Toward Cost-Oriented Forecasting of Wind Power Generation

Guo Li, *Student Member, IEEE*, and Hsiao-Dong Chiang, *Fellow, IEEE*

***Abstract*—Forecasting is considered to be one of the most cost-efficient solutions to integrating wind power into existing power systems. In some applications, unbiased forecasting is necessary, while in others, the forecasting value can be biased for optimal decision making. In this paper, we study optimal point forecasting problems under cost-oriented loss functions, which can lead to a forecasting process that is far more sensitive to the actual cost associated with forecasting errors. Theoretical points of optimal forecasting under different loss functions are illustrated, then a cost-oriented, boosted regression tree method is presented to formulate the optimal forecasting problem under study. Case studies using real wind farm data are conducted. A comparison between cost-oriented forecasting and traditional unbiased forecasting demonstrates the efficiency of the proposed method in maximizing benefits for the decision-making process.**

***Index Terms*—Optimal point forecasting, wind power, loss function, boosted regression tree, forecasting error.**

I. INTRODUCTION

RENEWABLE energy, such as wind and solar energy, is an important generation alternative to traditional means of power generation. In many countries, renewable energy has constituted a noticeable percentage of the total energy supply [1], [2]. However, the uncertain nature of renewable power generation leads to challenges in integrating it into existing electricity systems that were designed mainly for traditional power generation. The rapid expansion of renewable energy has created a demand for advanced techniques. Among them, renewable power generation forecasting is considered to be one of the most cost-efficient solutions. Accurate forecasting provides useful information for grid operation [3] and grid security evaluation [4], and plays a key role in applications such as the electricity market [5], unit commitment [6], and economic dispatch [7].

A large amount of research, focused on forecasting renewable power generation as accurately as possible, has resulted in unbiased forecasts with greater accuracy through using more sophisticated techniques. These techniques can be divided into three groups [8]: 1) physical models that simulate the converting process and calculate the output power directly [9], 2) statistical methods that analyze a historical power output time series [10], [11], and 3) machine learning methods that learn the relationship between power output and numerical inputs [12], [13]. A combination of these three methods can be used for accurate forecasting results [14].

While traditional unbiased point forecasting has captured much attention in some applications, such as power system security assessment, in other applications, forecasting results are used to guide optimal solutions for maximizing benefits. In these applications, as the cost of under-forecasting renewable power generation (the forecasting value is smaller than the real value) and over-forecasting (the forecasting value is larger than the real value) can be very different, the forecasting value can be biased for optimal decision making [15]. For example, the monthly mean supplemental reserve energy cost of regulation up is always higher than that of regulation down [16].

This cost difference has thus motivated the development of alternative approaches to obtaining optimal forecasting to guide the decision-making process. Some of these approaches focus on the use of probabilistic and interval forecasting, which can provide decision makers with complete information about future renewable power generation. In [5], a probabilistic forecast is used to derive optimal bidding strategies in the electricity market. In [17], probabilistic wind power forecasts are used to guide wind power trading decisions in the day-ahead and real-time markets. Quantile forecasting is used to guide optimal decisions with given nominal levels [18]. In [19], interval forecasting of wind power generation is developed to provide a solution to estimating and quantifying the potential impacts and risks facing system operation with prior wind penetration.

On the other hand, the effects of the loss function, used to form forecasts in the decision-making process, are studied. In [20], it is first proposed that traditional forecasting limits itself to quadratic loss functions, but in practice, loss functions are frequently non-symmetric. In [21], the theory of optimal forecasting under asymmetric loss functions is developed, and the related optimal predictor is characterized. The asymmetric loss function is used in many application domains, such as email spam-filtering problems [22], where the cost of losing a piece of good mail is much greater than receiving a few spams. In [23], a non-symmetric penalty function with different penalties for over-forecasting and under-forecasting is

Manuscript received March 16, 2016; revised July 11, 2016 and August 31, 2016; accepted September 18, 2016. Date of publication September 28, 2016; date of current version June 19, 2018. This work was supported in part by the National Nature Science Foundation of China under Grant 51337007, and in part by the National Science Foundation, USA, under Grant 1508986. Paper no. TSG-00340-2016.

G. Li is with the School of Electrical Engineering and Automation, Tianjin University, Tianjin 300072, China (e-mail: li_guo@tju.edu.cn).

H.-D. Chiang is with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY 14853 USA (e-mail: chiang@ece.cornell.edu).

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/TSG.2016.2614341

used to solve short-term load forecasting problems in power systems. This is actually a short-term load forecasting problem under an asymmetric loss function. In [24], it is proposed that in real-world solar energy forecasting scenarios, the loss function is connected to how energy producers, balancers, or traders apply the forecast and is often asymmetric.

In this paper, we propose a loss function-based forecasting method and focus on the optimal point forecasting problem of wind power generation, aiming to maximize the benefits resulting from the use of point forecasting. The structure of loss function used in our model-building process is piecewise so that the cost associated with forecasting errors can be accurately calculated. From an analytical viewpoint, the optimal point forecast is biased and impossible to solve analytically. Consequently, a boosted regression tree method is used to solve this problem.

The main contributions of this paper are:

1) To develop a point forecast for wind power generation problems using cost-oriented loss functions, which can incorporate the actual costs associated with forecasting errors into a model building and forecasting process. The proposed point forecasting does not need the additional assumptions or extensive simulation that interval forecasting (or probabilistic forecast) requires and is easy to handle or interpret by forecast users.

2) To characterize optimal forecasting under the quadratic loss function and cost-oriented loss function from a mathematical point of view.

3) To use a cost-oriented boosted regression tree method to perform the optimal point forecast, making it capable of handling the cost-oriented loss function efficiently.

The rest of this paper is organized as follows. In Section II, the formulation of forecasting problems and related optimal solutions are analyzed. In Section III, details of the cost-oriented boosted regression tree method are described. Finally, the test results and validation of the proposed approach are shown in Section IV.

II. Problem Description

Because of uncertainties in the process of wind power generation, it can be considered as a stochastic process. Point forecasting of wind power generation is issued at time t to estimate a single value of the stochastic process for a future time at $t+k$.

In this section, the definition of a traditional point forecast that aims at unbiased prediction is introduced first. Then, the definition of an optimal point forecast is presented. After analyzing the quadratic loss function, we show that traditional point forecasting is actually the optimal forecast under the quadratic loss function. Finally, we propose using the cost-oriented loss function in the forecasting process and then analyzing the related optimal forecast.

