Scheduled Operation of Wind Farm with Battery System Using Deep Reinforcement Learning

IEEJ TRANSACTIONS ON ELECTRICAL AND ELECTRONIC ENGINEERING
IEEJ Trans 2021; **16**: 687–695
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI:10.1002/tee.23348

Paper

Scheduled Operation of Wind Farm with Battery System Using Deep Reinforcement Learning

Mamoru Futakuchi, Member
Satoshi Takayama, Member
Atsushi Ishigame, Senior Member

With increasing amounts of wind power generation installed, the steep fluctuation of wind power generation output, called ramp events, causes serious problems for power system operation. Controlling fluctuations is an important issue for increasing the amount of wind power generation as a wind farm (WF) in the future. The authors reported the scheduled operation method of WF using a battery energy storage system (BESS) and forecast data of wind power generation output. In this paper, the authors propose a new scheduled operation method of WF. In particular, we propose the application of deep reinforcement learning to decide the output schedule of WF. Moreover, we compare the conventional method, the reinforcement learning method, and the deep reinforcement learning method in terms of the number of ramp events. In addition, we calculate the reducing effect of the storage capacity of BESS. © 2021 Institute of Electrical Engineers of Japan. Published by Wiley Periodicals LLC.

Keywords: wind generation; battery energy storage system; scheduled operation; reinforcement learning; deep learning; deep reinforcement learning

Received 22 July 2020; Revised 8 February 2021

1. Introduction

Renewable sources of energy such as wind have developed rapidly worldwide [1]. Installed wind power capacity has increased from 432 680 MW in 2015 to 486 790 MW in 2016 globally and from 3038 to 3234 MW in Japan over the same period [2]. Therefore, wind farms (WFs) are expected to be installed as sources of renewable and clean energy.

However, wind power is associated with uncertainty. The long-term fluctuation of wind power, called ramp events, occur, and they last for anywhere between tens of minutes to several hours. There are two basic types of ramp events: ramp-up and ramp-down. Ramp-up is characterized by increased levels of wind power, and they are caused by low-level jets, thunderstorms, wind gusts, or similar weather phenomena [3]. Ramp-down events are characterized by decreased wind power. Ramp events threaten the reliable and economic operation of power systems [4]. In recent years, this phenomenon has been studied by working in the fields of analysis [4–11], forecasting [12], and control [13–18]. In terms of analytical studies, time series measured at offshore WFs in Horns Rev., Denmark, have been compared to a time series simulated based on the power fluctuation model presented in Reference [5]. In References [4,6], the detection of the ramp using the optimized swing door algorithm has been discussed. In Reference [7], the authors focused on extreme ramp events and performed statistical extreme value analysis. In Japan, ramp events were first discussed by Ogimoto in Reference [8] and analyzed by Ikegami *et al.* [9], Yoshida *et al.* [10], and Oba *et al.* [11]. In terms of forecasting studies, in Reference [12], scenarios of wind power generation were generated using a neural-network-based stochastic process model and forecasts were made based on extracted features. Qi *et al.* proposed the suppressing method of ramp events based on bid strategy by using a competitive game strategy [13]. Moreover, many authors have been proposed scheduled operation methods using an energy storage system (ESS) [14–18].

When using ESS to adjust the output of renewable energy, state-of-charge (SOC) management is important. This is because output compensation cannot be performed when the ESS is fully charged or discharged. Therefore, in References [19,20] control strategies to protect battery-ESS (BESS) by maintaining SOC within a suitable range (i.e., between 30 and 70%) have been proposed. In Reference [21], a self-adaptive control strategy based on the fuzzy logic controller was adopted when SOC exceeds the warning limits. As another approach, a dispatch scheme is scheduled depending on SOC in References [22,23]. In general, the required BESS capacity increases in the absence of a SOC management scheme. Given that an additional large BESS is expensive, the implementation of a SOC management strategy should be considered.

In the conventional method [16], we have proposed setting the SOC target to 50% to respond flexibly to wind power forecast error.

However, there is still room for study on how to decide appropriate target values according to the situation. The SOC management method was the issue of previous research. If SOC of BESS can be managed properly, the required BESS capacity will be reduced compared to the management method that keeps SOC constant. However, it is very difficult to uniquely determine the appropriate SOC target value because it changes due to the influence of prediction error and grid code (GC).

[a] Correspondence to: Satoshi Takayama.
E-mail: takayama@eis.osakafu-u.ac.jp

Power System Research Group, Osaka Prefecture University, 1-1 Gakuen-cho, Nakaku, Sakai Osaka, 599-8531, Japan

In this paper, we propose a new scheduled operation method for WF using reinforcement learning. Reinforcement learning is a method in which an agent learns better behavior selection (decision-making) through interaction with the environment (controlled object) [24]. Since agents will be able to learn the optimal action selection through trial and error, the setting of SOC target value by autonomous learning can be expected.

Thus, in the model proposed herein, the purpose is achieved adequately by using reinforcement learning. However, reinforcement learning has the disadvantage that learning becomes unstable as the number of states increases. To solve this problem, deep reinforcement learning has proposed [25]. Deep reinforcement learning is a method that uses reinforcement learning and deep learning, and it has been reported that arcade games have surpassed human experts. Deep learning is a method concerning the structure of the human brain and has made it possible to extract high-level features from raw sensory data, leading to breakthroughs in speech recognition [26]. Therefore, deep reinforcement learning can expect autonomous learning using complex information. As mentioned above, many variables such as prediction error and GC have a complicated influence on the determination of the SOC target value. Therefore, deep reinforcement learning is appropriate to consider more variables and to deal with uncertainties.

In this paper, we consider the situation where the WF operator must notify the generation plan to the organization for cross-regional coordination of transmission operators (OCCTO) 1 h in advance. The goal of this study is to reduce the required capacity of BESS for suppressing ramp events under the institution.

