021.

[39] L. Yu, W. W. Xie, D. Xie, Y. L. Zou, D. Y. Zhang, Z. X. Sun, L. H. Zhang, Y. Zhang, and T. Jiang, "Deep reinforcement learning for smart home energy management," *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 2751–2762, Apr. 2020.

[40] Y. K. Liu, D. X. Zhang, and H. B. Gooi, "Optimization strategy based on deep reinforcement learning for home energy management," *CSEE Journal of Power and Energy Systems*, vol. 6, no. 3, pp. 572–582, Sep. 2020.

[41] S. Lee and D. H. Choi, "Federated reinforcement learning for energy management of multiple smart homes with distributed energy resources," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 1, pp. 488–497, Jan. 2022.

[42] G. Z. Zhang, W. H. Hu, D. Cao, W. Liu, R. Huang, Q. Huang, Z. Chen, and F. Blaabjerg, "Data-driven optimal energy management for a wind-solar-diesel-battery-reverse osmosis hybrid energy system using a deep reinforcement learning approach," *Energy Conversion and Management*, vol. 227, pp. 113608, Jan. 2021.

[43] X. D. Wang, Y. B. Liu, J. B. Zhao, C. Liu, J. Y. Liu, and J. Y. Yan, "Surrogate model enabled deep reinforcement learning for hybrid energy community operation," *Applied Energy*, vol. 289, pp. 116722, May 2021.

Jingxian Yang received a B.S. degree in Electrical Engineering and Automation from Automatization & Electric Engineering College, Lanzhou Jiaotong University, Lanzhou, China, in 2007. She is currently pursuing a Ph.D. degree in Electrical Engineering at Sichuan University. Her research interests include big data applications to the smart grid, optimization, and control in power systems.

Jichun Liu (SM'15) received B.E., M.S., and Ph.D. degrees in Electrical Engineering from Sichuan University, Chengdu, in 1997, 2000, and 2006, respectively. From 2008 to 2009, he worked as a postdoctoral researcher at the Illinois Institute of Technology in the area of electrical engineering. He is currently a Professor with the College of Electrical Engineering, Sichuan University, Chengdu, China. His main research interests include smart grids, power generation, and optimization of power systems.

Yue Xiang (SM'16) received B.S. and Ph.D. degrees from Sichuan University, Chengdu, China, in 2010 and 2016, respectively. From 2013 to 2014, he was a joint Ph.D. Student at the Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA and also a Visiting Scholar with the Department of Electronic and Electrical Engineering, University of Bath, Bath, U.K., in 2015. He is an Associate Professor at the College of Electrical Engineering, Sichuan University, China. His research interests include electric vehicles, power system planning and optimal operations, renewable energy integration, and smart grids.

Shuai Zhang received a B.S. degree in Mechanical Engineering and Automation in 2010 and an M.S. degree in Mechanical and Electronic Engineering in 2013 respectively from Southwest Petroleum University, Chengdu, Sichuan, China. He received a Ph.D. degree in Electrical Engineering at Sichuan University, Chengdu, Sichuan, China, in 2020. He is currently an Engineer with the State Grid Sichuan Electric Power Company Chengdu Power Supply Company, Chengdu, China. His research interests include renewable sources generation planning and dispatching, electric power market and economics, and energy storage devices.

Junyong Liu (M'16) received a Ph.D. degree from Brunel University, London, U.K., in 1998. He is currently a Professor with the College of Electrical Engineering, Sichuan University, Chengdu, China. He is the director of the Sichuan Province Key Smart Grid Laboratory. His research interests include power market, power system planning, operation, and computer applications.