

Research Article

Probabilistic Short-Term Wind Power Forecasting Using Sparse Bayesian Learning and NWP

Kaikai Pan, Zheng Qian, and Niya Chen

School of Instrument Science and Opto-Electronics Engineering, Beihang University, Beijing 100191, China

Correspondence should be addressed to Kaikai Pan; pankai@buaa.edu.cn

Received 13 October 2014; Accepted 24 April 2015

Academic Editor: Mehdi Khashei

Copyright © 2015 Kaikai Pan et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Probabilistic short-term wind power forecasting is greatly significant for the operation of wind power scheduling and the reliability of power system. In this paper, an approach based on Sparse Bayesian Learning (SBL) and Numerical Weather Prediction (NWP) for probabilistic wind power forecasting in the horizon of 1–24 hours was investigated. In the modeling process, first, the wind speed data from NWP results was corrected, and then the SBL was used to build a relationship between the combined data and the power generation to produce probabilistic power forecasts. Furthermore, in each model, the application of SBL was improved by using modified-Gaussian kernel function and parameters optimization through Particle Swarm Optimization (PSO). To validate the proposed approach, two real-world datasets were used for construction and testing. For deterministic evaluation, the simulation results showed that the proposed model achieves a greater improvement in forecasting accuracy compared with other wind power forecast models. For probabilistic evaluation, the results of indicators also demonstrate that the proposed model has an outstanding performance.

1. Introduction

As one of the most mature renewable energy sources, the generation of wind power has been growing rapidly worldwide over the last decade. The record shows that the global total installed wind generator capacity has exceeded 310 GW, and, in China, the current installed wind farm capacity is approximately 91.42 GW with a growth rate of 21.4% in 2013 [1]. However, since the wind is a variable and uncontrollable resource, the uncertainty of wind power generation is higher than those of traditional ones. The large-scale integration of wind power imposes difficulties in terms of power system operation and management [2]. Thus wind power forecasting is a significant process to provide a tool for the energy reserve scheduling.

Short-term wind power forecasts with a prediction horizon from 30 minutes to 72 hours are mainly used for decision-making areas such as Economic Dispatch (ED). The strategies for short-term wind power forecasting contain mainly statistical models and physical models. The common methods for the former have been developed based on single

methods or hybrids of several methods, such as the Persistence, Kalman filters [3], Autoregressive and Moving Average Model (ARMA) [4, 5], Artificial Neural Networks (ANN) [6, 7], Support Vector Machine (SVM) [8, 9], and Gaussian Processes (GP) [10]. Since wind varies rapidly with time, the statistical models which purely use historical wind speed data are effective only for very short-term forecasts (about 1–4 hours ahead). Physical models have a better performance for longer horizons and complicated terrain, because they adopt Numerical Weather Prediction (NWP) and improve its resolution to accurately predict a certain point (such as wind turbines at each) of weather parameters. However, this type of model also has limitations, such as the need of knowing physical properties and the huge amount of computation [11]. To overcome these limitations, some researchers present a prediction method for longer forecasting horizons by using NWP data and statistical methods [12]. Some of them used artificial intelligence algorithms as the final step of downscaling of NWP to obtain accurate predictions. In [13], a hybrid approach that integrates global Numerical Weather Prediction model with ANN is proposed. ANN is used to

carry out a downscaling of the NWP forecast. To improve the accuracy of the prediction system, different NWP forecasting models combined with different neural network banks are discussed in detail to compare the prediction results [14].

It should be noted that the majority of wind power forecasting tools at present can provide only deterministic values, which are also known as spot or point forecasts [15]. Such predictions cannot provide any information about the uncertainty of the predicted values, which limits their use in stochastic optimization and dispatching of wind power. Recently, some efforts have been made to build probabilistic wind power forecasting models. The problem is how to well represent the forecast errors of wind power. There are two main approaches to deal with this problem: the nonparametric approaches and the parametric approaches. The probabilistic models available today mainly concentrate on nonparametric approaches using quantile regression. A method to compute specified quantiles of the wind power generation is proposed in [16]. In addition, a nonparametric method that performs the interval estimation of wind speed and forecasts the wind power with the operation probability of wind turbine provides another modeling approach [17]. The main advantages of these approaches are that the shape of the distributions does not have to be assumed and any information about these distributions can easily be included. But these approaches have the drawback that can only provide predefined quantiles but not full probability density function (PDF). Otherwise, nonparametric approaches usually need more parameters than the parametric ones, which always means higher requirements on the computation and the amount of training data [15]. Several parametric approaches with shape assumption of predictive distribution have also been proposed. In recent researches, the forecast errors of wind power generation are usually assumed as a variable that obeys the law of Gaussian [18, 19] or Beta [20, 21] distribution. However, studies have shown that the shape of PDF of wind power forecast error varies with the forecast timescales and magnitudes. For very short-term wind power forecasting, Gaussian distribution is unsuitable to represent the error distributions at minute-level forecast timescales [22].

In this paper, we expand our previous work on short-term wind power forecasting. A novel probabilistic wind power forecasting approach is proposed based on Sparse Bayesian Learning (SBL) method. Real wind farm datasets including NWP data and measured data from a Supervisory Control and Data Acquisition (SCADA) system are analyzed and applied to forecast the wind power in the horizon of 1–24 hours. The major work and contributions of this paper are as follows:

- (1) As a parametric approach, the proposed approach based on SBL and NWP can improve the forecasting accuracy significantly for deterministic evaluation and provide full probabilistic information of wind power generation.
- (2) After analyzing the forecast data, the predicted wind speed data from NWP is corrected first using a separate model before wind power forecasting.

- (3) Selectors are used to determine the inputs of each model. And the historical wind speed data from a SCADA system is used as an additional input to the wind power forecasting model for 1–4-hour-ahead prediction since it is effective in this range of time horizons.
- (4) When using SBL method in each model, the modified-Gaussian kernel function is adopted and Particle Swarm Optimization (PSO) is used to optimize the parameters of the kernel functions.