The remainder of this paper is organized as follows. Section 2 presents an outline of this study: dispatching rules, scheduled operation, and a description of aggregated WFs with the BESS system. The mechanism of machine learning, such as reinforcement learning, deep learning, and deep reinforcement learning, is described in Section 3. The proposed scheduling model that decides the target value of SOC with machine learning is described in Section 4. Case studies based on numerical simulation using observed and forecast data from Tokai, Japan, are given in Section 5. Finally, a few concluding remarks are presented in Section 6.

2. Outline of this Study

2.1. System configuration A schematic diagram of aggregated WFs with BESS is shown in Fig. 1. WF aggregation leads to the smoothing effect, which mitigates WF power fluctuations. This effect related to geographical distribution results in low short-term fluctuations in wind power production [9,27]. In addition, the power flow at the junction between WF and BESS is monitored by the control center [1] for security.

2.2. Scheduled operation Scheduled operation is achieved using Scheduling Generation plans and real-time operating of BESS [28]. Figure 2 shows the schedule for scheduled operation. The balancing rule enforced in April 2016 includes 'the day before plan' to notify the generation plan of the next day on the previous day and 'the 1-h advance plan' to notify the generation plan 1 h before shown in Fig. 2. In this paper, we consider the 1-h advance plan. Therefore, the WF operator must notify the generation plan (The unit is kWh/30 min) 1.5 h in advance, considering the gate closing time (1 h) and the optimization calculation time (0.5 h). The WF operator prepares a generation plan every 30 min.

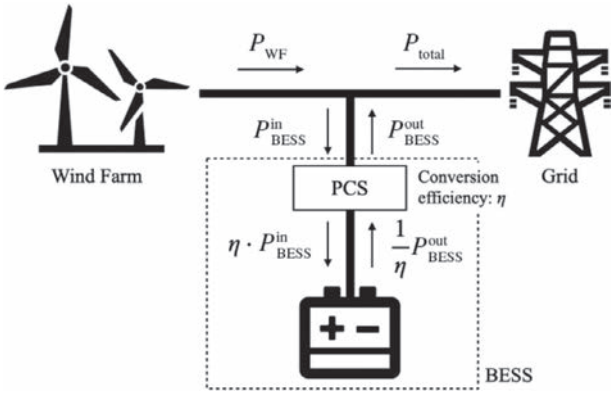


Fig. 1. Schematic of the aggregated WFs with BESS

In the scheduling phase, generation plans based on WF power forecasts are scheduled for submission to OCCTO. In this paper, a generation plan P^t_{plan}, which is the amount of energy generated in 30 min, is determined using the scheduling model. The scheduling model is described in Section 4. In the operating phase, BESS is used to adjust WF power considering the generation plans. Whereas a generation plan is scheduled at intervals of 30 min, WF operation is conducted every minute. The order of power of BESS can be optimized in advance [14] or determined sequentially depending on WF power [19–23]. The latter approach is adopted in this paper, which means the scheduled BESS power is not used, and P^t_{BESS} is calculated as follows:

$$P^t_{BESS} = P^t_{plan} - P^t_{WF} \tag{1}$$

2.3. Dispatching rules After interconnecting WFs to the power grid on a large scale, WF operators need to submit generation plans to the independent system operator, which means generation plans should be determined days-ahead or hours-ahead [22]. Similarly, the power transactions of renewable energy are expected to take place under the new institution after FIT in Japan. Under the new institution, WF operators must submit generation plans, which is the amount of energy generated in 30 min, to OCCTO, and operate WF power as specified in the generation plans. However, if WF operators cannot match WF power generation to the submitted generation plans, they must pay a penalty depending on the difference between the power generated and that specified in the generation plan. The energy imbalance is computed at intervals of 30 min. Under this institution, the available WF forecasts, which are 30-min averages of WF power over the last 24 h, are announced at intervals of 6 h.

In addition, WF operators must observe the GC to interconnect into the power grid. The GC is the maximum power ramp PR_{val} in a time interval ΔT [14,17]. In other words, adherence to the GC helps suppress ramp events: the combined power P_{total} (=WF power P_{WF} + BESS power P_{BESS}) remains within the GC. The definition of GC is as follows:

$$\max_{t'=t-\Delta T+1,\ldots,t} P^{t'}_{total} - \min_{t'=t-\Delta T+1,\ldots,t} P^{t'}_{total} \le PR_{val} \tag{2}$$

The maximum power ramp PR_{val} and the time interval ΔT are determined in each scenario. For instance, in Reference [14], ramp events are defined as 10% of the rated capacity over a 30-min