A. Traditional Point Forecasts

Point forecasts of wind power generation have been studied broadly. A definition is given in [18]. A traditional point forecast of wind power generation $\hat{y}_{t+k}$ corresponds to the conditional expectation of the stochastic process Y_{t+k}. Given information set Ω_t, which contains all of the data and knowledge to be used in characterizing the properties of Y_{t+k} up to time t, a traditional point forecast builds a forecasting model to estimate

$$\hat{y}_{t+k} = E[Y_{t+k}|\Omega_t], \tag{1}$$

where k is the lead time, and $t+k$ is the k-step-ahead forecast made at time t.

B. Optimal Point Forecast

The optimal point forecast is the one with the smallest loss on average, that is, the point forecast $\hat{y}_{t+k}$ is chosen to minimize

$$E\big[L\big(Y_{t+k}, \hat{y}_{t+k}|\Omega_t\big)\big] = \int_{-\infty}^{\infty} L\big(y_{t+k}, \hat{y}_{t+k}\big) f(y_{t+k}|\Omega_t) dy_{t+k}, \tag{2}$$

where $L(\cdot)$ is the loss function that is considered to calculate the costs associated with forecasting errors, y_{t+k} is all of the potential realizations of the stochastic process at time $t+k$, and $\hat{y}_{t+k}$ is the k-step-ahead forecast. $f(y_{t+k}|\Omega_t)$ is the conditional frequency function of y_{t+k}, given Ω_t.

C. Quadratic Loss Function and Optimal Forecast

The quadratic loss function is often considered as follows:

$$L_q\big(y, \hat{y}\big) = \big(y - \hat{y}\big)^2. \tag{3}$$

Under the quadratic loss function, substitute (3) into (2) and the conditional expectation becomes

$$E\Big[\big(Y_{t+k} - \hat{y}_{t+k}\big)^2\Big|\Omega_t\Big] = E\Big[(Y_{t+k} - M)^2|\Omega_t\Big] + E\Big[\big(M - \hat{y}_{t+k}\big)^2\Big|\Omega_t\Big]. \tag{4}$$

So, the optimal point forecast is given by the one that minimizes (4),

$$\hat{y}_{t+k} = M = E[Y_{t+k}|\Omega_t]. \tag{5}$$

Equation (5) is equal to (1), that is to say, the traditional point forecast is equal to the optimal point forecast under the quadratic loss function, as they are all conditional expectations of the stochastic process.

D. Cost-Oriented Loss Function and Optimal Forecast

Although the quadratic loss function is widely used, in some applications, it is not ideal. The cost-oriented loss function is proposed for these reasons:

1) Depending on the types and capacity of the renewable energy to be installed, the issues related to forecasting errors are different [25].

2) The users of forecasting results are different. Forecasting results can be used by renewable power producers or grid operators, and their benefits are different [26].

3) The costs of over-forecasting and under-forecasting are different. For example, the costs of regulation up and regulation down are different [16].

We consider a general cost-oriented loss function:

$$L_{co}(y,\hat{y}) = \begin{cases} l_1(y,\hat{y}), & if \quad -\infty < y-\hat{y} < \delta_1 \\ \cdots \\ l_i(y,\hat{y}), & if \quad \delta_{i-1} \le y-\hat{y} < \delta_i \\ \cdots \\ l_n(y,\hat{y}), & if \quad \delta_{n-1} \le y-\hat{y} < +\infty, \end{cases} \tag{6}$$

where δ is the breakpoint between segments.

The cost-oriented loss function $L_{co}(y,\hat{y})$ should satisfy three conditions [27]:

1) $L_{co}(0) = 0$. This means that there is no cost if the forecasting error is zero.

2) $\min L_{co}(y,\hat{y}) = 0$. This means that the cost is greater than or equal to zero.

3) $L_{co}(y,\hat{y})$ is monotonically non-decreasing as it moves away from zero. This means that, on each side of the origin, a bigger absolute value of the forecasting error leads to larger cost.

Under the cost-oriented loss function, the optimal point forecast is the one with the smallest cost on average. In (6), the cost associated with the forecasting error is incorporated into the loss function. So, by substituting (6) into (2), the optimal forecast is

$$\begin{aligned} &\min_{\hat{y}_{t+k}} E\left[L_{co}\left(Y_{t+k}, \hat{y}_{t+k}|\Omega_t\right)\right] \\ &\quad = \min_{\hat{y}_{t+k}} \int_{-\infty}^{\delta_1} l_1\left(y_{t+k}, \hat{y}_{t+k}\right) f(y_{t+k}|\Omega_t) dy_{t+k} \\ &\qquad + \cdots + \int_{\delta_{i-1}}^{\delta_i} l_i\left(y_{t+k}, \hat{y}_{t+k}\right) f(y_{t+k}|\Omega_t) dy_{t+k} \\ &\qquad + \cdots \int_{\delta_{n-1}}^{+\infty} l_n\left(y_{t+k}, \hat{y}_{t+k}\right) f(y_{t+k}|\Omega_t) dy_{t+k}. \end{aligned} \tag{7}$$

It is not possible to solve analytically for the optimal forecast in general, except for several simple loss functions, such as the LinLin loss function and the LinEx loss function [21].

The following theorem shows that the optimal forecast under the cost-oriented loss function is biased and that the bias is time-varying in general and depends on higher-order conditional moments.

Theorem 1 [21]: If $Y_{t+k}|\Omega_t$ has conditional mean $\mu_{t+k|t}$ and a vector of (possibly time-varying) conditional moments of order two and higher $\lambda_{t+k|t}$, and $L(e_{t+h})$ is any loss function defined on the k-step-ahead prediction error e_{t+h}, then the optimal prediction is of the form $\hat{y}_{t+k} = \mu_{t+k|t} + \alpha_{t+k|t}$, where $\alpha_{t+k|t}$ depends only on the loss function and $\lambda_{t+k|t}$.

As the optimal forecast under a cost-oriented loss function is a time-varying variance, the conditional mean in (5) is always a sub-optimal solution of (7). The manner in which a constant bias term is added to the conditional mean is also sub-optimal. So, a forecasting method that is capable of integrating the cost-oriented loss function into the model building and forecasting process is developed to capture the optimal point forecast in the next section.

III. Proposed Method

To evaluate the cost associated with the forecasting errors of wind generation power, the form of the cost-oriented loss function is piecewise and asymmetric most of the time, so it is not differentiable at the breakpoints.

Most of the former point forecast methods, such as multiple linear regression, ARIMA models, the back propagation neural network, and more, assume that the quadratic loss function is built in. Unlike cost-oriented loss functions, the quadratic loss function is differentiable for the whole range of $-\infty$ to ∞. Furthermore, these methods are specially designed to introduce the quadratic loss function in the model building process and then carry over in the forecasting process. Hence, these methods are ineffective to handle cost-oriented loss functions [23].

On the other hand, the cost-oriented loss functions are related to many factors, such as the status of the power systems, the users of forecasting results, and so on. Depending on these factors, the loss functions may change every day or even at different hours. When the loss functions change, new forecasting models should be built to capture the optimal point forecasts under these loss functions. As the loss functions may change frequently, the training time of the forecasting models should be shorter than the loss functions' changing time to make it practical.

