Alerting to Rare Large-Scale Ramp Events in Wind Power Generation

Yu Fujimoto, *Member, IEEE*, Yuka Takahashi, and Yasuhiro Hayashi, *Member, IEEE*

***Abstract*—Wind power is an unstable power source, as its output fluctuates drastically according to the weather. Such instability can cause sudden large-scale changes in output, called ramp events; the frequency of such events is relatively low throughout the year but they could negatively affect the supply–demand balance in a power system. This study focuses on an alerting scheme of wind power ramp events for a transmission system operator to support operational decisions on cold reserve power plants. The ramp alerting scheme is implemented from the viewpoint of supervised learning by using the prediction results of wind power output. In particular, the authors address the class imbalance problem, as the accuracy of ramp event prediction tends to be low because of the infrequency of such ramp events in the database used for learning. In this study, several data sampling strategies are proposed and implemented to overcome the problem in the ramp alert task. The effectiveness of the proposed data sampling framework is evaluated experimentally by predicting real-world wind power ramps, based on a dataset collected in Japan. The experimental results show that the proposed framework effectively improves the ramp alert accuracy by addressing the class imbalance problem.**

***Index Terms*—Class imbalance problem, machine learning, prediction, wind power ramp.**

NOTATION

Symbols

$\mathcal{B} = (\lvert\mathcal{B}\rvert - 1, \ldots, 0)$	A set of indices for the observations utilized in prediction.
$\mathcal{D}$	Data set.
$\boldsymbol{\varepsilon}_t = \hat{p}_t^{\mathcal{H}} - p_t^{\mathcal{H}*}$	Error sequence of wind power output forecasted at the time slice t.
h	Time slice of forecast horizon (30 min time resolution).
$\mathcal{H} = \{1, 2, \ldots, \lvert\mathcal{H}\rvert\}$	A set of forecast horizons (30 min time resolution).
$\langle\mathcal{H}\rangle = \{1, 2, \ldots, \lfloor\frac{\lvert\mathcal{H}\rvert}{2}\rfloor\}$	A set of forecast horizons (1 h time resolution).
$p_t \in [0, 1]$	Wind power output ([pu.h/30 min]) at the time slice t.
$\boldsymbol{p}_t^{\mathcal{H}} = (p_{t+1}, \ldots, p_{t+\lvert\mathcal{H}\rvert})$	Sequence of wind power outputs at the time slice t for forecast horizons in $\mathcal{H}$.
$\boldsymbol{p}_{\langle t\rangle}^{\langle\mathcal{H}\rangle} = (p_{\langle t+1\rangle}, \ldots, p_{\langle t+\lvert\mathcal{H}\rvert\rangle})$	Sequence of wind power outputs at the time slice $\langle t\rangle$ for hourly forecast horizons in $\langle\mathcal{H}\rangle$.
$\boldsymbol{p}_t^{\mathcal{B}*} = (p_{t-\lvert\mathcal{B}\rvert+1}^*, \ldots, p_t^*)$	Sequence of observed wind power outputs at the time slice t for the period $\mathcal{B}$.
$\Delta\boldsymbol{p}_t = (p_{t-\lvert\mathcal{B}\rvert+2}^* - p_{t-\lvert\mathcal{B}\rvert+1}^*, \ldots, \hat{p}_{t+\lvert\mathcal{H}\rvert} - \hat{p}_{t-\lvert\mathcal{H}\rvert-1})$	Sequence of the first-order difference in wind power outputs $(\boldsymbol{p}_t^{\mathcal{B}}, \boldsymbol{p}_t^{\mathcal{H}})$.
$R \in [0, 1]$	Ratio parameter for sampling.
$r_t^u, r_t^d \in \{0, 1\}$	Ramp up/down event occurred at the time slice t.
$\boldsymbol{r}_t^u, \boldsymbol{r}_t^d$	Sequence of ramp up/down events for forecast horizon $\mathcal{H}$; e.g., $\boldsymbol{r}_t^u = (r_{t+1}^u, \ldots, r_{t+\lvert\mathcal{H}\rvert}^u)$.
$t \in \{1, \ldots, T\}$	Time slice index (30 min time resolution).
$\mathcal{S}$	Data subset of $\mathcal{D}$.
$\langle t\rangle = \lfloor\frac{t}{2}\rfloor$	Time slice index (1h time resolution); hourly time index corresponding to t.
$\boldsymbol{x}_{\langle t\rangle} \in \mathbb{R}^P$	P-dimensional meteorological field forecasting result for the hourly time slice $\langle t\rangle$ used for power prediction.
$\psi(\cdot)$	Regressor for wind power outputs.
$\phi(\cdot)$	Classifier for wind power ramp events.
$\lfloor\cdot\rfloor$	Floor function.
$\lvert z\rvert$	Absolute value of a scalar z.
$\lvert\mathcal{Z}\rvert$	Cardinality of a set $\mathcal{Z}$.

Subscripts and Superscripts

$\cdot_t$	At the time slice t (30 min time resolution).
$\cdot_{t+h}$	h-ahead result at the time slice t (30 min time resolution).
$\cdot_{\langle t\rangle}$	At the time slice $\langle t\rangle$ (1 h time resolution); i.e., hourly value corresponding to the time slice t.
$\cdot^*$	Observed value.
$\bar{\cdot}$	Rough prediction result.
$\hat{\cdot}$	Predicted value.

Manuscript received August 10, 2017; revised February 23, 2018; accepted March 23, 2018. Date of publication April 4, 2018; date of current version December 14, 2018. Paper no. TSTE-00748-2017. *(Corresponding author: Yu Fujimoto.)*

Y. Fujimoto is with the Advanced Collaborative Research Organization for Smart Society, Waseda University, Tokyo 169-8555, Japan (e-mail: y.fujimoto@aoni.waseda.jp).

Y. Takahashi and Y. Hayashi are with the Department of Electrical Engineering and Bioscience, Waseda University, Tokyo 169-8555, Japan (e-mail: 01yukayukayuka20@ruri.waseda.jp; hayashi@waseda.jp).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSTE.2018.2822807

I. Introduction

WIND power generation systems have been introduced proactively in Japan [1], and are expected to contribute to improving the energy self-sufficiency rate, prevent global warming, and accelerate economic growth in related industries. However, sudden and unexpected fluctuations in wind power generation (called ramp events) [2] could have a significant effect on the supply-demand balance in a power system under a high amount of penetration. To mitigate such effects, prediction technology is a valuable tool for planning operation of alternative generators [3], [4] or energy storage devices [5], [6], and decision making in a market [7], [8]. For example, if a sudden large-scale decline in wind power generation could be predicted in advance, the transmission system operators could start up cold reserve power plants, such as aged thermal power generators that are not usually in operation. Note that the time required for the cold starting up of such power plants is generally several hours. Accordingly, ideally, such fluctuations would need to be predicted in good time to enable alternative generators being started up in time.

The development of methodology for predicting wind power output is considered an important research theme, not only for transmission system operators but also for wind power producers. In most initial research on this subject, such prediction has aimed to reduce the expected errors in the derived wind power output. Therefore, the performance has been evaluated from the viewpoint of the root mean squared error (RMSE) or mean absolute error (MAE) of the predicted wind power [9]. However, the relatively low average errors in the predicted wind power sequence does not always facilitate the accurate prediction of the timing of large-scale ramp events [10], [11]. For example, Hou [12] has shown that the RMSE between predicted power sequence and the observed sequence is composed of three components; i.e., the bias, the difference between standard deviations, and the dispersion between two sequences. Here, the dispersion intuitively implies the difference of the fluctuation timing corresponding to so-called phase errors [13]. This decomposition implies that the reduction in RMSE does not always indicate an improvement in prediction of the fluctuation timing directly. Accordingly, in recent years, various researchers have investigated this ramp prediction challenge (see [2], [14], [15]).

Kamath [16], for example, has analyzed historical wind power data to organize the characteristics of the ramp events in relation to time (duration) and intensity. Utilizing such historical data sets plays an important role in ramp event prediction; therefore various data driven frameworks have been studied [17]–[24]. Several studies have suggested that the application of machine learning methods [25], like auto regressive model [21], support vector machine [17], [18], hidden Markov model [19], random forest [17], [24], gradient boosted trees [23], wavelet transform [21], [26], and artificial neural networks [17], [21], [22], [27], particularly containing deep architectures [28]–[30], contributes to improving the accuracy of ramp event prediction. Another important point of view for predicting wind ramp events is utilizing the numerical weather prediction (NWP) model [31]–[35]. In relevant studies, the results of the NWP model have been used for the derivation of expected wind power output and predicting impending ramp events. Therefore, further improvements to the prediction accuracy have been attempted by adopting the data driven scheme, while utilizing the NWP results [19], [35].

The above-mentioned ramp prediction schemes can be classified roughly into two categories, namely, the direct scanning scheme and the separate prediction scheme. In the former scheme, the impending ramp events are identified based on the definition of a targeted ramp by directly scanning a time series subsequence of the predicted wind power output. A problem of the direct approach is that the fluctuation in the predicted wind power sequence naturally contains the errors in the predicted magnitude and the time of occurrence. For example, wind power generators basically output electric power within the range from zero to the rated power, so that the predicted results focusing on reduction of expected errors could have a tendency to be not extremely large or extremely small to avoid significantly large errors. Consequently, the fluctuation in the predicted subsequence could be smaller than the fluctuation in the actual output sequence. This tendency has been reported in the context of general wind power prediction scheme [36], [37]; for instance, the annual distributions of the ramp magnitude derived from actual wind power and various prediction results reported in [37] suggest this property. On the other hand, in the separate prediction scheme, the ramp events are predicted separately from the predicted wind power sequence; this scheme does not swallow the derived power generation sequence containing prediction errors and predicts the timing of large-scale fluctuations separately. The separate scheme is expected to facilitate early warning of the possible risk for the occurrence of large-scale ramp events such that the predicted wind power output sequences are hard to suggest.