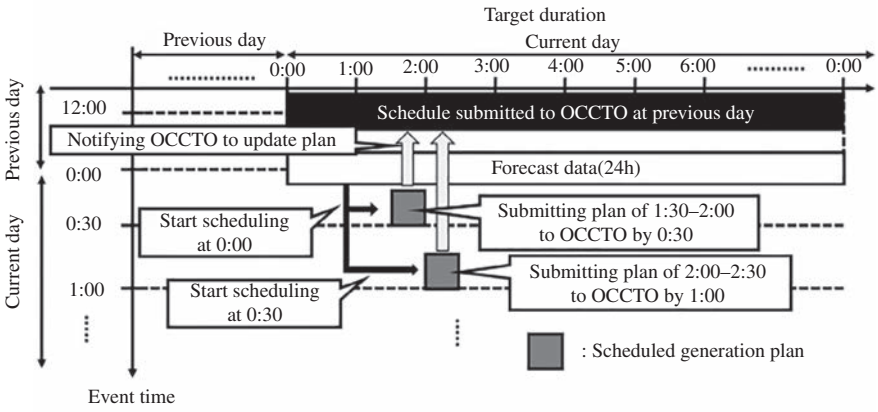


Fig. 2. Schedule for scheduled operation

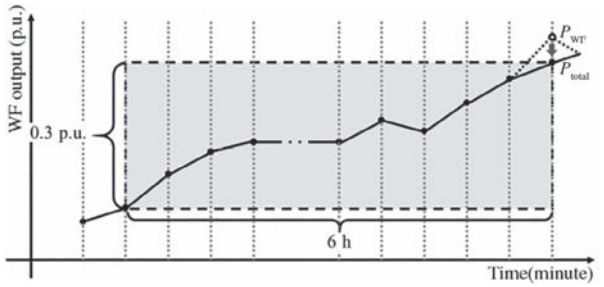


Fig. 3. Definition of grid code

time window, expressed as the rate of 0.1 p.u./30 min herein. In Reference [17], a ramp event is defined as a rate of 0.1 p.u./20 min assuming load frequency control and as a rate of 0.3 p.u./6 h assuming economic load dispatching control (EDC). In Reference [4], a ramp event is defined a change in power generation of greater than 20% of the installed wind capacity within a span of 4 h or less, that is, a rate of 0.2 p.u./4 h. In this paper, PR_{val} and ΔT are set as 0.3 p.u. and 6 h, respectively, assuming the correspondence in EDC with reference to [17] (shown in Fig. 3).

3. Machine Learning

3.1. Reinforcement learning Reinforcement learning is a method of machine learning in which the agent learns the optimal action by trial and error from the interactions with the environment. The mechanism of reinforcement learning is shown in Fig. 4. At first, the agent decides an action (a) based on the state (s) observed from the environment. Next, the environment rewards (r) the agent as a result of the action. Finally, the agent updates the Q value, value of selecting action (a) in the state (s), based on the reward. By repeating this series of flows, the agent searches for the optimal policy.

In this paper, we use profit sharing (PS) as the reinforcement learning algorithm. In PS, the Q value is stored in a tabular format as shown in an example in Fig. 4. Therefore, if continuous values are used for state and action, their discretization is required. The update of the Q value in the PS is performed as follows:

$$Q(s_i, a_i) \leftarrow Q(s_i, a_i) + C_{bid}[r(i) - Q(s_i, a_i)] \qquad (3)$$

Fig. 4. Reinforcement learning model

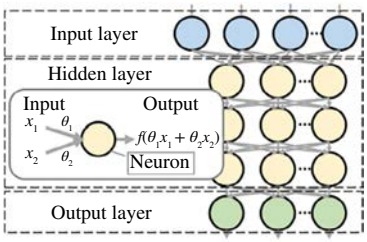


Fig. 5. Deep learning model

where, i $(0 < i < I)$ indicates the number of learning steps, I indicates the maximum value of learning step, $r(i)$ indicates a reward function, C_{bid} $(0 < C_{bid} < 1)$ indicates a selling value (Q value update ratio). In the PS, the search is performed until a reward is obtained from the environment, and then, the reward at each step is determined by the reward function, and the Q value is updated collectively.

3.2. Deep learning Deep learning is a technique using neural network (NN) that consists of a large-scale multiple layer of 'neurons' mimicking the human brain. As shown in Fig. 5, the neuron determines the output value by converting the sum of the input values x_i multiplied by the parameters θ_i using a non-linear function called the activation function f. NN is composed of an input layer, a hidden layer, and an output layer. When data is input to the NN, the outputs are sequentially determined from the input layer side to the output layer side. In deep learning, training data, mean collect answer data, is used for learning of NN together with input data. That is, the error of NN is calculated by comparing the training data and output data of NN. Therefore, the parameters in NN are updated so that the error becomes smaller.

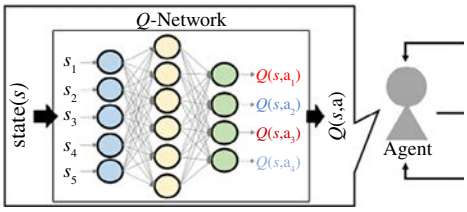


Fig. 6. Deep reinforcement learning model

3.3. Deep reinforcement learning Deep reinforcement learning is a type of reinforcement learning algorithm using deep learning. In deep reinforcement learning as shown in Fig. 6, the state is input to NN, and the Q value of each action is output. Then, a reward is applied to the training data, and a Q value updating algorithm is applied to the calculation of the error of NN for learning, thereby storing the Q value using the NN. Therefore, the discretization of input data is unnecessary, and continuous values can be used.

In this paper, we use deep Q-network with PS (DQN-PS) as the deep reinforcement learning algorithm which is proposed in Reference [29]. In the DQN-PS, a search is performed until a reward is obtained from the environment in the same way as the PS, and the reward at each step is determined by using the reward function. Then, the parameters of NN are updated based on the loss function as follows:

$$L_\theta = (r(t) - Q_\theta(s_t, a_t))^2 \tag{4}$$

Here, L_θ indicates the loss function, θ indicates the parameter of NN, Q indicates the Q value which is the result of output from the NN.

4. Proposed Method

4.1. Basic scheduling model In this section, the basic scheduling model is explained. The basic scheduling model has been proposed in Reference [16], whose purpose is to manage SOC while meeting GC for suppressing ramp fluctuations. The objective function and constraints are shown below.