To overcome these problems, a cost-oriented boosted regression tree method (COBRT) is used in this paper. This method is capable of handling cost-oriented loss function efficiently, and the model building time of this method is able to meet the requirements of real-world applications. It uses two techniques—regression trees and boosting—to incorporate the actual costs of regulation into the model building and forecasting process. The proposed method is discussed in detail next.

A. Regression Trees and Boosting

A regression tree is a non-parametric learning technique proposed in [28]. Unlike parametric learning algorithms, such as general linear model, neutral network, and support vector machine, a regression tree does not explicitly assume a specified parametric functional form in the learning process.

A regression tree uses a top-down and divide-conquer procedure to recursively partition the observations into smaller and smaller non-overlapping regions. The partitioning process is binary, and the observations in each region are assigned to have homogeneous responses. Then it fits the mean response for all of the observations in the same region as their output. The formulation of the regression tree is

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m), \tag{8}$$

where M is the number of regions, c_m is the mean response of the observations in each region, R_m is the partition of each region, and $I(x \in R_m)$ is an indicator function that equals 1 if $x \in R_m$ and 0 otherwise.

Given training data consisting of N observations (x_i, y_i) where $i = 1, \ldots, N$, each observation has p inputs and a response. The steps for building a regression tree are described as follows.

Step 1) Start with a single region containing all of the observations. Then search over all binary splits of all variables to find the jth splitting variable $x^{(j)}$ and split point s, which will reduce the error as much as possible. To find the best splitting variable and split point, the searching process solves

$$\min_{j,s}\left[\min_{c_1}\sum_{x_i\in R_1(j,s)}(y_i-c_1)^2+\min_{c_2}\sum_{x_i\in R_2(j,s)}(y_i-c_2)^2\right]. \quad (9)$$

Step 2) In each new region, step 1) is repeated to take the split and create two new regions.

Step 3) The splitting process is repeated until a stopping criterion is reached, for instance, when the minimum number of observations per region is reached.

Boosting is a method for improving model accuracy, originally developed for classification problems [29] and extended to regression problems [30]. Boosting is derived from the computational learning theory that weakly learnable and strongly learnable problems are equal [31]. It improves model performance by fitting many base learners (weakly learnable) and combining them together to obtain a more accurate model (strongly learnable). In the boosting process, base learners are sequentially applied to the training data at each iteration, and the training data are modified to increase the emphasis on observations modeled poorly by the existing base learners.

The boosting model is built in a forward stage-wise manner, sequentially adding base learners to the model without adjusting their parameters, one at a time, until all of them are in the model. The formulation of the boosted model is

$$F_M(x)=\sum_{m=0}^{M}h(x;a_m), \quad (10)$$

where M is the number of base learners, $h(x;a_m)$ is the base learner, and a_m is the parameter of the base learner.

B. Cost-Oriented Boosted Regression Tree

Using regression trees as base learners and boosting a cost-oriented loss function, we are able to incorporate the actual costs associated with the forecasting errors of renewable generation power into the model building and forecasting process.

Given a cost-oriented loss function (6), the optimal forecast is the solution to (7). Although it is not possible to solve (7) analytically, we are able to apply optimization in the function space to solve it. In the boosting process, instead of fitting current residuals, a regression tree that best fits the negative gradient vector of (6), which is a descent direction of (7) in function space, is added to the expansion at each step.

The steps of the method are described as follows:

Step 1) Initialize $F_0(x)=0$.

Step 2) For $m=1,\ldots,M$, do the following:

a) Compute the negative gradient of the current model:

$$\widetilde{y_i}=-\begin{cases}\left[\frac{\partial l_1(y,F(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{m-1}(x_i)}, & -\infty<y-F(x_i)<\delta_1\\ \ldots & \\ \left[\frac{\partial l_i(y,F(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{m-1}(x_i)}, & \delta_{i-1}\le y-F(x_i)<\delta_i\\ \ldots & \\ \left[\frac{\partial l_n(y,F(x_i))}{\partial F(x_i)}\right]_{F(x_i)=F_{m-1}(x_i)}, & \delta_{n-1}\le y-F(x_i)<+\infty,\end{cases}$$
$$i=1,\ldots,N. \quad (11)$$

b) Build a regression tree $T(x;\Theta_m)$ using $\{x_i, \widetilde{y_i}\}$ as training data, $i=1,\ldots,N$, where Θ_m is the structure of the regression tree.

c) Update the current model:

$$F_m(x)=F_{m-1}(x)+\rho_m T(x;\Theta_m), \quad (12)$$

where ρ_m is the step size.

Step 3) The model is built when all of the M regression trees are added sequentially.

Step 4) Predict renewable generation power using the generated model.

To prevent an overfitting problem, instead of the step size ρ_m, a shrinkage parameter λ [32], which is often a small positive number, is used in c) of step 2):

$$F_m(x)=F_{m-1}(x)+\lambda T(x;\Theta_m). \quad (13)$$

Typical values of the shrinkage parameter are 0.01 or 0.001. The flowchart of the cost-oriented boosted regression tree method is illustrated in Fig. 1.

IV. Test Results

In this section, the proposed method is tested on a wind power trading problem in electricity markets, with the aim of maximizing the benefit of wind power producers. The background of this problem is described first. Then, we evaluate the proposed method using real-world wind power generation data under different criteria. Finally, the performance and applicability of the proposed method is discussed.

A. Background

An electricity market typically includes three trading floors: the day-ahead market, the adjustment market, and the balancing market [33]. The day-ahead market takes place the day prior to energy delivery. Power producers submit production offers to this market. The market operator calculates power prices based on supply and demand for every hour the following day. The adjustment market is similar to the day-ahead market but is cleared closer to power delivery and covers a shorter trading horizon. The balancing market is used to ensure the correct frequency in the grid and to supply security. It is often operated by the transmission system operator (TSO) and provided by energy sources that are expensive but dispatchable, such as combined-cycle gas turbines.