In this study, we focused on the ramp alert task from the viewpoint of the separate prediction scheme. We proposed a framework for warning of ramp events, based on the supervised classifiers [25] that were learned from the predicted wind power sequence, derived by using the NWP results. This approach provided a simple but effective way for warning of the risk of a ramp occurrence. In particular, we addressed challenges, such as the ramp prediction accuracy, based on classification algorithms generally tending to be low, owing to the infrequency of such ramp events [24]. This tendency is known as the class imbalance problem [38]–[40] in the machine learning domain. To the best of our knowledge, the class imbalance problem in the ramp alert task has not been emphasized well, so that most of the previous studies have not explicitly addressed an intrinsic difficulty in predicting rare wind power ramp events.[1] In this study, we adopted several data undersampling/oversampling approaches [41], [42] to deal with the class imbalance problem relevant to the ramp alert task and propose a novel oversampling approach called the *error bootstrap oversampling* (*EBOS*)

[1]In recent years, Gupta *et al.* [23] has stated the difficulty of predicting rare ramp events, but no direct countermeasures have been taken. Our early work [24] also has reported the difficulty, though only a simple data-sampling strategy has been investigated.

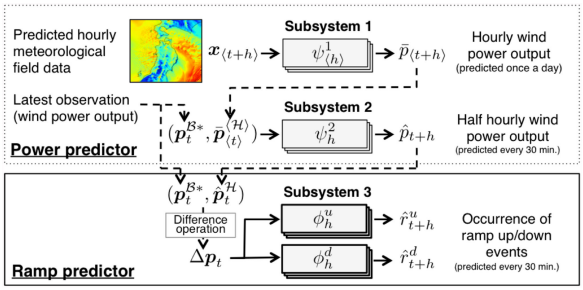


Fig. 1. Schematic image of ramp alert system. The system comprises the *power predictor* and the *ramp predictor*. In the power predictor, the half-hourly wind power output sequence, $\hat{p}_t^{\mathcal{H}} = (\hat{p}_{t+1}, \ldots, \hat{p}_{t+|\mathcal{H}|})$, is derived by using the predicted hourly meteorological field data, $\{x_{\langle t+h \rangle}\}$, and the latest wind power output information, $p_t^{\mathcal{B}*}$. In the ramp predictor, the occurrence of the ramp up/down events, $\hat{r}_{t+h}^{u}$ and $\hat{r}_{t+h}^{d}$, are predicted based on the power prediction results, $\hat{p}_t^{\mathcal{H}}$.

focusing on errors in wind power prediction results used for the ramp event prediction. We also introduce the hybridization [40] of these approaches and evaluate the data sampling approaches experimentally by comparing them with the existing approaches. The contributions made by this study are the following; (i) we explicitly discuss the class imbalance problem in a wind power ramp alert framework based on the separate prediction scheme, (ii) we propose a novel data sampling strategy *EBOS* for alleviating the class imbalance problem in learning ramp predictors, and (iii) we give evaluation results of the proposed ramp alert frameworks by comparison with the direct scanning scheme, based on real-world wind power data.

The rest of the paper is organized as follows. In Section II, we introduce formally the ramp alert task and explain our proposed ramp alert framework, based on wind power output prediction results. In Section III, we focus on the class imbalance problem in the ramp alert task and implement several data sampling approaches to address the problem. Our proposed approaches are compared with several existing methods related to ramp prediction accuracy by using the real-world data set in Section IV. Finally, in Section V, we provide our concluding remarks.

II. Ramp Alert Based on Separate Prediction Scheme Using Wind Power Prediction

A. Ramp Alert Framework

In this study, we focused on large-scale ramp events in wind power output; the occurrence frequency of such events is low throughout the year but they could affect the electric power system. In particular, we aimed to provide auxiliary information for system operations by predicting the occurrence of the ramp events, in addition to the prediction of ordinary wind power, which is called the ramp alert task in this study.

We introduced a hierarchical ramp alert scheme comprising three subsystems. Fig. 1 shows a schematic image of our proposed framework. In the first subsystem, ψ^1, the rough prediction of the hourly wind power output sequence was derived once daily by using the predicted hourly meteorological field data provided by the NWP model once a day; this subsystem outputs the expected wind power generation for each time slice corresponding to the snapshot of the hourly weather forecast. In the second subsystem, ψ^2, the sequence of half-hourly wind power output was predicted every 30 minutes by utilizing the latest wind power observations and the output of the first subsystem. A set of these two subsystems, ψ^1 and ψ^2, were called the *power predictors*, in this study. The third subsystem, ϕ, predicted the occurrence of the ramp up/down events for the target prediction horizons, based on the wind power output sequence derived from the power predictor. This subsystem, ϕ, was called the *ramp predictor*. Our procedure adjusted the prediction results every 30 minutes by utilizing the latest observations effectively, while using the predicted meteorological field data. Such data are useful for relatively long-term prediction but cannot be updated frequently. In the following subsections, we introduce briefly these subsystems for alerting to ramp events.

B. Rough Prediction of Hourly Wind Power Output

Let t be the index for indicating a time slice of 30 minutes and let $\langle t \rangle = \lfloor \frac{t}{2} \rfloor$ be the corresponding hourly time slice; e.g., if $t = 0$ implies 00:00 in a certain day, then epoch time $t = 5$ implies 02:30, and its corresponding $\langle t \rangle$ implies 02:00. In the first subsystem, the rough estimation result of the one-hour wind power output, $p_{\langle t+h \rangle}$, was predicted based on the predicted hourly meteorological field data, $\boldsymbol{x}_{\langle t+h \rangle}$, where t indicates the prediction timing and $h \in \mathcal{H}$ indicated the prediction horizon. We used the data set $\mathcal{D}_1 = \{(\boldsymbol{x}_{\langle t+h \rangle}, p^*_{\langle t+h \rangle})\}$, composed of pairs of $\boldsymbol{x}_{\langle t+h \rangle}$ and the observed hourly power output, $p^*_{\langle t+h \rangle}$, for learning the first subsystem; i.e.,

$$\psi^1_{\langle h \rangle} : \boldsymbol{x}_{\langle t+h \rangle} \mapsto \bar{p}_{\langle t+h \rangle} \quad \forall \langle h \rangle \in \langle \mathcal{H} \rangle, \tag{1}$$

where $\bar{p}_{\langle t+h \rangle}$ is the prediction result of $p_{\langle t+h \rangle}$. The model shown in (1) was independently learned for each prediction horizon, $\langle h \rangle \in \langle \mathcal{H} \rangle$, so that the hourly sequence $\bar{\boldsymbol{p}}_{\langle t \rangle}^{\langle \mathcal{H} \rangle} = (\bar{p}_{\langle t+1 \rangle}, \ldots, \bar{p}_{\langle t+|\mathcal{H}| \rangle})$ was obtained by concatenating the prediction results.

In our implementation, we used a popular nonlinear regression model, the so-called random forests (RF) [43],[2] which is known to work well, particularly when the number of explanatory variables for the prediction of $\bar{p}_{\langle t+h \rangle}$ is extremely large. At the learning step, the RF model $\psi^1(\cdot)$ was constructed from the viewpoint of minimizing the RMSE between the predicted and observed hourly wind power outputs in the given data set $\mathcal{D}_1$.

C. Prediction of Half-Hourly Wind Power Output Sequence

The second subsystem, ψ^2, aimed to obtain the estimation result of the half-hourly wind power output sequence, $\hat{\boldsymbol{p}}_t^{\mathcal{H}} = (\hat{p}_{t+1}, \ldots, \hat{p}_{t+|\mathcal{H}|})$, by using the data set, $\mathcal{D}_2 = \{(\boldsymbol{p}_t^{\mathcal{B}*}, \bar{\boldsymbol{p}}_{\langle t \rangle}^{\langle \mathcal{H} \rangle}, \boldsymbol{p}_t^{\mathcal{H}*})\}$, as follows:

$$\psi^2_h : (\boldsymbol{p}_t^{\mathcal{B}*}, \bar{\boldsymbol{p}}_{\langle t \rangle}^{\langle \mathcal{H} \rangle}) \mapsto \hat{p}_{t+h} \quad \forall h \in \mathcal{H}, \tag{2}$$

[2]The RF implemented in this study is basically the same as the learning architecture implemented in [17], [24]. A brief description of the mechanism is described in Appendix.

where $\boldsymbol{p}_t^{\mathcal{B}*} = (p^*_{t-|\mathcal{B}|+1}, \ldots, p^*_t)$ is the latest observed wind power output subsequence. In this subsystem, the half-hourly wind power output of the forecast horizon, h, i.e., p_{t+h}, was predicted based on the latest observation data sequence, $\boldsymbol{p}_t^{\mathcal{B}*}$, and the hourly wind power prediction results, $\bar{\boldsymbol{p}}^{\langle\mathcal{H}\rangle}_{\langle t\rangle}$. The prediction task given in (2) was modeled by the RF regressor. The aim of this step was to achieve the update of the wind power forecast result every 30 minutes. This was achieved by utilizing the latest observed information, $\boldsymbol{p}_t^{\mathcal{B}*}$, and deriving a half-hourly prediction sequence by correcting the statistical errors involved in the rough hourly prediction results, $\bar{\boldsymbol{p}}^{\langle\mathcal{H}\rangle}_{\langle t\rangle}$.