This paper is organized as follows. Section 2 describes the properties of meteorological variables in NWP. Section 3 expresses the SBL algorithm applied to build the models. Section 4 constructs the NWP wind speed correction model and the probabilistic wind power forecast model and gives the detailed optimization method. The case study and simulation results are presented and analyzed in Section 5. Finally, Section 6 summarizes the main conclusions.

2. Data Analysis

2.1. NWP Model. Numerical Weather Prediction (NWP) model uses atmospheric hydrodynamics and thermodynamics to predict the future weather under initial values and boundary conditions. With a high accuracy, NWP based on high-speed computers can describe the atmosphere in detail and provide the forecast of weather from different grid points and layers. NWP models can be divided into global models or regional ones. In our research, we use the Weather Research and Forecasting (WRF) NWP model. WRF is a new regional NWP model created jointly by more than 150 research institutions in the United States and other countries [23].

Short-term wind power forecasting needs the predictions from WRF model with a high spatial resolution. To increase the resolution, there are two steps: firstly, the accurate coordinate point of a wind turbine in a wind farm is inferred from a weather observation point; secondly, the observed altitude of meteorological data is transformed to the hub height of the wind turbine. A 3-layer nested grid is built and the NWP predicted data can be abstracted from the corresponding point in the 3rd layer. The data includes several kinds of meteorological variables: wind speed, wind direction, temperature, air pressure, and humidity. The data is provided at 20:00 GMT every day with an interval of 10 minutes for the following 72 hours. As we focus on probabilistic short-term wind power forecasting in the horizon of 1–24 hours, the NWP data is hourly sampled.

The other available data is the hourly measured data (wind speed and wind power) from the Supervisory Control and Data Acquisition (SCADA) system.

2.2. Error Analysis of NWP Speed Data. It is clear that wind power mainly depends on actual wind speed and the most important task for wind power forecasting is to get accurate wind speed prediction. However, there exists error between the NWP predicted wind speed and the measured wind speed. Figure 1 shows the approximately linear relationship

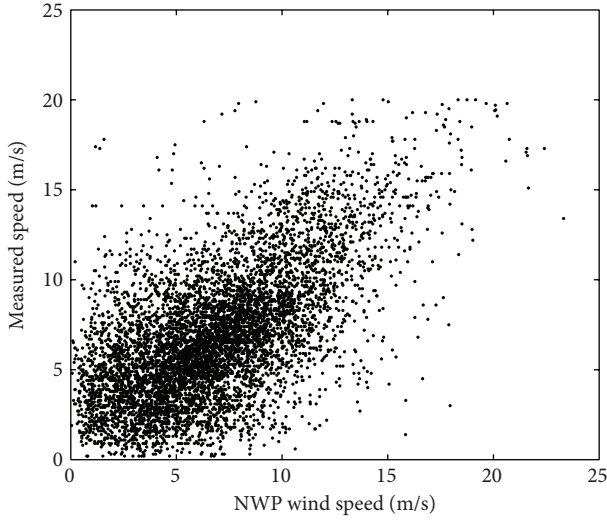


FIGURE 1: NWP predicted wind speed and the measured wind speed.

between the predicted values and the measured values of wind speed, with a large fluctuation existing. Therefore, the statistical analysis of wind speed prediction error is necessary. Statistical analysis is to describe the average deviation between the predicted values and measured values over a period of time.

The Root Mean Square Error (RMSE) is the most commonly used error evaluation indicator. RMSE can be decomposed into two parts: the bias and the variance of the error. These measures are defined as follows:

$$e_t = \hat{y}_t - y_t, \quad (1)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2}, \quad (2)$$

$$\text{RMSE}^2 = \text{Bias}^2 + \text{Sde}^2, \quad (3)$$

$$\text{Bias} = \bar{e}, \quad (4)$$

$$\text{Sde} = \sigma(e), \quad (5)$$

where \hat{y}_t denotes the predicted value of wind speed at time t , y_t denotes the measured value for the same period, n denotes the number of prediction data, \bar{e} denotes the difference between the mean values of \hat{y} and y , and $\sigma(e)$ denotes the standard deviation, reflecting the fluctuation of error around Bias.

From (3)–(5), we can see that RMSE is decomposed into Bias and Sde. As shown in Table 1, there exists large error in the original NWP wind speed data. This proves that it is necessary to correct the error of NWP wind speed prediction first in wind power forecasting using NWP. The detailed correction method is illustrated in Section 4.2.

TABLE 1: Error analysis of NWP wind speed prediction data.

Location	RMSE (m/s)	Bias (m/s)	Sde (m/s)	Average relative error
WF-G	3.152	0.503	3.112	25.64%
WF-J	2.381	0.964	2.177	32.67%

3. Sparse Bayesian Learning

Recently, Bayesian methods have shown some unique advantages in statistical modeling and data analysis with uncertainty and variability [24]. Bayesian methods are adopted in modeling distributions in the application of forecasting [24].

Sparse Bayesian Learning (SBL), known as relevance vector machine formerly, is proposed by Tipping firstly as a new supervised learning method [25]. Using the Lagrange coefficient vector as the weights parameters and introducing prior probability to establish corresponding saliency framework, Tipping removed the irrelevant point to obtain a sparse model.

Compared with SVM which has similar function structure, SBL has obvious advantages. SBL is conducted under Bayesian framework and can be used in probabilistic forecasting. For one training sample set, SBL is sparser than that of SVM, representing the main feature of data. And in SVM, we have to estimate the possible values of the error trade-off parameter and the insensitivity parameter which is wasteful of computation. SBL also overcomes the limitation that the kernel function must meet the Mercer condition and provides probabilistic regression [26]. Of course, like other intelligent methods, SBL has drawbacks when used in wind power forecasting and the main one is the Gaussian assumption of the underlying distribution. The influence of Gaussian assumption for hour-level forecast timescales will be considered in further research.

SVM has been used popularly in wind speed and power forecasting. But the application of SBL present is mainly in image classification, text recognition, and timing analysis [27]. The detail of SBL can be found from [25], so here we will provide a brief description.