In the electricity markets, wind power producers participate in different trading floors. They trade energy in the day-ahead and adjustment markets based on their forecasts first, then participate in the balancing markets to cover the deviations

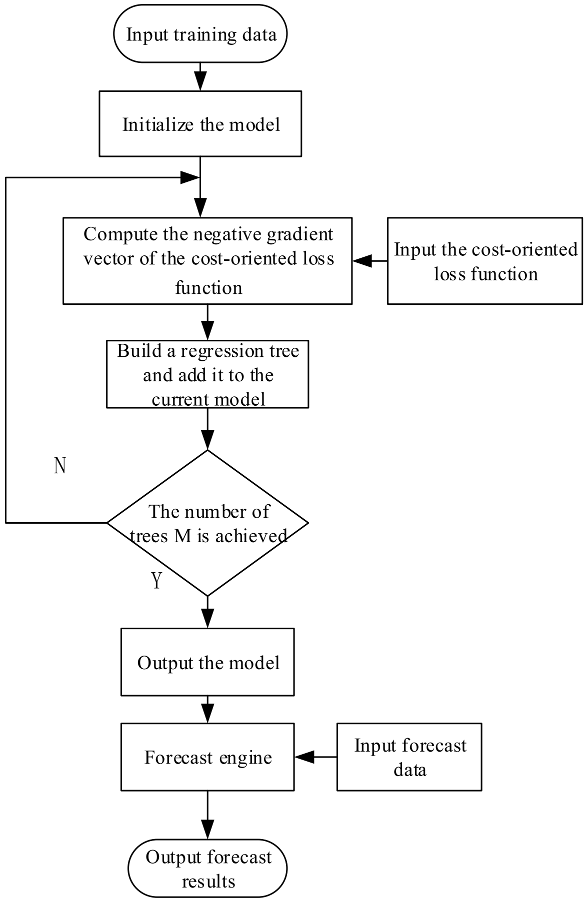


Fig. 1. Flowchart of the proposed method.

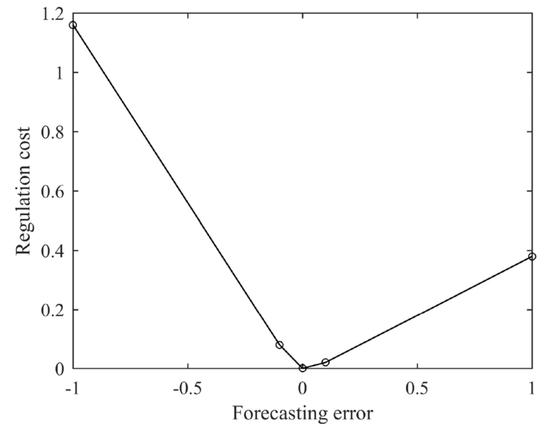


Fig. 2. The relationship between the regulation cost and forecasting errors.

from the production pattern settled in the markets. They are forced to pay the imbalance costs between predicted wind power production and actual production. The regulation prices for positive and negative imbalances depend on the regulation mechanism and are generally asymmetric.

The decision-making process of wind power producers seeks to maximize their revenue in the electricity markets. The revenue of a wind power producer includes income from the selling of actual wind power in the day-ahead and adjustment markets, minus the costs for regulation in the balancing market. It is demonstrated in [5] that maximizing the revenue is equal to minimizing the costs for regulation, since the income is dependent only on actual wind power, and the contracted energy merely appears in the regulation costs.

As the regulation prices for positive and negative imbalances are generally different, the optimal forecast is biased. By incorporating the regulation costs associated with forecasting errors into the loss function, we are able to capture optimal forecasts to maximize the revenue of wind power producers in the electricity markets.

B. Data Description

To make the test results comparable and reproducible, publicly available data, the power generation of wind farms from GEFCOM2012-WF [34] is used to conduct the case study.

The data set contains three years of historical data, from 2009/7/1 to 2012/6/28, for seven wind farms. The historical data include both wind power generation measurements and meteorological forecasts of the wind components. The period between 2009/7/1 and 2010/12/31 is a model identification and training period. The target is to forecast the hourly wind power generation for seven wind farms from 2011/1/1 to 2012/6/28.

The meteorological forecasts of the wind components contain the forecasts of zonal and meridional components of surface winds (u and v), wind speed (s), and wind direction. They are issued every 12 h, with a forecast horizon of 48 h and an hourly temporal resolution. The wind power measurements are normalized to take values between 0 and 1 to keep the original characteristics of the wind farms unrecognizable.

On the other hand, it is assumed that estimating the regulation prices for positive and negative imbalances is possible. One method of estimating regulation prices is illustrated in [35]. As described in [18], when settled through the balancing market, the costs of underestimating and overestimating renewable generation power are different. Furthermore, the costs of regulation may rise substantially and suddenly during a phase when the cheapest reserves have already been used and the more expensive new reserves must be allocated [36].

Based on the data and examples of the imbalance costs given in [5] and [36], we consider the regulation price for a forecasting error in the range $(-\infty, -0.1)$ to be 1.2, $[-0.1, 0)$ is 0.8, $[0, 0.1)$ is 0.2, and $[0.1, \infty)$ is 0.4. As the wind power measurements in the data set are normalized to take values between 0 and 1, the regulation price is unitless.

So, the regulation costs can be calculated as follows:

$$
\begin{aligned}
&e = y - \hat{y},\\
&C(e) = \begin{cases} -1.2e - 0.04, & \textit{if} \quad -\infty < e < -0.1 \\ -0.8e, & \textit{if} \quad -0.1 \le e < 0 \\ 0.2e, & \textit{if} \quad 0 \le e < 0.1 \\ 0.4e - 0.02, & \textit{if} \quad 0.1 \le e < +\infty, \end{cases}
\end{aligned}
\tag{14}
$$

where e is the forecasting error, y is the real wind power generation value, $\hat{y}$ is the forecasting value, and C is the regulation cost. The relationship between the regulation costs and forecasting errors is shown in Fig. 2.

Using the regulation costs as the loss function of the proposed method, we are able to capture optimal forecasts to maximize the revenue of wind power producers in the electricity markets.

C. Evaluation Criteria

To evaluate the overall forecasting performance of the proposed method, several different evaluation criteria are used.

1) Total cost associated with forecasting errors (TCFE). The TCFE is a measure of cost by using forecasting. It is an extension of the evaluation criteria proposed in [16] and [37]. The TCFE can be calculated as follows:

$$TCFE = \sum_{i=1}^{n} C(e_i) \cdot e_i, \tag{15}$$

where e_i is the forecasting error of the ith test sample, and n is the number of total test samples.

2) Root mean square error (RMSE). The RMSE is a widely used measure of accuracy. It can be calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y - \hat{y}\right)^2}. \tag{16}$$

Moreover, two additional criteria are used to analyze the error distribution of the forecasting results.

1) Skewness. To measure the asymmetry of the probability distribution of a forecasting error, skewness is used as a criterion. It measures the amount of asymmetry around the mean in a distribution. A zero value of skewness indicates that the distribution is symmetric around the mean. The larger the absolute size of the skewness, the more asymmetrical the distribution. A positive value indicates a right tail distribution, and a negative value indicates a left tail distribution. This can be calculated as follows:

$$\hat{S} = \frac{\frac{1}{n} \cdot \sum_{i=1}^{n} (e_i - \bar{e})^3}{\left[\frac{1}{n} \cdot \sum_{i=1}^{n} (e_i - \bar{e})^2\right]^{\frac{3}{2}}}, \tag{17}$$

where $\bar{e}$ is the mean of forecasting errors for all test samples.