D. Alerting to Ramp Events Based on Classifiers

Now, let $r_t^u \in \{0,1\}$ and $r_t^d \in \{0,1\}$ be the binary variables for indicating ramp up/down events at the time slice t, where 1 indicates the occurrence and 0 indicates non-occurrence. The third subsystem, ϕ, comprised a set of classifiers for alerting to ramps and used the data set $\mathcal{D}_3 = \{(\boldsymbol{p}_t^{\mathcal{B}*}, \hat{\boldsymbol{p}}_t^{\mathcal{H}}, \boldsymbol{r}_t^{u*}, \boldsymbol{r}_t^{d*})\}$; we firstly derived the first-order difference in the wind power sequence $(\boldsymbol{p}_t^{\mathcal{B}*}, \hat{\boldsymbol{p}}_t^{\mathcal{H}})$,

$$\begin{aligned}&\Delta\boldsymbol{p}_t\\ &= (p^*_{t-|\mathcal{B}|+2} - p^*_{t-|\mathcal{B}|+1}, \ldots, \hat{p}_{t+1} - p^*_t, \ldots, \hat{p}_{t+|\mathcal{H}|} - \hat{p}_{t-|\mathcal{H}|-1}),\end{aligned} \tag{3}$$

and subsequently used it for predicting the occurrence of ramp events, as follows:

$$\begin{aligned}\phi_h^u &: \Delta\boldsymbol{p}_t \mapsto \hat{r}^u_{t+h} \quad \forall h \in \mathcal{H},\\ \phi_h^d &: \Delta\boldsymbol{p}_t \mapsto \hat{r}^d_{t+h} \quad \forall h \in \mathcal{H}.\end{aligned} \tag{4}$$

In this subsystem, the RF classifiers were modeled to alert ramp up/down events for each forecast horizon $h \in \mathcal{H}$ individually. Note that ϕ_h^u was independently learned the ramp up events for each horizon, $h \in \mathcal{H}$, so that the 30-minute sequence $\hat{\boldsymbol{r}}_t^u = (\hat{r}^u_{t+1}, \ldots, \hat{r}^u_{t+|\mathcal{H}|})$ was derived by concatenating the results. The ramp down sequence $\hat{\boldsymbol{r}}_t^d = (\hat{r}^d_{t+1}, \ldots, \hat{r}^d_{t+|\mathcal{H}|})$ was derived in a similar way.

III. Class Imbalance Problem in Ramp Alert Task

In this section, we focus particularly on the ramp predictor and discuss a problem in ramp prediction. The classifiers shown in (4) were expected to learn the relationship between the predicted power sequence and the ramp occurrence. However, ramp events of wind power generation do not occur frequently throughout the year. Therefore, it was a difficult learning task to predict the occurrence of such relatively rare events based on a data set. Accordingly, the accuracy of the ramp alert task based on naive implementation of the classifiers could be significantly low. In the machine learning community, such issues have been known to occur generally in classification problems and are referred to as the class imbalance problem [40], [41].

A. Basic Idea of Data Sampling

In this study, we expect to improve the classifier-based ramp alert accuracy from the viewpoint of alleviating the class imbalance by adjusting sample sizes used for learning ramp predictors. For the sake of explanation, we focus on ramp up events in this section, although the same argument holds for ramp down events. Let N be the number of the target sampling size, and $\mathcal{S}_h^1$ and $\mathcal{S}_h^0$ be sets of ramp occurrences and non-occurrences derived from the database, $\mathcal{D}_3$, i.e.,

$$\begin{aligned}\mathcal{S}_h^1 &= \{(\Delta\boldsymbol{p}_t, r^{u*}_{t+h}) | r^{u*}_{t+h} = 1\}, \text{ and}\\ \mathcal{S}_h^0 &= \{(\Delta\boldsymbol{p}_t, r^{u*}_{t+h}) | r^{u*}_{t+h} = 0\}.\end{aligned} \tag{5}$$

Since we focus on the ramp events occurred only a few times a year, the sample sizes had the relationship $|\mathcal{S}_h^1| \ll |\mathcal{S}_h^0|$. The basic idea of data sampling strategy is simple; learning data $\mathcal{S}_h^1$ or/and $\mathcal{S}_h^0$ are reconstructed by resampling their instances, so that $|\mathcal{S}_h^1|$ and $|\mathcal{S}_h^0|$ equals to N. Here, we introduce the following sampling approaches.

1) Random Undersampling: The *random undersampling* (*RUS*) approach [40] is a simple but effective method to alleviate the class imbalance. As shown in Fig. 2(a) schematically, samples involved in majority class $\mathcal{S}_h^0$ are randomly undersampled until the majority class contains $N = |\mathcal{S}_h^1|$ samples in the *RUS* approach. The procedure is shown in Algorithm 1.

2) Random Oversampling: The *random oversampling* (*ROS*) approach [40] is another simple method aiming to balance class distribution; samples involved in minority class $\mathcal{S}_h^1$, i.e., the data subset composed of ramp occurrence, are randomly resampled as many as $N = |\mathcal{S}_h^0|$ as shown in Fig. 2(b). Algorithm 2 shows the procedure of the *ROS* approach.

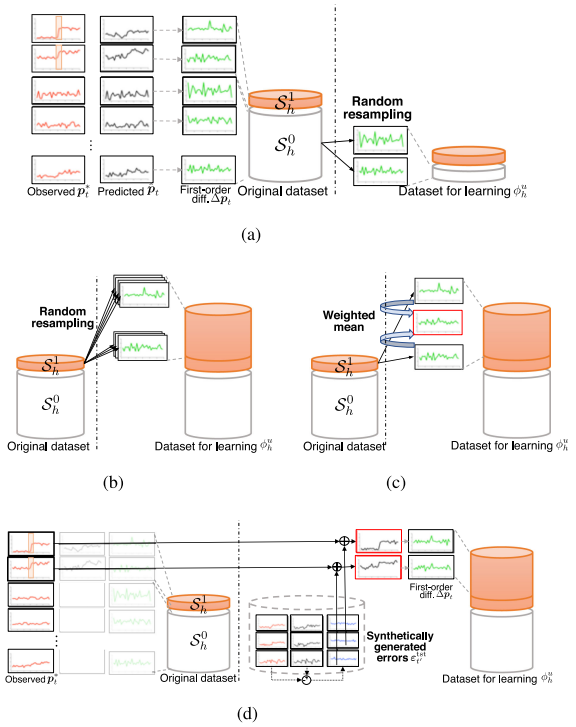


Fig. 2. Schematic image of data sampling methods; red curve shows the observed power sequence, black curve shows the predicted power sequence, green curve shows the first-order difference of the predicted sequence i.e., $\Delta\boldsymbol{p}_t$, and blue curve shows the synthetically generated errors $\boldsymbol{\varepsilon}_t$. (a) RUS. (b) ROS. (c) SMOTE. (d) EBOS.

Algorithm 1: *RUS* Implementation for **Under sampling(.)**.

Input: $\mathcal{S}_h^0$, N.
  while $|\mathcal{S}_h^0| > N$ **do**
    Randomly remove a datum from $\mathcal{S}_h^0$.
  end while
Output: $\mathcal{S}_h^0$.

Algorithm 2: *ROS* Implementation for **Oversampling(.)**.

Input: $\mathcal{S}_h^1$, N.
  while $N > |\mathcal{S}_h^1|$ **do**
    Randomly duplicates a datum in $\mathcal{S}_h^1$.
  end while
Output: $\mathcal{S}_h^1$.

Algorithm 3: *SMOTE* Implementation for **Oversampling(.)**.

Input: $\mathcal{S}_h^1$, N.
  while $N > |\mathcal{S}_h^1|$ **do**
    Randomly focus on a datum $\Delta \boldsymbol{p}_t$ from $\mathcal{S}_h^1$.
    Derive K-nearest neighbors $\{\Delta \boldsymbol{p}_{t_1}, \ldots, \Delta \boldsymbol{p}_{t_K}\}$ of $\Delta \boldsymbol{p}_t$.
    Select a random number $c \in (0, 1)$.
    Generate K new samples

$$\Delta \boldsymbol{p}_t^{\mathrm{new}} = \Delta \boldsymbol{p}_t + c(\Delta \boldsymbol{p}_{t_k} - \Delta \boldsymbol{p}_t) \quad (k = 1, \ldots, K), \tag{6}$$

    and add to $\mathcal{S}_h^1$.
  end while
Output: $\mathcal{S}_h^1$.

3) Synthetic Minority Oversampling Technique: The *synthetic minority oversampling technique* (*SMOTE*) [42] is one of the other well-known oversampling approaches. In the context of ramp prediction, the *SMOTE* utilizes the nearest neighbors of a selected sample belonging to the minority class. It randomly generates the synthetic input data for ramp occurrence by deriving the weighted mean of the original and a neighbor. Algorithm 3 shows the procedure of the *SMOTE* approach. The synthetic sample, $\Delta \boldsymbol{p}_t^{\mathrm{new}}$, was generated by the weighted average of the surrounding minority class samples as shown in (6) in Algorithm 3. Fig. 2(c) shows a schematic image of the SMOTE.