3.1. Method Specification. SBL that Tipping described is as follows: Given a dataset $\{(\mathbf{x}_n, t_n), n = 1, 2, \dots, N\}$, \mathbf{x}_n denotes an input vector and t_n denotes a scalar-valued target. Tipping assumes t_n are samples distributed independently identically with additive noise ε_n which is assumed to be a random variable with Gaussian distribution of mean zero and variance σ^2

$$t_n = y(\mathbf{x}_n; \mathbf{w}) + \varepsilon_n, \quad (6)$$

$$p(t_i) = N(t_i | y(\mathbf{x}_i; \mathbf{w}), \sigma^2).$$

Considering the function structure of SVM, a general basis function $\Phi_i(\mathbf{x})$ is identified in SBL. According to the assumption of t_n , the likelihood of the complete dataset can be expressed as

$$p(\mathbf{t} | \mathbf{w}, \sigma^2) = (2\pi\sigma^2)^{-N/2} \exp\left(-\frac{\|\mathbf{t} - \Phi\mathbf{w}\|^2}{2\sigma^2}\right), \quad (7)$$

where $\mathbf{w} = [w_0, w_1, w_2, \dots, w_N]^T$ is the weight vector and Φ is $[\Phi(\mathbf{x}_1), \Phi(\mathbf{x}_2), \dots, \Phi(\mathbf{x}_N)]^T$.

The kernel method is usually used for the general basis function: $\Phi_i(\mathbf{x}) \equiv K(\mathbf{x}, \mathbf{x}_i)$, $\Phi(\mathbf{x}_i) = [1, K(\mathbf{x}_i, \mathbf{x}_1), K(\mathbf{x}_i, \mathbf{x}_2), \dots, K(\mathbf{x}_i, \mathbf{x}_N)]^T$, and $K(\mathbf{x}, \mathbf{x}_i)$ represents the kernel function. The way to optimize the kernel function will be an important work presented in Section 4.2 of this paper.

3.2. Bayesian Inference of the Parameter. Using maximum likelihood estimation directly to calculate \mathbf{w} and σ^2 in (3) will result in the fact that most values of \mathbf{w} are nonzero. In SVM, overfitting occurs when there are too many nonzero values of \mathbf{w} . To avoid this, Tipping set the prior distribution of w_i as Gaussian distribution $N(0, \alpha_i^{-1})$ and further set the prior distribution of α_i and σ^2 as Gamma distribution $\Gamma(0, 0)$. α_i is a hyperparameter associated independently with w_i .

After the probability distribution given above, the posterior distribution of \mathbf{w} , α , and σ^2 can be inferred under the Bayesian theorem

$$p(\mathbf{w} | \mathbf{t}, \alpha, \sigma^2) = \frac{p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \alpha)}{\int p(\mathbf{t} | \mathbf{w}, \sigma^2) p(\mathbf{w} | \alpha) d\mathbf{w}}. \quad (8)$$

The distribution of \mathbf{w} can be calculated from (3): the mean value of \mathbf{w} is $E = (\sigma^{-2} \Phi^T \Phi + \mathbf{A})^{-1}$, and the variance of \mathbf{w} is $\mu = \sigma^{-2} E \Phi^T \mathbf{t}$, where \mathbf{A} is $\text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$. Accordingly, the problem is transformed to estimate the optimal values of the hyperparameters α and σ^2 :

$$\begin{aligned} (\alpha_{\text{MP}}, \sigma_{\text{MP}}^2) &= \arg \max_{\alpha, \sigma^2} p(\alpha, \sigma^2 | \mathbf{t}) \\ &= \arg \max_{\alpha, \sigma^2} p(\mathbf{t} | \alpha, \sigma^2), \end{aligned} \quad (9)$$

where α_{MP} and σ_{MP}^2 are the maximum likelihood estimations of α and σ^2 , respectively.

From (9), it can compute α_{MP} and σ_{MP}^2 by equating the differentiation over α and σ^2 of the function to zero. In real application, the initial values of α and σ^2 would be given first, and α_{MP} and σ_{MP}^2 can be approximated through iteration. After the process of iterative reestimation, most of α_i tends to infinity, where corresponding w_i tends to zero. That is why SBL is a sparse model.

3.3. Probabilistic Forecasting. The distribution of the target value t_* for a new input vector \mathbf{x}_* can be written as

$$\begin{aligned} p(t_* | \mathbf{t}) \\ = \int p(t_* | \mathbf{w}, \alpha, \sigma^2) p(\mathbf{w}, \alpha, \sigma^2 | \mathbf{t}) d\mathbf{w} d\alpha d\sigma^2. \end{aligned} \quad (10)$$

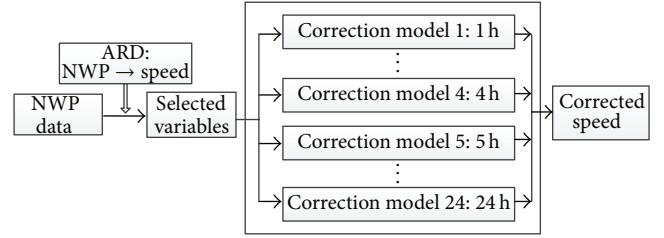


FIGURE 2: Structure of NWP wind speed correction model.

$P(\alpha, \sigma^2 | t)$ can be approximated by $\delta(\alpha - \alpha_{\text{MP}}) \delta(\sigma^2 - \sigma_{\text{MP}}^2)$, where $\delta(\alpha - \alpha_{\text{MP}})$ and $\delta(\sigma^2 - \sigma_{\text{MP}}^2)$ are both delta functions. The new target value t_* can be calculated as follows:

$$\begin{aligned} p(t_* | \mathbf{t}) &= N(t_* | y_*, \sigma_*^2), \\ y_* &= \mu^T \Phi(\mathbf{x}_*), \\ \sigma_*^2 &= \sigma_{\text{MP}}^2 + \Phi(\mathbf{x}_*)^T \Sigma \Phi(\mathbf{x}_*). \end{aligned} \quad (11)$$

4. Modeling Process

4.1. Forecasting Model Structure. The NWP data includes wind speed, wind direction, temperature, air pressure, and humidity. In general, the wind power generation mainly depends on actual wind speed. However, we do not know whether the other variables (wind direction, temperature, air pressure, and humidity) also play an important role. It is necessary to select the most beneficial training data for accurate prediction. According to our previous work presented in [28], the ARD toolbox in Netlab and the results in [28] are used to select appropriate inputs for each model.