2) Percentage of over-forecasting (OF%) and under-forecasting (UF%). Since skewness measures the amount of asymmetry around the mean, OF% and UF% are used to measure the amount of asymmetry around zero. OF% indicates the percentage of over-forecasting samples in the test set, and UF% indicates the percentage of under-forecasting samples. They can be calculated as follows:

$$OF\% = \frac{n_{of}}{n} \times 100\%,$$
$$UF\% = \frac{n_{uf}}{n} \times 100\%, \tag{18}$$

where n_{of} is the number of over-forecasting samples, and n_{uf} is the number of under-forecasting samples.

D. Comparison and Parameter Settings

For the purpose of comparison, we benchmark the proposed COBRT method with two additional wind power trading strategies in the electricity markets.

1) Least-squares boosted regression tree (LSBRT) method. This method uses the quadratic loss function and results in traditional unbiased forecasts.

2) Adding a constant bias term to the forecasting results of the LSBRT (LSBRT-B). A useful manner of taking into account cost-oriented loss functions is to add a constant bias term to the unbiased forecasts [20].

Let $\hat{F}_c(x)$ be the empirical cumulative distribution function of the forecasting errors. The constant bias term α_0 is found by solving:

$$\hat{F}_c(\alpha_0) = \frac{C_{over}}{C_{over} + C_{under}}, \tag{19}$$

where C_{over} is the cost of over-forecasting and C_{under} is the cost of under-forecasting. $\hat{F}_c(x)$, C_{over}, and C_{under} are all estimated using the distribution of training errors.

All of these methods are implemented in the MATLAB computing environment on a laptop computer with Intel Core i5-3230M CPU @ 2.6 GHz and 8.00 GB RAM.

In the model-building process, a sub-model is built for each farm and every 12-hour forecast interval. As there are seven wind farms with a forecast horizon of 48h, 28 sub-models are built.

To make the comparison fair, we use the same features and parameters in the training process. This may be not optimal for both methods, and in real-world applications, some strategies, such as cross-validation, are needed to choose the features and parameters. The features used are the forecasting-issued time (year, month, day, and hour), wind information (u, v, and s) of the target wind farm for three hours, and wind information (u, v, and s) for all of the other six wind farms for one hour. The dimension of features is $4+3\times3+3\times6\times1 = 31$. The number of regression trees is 30 000, and the shrinkage parameter is 0.001. It should be mentioned that the least-squares boosted regression trees received 0.15175 total RMSEs for the whole test data, which is a promising result compared with the top entries of GEFCOM2012 [14].

The average training time of the COBRT is 754.5 seconds and the LSBRT is 756.3 seconds. In most applications, forecasting processes are conducted with a time interval of at least 15 minutes, so the COBRT method is practical in these applications, even when the loss functions change frequently.

E. Results and Discussions

1) Residual plots. Using residual plots, the differences in the forecast results between the traditional LSBRT, LSBRT-B, and the proposed COBRT are shown separately in Fig. 3, Fig. 4, and Fig. 5. A residual plot is a graphical technique used in statistics that attempts to show the actual data, the forecasting value, and the residuals in a single graph. In Fig. 3, there are no obvious patterns in the residuals. The forecasting value is located around the real value. In other words, the possibility of overestimating or underestimating by traditional unbiased forecasting is almost the same. In Fig. 4, as a constant bias term is added to the forecasting results of the LSBRT, more residuals are above zero. In Fig. 5, obvious biased patterns can be seen in the residuals. The residuals are above zero most of the time. The forecasting value is nearly

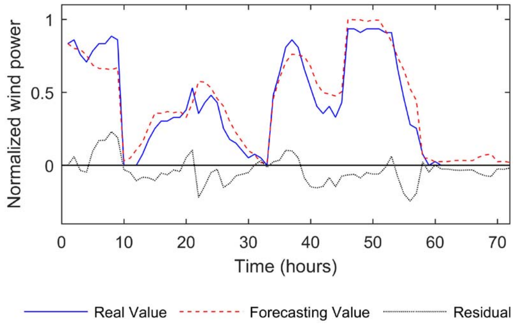


Fig. 3. Residual plot of a forecasting example using the traditional LSBRT.

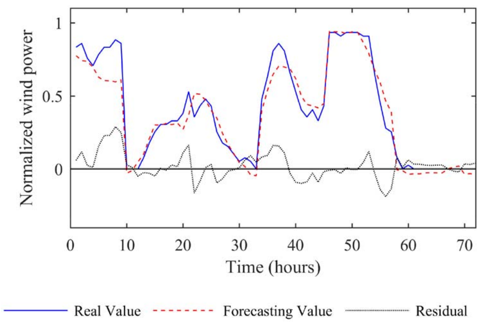


Fig. 4. Residual plot of a forecasting example using the LSBRT-B.

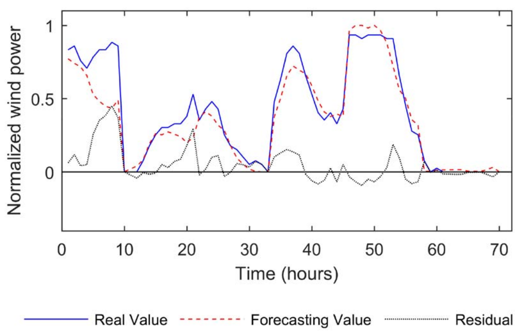


Fig. 5. Residual plot of a forecasting example using the proposed COBRT.

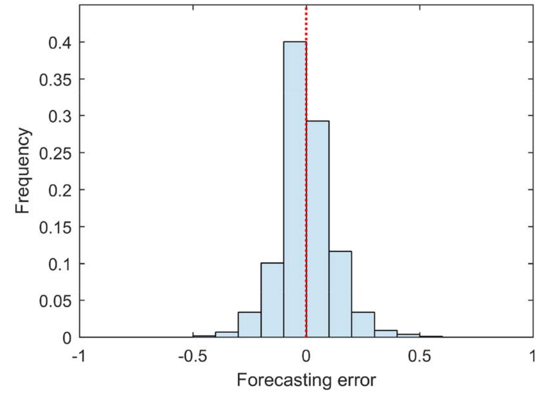


Fig. 6. Histogram of forecasting errors using the LSBRT.

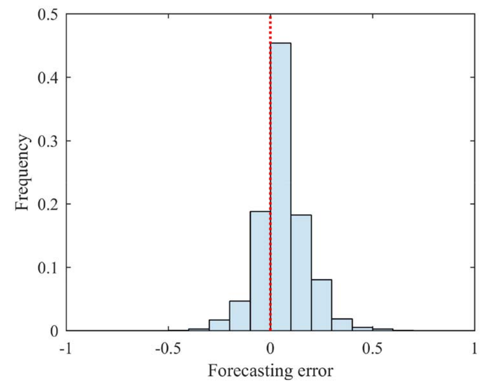


Fig. 7. Histogram of forecasting errors using the LSBRT-B.