4) Error Bootstrap Oversampling: The above-mentioned oversampling approaches, *ROS* and *SMOTE*, have been proposed for general classification problems. However, we premise to utilize the prediction result of the wind power output $\hat{\boldsymbol{p}}_t^{\mathcal{H}}$ to predict ramp events, as shown in (3) and (4). Considering the mechanism of wind power prediction in a ramp alert task, we proposed a novel oversampling approach, called the *error bootstrap oversampling* (*EBOS*) approach in this study, to alleviate the class imbalance problem. Algorithm 4 shows the procedure of the proposed *EBOS* approach. This approach enabled increasing various synthetic samples of the ramp occurrence for learning classifiers such as *SMOTE*. Furthermore, the intensity of the obtainable fluctuation under the ramp occurrence was expected to be reproduced by utilizing bootstrap oversampling of the plausible errors involved in the predicted wind power output sequence. In the *EBOS* approach, we focused on the relationship between the predicted power generation, $\hat{p}_{t+h}$, and the error, $\hat{p}_{t+h} - \hat{p}^{\mathrm{tst}}_{t'+h}$, and we used the sampling weight defined by the Laplacian kernel function as (7). This weight was utilized for selecting an error sequence that had the similar predicted power at the target horizon, h, so as to provide plausible synthetic samples. Fig. 2(d) shows an image of the EBOS; the input used for learning the classifiers is derived from the prediction result of the wind power output sequence.

Algorithm 4: *EBOS* Implementation for **Oversampling(.)**.

Input: $\mathcal{S}_h^1$, N, $\mathcal{D}_1$, h.
  $\{\varepsilon_t^{\mathrm{tst}}\} \leftarrow \mathbf{generateSyntheticErrors}(\mathcal{D}_1)$:
    Randomly divide $\mathcal{D}_1$ into $\mathcal{D}_1^{\mathrm{lrn}}$ and $\mathcal{D}_1^{\mathrm{tst}}$.
    Learn the predictors ψ^1 and ψ^2 by using $\mathcal{D}_1^{\mathrm{lrn}}$.
    Derive $\hat{\boldsymbol{p}}_t^{\mathcal{H}\mathrm{tst}}$ under $\mathcal{D}_1^{\mathrm{tst}}$ based on ψ^1 and ψ^2.
    Calculate $\boldsymbol{\varepsilon}_t^{\mathrm{tst}} = \hat{\boldsymbol{p}}_t^{\mathcal{H}\mathrm{tst}} - \boldsymbol{p}_t^{\mathcal{H}*\mathrm{tst}}$ for all $\hat{\boldsymbol{p}}_t^{\mathcal{H}\mathrm{tst}}$ in $\mathcal{D}_1^{\mathrm{tst}}$.
  while $N > |\mathcal{S}_h^1|$ **do**
    Randomly pickup $\hat{\boldsymbol{p}}_t^{\mathcal{H}}$ from $\mathcal{S}_h^1$.
    Calculate the weight

$$w_h(\hat{\boldsymbol{p}}_{t'}^{\mathcal{H}\mathrm{tst}} | \hat{\boldsymbol{p}}_t^{\mathcal{H}}) \propto \exp(-|\hat{p}_{t+h} - \hat{p}^{\mathrm{tst}}_{t'+h}|), \tag{7}$$

    for all $\{\hat{\boldsymbol{p}}_{t'}^{\mathcal{H}\mathrm{tst}}\}$.
    Randomly pick up the index t' according to the weight $w_h(\hat{\boldsymbol{p}}_{t'}^{\mathcal{H}\mathrm{tst}} | \hat{\boldsymbol{p}}_t^{\mathcal{H}})$.
    Generate a synthetic sample $\Delta \boldsymbol{p}_t^{\mathrm{new}}$ from

$$\hat{\boldsymbol{p}}_t^{\mathcal{H}\mathrm{new}} = \boldsymbol{p}_t^{\mathcal{H}*} + \boldsymbol{\varepsilon}_{t'}^{\mathrm{tst}}, \tag{8}$$

    and add to $\mathcal{S}_h^1$.
  end while
Output: $\mathcal{S}_h^1$.

B. Learning Ramp Predictors Based on Oversampled Data

The idea of data sampling works well for alleviating class imbalance problem in general classification problems and it is also expected to improve the accuracy of ramp prediction in our context. Note that the alleviation effect of the class imbalance problem depends largely on the suitability between the target data and the concrete approach, so that it is difficult to discuss the universal superiority of the specific data sampling approach. However, here we briefly describe the expected characteristics of the several oversampling methods introduced in the previous subsection by focusing on the learning framework of ramp predictors introduced in Section II.A.

Firstly, the ROS simply increases minority data under the ramp occurrence by resampling the wind power fluctuation sequence $\Delta \boldsymbol{p}_t$ randomly from the existing samples. Though the ROS alleviates the class imbalance problem while emphasizing situations where ramp events occur, this method does not generate various data having new features to distinguish from situations of ramp non-occurrence. Meanwhile, the SMOTE

Algorithm 5: Hybrid Sampling Scheme.

Input: $\mathcal{D}_3$, h, R.
Divide $\{(\Delta \boldsymbol{p}_t, r^{u*}_{t+h}); t\}$ into $\mathcal{S}^1_h$ and $\mathcal{S}^0_h$ based on (5).
Set $N = \lfloor R|\mathcal{S}^0_h| + (1-R)|\mathcal{S}^1_h| \rfloor$.
$\mathcal{S}^0_h \leftarrow \mathbf{undersampling}(\mathcal{S}^0_h, N)$.
$\mathcal{S}^1_h \leftarrow \mathbf{oversampling}(\mathcal{S}^1_h, N, ...)$.
Output: $\mathcal{S}^0_h \cup \mathcal{S}^1_h$ for learning ϕ^u_h.

synthetically generates data with new characteristics by weighted average of the pair of wind power fluctuation sequences. This method is a promising approach to generate plausible data observable under the ramp occurrence to mitigate the class imbalance problem. However, $\Delta \boldsymbol{p}_t$ represents a transition of fluctuations in predicted wind power, so that the fluctuations in the synthetic sample $\Delta \boldsymbol{p}^{\text{new}}_t$ given in (6) of Algorithm 3 tend to be small, since the weighted average of arbitrary two values $v_1, v_2 \in \mathbb{R}$ has the following property,

$$|v_1 + c(v_2 - v_1)| \leq \max(|v_1|, |v_2|), \tag{9}$$

where $c \in (0, 1)$. In prediction of ramp events based on the fluctuations in the predicted power sequence, the SMOTE may not be effective; because this approach cannot create new samples representing plausible large fluctuations. The proposed approach, the EBOS, samples error sequences of the predicted power derived from the database, and generates synthetic prediction results $\hat{\boldsymbol{p}}^{\mathcal{H}\text{new}}_t$ given in (8) of Algorithm 4 by adding plausible error sequences to an observation under the ramp occurrence. This method is expected to generate various types of plausible predicted power fluctuation sequences for actual ramp events in order to alleviate the class imbalance problem. Note that the EBOS may generate sequences with various magnitude of fluctuations over the original fluctuation sequences under the ramp occurrence; this is the main difference between SMOTE and EBOS.

C. Hybridization of Under/Oversampling Approaches

In the imbalanced class setting, hybrid sampling approaches [40] composed of the pair of under/oversampling methods often achieve better classification than individual approaches. In the learning step of ramp predictors, we can also introduce the hybrid sampling scheme. Let

$$N = \lfloor R|\mathcal{S}^0_h| + (1-R)|\mathcal{S}^1_h| \rfloor \tag{10}$$

be the number of the target sampling size, where $R \in [0, 1]$ is the sampling parameter. In the hybrid approach, the samples involved in majority class $\mathcal{S}^0_h$ are undersampled and those involved in minority class $\mathcal{S}^1_h$ are oversampled until both classes have N samples. Note that the hybrid approach reduced to the undersampling scheme under $R = 0$ and the oversampling scheme under $R = 1$. The procedure of hybrid sampling is shown in Algorithm 5.

In this study, the three types of hybridization are defined, i.e., "*RUS+ROS*", "*RUS+SMOTE*", and "*RUS+EBOS*". For example, in *RUS+ROS*, the *RUS* method is implemented at the undersampling phase and the *ROS* method is implemented at the oversampling phase. At the undersampling phase, the *RUS* approach randomly eliminated samples involved in the majority class $\mathcal{S}^0_h$ so that $|\mathcal{S}^0_h| = N$ holds where N is given as (10). At the oversampling phase, the *ROS* approach randomly resampled the minority class $\mathcal{S}^1_h$ as many as N. In our implementation, the sampling parameter R was tuned individually for each h from the viewpoint of expected accuracy of ramp prediction at the horizon h, based on 10-fold cross-validation [25].

IV. Numerical Experiments

In this section, we evaluate our ramp alert scheme by using a real-world data set. First, we explain our experimental setup and briefly refer to the performance of our wind power predictor for the input of the ramp predictor. Subsequently, we compare some methods for dealing with the class imbalance problem in the ramp alert task, and we discuss the effectiveness of the proposed method for improving the accuracy of the ramp predictor.