From large body of empirical analysis and previous work, we know that the character of wind changes with diurnal cycle, and the NWP data shows different properties depending on the forecast horizon of 1–24 hours. On the other hand, a suitable number of training samples are needed in SBL which has bad performance in large training dataset case. Therefore, the training dataset (NWP and historical data) was separated into 24 subsets according to the horizon of 1–24 hours to build models.

In addition, when using SBL in each model, the selection of kernel function and the optimization of its parameters have a significant impact on the model performance. In this paper, the application of SBL is improved by adopting modified-Gaussian kernel function and Particle Swarm Optimization (PSO) to optimize the parameters of the kernel functions.

According to Section 2, there are some systematic and stochastic biases present in the original NWP wind speed data. To improve the accuracy of the wind power forecast, it is important to get accurate wind speed prediction. Therefore, a wind speed correction model is applied first in this paper to correct the error in NWP wind speed prediction.

Figure 2 shows the detailed structure of the NWP wind speed correction model. ARD is used to determine which variables in NWP are relevant to the measured wind speed.

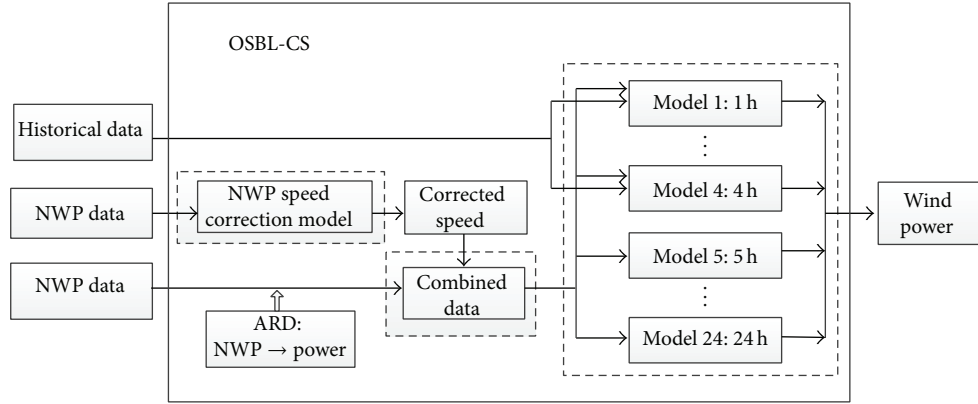


FIGURE 3: Structure of the OSBL-CS model.

Correction models 1–24 based on optimized SBL (OSBL) give the result of corrected speed for 1–24-hour-ahead prediction.

According to the analysis above, the whole modeling process of the probabilistic short-term wind power forecasting model based on SBL and NWP is illustrated in Figure 3. We donate our method of probabilistic wind power forecasting in this paper as OSBL-CS (meaning based on optimized SBL and corrected speed).

As shown in Figure 3, firstly, the corrected wind speed is obtained through the correction model. And then, the original NWP speed is replaced with this corrected wind speed. Secondly, ARD is also applied to choose variables which are relevant to wind power. As mentioned before, the historical data of wind speed is useful for 1–4-hour-ahead wind power forecasting and therefore it is included as an additional input for models in the horizon of 1–4 hours. Finally, with these combined data as inputs, a separate model for each of the 24 hourly horizons is built for probabilistic wind power forecasting.

4.2. Optimization of Kernel Function. Kernel function implies the similarity of input data, and the parameters of kernel functions have a “zoom-in” or “zoom-out” effect on this scale. Both of them have a significant impact on the performance of the model.

Gaussian kernel function is usually adopted in kernel function based methods and has been proven to be effective [29]:

$$K(\mathbf{x}, \mathbf{x}_c) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_c\|^2}{2\sigma^2}\right), \quad (12)$$

where σ is the width of kernel function.

In recent researches using kernel-based methods, the entire model always uses only one scale parameter σ in the application of the Gaussian kernel function and uses cross-validation method to find the optimal value of σ . However, for multidimensional signal having complex frequency components, it tends to be difficult to learn the model accurately if there is only one fixed scale parameter. The inputs of each model in this paper have multidimensional variables

(selected from wind speed, wind direction, temperature, air pressure, and humidity). To solve this problem, we modify the original Gaussian kernel function and introduce a number of scale parameters. Based on the form of Gaussian kernel function, the kernel function becomes

$$K(\mathbf{x}_m, \mathbf{x}_n) = \exp\left\{-\sum_i^d \frac{1}{r_i^2} (x_{mi} - x_{ni})^2\right\}, \quad (13)$$

where d denotes the number of variables comprised in the input vector and (r_1, \dots, r_d) denote the corresponding scale parameters.

Selection of kernel parameters also plays a crucial role in the prediction results. For (10), due to an increase in the number of kernel parameters, it is difficult to find the optimal values of (r_1, \dots, r_d) using cross-validation method. We adopted Particle Swarm Optimization (PSO) to search for the optimal kernel scale parameters. Compared to a similar method Genetic Algorithm (GA), PSO can reach the global optimum more quickly, and there is no need to adjust so many parameters. The detailed method of PSO can be found in [30].

The set of parameters in PSO is relatively fixed. The number of particles for each parameter we set is 30 and the adaptive function of particles is Root Mean Square Error (RMSE), as defined in Section 2.2. The maximum number of iterations is 100. The speed and location of particles are updated by the following functions:

$$\begin{aligned} V_{ij}^t &= w^t V_{ij}^{t-1} + c_1 r_1 (P_{best,i,j}^{t-1} - X_{ij}^{t-1}) \\ &\quad + c_2 r_2 (G_{best,i,j}^{t-1} - X_{ij}^{t-1}), \\ X_{ij}^t &= X_{ij}^{t-1} + V_{ij}^t, \end{aligned} \quad (14)$$

where V_{ij}^t and X_{ij}^t are speed and location of the i th particle in the t th iteration in j -dimension; P_{best} and G_{best} are, respectively, the individual best location and group best location; r_1 and r_2 are uniform random numbers $[0, 1]$; c_1 and c_2 are accelerating factors; w^t is inertia weight factor.