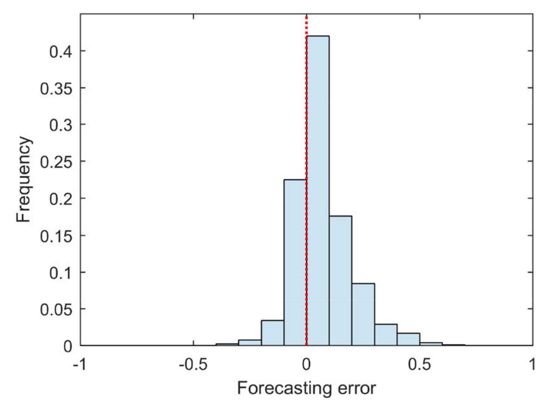


Fig. 8. Histogram of forecasting errors using the COBRT.

always smaller than the real value. That is to say, using the COBRT method, the forecast value is underestimated most of the time.

2) Overall performance. These methods are applied to all of the seven wind farms; the overall performance for each wind farm and the totals are shown in Table I. From this table, it can be seen that the values of TCFEs are always smaller using the COBRT method than when using the LSBRT and LSBRT-B methods. The total TCFE is 3449.2 using the LSBRT method, 2760.2 using the LSBRT-B, and 2653.7 using the COBRT. Compared with the LSBRT, the total reduction of the seven wind farms is 23.06% using the COBRT. That is to say, using the proposed COBRT method in the decision-making process, the optimal solution that maximizes the benefits can be obtained. This is because the forecast value using the COBRT is biased and tends to be underestimated most of the time, as illustrated in Fig. 5. The LSBRT-B method can also reduce the values of TCFEs significantly. However, as the optimal forecast under a cost-oriented loss function is a time-varying variance, this method is sub-optimal [21].

At the same time, the values of RMSEs are always smaller using the LSBRT method than using the COBRT method. That is to say, the LSBRT results in more accurate forecasting values. From the formulation of the RMSE in (16) and the quadratic loss function in (3), it can be seen that this is because the LSBRT is built to minimize RMSEs.

3) Error distributions of forecasting results. Error distributions for the LSBRT and COBRT methods are shown in Fig. 6, Fig. 7, and Fig. 8. For the LSBRT, although an unbiased loss function is used, the skewness is not zero, which means the error distribution is a bit asymmetric. 44.82% of

TABLE I
OVERALL PERFORMANCE OF THE LSBRT, LSBRT-B, AND COBRT METHODS

		Wind farm 1	Wind farm 2	Wind farm 3	Wind farm 4	Wind farm 5	Wind farm 6	Wind farm 7	Total
LSBRT	TCFE	437.6	526.8	591.7	464.1	546.9	432.5	449.6	3449.2
	RMSE	0.1380	0.1512	0.1736	0.1465	0.1684	0.1356	0.1449	0.1518
	$\hat{S}$	0.0997	0.2404	0.2665	-0.1327	0.3034	0.0681	-0.0081	0.1424
	UF%	47.01%	43.02%	45.14%	46.49%	41.76%	47.46%	42.90%	44.82%
	OF%	50.79%	54.61%	53.93%	50.59%	57.25%	52.11%	53.03%	53.19%
LSBRT-B	TCFE	352.2	408.9	472.0	384.8	429.0	349.8	363.5	2760.2
	RMSE	0.1551	0.1633	0.1871	0.1632	0.1806	0.1499	0.1596	0.1660
	$\hat{S}$	0.1070	0.2536	0.2842	-0.1220	0.3211	0.0837	0.0099	0.1651
	UF%	72.99	69.34	71.00	74.91	69.17	73.98	73.93	72.17
	OF%	27.01	30.66	29.00	25.09	30.83	26.01	26.07	27.82
COBRT	TCFE	357.7	386.0	450.2	367.5	399.9	342.0	350.4	2653.7
	RMSE	0.1734	0.1821	0.2073	0.1784	0.2029	0.1685	0.1794	0.1851
	$\hat{S}$	0.8234	0.9527	0.8128	0.6041	1.0492	0.8722	0.8300	0.8803
	UF%	68.00%	63.64%	66.71%	67.27%	65.52%	70.39%	62.46%	66.28%
	OF%	27.99%	31.37%	30.85%	26.64%	30.34%	27.98%	27.87%	29.01%

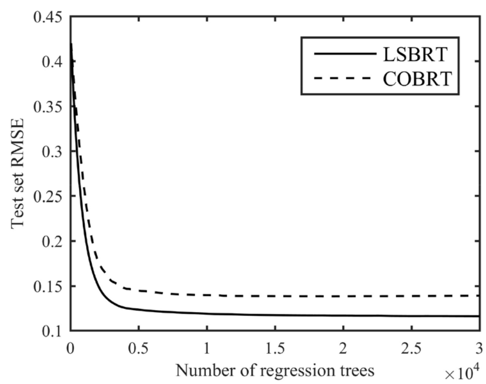


Fig. 9. The values of RMSEs after each regression tree is added.

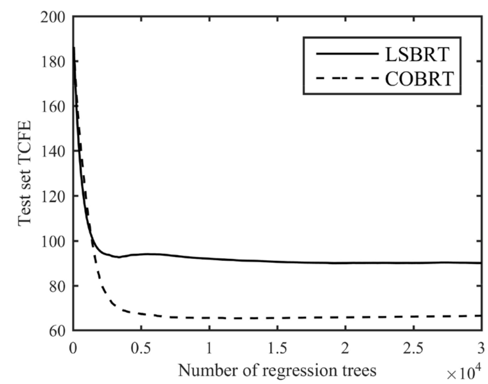


Fig. 10. The values of TCFEs after each regression tree is added.

the total test samples are under-forecasted, and 53.19% are over-forecasted. This is due to the form of the wind farm power curve [38]; for example, at the cut-out wind speed, the wind power may suddenly switch to zero, which leads the model to over-forecast. After adding a constant bias term to the forecasting results of the LSBRT, more samples are under-forecasted. 72.17% of the total test samples are under-forecasted whereas 27.82% are over-forecasted.

Compared with the LSBRT, the COBRT has larger positive skewness. That is to say, the forecasting errors follow a right tail distribution. 66.28% of the total test samples are under-forecasted and 29.01% are over-forecasted. That is, in a cost-oriented loss function, positive errors and negative errors are treated with a more flexible approach that can incorporate the actual costs associated with forecasting errors.