A. Experimental Setup

In this evaluation, we used real-world wind power profiles at 30-minute intervals, observed in Tohoku region, Japan; the profiles consists of the sum of outputs of twenty wind farms in this region and the total rated power output of the wind farms is 440 MW. The data set comprised observations collected from April 1st 2012 to March 31st 2013. We focused on the half-hourly wind power output of the Tohoku area from thirty minutes to 48 hours ahead, i.e., $\mathcal{H} = \{1, \ldots, 96\}$, and we used the observed wind power outputs of the previous 17 hours, i.e., $\mathcal{B} = \{33, \ldots, 0\}$, for our prediction. In addition, we focused on the ramp events corresponding to the forecast horizons $h \in \mathcal{H}$. We used the predicted hourly meteorological data of the corresponding area. These data were provided by a regional climate model, called NuWFAS (Numerical Weather Forecasting and Analysis System) [44], which has been developed by the Central Research Institute of Electric Power Industry (CRIEPI) based on the Weather Research and Forecasting (WRF) model [45]. The spatial resolution of the meteorological data was 3 km. We focused particularly on the forecast result of the hourly wind speed at an altitude of 60 m, provided once a day at 17:00 (JST) for the next 76 h, and we extracted the subarea within 10 km of the existing wind farms. The dimension of the input vector, $\boldsymbol{x}_{\langle t+h \rangle}$, was 820.

The evaluation procedure is given as follows. We extracted one month from the original data set for validation, trained the predictors by using the remaining eleven months, and evaluated the predictors by using the validation set. This procedure was repeated for all twelve months. In this numerical experiment, the prediction accuracy of a certain month implied the evaluation result of the held-out data subset by focusing on the corresponding month. Since our ramp alert setup comprised the supervised learning framework, the proposed scheme could be applied to various ramp definitions [2] under the situation that the ramp events are identified in the historical data set of the wind power sequence. In this study, we targeted the total wind power output in the area, and we focused on the fluctuations simultaneously satisfying a ramp rate of more than 30[%] of the total capacity of

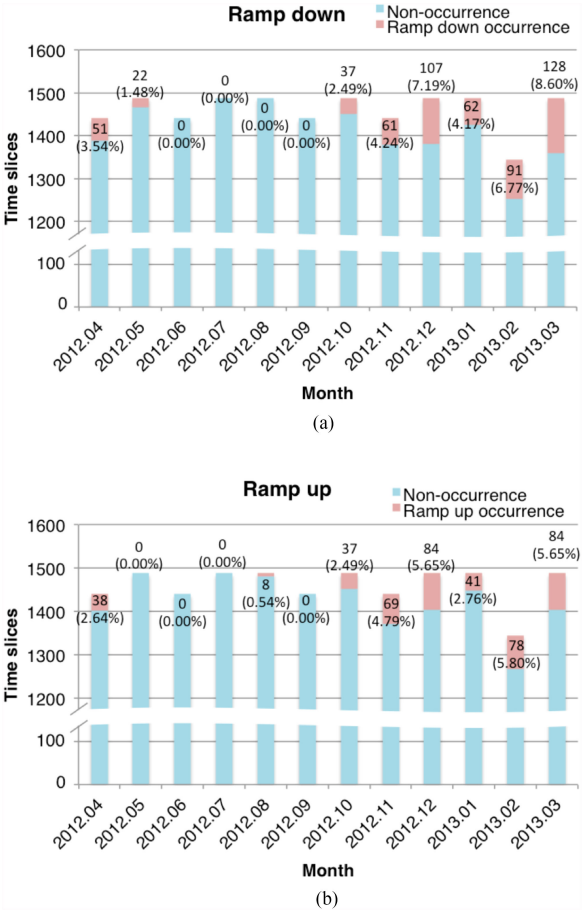


Fig. 3. Number of large-scale ramp events in the data set. (a) Number of ramp down events. (b) Number of ramp up events.

the wind farms per six hours and averagely more than 5[%] per hour [46]. Fig. 3 shows the number of large-scale ramp events in this data set. The figure shows the seasonality in the occurrence of the ramp events in Japan. Note that the occurrence rate of these ramp events was less than 4% throughout the year; therefore, the class imbalanced problem would critically affect the accuracy of the ramp prediction.

We used the mean absolute errors (MAE) for the wind power prediction, i.e.,

$$\mathrm{MAE}_h = \frac{1}{T}\sum_t |\hat{p}_{t+h} - p^*_{t+h}| \quad (h \in \mathcal{H}), \tag{11}$$

and we used indices called the precision, the recall, and the critical success index (CSI) for evaluation of the ramp predictor [19]. For example, the prediction results for a ramp up event for the prediction horizon h was evaluated as follows:

$$\mathrm{Prec}^u_h = \frac{\sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}}{\sum_t \hat{r}^u_{t+h}} \quad (h \in \mathcal{H}), \tag{12}$$

$$\mathrm{Rec}^u_h = \frac{\sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}}{\sum_t r^{u*}_{t+h}} \quad (h \in \mathcal{H}), \tag{13}$$

$$\begin{aligned}\mathrm{CSI}^u_h &= \frac{\sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}}{T - \sum_t (1 - r^{u*}_{t+h})(1 - \hat{r}^u_{t+h})} \\ &= \frac{\sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}}{\sum_t r^{u*}_{t+h}(1-\hat{r}^u_{t+h}) + \sum_t (1-r^{u*}_{t+h})\hat{r}^u_{t+h} + \sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}} \\ &= \frac{\sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}}{\sum_t r^{u*}_{t+h} + \sum_t \hat{r}^u_{t+h} - \sum_t r^{u*}_{t+h}\hat{r}^u_{t+h}} \quad (h \in \mathcal{H}).\end{aligned} \tag{14}$$

The precision given in (12) is also known as the forecast accuracy and indicates the ratio of actual ramp events among the predicted ramp events. The recall given in (13) is also known as the ramp capture or the hit percentage; it indicates the ratio of correct prediction among the actual ramp events. The CSI given in (14) is also known as the threat score and indicates the ratio of the number of correct prediction to the number of whole events except for the correct rejection events. These indices take values in the interval $[0, 1]$ where the larger value implies the better prediction[3]. Note that the prediction of the ramp down event was similarly evaluated by using r^{d*}_{t+h} and $\hat{r}^d_{t+h}$ instead of r^{u*}_{t+h} and $\hat{r}^u_{t+h}$ in (12)–(14).

B. Evaluation of Wind Power Prediction

Firstly, we briefly evaluated our wind power prediction results derived every 30 minutes, from the viewpoint of the prediction error. Fig. 4(a) shows the monthly MAEs of the wind power predictor for each prediction horizon $h \in \mathcal{H}$. The prediction results were related to seasonality, i.e., the error was relatively small in summer (July–September), when the frequency of ramp events was low. However, it tended to be high in autumn and winter (October–March), when the frequency of the ramp events was relatively high. Although prediction errors tended to increase monotonically with respect to the prediction horizon, the wind power prediction scheme prevented quite large errors at the far horizon. Fig. 4(b) also shows the power output prediction results of the simple persistence model [9] for reference, whose prediction result was derived as $\hat{p}_{t+h} = p^*_t$. In comparing these figures, our power predictor showed the advantage of the relative prediction accuracy.

C. Evaluation of Ramp Event Prediction

Subsequently, we evaluated the accuracy of the ramp alert by comparing nine approaches, as shown in Table I. Note that the *direct* scanning approach is one of the straightforward ways for predicting ramp events; this approach identifies ramp events by directly scanning the predicted power sequence according to the ramp definition without using ramp predictors. We also evaluate the *naive* approach which naively utilizes the ramp predictor without taking any measures against the class imbalance problem.

Fig. 5 shows the relationship between the annual CSI for ramp prediction and the prediction horizon h. In Fig. 5(a), the

[3]When a denominator in (12)–(14) takes zero, the corresponding numerator also takes zero. In this paper, we regard $0/0$ as zero in accordance with the conventional treatment of these metrics, though it is not well-defined mathematically.

TABLE I
RAMP ALERT APPROACHES COMPARED IN THE NUMERICAL EXPERIMENT

Scheme	Approach		Description
Direct scanning	*Direct*		Identifying ramp events by directly scanning the predicted power sequence according to the ramp definition without using ramp predictor.
Separate prediction	*Naive*		Naively utilizing ramp predictor without taking any measures against the class imbalance problem.
	undersampling:	*RUS**	Utilizing the undersampling method ($R = 0$).
	oversampling:	*ROS**, *SMOTE**, *EBOS*	Utilizing the oversampling methods ($R = 1$).
	hybrid sampling:	*RUS+ROS*, *RUS+SMOTE*, *RUS+EBOS*	Utilizing the hybrid sampling approaches by using tuned R.

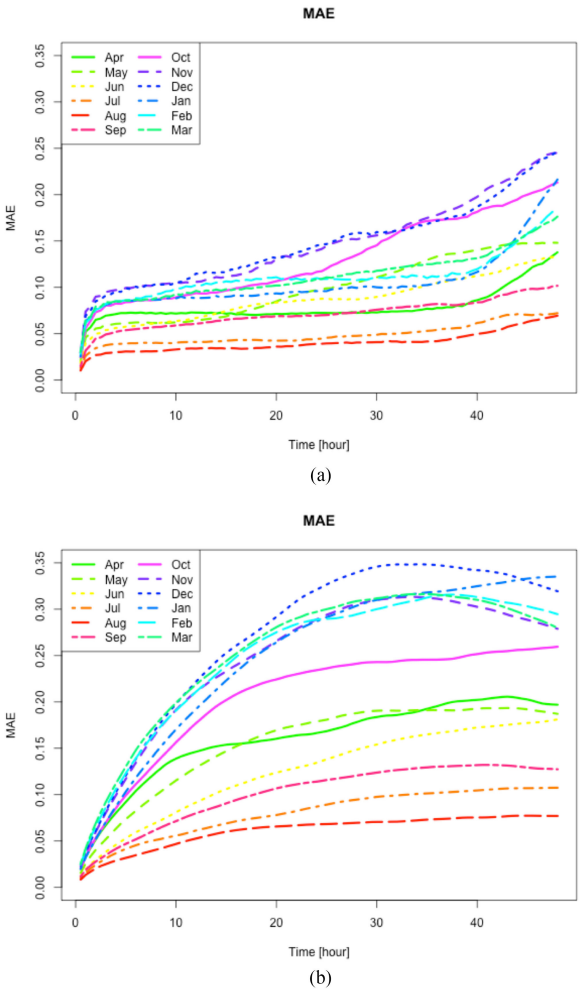


Fig. 4. Relationship between monthly mean absolute errors of the wind power prediction and the prediction horizon. (a) Results of our proposed power predictor composed of RF regressors. (b) Results of persistence model.