TABLE 2: Information of wind farms.

Location	Height (m)	Installed capacity (kW)	Number
WF-G	65	1500	33
WF-J	80	1500	67

5. Experimental Validation

5.1. Case Study Description. In order to evaluate our approach, two real-world datasets based on wind farms are used. The first wind farm is located in Gansu Province, in western China, denoted by WF-G. WF-G is located in high altitude area, where the wind is strong and the wind direction is relatively single. The second is located in Jiangsu Province, in eastern China, denoted by WF-J. The humidity of WF-J is relatively high. It is a good way to validate the prediction performance of the proposed approach under different weather conditions.

As shown in Table 2, WF-G has 33 wind turbines in one group, and WF-J is a larger wind farm with 67 wind turbines in one group. The installed capacity of single wind turbine in both wind farms is 1500 kW. The time span of the NWP dataset and measured dataset is from April 2011 to May 2013. We use the data in the whole year round from April 1, 2011, to April 1, 2012, as the training set and the remaining data as the test set.

5.2. Forecasting Accuracy Evaluation. Before evaluating and comparing the models, we should introduce several indicators. To the probabilistic forecasting based on SBL, one aspect is to evaluate the error of forecasted expectation values (deterministic evaluation) and the other is to compare the difference between the wind power predictive distribution and the actual one (probabilistic evaluation). These indicators are listed as follows.

(1) The Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Normalized Mean Absolute Percentage Error (NMAPE) are used as deterministic evaluation indicators. RMSE has been defined as (2); MAE and NMAPE are defined as

$$\begin{aligned} \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |\hat{y}_t - y_t|, \\ \text{NMAPE} &= \frac{1}{n} \sum_{i=1}^n \left| \frac{\hat{y}_t - y_t}{C} \times 100 \right|, \end{aligned} \quad (15)$$

where \hat{y}_t denotes the predicted value at time t , y_t denotes the measured value for the same period, n denotes the number of prediction data, and C is the installed capacity of wind farm.

(2) The reliability evaluation (RE) criterion [31] is employed to evaluate the probabilistic correctness of the predictive distribution:

$$\text{RE} = \left[\frac{\xi^{(1-\alpha)}}{N} - (1-\alpha) \right] \times 100\%, \quad (16)$$

TABLE 3: ARD results of relevance values on WF-G and WF-J.

Location	Wind speed	Wind direction	Temperature	Air pressure	Humidity
WF-G	0.3882	0.0463	2.2647	3.5917	0.9233
WF-J	0.1203	0.0014	3.2269	0.0383	0.3569

where N is the number of test samples and $\xi^{(1-\alpha)}$ is the number of times that actual test samples lie within the α -confidence prediction interval. To ensure high reliability, RE should be as close to zero as possible; that is, the smaller the $|\text{RE}|$, the better.

In addition, the Skill Score (SC) criterion is employed to evaluate the quality of probabilistic forecasting as a whole:

$$\text{SC} = \frac{1}{N} \sum_{n=1}^N \sum_{i=1}^{N_\beta} (\xi^{(\beta_i)} - \beta_i) (y_n - \hat{q}^{(\beta_i)}), \quad (17)$$

where N_β is the number of quantiles and $\hat{q}^{(\beta_i)}$ is the quantile with proportion β_i in predictive distribution. $\xi^{(\beta_i)}$ is an indicator for $\hat{q}^{(\beta_i)}$, being equal to 1 if $y_n < \hat{q}^{(\beta_i)}$ and being zero otherwise. In general, the SC is a minus, and the higher the SC, the better.

For the deterministic evaluation, the RMSE, MAE, and NMAPE are popular indicators [32, 33]. According to the specified accuracy requirement of Chinese government, NMAPE should be less than 10% for 1–4-hour forecast and less than 20% for 5–24-hour forecast. Thus, NMAPE is used as the main indicator in deterministic evaluation.

Evaluation of the quality of the probabilistic forecasting results is quite different from common statistical approach. The RE, the SC, and the Sharpness Evaluation (SE) are main indicators introduced in [34] for probabilistic forecasting of wind power. The method in this paper has the underlying Gaussian assumption of the distribution shape. We mainly use the RE and the SC as the indicators, since the SE directly relates to the variance σ^2 ; the smaller the σ^2 , the sharper and the better.

5.3. Simulation Results

5.3.1. Effectiveness of NWP Wind Speed Correction. The ARD toolbox is employed to choose inputs of the model. The relevance values of each NWP variable to the error of measured wind speed and NWP predicted speed are listed in Table 3 [28]. According to the results, we choose wind speed, temperature, and humidity as inputs of each correction model for WF-J and add air pressure for WF-G.

To illustrate the effectiveness of the NWP speed correction model, we introduce a common correction method based on Linear Regression (LR) as a benchmark. LR is a statistical method using the least squares approach to model the relationship between one or more explanatory variables and the dependent variable. For the horizon of 1–24 hours, a comparison of corrected error using LR and OSBL is displayed in Figure 4.