[objective function]

$$\begin{aligned} Min \quad & \omega_1 \textstyle\sum_{t=1}^{N} \{|\mathrm{SOC}(t) - \mathrm{SOC}_{target}|\} \\ & + \omega_2 \textstyle\sum_{t=1}^{N} \{Z^{out}(t) + Z^{in}(t)\} \\ & + \omega_3 \textstyle\sum_{t=1}^{N} \{Z^{GC+}(t) + Z^{GC-}(t)\} \end{aligned}$$

[constraints]

$$\begin{aligned} & \frac{30}{60}\{P_{\mathrm{total}}(t) - P_{\mathrm{WF}}(t)\} \\ & = \frac{30}{60}\{P^{\mathrm{out}}_{\mathrm{BESS}}(t) - P^{\mathrm{in}}_{\mathrm{BESS}}(t) + Z^{\mathrm{out}}(t) - Z^{\mathrm{in}}(t)\} \end{aligned} \tag{6}$$

$$P^{\mathrm{in}}_{\mathrm{BESS}}(t) * P^{\mathrm{out}}_{\mathrm{BESS}}(t) = 0 \tag{7}$$

$$0 \le P^{\mathrm{in}}_{\mathrm{BESS}}(t) \le \mathrm{kW} \tag{8}$$

$$0 \le P^{\mathrm{out}}_{\mathrm{BESS}}(t) \le \mathrm{kW} \tag{9}$$

$$\mathrm{P_{tmax}}(t) = \max\{\mathrm{P_{total}}(t-1, t-2, \ldots, t-12), 0.3\} \tag{10}$$

$$\mathrm{P_{tmin}}(t) = \min\{\mathrm{P_{total}}(t-1, t-2, \ldots, t-12), 0.7\} \tag{11}$$

$$\mathrm{P_{tmax}}(t) - (0.3 + Z^{GC-}(t)) \le \mathrm{P_{total}}(t) \tag{12}$$

$$\mathrm{P_{total}}(t) \le \mathrm{P_{tmin}}(t) + (0.3 + Z^{\mathrm{GC+}}(t)) \tag{13}$$

$$\begin{aligned} & \mathrm{SOC}(t) = SOC(t-1) + \\ & \frac{30}{60}\left(\eta * P^{\mathrm{in}}_{\mathrm{BESS}}(t) - \frac{1}{\eta} P^{\mathrm{out}}_{\mathrm{BESS}}(t)\right) / \mathrm{kWh} \end{aligned} \tag{14}$$

$$P^{\mathrm{in}}_{\mathrm{BESS}}(t) \le M * \{1 - \mathrm{SOC}(t)\} \tag{15}$$

$$P^{\mathrm{out}}_{\mathrm{BESS}}(t) \le M * \mathrm{SOC}(t) \tag{16}$$

$$|P_{\mathrm{total}}(t-1) - P_{\mathrm{total}}(t)| \le 0.05 \tag{17}$$

$$Z^{\mathrm{out}}(t), Z^{\mathrm{in}}(t), Z^{\mathrm{GC+}}(t), Z^{\mathrm{GC-}}(t) \ge 0 \tag{18}$$

where,

N	the number of planning (20 in this paper)
t	time (30 min is 1 unit in this paper)
SOC	SOC of BESS
$\mathrm{SOC_{target}}$	the target of SOC
Z^{out}, Z^{in}	additional compensation power
$Z^{\mathrm{GC+}}$, $Z^{\mathrm{GC-}}$	GC mitigation amount
ω_1, ω_2, ω_3,	weight coefficients
P_{total}	the generation plan
P_{WF}	WFs power
$P^{\mathrm{out}}_{\mathrm{BESS}}$	discharge power of BESS on the system side of the PCS.
$P^{\mathrm{in}}_{\mathrm{BESS}}$	charge power of BESS on the system side of the PCS.
η	conversion efficiency
kW	power capacity of BESS
P_{tmax}	maximum supply output in the past 6 h
P_{tmin}	Minimum supply output in the past 6 h
kWh	the energy capacity of BESS
M	optimization parameter
P_{submit}	submitted plan

The first clause of the objective function (5) for managing SOC aims to reduce the gap between SOC(t) and SOC target $\mathrm{SOC_{target}}$, determined by deep reinforcement learning and will be described later in Section 4.3. The SOC value is divided by 100 to normalize SOC(t)$\in$ [0, 1]. The second one and third one is respectively for minimizing the amount of additional compensation power and the amount of GC mitigation. The purpose of this paper is to minimize deviations from the GC in the schedule operation of WF. Under the current system in Japan, imbalances are handled by price settlement, not by penalty. In other words, it is impossible to know the penalty unit price in advance. Therefore, in this paper, the purpose of the scheduled operation is to minimize the imbalance (p.u.-h). Therefore, we formulated to create a generation plan even if the storage capacity of BESS is insufficient or the GC cannot be achieved at the planning stage. Where, weight coefficients ω_2, ω_3 are usually given to be 0. The formulation (6) shows the

power balance between the combined power and the generation plan. The BESS cannot charge and discharge simultaneously. Thus, the BESS power constraint is described in (7). Moreover, since the BESS has rated output power, it is described as (8) and (9). $P^{\text{in}}_{\text{BESS}}$ and $P^{\text{out}}_{\text{BESS}}$ indicate the charge/discharge power on the system side of the PCS as shown in Fig. 1. kW indicates the power capacity on the AC side of the PCS. Observing the GC helps suppress ramp events. In other words, controlling the combined power P_{total} within the GC is necessary. The formulation of GC is expressed as (10)–(13). Given the energy capacity and the conversion efficiency, the energy state of the BESS can be calculated by (14). In addition, the BESS cannot be charged when $\text{SOC}(t) = 1$ or cannot be discharged when $\text{SOC}(t) = 0$. BESS power is restrained to avoid overcharging or over-discharging in (15) and (16). M is an optimization parameter and an adequately large constant. The formulation (17) is a constraint equation that suppresses short-period fluctuations caused by a shift between frames. This report prohibits the fluctuation between frames of 5% or more of the rated WF output. The real-time operation of WFs is conducted based on the 1.5-h-ahead generation plan $P_{\text{total}}(4)$. Note that the additional compensation amounts $Z^{\text{out}}(t)$, $Z^{\text{in}}(t)$, and the GC relaxation amount $Z^{\text{GC}+}(t)$, $Z^{\text{GC}-}(t)$ take non-negative values as shown in (18). The generation plan is determined by solving the above optimization calculation of the objective function (5) and the constraints from (6) to (18).

4.2. Correction of prediction error The generation plans requires several hours of prediction WF output, but the prediction information includes an error. In this paper, we correct the prediction error using deep learning. Specifically, the prediction error is corrected by using the predicted value and the past prediction error as the input data of NN and the measured value as the training data.