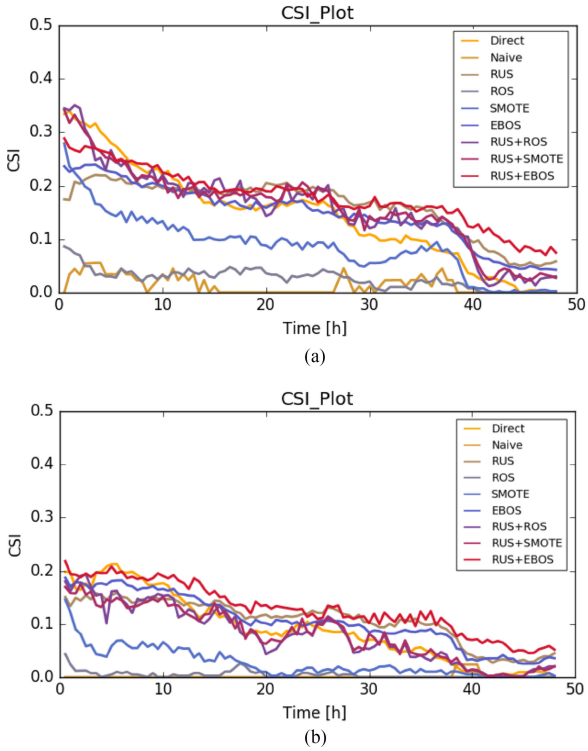


Fig. 5. Relationships between yearly CSI of ramp prediction and the prediction time horizon; (a) results of ramp down, and (b) results of ramp up.

direct approach shows relatively satisfactory results, particularly at the short prediction horizons. However, the CSI of the *naive* approach was remarkably low, and the result suggested the difficulty of the separate prediction scheme for alerting to rare ramp events. In addition, the experimental result showed that our proposed hybrid sampling approaches *RUS+ROS*, *RUS+SMOTE*, and *RUS+EBOS* were superior on average to the *direct* approach at the long prediction horizons. Fig. 5(b) also shows the prediction results of the ramp up events. The results showed a tendency similar to that of the ramp down prediction. In this instance, the proposed hybrid approach *RUS+EBOS* showed relatively satisfactory results for most prediction horizons.

Fig. 6 shows the monthly CSI of six-hours and twelve-hours ahead ramp up prediction for October 2012 [see Fig. 6(a) and (b)] with a moderate number of ramp events, and March 2013 [see Fig. 6(c) and (d)] with a considerable number of ramp events. As shown in the figure, the CSI values of the *direct* method were relatively accurate for the prediction horizon of six-hours ahead (i.e., $h = 12$) but the accuracy declined at twelve-hours ahead (i.e., $h = 24$). In contrast, the CSI values of the *naive* method were extremely low even at $h = 12$. This figure also implied specifically that *RUS* and the proposed *EBOS* contributed to improving the ramp alert accuracy. A hybrid of these methods, i.e., *RUS+EBOS*, improved the accuracy drastically, even at twelve-hours ahead. This result suggested an important advantage of the proposed sampling strategy in alerting to rare ramp events.

Fig. 7 shows the precision-recall plot of the ramp up alerting result of March 2013 at $h = 12$, which corresponded to Fig. 6(c). The figure shows that the *SMOTE* approach achieved a significantly high precision value, while it had a low recall value. The result implied the relatively considerable number of the unpredicted ramp events. However, the *RUS* and the proposed *EBOS*

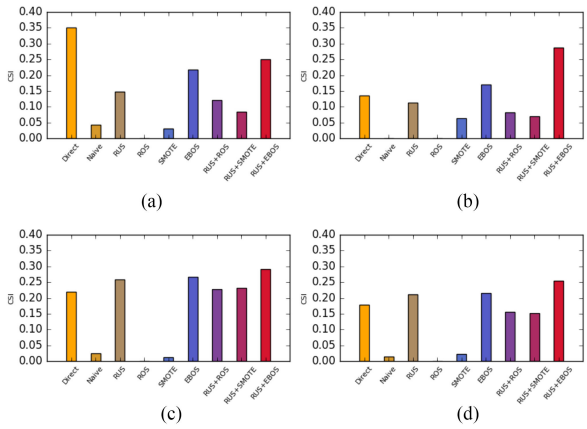


Fig. 6. Monthly CSI for ramp up alert of October 2012 (moderate number of ramp events) and March 2013 (relatively many ramp events). We focus on the prediction horizon of $h = 12$ (i.e., six hours ahead) and $h = 24$ (i.e., 12 hours ahead). (a) October 2012 (6 hours ahead). (b) October 2012 (12 hours ahead). (c) March 2013 (6 hours ahead). (d) March 2012 (12 hours ahead).

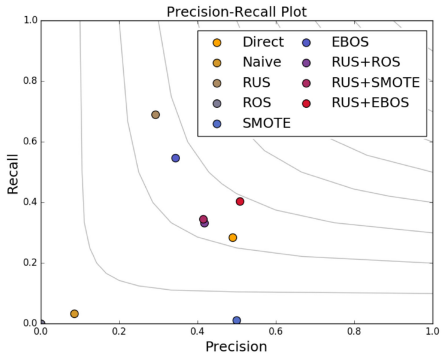


Fig. 7. Plot of precision and recall for ramp up alert of March 2013. We focused on the prediction horizon of $h = 12$ (i.e., six-hours ahead). The curves in the figure represent the contour lines of the CSI.

approaches achieved relatively high recall values, while retaining the precision values. Note that the *RUS+EBOS* approach retained both the precision and recall values; consequently, the forth-coming ramp events could be alerted meaningfully. Similar results were obtained for other seasons, other horizons, and for ramp down events.

Fig. 8 shows a typical example of the wind power prediction results for each prediction horizon, $h \in \mathcal{H}$, and the corresponding ramp events. In the observed wind power output, ramp up events existed at 8.0–12.5 h and 39.0–41.5 h, and ramp down events at 14.5–24.0 h and 47.5–48.0 h. The dynamics of the wind power output sequence appeared well predicted on average by our power predictor, but the prediction of fluctuation tended to be smaller than it actually was. In the *direct* approach, the ramp up event occurring at 8.0–12.5 h could not be alerted owing to this tendency. In addition, the *naive* ramp prediction approach had completely failed to identify any ramp events. However, the proposed *RUS+EBOS* approach achieved relatively precise prediction, which predicted the forthcoming ramp events.

These experimental results implied that the class imbalance problem caused difficulty in realizing the ramp alert task. We therefore suggest that the implementation of an appropriate

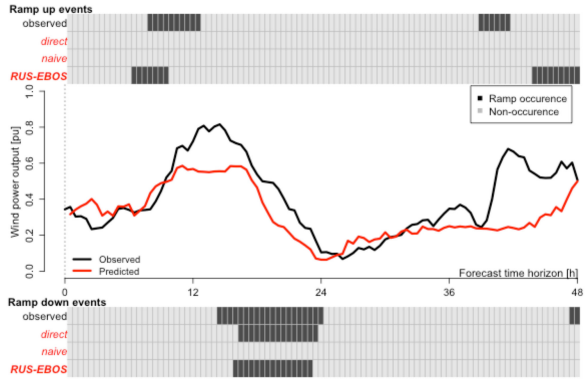


Fig. 8. An example of ramp alert. Black and red lines represent the observed and predicted wind power outputs. Black boxes at the top/bottom of the figure represent the ramp up/down events at the corresponding time slices. Note that the ramp events predicted by the *direct*, *naive*, and *RUS+EBOS* approaches were derived by using predicted wind power output (red line).

countermeasure, such as the *RUS+EBOS*, could contribute to drastically improving the accuracy of the ramp alert task.

V. Conclusion

In this study, we focused on the ramp alert task by using the prediction results of wind power output. In particular, we addressed the class imbalance problem in the ramp alert task and we proposed remedies to improve the accuracy. These include the error bootstrap oversampling (*EBOS*) technique and several hybrid data sampling approaches. The proposed data sampling strategy was expected to be effective for predicting the rare ramp events explicitly labeled in the historical data set. The experimental results, based on a real-world wind power generation data set, indicated that the proposed methods, particularly the *RUS+EBOS* method, drastically improved the ramp alert accuracy, even at the long prediction horizons.

We addressed the problem in the construction of a ramp predictor, based on the predicted result of the wind power output sequence. However, several significant topics were not mentioned in this study. One of the important aspects for improving the ramp alert accuracy is error reduction in the power predictor; e.g., application of the deep learning framework [47] can be expected as an effective approach to realize highly accurate predictors [28]–[30]. Another significant aspect is handling the uncertainty in wind power/ramp prediction [36], [48]–[50], particularly for real-world system operations. Accordingly, we are planning to conduct further research to realize an appropriate ramp alert system.