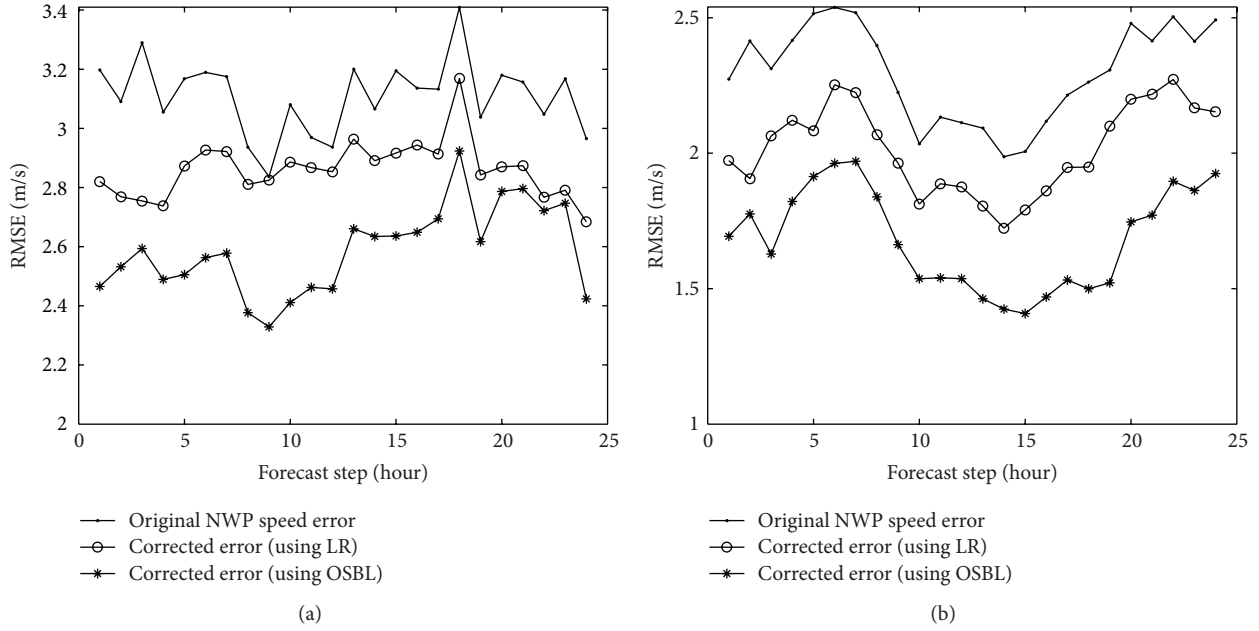


FIGURE 4: Comparison of corrected speed error. (a) RMSE values of speed in WF-G. (b) RMSE values of speed in WF-J.

TABLE 4: Error analysis of NWP wind speed prediction data.

Location	Data	RMSE (m/s)	MAE (m/s)	RMSE improvement
WF-G	NWP speed	3.1098	2.4069	—
	Corrected speed (LR)	2.7653	2.1310	11.07%
	Corrected speed (OSBL)	2.4774	1.9294	20.34%
WF-J	NWP speed	2.3506	1.7891	—
	Corrected speed (LR)	1.9244	1.4332	18.13%
	Corrected speed (OSBL)	1.5737	1.1896	33.05%

Figure 4 shows that the corrected results in both farms are much better than the original NWP data, especially using the proposed NWP speed correction model in this paper.

Table 4 shows the RMSE and MAE results of the whole correction process on test datasets. From Table 4, the correction model in this paper successfully reduced the prediction error. The improvement of accuracy based on RMSE is 20.36% for WF-G and 33.05% for WF-J.

5.3.2. Deterministic Evaluation of Wind Power Forecasting.

This section aims to evaluate the forecasted expectation values using the probabilistic model OSBL-CS. According to the ARD results, in addition to the corrected wind speed, we also choose temperature, air pressure, and humidity as inputs of wind power forecasting model for WF-G and choose wind direction and temperature as inputs for WF-J, which is reasonable for the different weather conditions in the two farms.

In order to validate its performance, some methods commonly used for deterministic wind power forecasting are adopted as the benchmarks: the Persistence method, MLP neural networks, SVM, and original SBL. The Persistence method simply uses the current value as the forecast. The

MLP neural networks have been applied to researches on short-term wind power forecasting widely. Like OSBL-CS, the model using the method of MLP also corrects wind speed first and adds historical data for 1–4 hours ahead, denoted by MLP-CS. The first part of MLP-CS contains 9 hidden layer neurons; the second part using the corrected wind speed as inputs contains 7 hidden neurons.

SVM has similar function structure to SBL; hence a SVM based model which has the same wind power forecasting process with OSBL-CS is built for comparison. If we use the modified-Gaussian kernel function proposed in this paper, it is difficult to estimate so many additional parameters accurately and it becomes meaningless in practical application of SVM. Therefore, standard Gaussian kernel function is adopted in the model based on SVM, and cross-validation method is used for parameters optimization. Furthermore, an original SBL based model using Gaussian kernel function which is the same as SVM-CS is built to compare with the improved model of OSBL-CS.

The deterministic results of 24-hour-ahead wind power forecast in each wind farm are shown in Figure 5. At the prediction interval of 1–4 hours, these methods all have a good performance. It shows that the additional inputs