4.3. Proposed method In this paper, it is assumed that the environment is a power system operator such as OCCTO and the agent is a WF operator. The WF operator determines the target SOC, creates a power generation plan from the optimization calculation, and executes the scheduled operation by controlling BESS. In the proposed method, deep reinforcement learning is applied to the determination of the target SOC and the result of the scheduled operation is learned as a reward. Figure 7 shows the flow chart of the proposed method. First, the predicted WF output is corrected using deep learning from past predicted information and measured information. Next, the target SOC is determined by deep reinforcement learning using multiple information including the corrected predicted WF output. Finally, the generation plan is determined by the optimization calculation considering the target SOC and submitted. After that, based on the submitted plan, the control of BESS is performed, and it is confirmed whether the ramp fluctuation suppression has been achieved. If no ramp event occurs in the power supplied to the grid, the search is continued, and if it occurs, learning is performed according to an algorithm. By determining the target value of the optimization calculation by learning in this manner, it is possible to reduce the deviation due to insufficient capacity while guaranteeing a plan for achieving the GC.

4.4. Applying reinforcement learning This section describes the determination method of the target SOC using

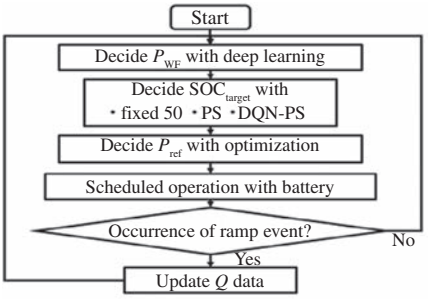


Fig. 7. The flow of scheduled operation

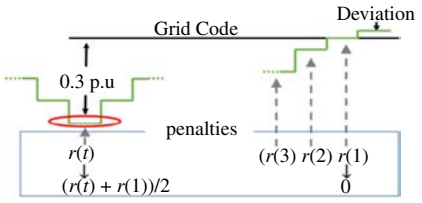


Fig. 8. Image of penalties

reinforcement learning. To apply reinforcement learning, it is necessary to define the state, action, and reward function respectively. For the state, the target SOC before the change, the predicted WF outputs, the current SOC, and the degree of the prediction error are used. For predicted WF outputs including trends for several hours, we will use Autoencoder, a method of deep learning in Reference [30], to extract its features and discretize them. As the degree of the prediction error, the value of the prediction error and the tendency of the prediction error are used. The action is the target SOC, and the reward function is calculated as follows:

$$r(t) = -10 * P * (t - T + 36)/36 \tag{19}$$

Here, P is a penalty that increases as the combined output deviates from the GC. As shown in Fig. 8, in the formulation (19), a penalty is determined according to the length of the departure time, and learning of an action to reduce the deviation is encouraged. In addition, since the time of the available predictive information at the time of determining the action is 18 h, a penalty is given to the action up to 18 h before. Furthermore, as shown in Fig. 3, the GC depends on the past generation plan, so penalties are also imposed on the action that caused the deviation to occur (in most cases, the action on the generation plan 6 h ago). On the other hand, actions, when the power generation plan is blocked by the GC, are exempt from penalties and prevent erroneous learning.

4.5. Applying deep reinforcement learning In this section, we describe a target SOC determination method that uses the proposed method, deep reinforcement learning. In deep reinforcement learning, it is necessary to define states, actions, reward functions, and loss functions. For the state, the SOC, the power generation plan, the WF output, and the prediction error for the past 24 h, the target SOC for the past 8 h and the predicted WF output for the next 18 h are used without discretization. The behavior and reward function are set the same as in Section 4.4, and (4) is used for the loss function.

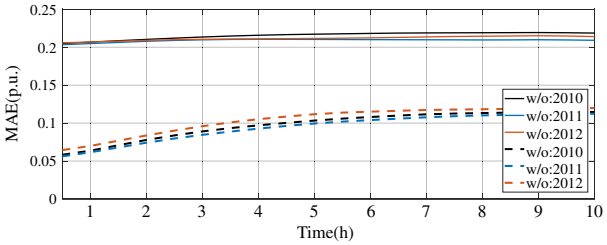


Fig. 9. The result of error correction

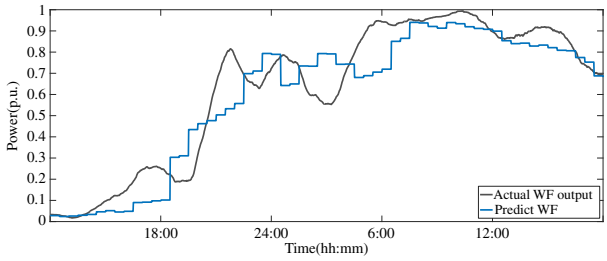


Fig. 10. Actual output and predicted output of WF

Table I. Number of deviations from grid code when the correction method is applied to the constant method

	Number of deviations (times)		
Method	2010	2011	2012
w/o correction	35 643	38 071	41 600
w/ correction	16 609	20 466	20 869