Appendix
Random Forests

We briefly introduce the construction procedure of random forests (RF) implemented in our scheme. The method is based on the framework of bootstrap aggregating [51] and decision trees [52] (please refer [43] for detail).

Firstly, we focus on the regression framework. Let $\boldsymbol{x} = (x^1, \ldots, x^P) \in \mathbb{R}^P$ be the vector of P input variables used for the prediction, $y \in \mathbb{R}$ be the corresponding output, $\mathcal{P} =$

$\{1, \ldots, P\}$ be the variable index set, and

$$\mathcal{D} = \{(\boldsymbol{x}_n, y_n) \mid n \in \mathcal{N}\},$$

be the dataset used for training the RF predictor ψ where $\mathcal{N} = \{1, \ldots, N\}$ is the sample index set for N samples. In the learning phase, we focus on the data subset $\mathcal{N}' \subset \mathcal{N}$ which is randomly selected from the training set $\mathcal{D}$ so as to generate new dataset

$$\mathcal{D}' = \{(\boldsymbol{x}_n, y_n) \mid n \in \mathcal{N}'\}.$$

The component of the RF, a decision tree, is grown by recursively splitting $\mathcal{D}'$ to subsets

$$\mathcal{D}'_-(p,\theta) = \{\mathcal{D}' | x_n^p < \theta\}, \quad \text{and} \quad \mathcal{D}'_+(p,\theta) = \{\mathcal{D}' | x_n^p \geq \theta\},$$

by focusing on randomly selected variable index subset $\mathcal{P}' \subset \mathcal{P}$; subsets are derived by selecting $p \in \mathcal{P}'$ and θ that maximizes the decrease of deviance,

$$\{\hat{p}, \hat{\theta}\} = \underset{\{p \in \mathcal{P}', \theta\}}{\operatorname{argmax}} \sum_{n \in \mathcal{N}(\mathcal{D}')} (y_n - \mu(\mathcal{D}'))^2 - \sum_{n \in \mathcal{N}(\mathcal{D}'_-(p,\theta))} (y_n - \mu(\mathcal{D}'_-(p,\theta)))^2 - \sum_{n \in \mathcal{N}(\mathcal{D}'_+(p,\theta))} (y_n - \mu(\mathcal{D}'_+(p,\theta)))^2,$$

where $\mathcal{N}(\mathcal{A})$ is the index subset involved in set $\mathcal{A}$ and $\mu(\mathcal{A})$ is the average of outputs involved in $\mathcal{A}$, i.e.,

$$\mu(\mathcal{A}) = \frac{\sum_{n \in \mathcal{N}(\mathcal{A})} y_n}{|\mathcal{A}|}.$$

The recursive splitting procedure is described as hierarchical binary decision rules composing a tree structure, so that the original data set $\mathcal{D}$ is partitioned into L subsets $\mathcal{A}_l$ $(l \in 1, \ldots, L)$ where L is the number of leaves in the decision tree, i.e., the number of subsets partitioned by the decision tree. The RF regression model is composed of M decision trees $\{\psi^{(m)}\}$; each decision tree outputs

$$\psi^{(m)}(\boldsymbol{x}) = \mu(\mathcal{A}_l^{(m)} \ni \boldsymbol{x}),$$

corresponding to the arbitrary $\boldsymbol{x}$ where $\mathcal{A}_l^{(m)}$ indicates the subset partitioned by the m-th decision tree. The final prediction result of the RF corresponding to $\boldsymbol{x}$ is given as

$$\hat{y} = \psi(\boldsymbol{x}) = \frac{1}{M} \sum_{m=1}^{M} \psi^{(m)}(\boldsymbol{x}).$$

The procedure is almost the same for the binary classification task. Let $c \in \{0, 1\}$ be the output corresponding to the given input $\boldsymbol{x} \in \mathbb{R}^P$, and

$$\mathcal{D} = \{(\boldsymbol{x}_n, c_n) \mid n \in \mathcal{N}\},$$

be the dataset for training the RF predictor ϕ. A decision tree in the RF is grown by recursively by focusing on randomly selected $\mathcal{N}' \subset \mathcal{N}$ and $\mathcal{P}' \subset \mathcal{P}$. The partitioning parameters $p \in \mathcal{P}'$ and θ are selected by maximizing the decrease of impurity,

$$\{\hat{p}, \hat{\theta}\} = \underset{\{p \in \mathcal{P}', \theta\}}{\operatorname{argmax}}\, g(\mathcal{D}') - g(\mathcal{D}'_-(p,\theta)) - g(\mathcal{D}'_+(p,\theta)),$$

where $g(\mathcal{A})$ is the Gini index of the subset $\mathcal{A}$, i.e.,

$$g(\mathcal{A}) = 1 - \left(\frac{\sum_{n \in \mathcal{N}(\mathcal{A})} c_n}{|\mathcal{A}|}\right)^2 - \left(\frac{\sum_{n \in \mathcal{N}(\mathcal{A})} (1 - c_n)}{|\mathcal{A}|}\right)^2.$$

The RF classification model is composed of M decision trees $\{\phi^{(m)}\}$; each tree outputs

$$\phi^{(m)}(\boldsymbol{x}) = \begin{cases} 0 & \left(\sum_{n \in \mathcal{N}(\mathcal{A}_l^{(m)})} c_n < |\mathcal{A}_l^{(m)}|/2\right) \\ 1 & \left(\sum_{n \in \mathcal{N}(\mathcal{A}_l^{(m)})} c_n \geq |\mathcal{A}_l^{(m)}|/2\right), \end{cases}$$

for given $\boldsymbol{x}$. The final prediction result of the RF for $\boldsymbol{x}$ is given according to the majority vote of M decision trees as

$$\hat{c} = \phi(\boldsymbol{x}) = \begin{cases} 0 & \left(\sum_{m=1}^{M} \phi^{(m)}(\boldsymbol{x}) < M/2\right) \\ 1 & \left(\sum_{m=1}^{M} \phi^{(m)}(\boldsymbol{x}) \geq M/2\right). \end{cases}$$

ACKNOWLEDGMENT

This paper is based on results obtained from *R&D Project on Grid Integration of Variable Renewable Energy Mitigation Technologies on Output Fluctuations of Renewable Energy Generations in Power Grid* commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

REFERENCES

[1] Y. Ueda, M. Takamoto, and T. Saito, "Target & roadmap for Japan wind power," presented at the Int. Conf. Exhib. Grand Renewable Energy, Tokyo, Japan, 2014.

[2] C. Ferreira, J. Gama, L. Matias, A. Botterud, and J. Wang, "A survey on wind power ramp forecasting," Argonne Nat. Lab., Lemont, IL, USA, Tech. Rep. ANL/DIS-10-13, 2010.

[3] C. S. Reddy, K. Balaraman, A. Patil, and K. Vasudevan, "Wind power ramp forecasting for stochastic unit commitment in smart grid environment," in *Proc. IEEE Innov. Smart Grid Technol.—Asia*, 2013, pp. 1–6.

[4] M. Cui, J. Zhang, H. Wu, and B. M. Hodge, "Wind-friendly flexible ramping product design in multi-timescale power system operations," *IEEE Trans. Sustain. Energy*, vol. 8, no. 3, pp. 1064–1075, Jul. 2017.

[5] Y. Gong, Q. Jiang, and R. Baldick, "Ramp event forecast based wind power ramp control with energy storage system," *IEEE Trans. Power Syst.*, vol. 31, no. 3, pp. 1831–1844, May 2016.

[6] M. Ito, Y. Fujimoto, M. Mitsuoka, H. Ishii, and Y. Hayashi, "Control methods for an energy storage system when wind power output deviates from grid code," *J. Int. Council Elect. Eng.*, vol. 7, no. 1, pp. 159–165, 2017.

[7] R. Chen, J. Wang, A. Botterud, and H. Sun, "Wind power providing flexible ramp product," *IEEE Trans. Power Syst.*, vol. 32, no. 3, pp. 2049–2061, May 2017.

[8] Q. Wang and B. M. Hodge, "Enhancing power system operational flexibility with flexible ramping products: A review," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 1652–1664, Aug. 2017.

[9] A. Costa, A. Crespo, J. Navarro, G. Lizcano, H. Madsen, and E. Feitosa, "A review on the young history of the wind power short-term prediction," *Renewable Sustain. Energy Rev.*, vol. 12, pp. 1725–1744, 2008.

[10] M. Lange, "On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors," *J. Solar Energy Eng.*, vol. 127, no. 2, pp. 177–184, 2005.

[11] Y. Takahashi, Y. Fujimoto, and Y. Hayashi, "Prediction of wind power output for alerting ramp events," in *Proc. 10th Int. Renewable Energy Storage Conf.*, 2016, pp. 512–515.

[12] D. Hou, E. Kalnay, and K. K. Droegemeier, "Objective verification of the SAMEX '98 ensemble forecasts," *Monthly Weather Rev.*, vol. 129, no. 1, pp. 73–91, 2001.

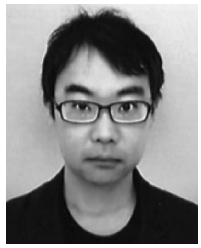
[13] L. L. Takacs, "A two-step scheme for the advection equation with minimized dissipation and dispersion errors," *Monthly Weather Rev.*, vol. 113, pp. 1050–1065, 1985.

[14] T. Ouyang, X. Zha, and L. Qin, "A survey of wind power ramp forecasting," *Energy Power Eng.*, vol. 5, pp. 368–372, 2013.