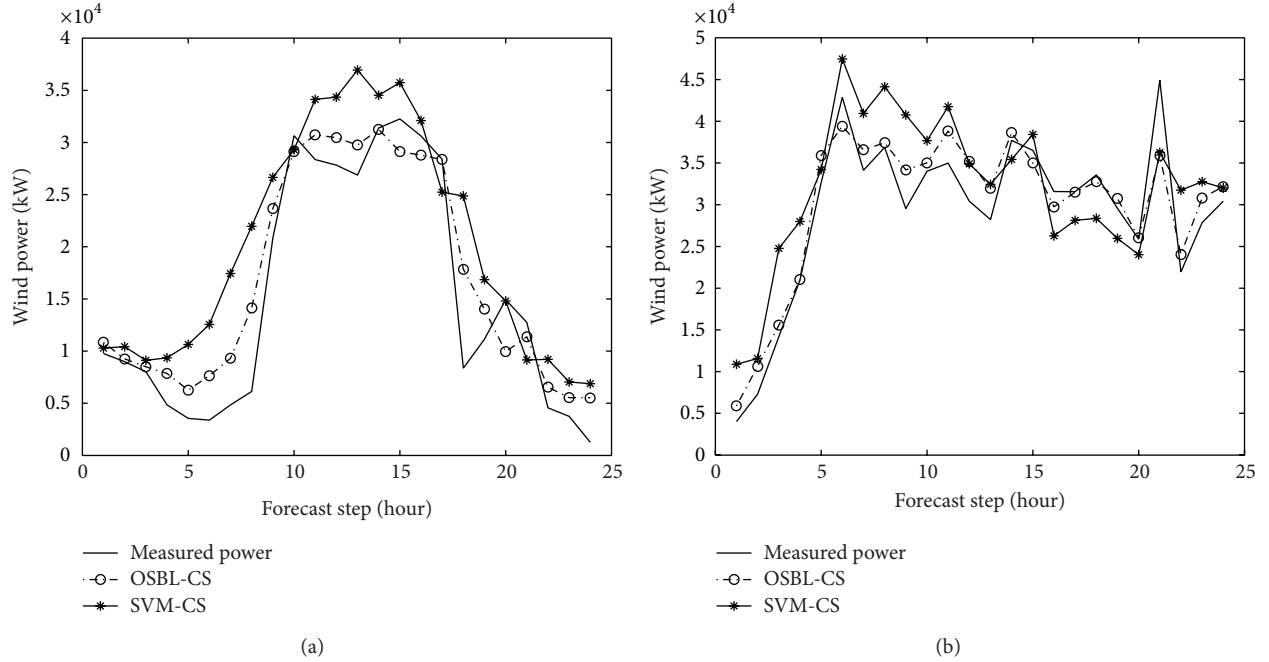


FIGURE 5: 24-hour-ahead wind power forecast. (a) Forecast results of WF-G. (b) Forecast results of WF-J.

TABLE 5: Power forecast error of WF-G (33 wind turbines, 49.5 MW).

Model	RMSE (MW)	MAE (MW)	NMAPE	NMAPE improvement
Persistence	10.88	7.90	15.97%	—
MLP-CS	8.46	6.09	12.30%	22.98%
SVM-CS	8.12	5.74	11.60%	27.36%
SBL-CS	8.04	5.66	11.43%	28.43%
OSBL-CS	7.86	5.41	10.93%	31.52%

TABLE 6: Power forecast error of WF-J (67 wind turbines, 100.5 MW).

Model	RMSE (MW)	MAE (MW)	NMAPE	NMAPE improvement
Persistence	26.34	18.22	18.13%	—
MLP-CS	17.32	12.04	11.98%	33.92%
SVM-CS	17.27	12.00	11.94%	34.14%
SBL-CS	16.46	11.42	11.36%	37.34%
OSBL-CS	15.29	10.58	10.53%	41.92%

of historical data for each model successfully reduced the prediction error. With the increase of the forecast horizon, the forecast error also increases. The predictive wind power curve of the OSBL-CS model basically captures the actual wind power curve in both wind farms. In some moments, the wind power generation changes dramatically, and the prediction effect of the OSBL-CS model becomes a bit worse. But on the whole, the OSBL-CS model has a better performance compared with the SVM model.

Tables 5 and 6 show the calculation results on the whole test sets of RMSE, MAE, and NMAPE in each wind farm.

The proposed OSBL-CS model has a better performance than the other models. Compared to the Persistence method, the improvement of accuracy is 31.52% for test dataset of WF-G and 41.92% for WF-J. Compared with other artificial intelligence methods, the SBL based models (SBL-CS and OSBL-CS) have better forecast accuracy. Compared to the original SBL-CS model, the OSBL-CS model has improved the accuracy of 2.65% for WF-G and 7.36% for WF-J. According to the above analysis, this proposed model has the feature that the improvement of performance in WF-J is higher than the one in WF-G. And the OSBL-CS model has lower average forecast error in WF-J. That is mainly because WF-G is located in the windy area and the wind speed changes more dramatically.

The NMAPE values in Tables 4 and 5 embody the overall forecast performance. As the specified requirement of NMAPE by Chinese government for 1–4-hour and 5–24-hour forecast, the detailed results of NMAPE in each forecast hour for WF-G are calculated, as shown in Figure 6. It can be seen from this figure that NMAPE according to OSBL-CS model is lower than the others and can satisfy the government's requirement successfully.

We should note that the forecast accuracy of OSBL-CS is not so competitive for 1-hour-ahead forecast, as the NMAPE values of the Persistence model and the MLP-CS model are a bit smaller. It suggests that, for very short-term wind power forecasting, the OSBL-CS could not outperform others universally. But the other points in Figure 6 show a significant improvement. From the above, we can conclude that the OSBL-CS is a better model for deterministic wind power forecasting compared with the other four models in this paper.

5.3.3. Probabilistic Evaluation of Wind Power Forecasting. To fully validate the performance of the OSBL-CS model,

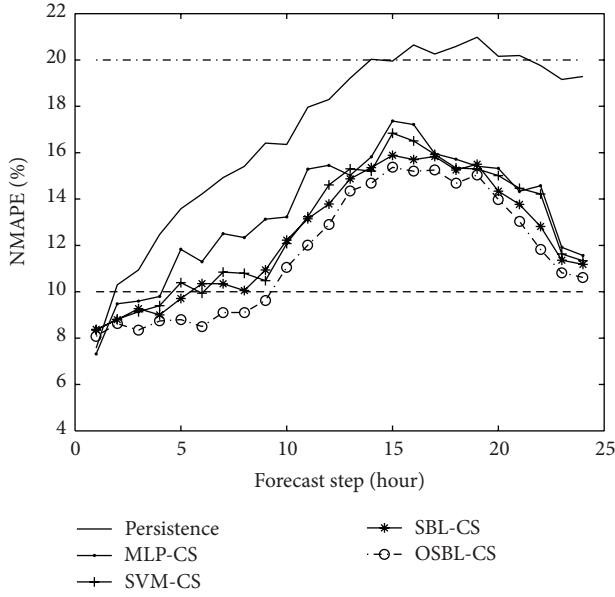


FIGURE 6: NMAPE values of wind power forecasting (WF-G).

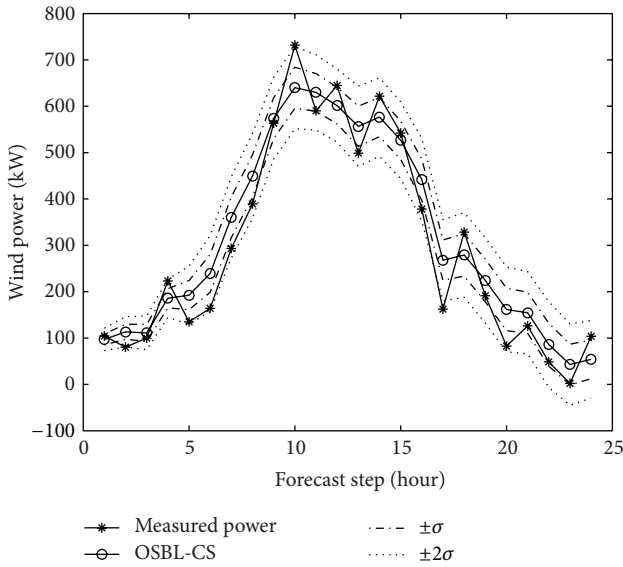


FIGURE 7: Forecast results containing the information of uncertainty.

another aspect is to evaluate the wind power predictive distribution.