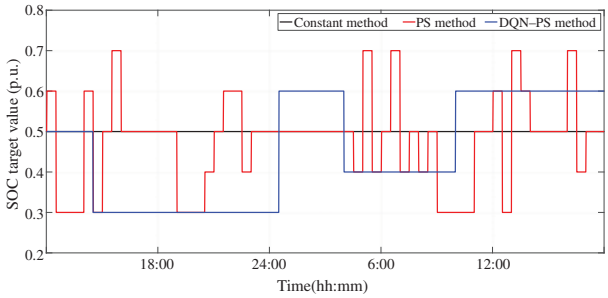


Fig. 11. SOC target value by each method

5. Case Studies

5.1. Conditions The WF outputs is calculated from the 1-min value data of the wind speed of the ground weather observation measured by Japan Meteorological Business Support Center and the power curve of the wind power generation system 'HTW2.0-86'. The predicted WF output is calculated from the hourly wind speed value announced by the GPV weather forecast. In addition, we used wind speed data from 2009 to 2012 at Minamichita and Irako sites in Aichi Prefecture. In this paper, we examined the small-scale WF with two wind power generators. We set the rated capacity of WF to 4 MW. Because the small-scale WF is greatly affected by the GC due to the magnitude of output fluctuation. Therefore, all power variables were normalized by 4 MW. The power capacity of BESS is 0.4 p.u. and the storage capacity of BESS is 10 h. The correction of prediction error is learned based on the 2009 data, and is applied to the 2010 and subsequent prediction information and evaluated. For learning in reinforcement learning and deep reinforcement learning, training data are used for 2010 and 2011, and the learning results are evaluated by applying them to 2012 data. The number of trails of learning in 1 year is 1800 for both reinforcement learning and deep reinforcement learning. In addition, the conventional method to be compared is a method in which the target value is fixed at 0.5 p.u. Moreover, real-time WF operation is conducted without control delay of the BESS—response delay of charge and discharge, measurement lag, and communication lag in the OWF phase—and the BESS is used to compensate for WF power directly, as expressed in (2).

5.2. Scheduled operation Figure 9 shows the results of the annual mean absolute error (MAE) for each forecast time. It can be seen that the MAE value was reduced all year and all-time by correcting the forecast information. In particular, after the correction, the MAE value is reduced when the prediction is close.

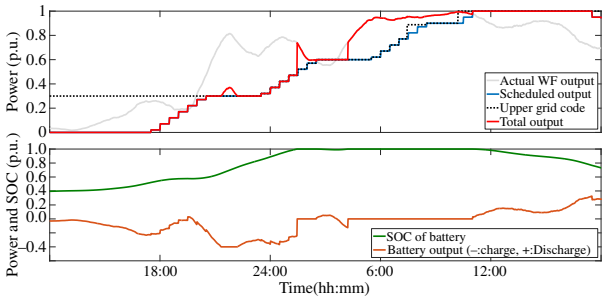
Table I shows the results of the scheduled operation using the conventional constant method. The correction of the predicted WF output reduces the number of deviations throughout the year.

Figure 10 shows the wind power output and WF output forecast during the time period when the proposed method is evaluated. As shown in Fig. 10, we examined the ramp-up in which the WF output rises sharply as a case study.

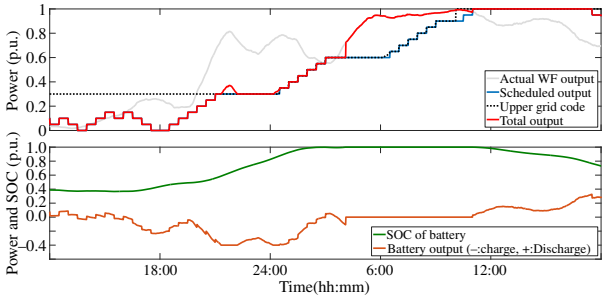
Figure 11 shows the SOC target values set by each method. As shown in Fig. 11, in the PS method, the SOC target value is frequently changed due to the influence of the discretization of variables for applying reinforcement learning. On the other hand, the state can be considered as a continuous value in the DQN-PS method. Therefore, frequent target value changes have been resolved.

Figure 12 shows the result of the scheduled operation on March 19, 2012, as an example. It should be noted that the form of the GC in Fig. 12(a)–(c) are different because the GC changes according to the past WF output shown in Fig. 3. In the constant method shown in Fig. 12(a), a generation plan was made to charge the storage of BESS greatly until 18:00 to bring the SOC closer to the target value of 0.5 p.u. As a result, the storage of BESS became fully charged from around 2:00, and the combined output deviated from GC until 11:00. On the other hand, in the method using PS in Fig. 12(b), by lowering the target SOC at around 15:00, it was possible to respond to the ramp fluctuations from 18:00 with less SOC than the constant method, so the time to full charge was reduced by about 1 h. Furthermore, in the method using DQN-PS shown in Fig. 12(c), by setting the target SOC to 0.3 p.u. for a longer time than the PS method, the generation plan is increasing around 18:00 with the increase in WF. As a result, it can be seen that full charge was avoided, and combined output is as planned without any deviation during all hours except for deviation due to insufficient power capacity of BESS.

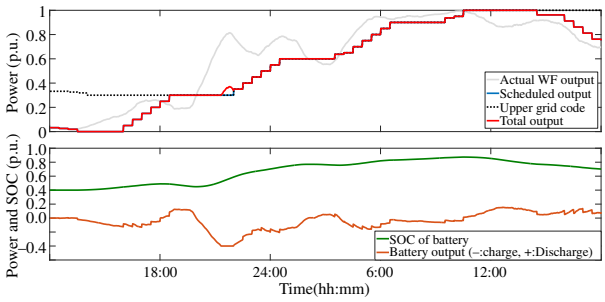
Table II shows the results of scheduled operation in 2012 when learning and changing the target value. Compared with the constant method, it can be confirmed that the number of deviations in PS is worse. This seems to be because PS, select the usage information to



(a)



(b)



(c)

Fig. 12. The result of the scheduled operation: (a) Constant method (b) PS method (c) DQN-PS method

discretize the input information and avoid an increase in memory, is lacking the necessary information for learning, and the number of cases where PS is worse than constant methods is increasing. On the other hand, it can be seen that DQN-PS can reduce the number of deviations compared to the constant method. In addition, the effectiveness of the proposed method can be confirmed from the viewpoint of imbalance, which is the power balance difference between the generation plan and the combined output. This is because the input information can be used as a continuous value as a state, so that judgment can be made based on more information, and because it can be easily applied to unlearned data, so that it learns the relationship between the state and the Q value instead of learning the Q value itself. Moreover, it can be seen that the effect of reducing the number of deviations is enhanced by learning not only in 1 year but also in multiple years for both PS and DQN-PS. Therefore, it is thought that learning more about various states leads to the improvement of the ability to respond to unknown data.