4) Effects of the loss functions. The LSBRT and COBRT are forecasting models based on boosting and regression trees. The difference between them is the loss functions they use.

We test the performance of each model using test data in the model building process. Fig. 9 and Fig. 10 illustrate the effect of loss functions on the test results. In Fig. 9, the RMSE values of each forecasting model in the test set are displayed after each regression tree is added. In Fig. 10, the values of the TCFEs are displayed. By comparing these two figures, it can be obtained that the LSBRT is built to minimize RMSEs and the COBRT is built to minimize TCFEs.

Using different loss functions, the LSBRT and COBRT change the contributions of each regression tree and lead to different forecasting results. The LSBRT uses quadratic loss functions, and each regression tree in the model fits a decent direction of the quadratic loss function in function space. The COBRT uses cost-oriented loss functions, so each regression tree in the model fits a decent direction of the cost-oriented loss function.

F. Some Further Results and Discussions

In real-world electricity markets, regulation prices are different from one quarter to another, or even from one month to another. Furthermore, since the types of energy and network structures are different, the behaviors of electricity markets also vary from one to another. So, the prices for upward and

TABLE II
PERFORMANCE OF THE FORECASTING MODELS UNDER DIFFERENT REGULATION PRICES

Regulation prices		[2.4, 1.6, 0.2, 0.4]	[0.6, 0.4, 0.2, 0.4]	[0.4, 0.2, 0.8, 1.2]	[0.4, 0.2, 1.6, 2.4]	[0.4, 0.2, 0.4, 0.6]
LSBRT	TCFE	6075.3	2136.2	3554.7	6352.8	2155.6
LSBRT-B	TCFE	3654.1	2062.5	2982.3	4089.2	2114.4
	RMSE	0.1960	0.1528	0.1692	0.1917	0.1564
	$\hat{S}$	0.1938	0.1437	0.1022	0.0893	0.1156
	UF%	84.99	53.46	25.44	18.24	33.41
	OF%	15.01	46.54	74.57	81.76	66.59
COBRT	TCFE	3426.8	1990.2	2837.4	3887.8	2042.9
	RMSE	0.2104	0.1603	0.1668	0.1877	0.1533
	$\hat{S}$	1.0666	0.4081	-0.2783	-0.4229	-0.0246
	UF%	72.54%	56.73%	28.07%	21.23%	37.84%
	OF%	22.22%	39.82%	71.12%	78.21%	60.72%

downward regulation can be different in different periods or different electricity markets.

This leads to the change in loss function when the proposed method is applied to maximize the benefit of wind power production. So, we test the performance of the forecasting models under different regulation prices. The test results are shown in Table II. The regulation prices are described in the same order as in (14). For example, [2.4, 1.6, 0.2, 0.4] means the regulation price for a forecasting error in the range $(-\infty, -0.1]$ is 2.4, $[-0.1, 0)$ is 1.6, $[0, 0.1)$ is 0.2, and $[0.1, \infty)$ is 0.4. Only the TCFE criterion of the LSBRT is shown in Table II, since the other performances remain the same as in Table I.

Table II demonstrates that different loss functions lead to different error distributions. When the regulation prices are highly asymmetric, such as [2.4, 1.6, 0.2, 0.4] and [0.4, 0.2, 1.6, 2.4], more costs can be reduced using the COBRT. However, this also leads to the increase in RMSEs.

So, in practice, when the regulation prices are highly asymmetric, a strategy that considers both RMSEs and TCFEs is needed.

V. CONCLUSION

We have proposed a cost-oriented forecasting formulation. Using this formulation, one can treat forecasting problems with flexible loss functions, which can incorporate the actual costs associated with forecasting errors into the model building and forecasting process. It is shown that an optimal forecast using the cost-oriented loss function is biased and that the bias is time-varying in general and depends on higher-order conditional moments. A cost-oriented boosted regression tree method is proposed to perform the optimal point forecast, which can guide the decision-making process efficiently.

Test results using real wind farm data indicate that the proposed cost-oriented formulation and the proposed method are effective when overestimating and underestimating lead to different costs. In the test case, compared with the traditional unbiased point forecasting method, the cost-oriented forecasting method can reduce 23.06% of the total cost, which is a promising result.

It should be noted that the proposed method can also be used to deal with a wide range of forecasting problems. Depending on the process characteristics of interest to forecast users, loss functions can be varied, and the related optimal forecasts can differ accordingly. In the future, we will extend optimal forecasting problems to considering the influences of forecasted renewable power generation on power networks, for example, power system voltage stability and power system congestion.

REFERENCES

[1] T. Stetz *et al.*, "Twilight of the grids: The impact of distributed solar on Germany's energy transition," *IEEE Power Energy Mag.*, vol. 13, no. 2, pp. 50–61, Mar./Apr. 2015.

[2] L. Jiang *et al.*, "Growth in wind and sun: Integrating variable generation in China," *IEEE Power Energy Mag.*, vol. 13, no. 6, pp. 40–49, Nov./Dec. 2015.

[3] N. Gast, D.-C. Tomozei, and J.-Y. Le Boudec, "Optimal generation and storage scheduling in the presence of renewable forecast uncertainties," *IEEE Trans. Smart Grid*, vol. 5, no. 3, pp. 1328–1339, May 2014.

[4] R. Karki, P. Hu, and R. Billinton, "A simplified wind power generation model for reliability evaluation," *IEEE Trans. Energy Convers.*, vol. 21, no. 2, pp. 533–540, Jun. 2006.

[5] P. Pinson, C. Chevallier, and G. N. Kariniotakis, "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1148–1156, Aug. 2007.

[6] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Trans. Power Syst.*, vol. 23, no. 3, pp. 1319–1327, Aug. 2008.

[7] A. Botterud *et al.*, "Demand dispatch and probabilistic wind power forecasting in unit commitment and economic dispatch: A case study of Illinois," *IEEE Trans. Sustain. Energy*, vol. 4, no. 1, pp. 250–261, Jan. 2013.

[8] B. Ernst *et al.*, "Predicting the wind," *IEEE Power Energy Mag.*, vol. 5, no. 6, pp. 78–89, Nov./Dec. 2007.

[9] L. Landberg, "A mathematical look at a physical power prediction model," *Wind Energy*, vol. 1, no. 1, pp. 23–28, 1998.

[10] J. L. Torres, A. García, M. De Blas, and A. De Francisco, "Forecast of hourly average wind speed with ARMA models in Navarre (Spain)," Solar *Energy*, vol. 79, no. 1, pp. 65–77, 2005.

[11] J. P. S. Catalao, H. M. I. Pousinho, and V. M. F. Mendes, "Hybrid wavelet-PSO-ANFIS approach for short-term wind power forecasting in Portugal," *IEEE Trans. Sustain. Energy*, vol. 2, no. 1, pp. 50–59, Jan. 2011.