[15] C. Gallego-Castillo, A. Cuerva-Tajero, and O. Lopez-Garcia, "A review on the recent history of wind power ramp forecasting," *Renewable Sustain. Energy Rev.*, vol. 52, pp. 1148–1157, 2015.

[16] C. Kamath, "Understanding wind ramp events through analysis of historical data," in *Proc. IEEE PES Transmiss. Distrib. Conf. Expo.*, 2010, pp. 1–6.

[17] H. Zheng and A. Kusiak, "Prediction of wind farm power ramp rates: A data-mining approach," *J. Solar Energy Eng.*, vol. 131, no. 3, pp. 31011–31018, 2009.
[18] H. Zareipour, D. Huang, and W. Rosehart, "Wind power ramp events classification and Forecasting: A data mining approach," in *Proc. IEEE Power Energy Soc. Gen. Meet.*, 2011, pp. 1–3.
[19] C. A. Ferreira, J. Gama, V. Santos Costa, V. Miranda, and A. Botterud, "Predicting ramp events with a stream-based HMM framework," in *Discovery Science: 15th International Conference*, J.-G. Ganascia, P. Lenca, and J.-M. Petit, Eds. Berlin, Germany: Springer, 2012, pp. 224–238.
[20] A. Florita, B.-M. Hodge, and K. Orwig, "Identifying wind and solar ramping events," in *Proc. IEEE Green Technol. Conf.*, 2013, pp. 147–152.
[21] C. Gallego, A. Costa, Á. Cuerva, L. Landberg, B. Greaves, and J. Collins, "A wavelet-based approach for large wind power ramp characterisation," *Wind Energy*, vol. 16, pp. 257–278, 2013.
[22] M. Cui, D. Ke, Y. Sun, D. Gan, J. Zhang, and B. M. Hodge, "Wind power ramp event forecasting using a stochastic scenario generation method," *IEEE Trans. Sustain. Energy*, vol. 6, no. 2, pp. 422–433, Apr. 2015.
[23] S. Gupta, N. A. Shrivastava, A. Khosravi, and B. K. Panigrahi, "Wind ramp event prediction with parallelized gradient boosted regression trees," in *Proc. Int. Joint Conf. Neural Netw.*, 2016, pp. 5296–5301.
[24] Y. Takahashi, Y. Fujimoto, and Y. Hayashi, "Forecast of infrequent wind power ramps based on data sampling strategy," *Energy Procedia*, vol. 135, pp. 496–503, 2017.
[25] C. M. Bishop, *Pattern Recognition and Machine Learning*. Berlin, Germany: Springer, 2006.
[26] C. Gallego, A. Cuerva, and A. Costa, "Detecting and characterising ramp events in wind power time series," *Proc. J. Phys., Conf. Ser.*, vol. 555, 2014, Art. no. 012040.
[27] M. Dorado-Moreno, L. Cornejo-Bueno, P. Gutiérrez, L. Prieto, C. Hervás-Martínez, and S. Salcedo-Sanz, "Robust estimation of wind power ramp events with reservoir computing," *Renewable Energy*, vol. 111, pp. 428–437, 2017.
[28] C. Y. Zhang, C. L. Chen, M. Gan, and L. Chen, "Predictive deep Boltzmann machine for multiperiod wind speed forecasting," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1416–1425, Oct. 2015.
[29] H. Wang, G. Li, G. Wang, J. Peng, H. Jiang, and Y. Liu, "Deep learning based ensemble approach for probabilistic wind power forecasting," *Appl. Energy*, vol. 188, pp. 56–70, 2017.
[30] K. Higashiyama, Y. Fujimoto, and Y. Hayashi, "Feature extraction of numerical weather prediction results toward reliable wind power prediction," in *Proc. 7th IEEE PES Innov. Smart Grid Technol. Eur.*, 2017, pp. 1–6.
[31] N. Cutler, M. Kay, K. Jacka, and T. S. Nielsen, "Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT," *Wind Energy*, vol. 10, no. 5, pp. 453–470, 2007.
[32] K. T. Bradford, R. L. Carpenter, and B. L. Shaw, "Forecasting southern plains wind ramp events using the WRF model at 3-km," in *Proc. 9th Amer. Meteorological Soc. Student Conf.*, 2010.
[33] A. Bossavy, R. Girard, and G. Kariniotakis, "Forecasting ramps of wind power production with numerical weather prediction ensembles," *Wind Energy*, vol. 16, no. 1, pp. 51–63, 2013.
[34] A. Couto, P. Costa, L. Rodrigues, V. Lopes, and A. Estanqueiro, "Impact of weather regimes on the wind power ramp forecast," *IEEE Trans. Sustain. Energy*, vol. 6, no. 3, pp. 934–942, Jul. 2015.
[35] Y. Li *et al.*, "Separate wind power and ramp predictions based on meteorological variables and clustering method," in *Proc. IEEE 6th Int. Conf. Power Syst.*, 2016, pp. 1–6.
[36] Y. Takahashi, Y. Fujimoto, and Y. Hayashi, "Interval forecasting for wind power generation based on random forests," in *Proc. Int. Conf. Elect. Eng.*, 2015, pp. 1–6.
[37] J. Zhang, M. Cui, B.-M. Hodge, A. Florita, and J. Freedman, "Ramp forecasting performance from improved short-term wind power forecasting over multiple spatial and temporal scales," *Energy*, vol. 122, pp. 528–541, 2017.
[38] N. Japkowicz and S. Stephen, "The class imbalance problem: A systematic study," *Intell. Data Anal.*, vol. 6, no. 5, pp. 429–449, 2002.
[39] R. Longadge, S. S. Dongre, and L. Malik, "Class imbalance problem in data mining: review," *Int. J. Comput. Sci. Netw.*, vol. 2, no. 1, pp. 83–87, 2013.
[40] S. M. A. Elrahman and A. Abraham, "A review of class imbalance problem," *J. Netw. Innov. Comput.*, vol. 1, pp. 332–340, 2013.
[41] N. Japkowicz, "Learning from imbalanced data Sets: A comparison of various strategies," in *Proc. AAAI Workshop Learn. Imbalanced Data Sets*, vol. 68, 2000, pp. 10–15.
[42] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-samping technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, 2002.
[43] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.
[44] A. Hashimoto and H. Hirakuchi, "Enhancement and accuracy evaluation of numerical weather forecasting and analysis system (NuWFAS) for the Hokkaido island," (in Japanese), Central Res. Inst. Elect. Power Ind., Tokyo, Japan, Tech. Rep. N09024, 2010.
[45] J. G. Powers *et al.*, "The weather research and forecasting (WRF) model: Overview, system efforts, and future directions," *Bull. Amer. Meteorological Soc.*, vol. 98, pp. 1717–1737, 2017.
[46] The New Energy and Industrial Technology Development Organization of Japan, Kawasaki, Japan, "R&D project on grid integration of variable renewable energy mitigation technologies on output fluctuations of renewable energy generations in power grid (FY2014-FY2018)," 2017. Online. Available: http://www.nedo.go.jp/library/seika/shosai_201706/20170000000402.html, Accessed on: May 23, 2018.
[47] Y. Lecun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[48] J. W. Taylor, "Probabilistic forecasting of wind power ramp events using autoregressive logit models," *Eur. J. Oper. Res.*, vol. 259, no. 2, pp. 703–712, 2017.
[49] M. Cui *et al.*, "Probabilistic wind power ramp forecasting based on a scenario generation method," in *Proc. IEEE Power Energy Soc. General Meet.*, Chicago, IL, USA, 2017.
[50] M. Cui, J. Zhang, Q. Wang, V. Krishnan, and B.-M. Hodge, "A data-driven methodology for probabilistic wind power ramp forecasting," *IEEE Trans. Smart Grid*, to be published, doi: 10.1109/TSG.2017.2763827.
[51] L. Breiman, "Bagging predictors," *Mach. Learn.*, vol. 24, pp. 123–140, 1996.
[52] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. New York, NY, USA: Taylor & Francis, 1984.



Yu Fujimoto (M'17) received the Ph.D. degree in engineering from Waseda University, Tokyo, Japan, in 2007. He is an Associate Professor with the Advanced Collaborative Research Organization for Smart Society, Waseda University. His primary areas of interest include machine learning and statistical data analysis. His current research interests include data mining in energy domains especially for controlling power in smart grids, and statistical prediction of the power fluctuation under the large introduction of renewable energy sources. He is a member of the Information Processing Society of Japan.

Yuka Takahashi received the B.Eng. and M.Eng. degrees from Waseda University, Tokyo, Japan, in 2015 and 2017, respectively. Her research focuses on forecasting of wind power output.

Yasuhiro Hayashi (M'91) received the B.Eng., M.Eng., and D.Eng. degrees from Waseda University, Tokyo, Japan, in 1989, 1991, and 1994, respectively. In 1994, he became a Research Associate with Ibaraki University, Mito, Japan. In 2000, he became an Associate Professor with the Department of Electrical and Electronics Engineering, Fukui University, Fukui, Japan. He has been with Waseda University as a Professor with the Department of Electrical Engineering and Bioscience since 2009; and as a Director of the Research Institute of Advanced Network Technology since 2010. Since 2014, he has been a Dean of the Advanced Collaborative Research Organization for Smart Society, Waseda University. His current research interests include optimization of distribution system operation and forecasting, operation, planning, and control concerned with renewable energy sources and demand response. Prof. Hayashi is a member of the Institute of Electrical Engineers of Japan and a Regular Member of CIGRE SC C6 (Distribution Systems and Dispersed Generation).