Table II. The number of deviations and imbalance from the grid code in each method in 2012

Method		Number of deviations (times)	Imbalance (p.u.-h)
Constant method		20 869	80.48
PS	2011	22 670	88.87
	2010 + 2011	22 652	83.37
DQN-PS	2011	15 430	64.94
	2010 + 2011	14 894	59.52

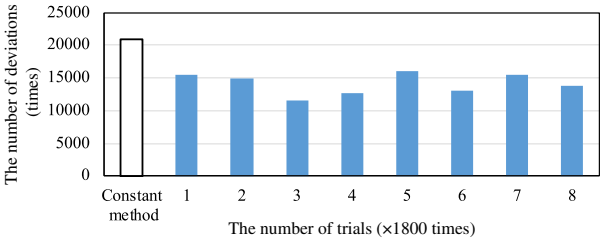


Fig. 13. The number of deviations from the grid code for the number of trials using DQN-PS method

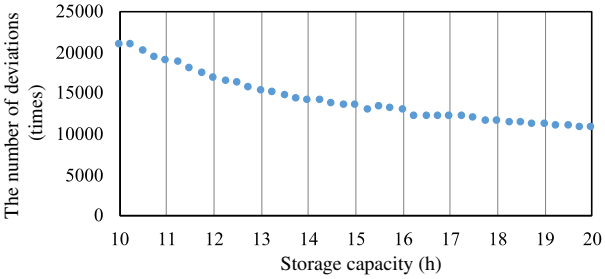


Fig. 14. The number of deviations from the grid code for the storage capacity of BESS when constant method is applied

Figure 13 shows the number of deviations when the number of trials in learning is increased in DQN-PS. It is not possible to confirm cases that exceed the constant method, but it can be seen that the number of deviations repeatedly changes with an increase in the number of trials. This seems to be caused by over-learning caused by repeated learning. At present, it seems very difficult to determine the optimal number of trials and this is an issue for the future.

Figure 14 shows the number of deviations for a year when the storage capacity of BESS is changed in the constant method. As shown in Fig. 14, the number of deviations decreases as the storage battery capacity increases. This is because the power deviating from the GC can be charged and discharged more by increasing the storage battery capacity.

Table III shows the required storage capacity of BESS in the DQN-PS method and the constant method. In Table III, the required storage capacity of BESS is evaluated based on the maximum value, average value, and minimum value of the number of deviations in the DQN-PS method shown in Fig. 13. As shown in Table III, the required storage capacity of BESS for the Constant method were 12.5, 14.25, and 17.75 h, respectively.

Table III. The result of reducing storage capacity

Deviation of DQN-PS		The required capacity of the constant method (h)	The required capacity of the DQN-PS method (h)	BESS reduction effect (h)
Maximum	16 021	12.50	10.00	2.50
Average	14 097	14.25	10.00	4.25
Minimum	11 494	17.75	10.00	7.75

On the other hand, it for the DQN-PS method are all 10 h as shown in Section 5.1. These results show that the Constant method requires the storage capacity of BESS of approximately 2.5–7.75 h to achieve the same result as the proposed DQN-PS method. It can be confirmed that the storage capacity of BESS can be reduced by the DQN-PS method.

6. Conclusion

A machine learning method for scheduled WFs operation to control ramp events is proposed in this paper. By correcting the predicted information using deep leaning and deciding a target SOC using deep reinforcement learning, the scheduled operation for controlling ramp events is achieved. In addition, by using multiple information for learning, the effectiveness of the proposed method is improved. A comparison of the conventional method, reinforcement learning method, and deep reinforcement learning method through numerical simulation highlighted the effectiveness of the proposed scheduling model. Finally, the required storage capacity's reduction of the proposed method can be expected. Future tasks include studying how to judge the appropriate number of learning trials, confirming the effects of rationality and timing, and comparing it with other optimization methods.

References

(1) Kabouris J, Kanellos FD. Impacts of large-scale wind penetration on designing and operation of electric power system. *IEEE Transactions on Sustainable Energy* 2010; **1**(2):107–114.

(2) Global Wind Energy Council, *Global wind report 2016 – Annual market update*. Accessed at May 2, 2017 from http://files.gwec.net/files/GWR2016.pdf

(3) Ferreira C, Gama J, Matias L, Botterud A, Wang J. *A survey on wind power RAMP forecasting, Argonne National lab. (ANL), DuPage Country*, IL, USA, Tech. Rep. ANL/DIS-10-13, December 2010.

(4) Cui M, Zhang J, Florita AR, Hodge BM, Ke D, Sun Y. An optimized swinging door algorithm for identifying wind ramping events. *IEEE Transactions on Sustainable Energy* 2016; **7**(1):150–162.

(5) Sorensen P, Cutululis NA, Vigueras-Rodriguez A, Jensen LE, Hjerrild J, Donovan MH, Madsen H. Power fluctuations from large wind farms. *IEEE Transactions on Power Apparatus and Systems* 2007; **22**(3):958–965.

(6) Sevlian R, Rajagopal R. Detection and statistics of wind power ramps. *IEEE Transactions on Power Apparatus and Systems* 2013; **28**(4):3610–3620.

(7) Ganger D, Zhang J, Vittal V. Statistical characterization of wind power ramps via extreme value analysis. *IEEE Transactions on Power Apparatus and Systems* 2014; **29**(6):3118–3119.

(8) Ogimoto K. Generation forecast of renewable energy and system technology. *IEEJ Transactions on Power and Energy* 2014; **134**(6):473–476 (in Japanese).

(9) Ikegami T, Kataoka K, Ogimoto K, Saitou T. Development of wind power data for power supply-demand analysis and analysis of long-term wind power variability. *IEEJ Transactions on Power and Energy* 2014; **134**(3):236–247 (in Japanese).