[12] Q. Xu *et al.*, "A short-term wind power forecasting approach with adjustment of numerical weather prediction input by data mining," *IEEE Trans. Sustain. Energy*, vol. 6, no. 4, pp. 1283–1291, Oct. 2015.

[13] J. Shi *et al.*, "Hybrid forecasting model for very-short term wind power forecasting based on grey relational analysis and wind speed distribution features," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 521–526, Jan. 2014.

[14] D. Lee and R. Baldick, "Short-term wind power ensemble prediction based on Gaussian processes and neural networks," *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 501–510, Jan. 2014.

[15] J. Tastu, P. Pinson, P.-J. Trombe, and H. Madsen, “Probabilistic forecasts of wind power generation accounting for geographically dispersed information,” *IEEE Trans. Smart Grid*, vol. 5, no. 1, pp. 480–489, Jan. 2014.

[16] A. Fabbri, T. G. S. Roman, J. R. Abbad, and V. H. M. Quezada, “Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market,” *IEEE Trans. Power Syst.*, vol. 20, no. 3, pp. 1440–1446, Aug. 2005.

[17] A. Botterud *et al.*, “Wind power trading under uncertainty in LMP markets,” *IEEE Trans. Power Syst.*, vol. 27, no. 2, pp. 894–903, May 2012.

[18] J. M. Morales, *Integrating Renewables in Electricity Markets: Operational Problems* (International Series in Operations Research & Management Science). New York, NY, USA: Springer, 2014.

[19] C. Wan, Z. Xu, P. Pinson, Z. Y. Dong, and K. P. Wong, “Optimal prediction intervals of wind power generation,” *IEEE Trans. Power Syst.*, vol. 29, no. 3, pp. 1166–1174, May 2014.

[20] C. W. J. Granger, “Prediction with a generalized cost of error function,” *J. Oper. Res. Soc.*, vol. 20, no. 2, pp. 199–207, 1969.

[21] P. F. Christoffersen and F. X. Diebold, “Optimal prediction under asymmetric loss,” *Econom. Theory*, vol. 13, no. 6, pp. 808–817, 1997.

[22] J. He and B. Thiesson, “Asymmetric gradient boosting with application to spam filtering,” in *Proc. 4th Conf. Email Anti-Spam (CEAS)*, Mountain View, CA, USA, 2007.

[23] H. Kebriaei, B. N. Araabi, and A. Rahimi-Kian, “Short-term load forecasting with a new nonsymmetric penalty function,” *IEEE Trans. Power Syst.*, vol. 26, no. 4, pp. 1817–1825, Nov. 2011.

[24] I. Khabibrakhmanov, S. Lu, H. F. Hamann, and K. Warren, “On the usefulness of solar energy forecasting in the presence of asymmetric costs of errors,” *IBM J. Res. Develop.*, vol. 60, no. 1, pp. 7:1–7:6, Jan./Feb. 2016.

[25] Y. Sasaki, N. Yorino, and L. Ma, “A day-ahead regional PV generation forecast applied to micro EMS,” in *Proc. 13th Int. Workshop Elect. Power Control Centers (EPCC)*, Bled, Slovenia, May 2015.

[26] P. H. Nguyen, W. L. Kling, and P. F. Ribeiro, “A game theory strategy to integrate distributed agent-based functions in smart grids,” *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 568–576, Mar. 2013.

[27] C. W. J. Granger, “Outline of forecast theory using generalized cost functions,” *Spanish Econ. Rev.*, vol. 1, no. 2, pp. 161–173, 1999.

[28] L. Breiman, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth, 1984.

[29] Y. Freund and R. E. Schapire, “A decision-theoretic generalization of on-line learning and an application to boosting,” *J. Comput. Syst. Sci.*, vol. 55, pp. 119–139, Aug. 1997.

[30] J. H. Friedman, “Greedy function approximation: A gradient boosting machine,” *Ann. Stat.*, vol. 29, no. 5, pp. 1189–1232, 2001.

[31] R. E. Schapire, “The strength of weak learnability,” in *Proc. 2nd Annu. Workshop Comput. Learn. Theory*, 1989, p. 383.

[32] J. Elith, J. R. Leathwick, and T. Hastie, “A working guide to boosted regression trees,” *J. Animal Ecol.*, vol. 77, no. 4, pp. 802–813, 2008.

[33] A. J. Conejo, M. Carrión, and J. M. Morales, *Decision Making Under Uncertainty in Electricity Markets*. New York, NY, USA: Springer, 2010.

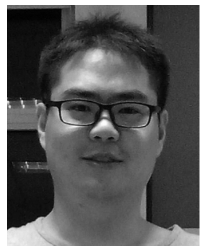
[34] T. Hong, P. Pinson, and S. Fan, “Global energy forecasting competition 2012,” *Int. J.* Forecast., vol. 30, no. 2, pp. 357–363, 2014.

[35] Y.-M. S. Drenant, “Wind power predictions analysis—Part 1: TenneT imbalance price system. Development of a model for TenneT imbalance price system,” Energy Res. Center Netherlands, Petten, Netherlands ECN Tech. Rep. ECN-I-02-010, 2002.

[36] T. Ackermann, *Wind Power in Power Systems*, 2nd ed. Hoboken, NJ, USA: Wiley, 2012.

[37] H. F. Ravn, “Short term wind power prognosis with different success criteria,” in *Proc. IEEE/PMAPS 9th Conf. Probabilistic Methods Appl. Power Syst.*, Stockholm, Sweden, Jun. 2006. pp. 1–5.

[38] J. P. S. Catalão, *Electric Power Systems: Advanced Forecasting Techniques and Optimal Generation Scheduling*. Boca Raton, FL, USA: CRC Press, 2012.



Guo Li (S’16) is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Automation, Tianjin University, Tianjin, China.

His current research interests include data analytics and related applications in power systems engineering.

Hsiao-Dong Chiang (M’87–SM’91–F’97) received the Ph.D. degree in EECS from the University of California at Berkeley in 1986. He is a Professor with the School of Electrical and Computer Engineering, Cornell University in Ithaca, NY, USA. He has published over 380 papers in refereed journals and conference proceedings. He has authored of two books entitled *Direct Methods for Stability Analysis of Electric Power Systems: Theoretical Foundation, BCU Methodologies, and Applications* (Wiley, 2011), and (with L. F. C. Alberto) *Stability Regions of Nonlinear Dynamical Systems: Theory, Optimal Estimation, and Applications* (Cambridge Press, 2015). He holds 17 U.S. and overseas patents and several consultant positions. His current research interests include nonlinear system theory, nonlinear computation, and nonlinear optimization and their practical applications. He was an Associate Editor for the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS from 1990 to 1991 and from 1993 to 1995.