(10) Yoshida K, Hayashi N, Ohba M, Nohara D, Ogasawara N, Okaba M, Mori Y, Ogimoto K, Kataoka K. Analysis of meteorogical factors of wind power ramps in Hokkaido and Tohoku area. *IEEJ Transactions on Power and Energy* 2017; **137**(1):71–78 (in Japanese).

(11) Ohba M, Kadokura S, Nohara D. Impacts of synoptic circulation patterns on wind power ramp events in East Japan. *Renewable Energy* 2016; **96**:591–602.

(12) Cui M, Ke D, Sun Y, Gan D, Zhang J, Hodge BM. Wind power ramp event forecasting using a stochastic scenario generation method. *IEEE Transactions on Sustainable Energy* 2015; **6**(2):422–433.

(13) Qi Y, Liu Y. Wind power ramping control using competitive game. *IEEE Transactions on Sustainable Energy* 2016; **7**(4):1516–1524.

(14) Gong Y, Jiang Q, Baldick R. Ramp event forecast based wind power ramp control with energy storage system. *IEEE Transactions on Power Apparatus and Systems* 2016; **31**(3):1831–1844.

(15) Lee D, Kim J, Baldick R. Stochastic optimal control of the storage system to limit ramp rates of wind power output. *IEEE Transactions on Smart Grid* 2013; **4**(4):2256–2265.

(16) Yoshida K, Negishi S, Takayama S, Ishigame A. Scheduled operation of wind farm with battery energy storage system considering ramp events. *Proc. 23rd Int. Conf. On Elect. Eng*., Weihai, 2017; 143–148.

(17) Ito M, Fujimoto Y, Mitsuoka M, Ishii H, Hayashi Y. Control methods for an energy storage system when wind power output deviates from grid code. *Journal of International Council on Electrical Engineering* 2017; **7**(1):159–165.

(18) Hamamoto A, Hara R, Kita H. Study on suppression of wind farms outputs fluctuation by a HP/BG heat supply system. *IEEJ Transactions on Power and Energy* 2017; **137**(6):446–452 (in Japanese).

(19) Jiang QY, Gong YZ, Wang HJ. A battery energy storage system dual-layer control strategy for mitigating wind farm fluctuation. *IEEE Transactions on Power Apparatus and Systems* 2013; **28**(3):3263–3273.

(20) Teleke S, Baran ME, Huang AQ, Bhattacharya S, Anderson L. Control strategies for battery energy storage for wind farm dispatching. *IEEE Transactions on Energy Conversion* 2009; **24**(3):725–732.

(21) Zhang F, Meng K, Xu Z, Dong Z, Zhang L, Wan C, Liang J. Battery ESS planning for wind smoothing via variable-interval reference modulation and self-adaptive SOC control strategy. *IEEE Transactions on Sustainable Energy* 2017; **8**(2):695–707.

(22) Luo F, Meng K, Dong ZY, Zhen Y, Chen Y, Po Wong K. Coordinated operational planning for wind farm with battery energy storage system. *IEEE Transactions on Sustainable Energy* 2015; **6**(1):253–262.

(23) Li Q, Choi SS, Yuan Y, Yao DL. On the determination of battery energy storage capacity and short-term power dispatch of a wind farm. *IEEE Transactions on Sustainable Energy* 2011; **2**(2):148–158.

(24) Mnih V, Kavukcuoglu K, Silver D, Graves A, Antonoglou I, Wierstra D, Riedmiller M. Playing Atari with deep reinforcement learning, *NIPS Deep Learning Workshop*. 2013

(25) Sutton RS, Barto AG. *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press; 2018.

(26) Dahl GE, Yu D, Deng L, Acero A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 2012; **20**(1):30–42.

(27) Ackermann T. Basic integration issues related to wind power. In *Wind Power in Power Systems*. John Wiley & Sons: Chichester, England; 2005; 40–46.

(28) Takayama S, Hara R, Kita Y, Ito T, Ueda Y, Saito Y, Takitani K, Yamaguchi K. Scheduled operation of PV power station considering solar radiation forecast error. *IEEJ Transactions on Power and Energy* 2011; **131**(3):304–312 (in Japanese).

(29) Miyazaki K. Proposal of a deep Q-network with profit sharing. *Procedia Computer Science* 2018; **123**:302–307.
(30) Hinton GE, Osindero S, Teh Y. A fast learning algorithm for deep belief nets. *Neural Computation* 2006; **18**(7):1527–1554.

Mamoru Futakuchi (Member) received the master's degree in engineering from Osaka Prefecture University, Osaka, Japan, in 2020. He worked on the design of energy management systems integrated with wind power generation system. In 2020, he joined Toshiba Corporation, Tokyo, Japan, where he is currently involved in the research and development of power electronics products. Mr. Futakuchi is also a member of the Institute of Electrical Engineers of Japan (IEEJ).

Satoshi Takayama (Member) received the Ph.D. degree in electrical engineering from Hokkaido University, Hokkaido, Japan, in 2011. Since 2011, he has been a Faculty with the Department of Electrical Engineering, Osaka Prefecture University, where he is currently a Senior Lecturer. His research interests include power and energy systems, system optimization, and machine learning applications. Dr. Takayama is also a member of the Institute of Electrical Engineers of Japan and the Institute of Electrical Installation Engineers of Japan, and so on.

Atsushi Ishigame (Senior Member) received the Ph.D. degree in electrical engineering from Osaka Prefecture University, Osaka, Japan, in 1993. Since 1993, he has been a Faculty with the Department of Electrical Engineering, Osaka Prefecture University, where he is currently a Professor. From 1995 to 1996, he was a Visiting Scholar with the School of Electrical Engineering, Cornell University. His research interests include power and energy systems, system optimization, and control applications. Dr. Ishigame is also a member of the Institute of Electrical Engineers of Japan and the Institute of Electrical Installation Engineers of Japan.