An example of 24-hour-ahead wind power forecast for a wind turbine in WF-J containing the information of uncertainty (the σ and the 2σ confidence intervals) is shown in Figure 7. As shown in this figure, the actual wind power values are mostly included in the $\pm\sigma$ and $\pm 2\sigma$ range of the predicted values. The value of σ is small in the horizon of 1–4 hours, which indicates that the Sharpness Evaluation values are also small in such prediction period, and the value of σ becomes large especially when the measured power fluctuates obviously.

TABLE 7: RE values for probabilistic forecasting results in WF-G.

α	RE					
	1 h	2 h	3 h	4 h	5 h	6 h
0.1	2.65%	1.68%	2.11%	1.94%	2.47%	2.33%
0.05	1.69%	-0.23%	0.59%	1.43%	1.20%	-1.01%
0.01	-1.82%	-1.95%	-0.30%	-1.46%	-1.63%	-2.08%
	7 h	8 h	9 h	10 h	11 h	12 h
0.1	1.86%	2.77%	3.19%	2.34%	4.12%	3.68%
0.05	-0.95%	0.86%	0.94%	1.06%	-2.13%	1.75%
0.01	-1.42%	-2.19%	-1.24%	-2.07%	-2.58%	-2.93%
	13 h	14 h	15 h	16 h	17 h	18 h
0.1	2.67%	1.61%	3.29%	4.34%	3.35%	2.93%
0.05	-1.23%	1.48%	1.58%	1.67%	-1.76%	1.62%
0.01	-2.39%	-1.85%	-2.46%	-2.40%	-2.46%	-2.69%
	19 h	20 h	21 h	22 h	23 h	24 h
0.1	3.54%	4.26%	3.68%	2.55%	4.01%	3.87%
0.05	1.21%	-2.41%	1.96%	2.14%	1.62%	1.69%
0.01	-1.95%	-2.58%	-2.35%	-2.24%	-2.07%	-2.66%

In order to evaluate its probabilistic forecasting performance, quantitative analysis is still needed. The reliability represents the statistical consistency of predictive distribution and observations. The reliability evaluation plays the role of basic evaluation [35]. By definition (16), for high reliability, the actual test samples are expected to lie within the α -confidence prediction interval. We choose the nominal confidence levels as 90%, 95%, and 99% (while α is 0.1, 0.05, and 0.01, resp.) and calculate the RE values in the forecast horizon of 1–24 hours, as shown in Table 7.

According to Table 7, in the horizon of 1–24 hours, the RE values for the probabilistic forecasting results using the proposed OSBL-CS model are small, especially for the confidence levels of 95% (α is 0.05) and 99% (α is 0.01). The most absolute RE values at confidence level of 95% are less than 2% and less than 3% at 99%. Particularly, the RE values corresponding to the horizon of 1–10 hours are closer to zero, which means that the predictive distributions match the observations better. And the OSBL-CS model has a good performance of probabilistic forecasting in this horizon. Overall, the RE values obtained by the Gaussian distribution model OSBL-CS show high reliability of the predictive distribution.

The Skill Score (SC) criterion provides comprehensive information on probabilistic forecasting. Since the forecast results obey the law of Gaussian distribution, for (17), we choose seven quantiles according to $\pm\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ points around expectation, and the corresponding β_i is 0.0013, 0.0228, 0.1587, 0.5, 0.8413, 0.9772, and 0.9987. The SC values are calculated in the horizon of 1–24 hours for the probabilistic forecasting results given by OSBL-CS model and the original SBL-CS model is also used as a comparison in both farms WF-G and WF-J. The SC values are contained in Tables 8 and 9.

According to Tables 8 and 9, the performance of the OSBL-CS model is clearly better compared with the SBL-CS

TABLE 8: SC values for probabilistic forecasting results in WF-G.

Model	SC					
	1 h	2 h	3 h	4 h	5 h	6 h
SBL-CS	-0.78	-0.83	-0.79	-0.76	-0.82	-0.90
OSBL-CS	-0.68	-0.72	-0.73	-0.69	-0.68	-0.76
	7 h	8 h	9 h	10 h	11 h	12 h
SBL-CS	-0.86	-0.91	-0.90	-1.01	-1.00	-0.96
OSBL-CS	-0.78	-0.74	-0.77	-0.83	-0.80	-0.82
	13 h	14 h	15 h	16 h	17 h	18 h
SBL-CS	-0.95	-0.97	-1.04	-1.00	-0.93	-0.89
OSBL-CS	-0.82	-0.86	-0.87	-0.79	-0.73	-0.74
	19 h	20 h	21 h	22 h	23 h	24 h
SBL-CS	-0.88	-0.98	-0.96	0.87	0.88	0.88
OSBL-CS	-0.77	-0.84	-0.80	-0.75	-0.79	-0.74

TABLE 9: SC values for probabilistic forecasting results in WF-J.

Model	SC					
	1 h	2 h	3 h	4 h	5 h	6 h
SBL-CS	-1.91	-1.90	-1.88	-1.89	-2.03	-1.99
OSBL-CS	-1.64	-1.60	-1.57	-1.62	-1.71	-1.69
	7 h	8 h	9 h	10 h	11 h	12 h
SBL-CS	-2.12	-1.94	-1.94	-1.95	-2.03	-2.11
OSBL-CS	-1.77	-1.66	-1.63	-1.65	-1.69	-1.73
	13 h	14 h	15 h	16 h	17 h	18 h
SBL-CS	-2.09	-2.10	-2.04	-1.95	-1.95	-2.05
OSBL-CS	-1.71	-1.75	-1.68	-1.63	-1.64	-1.69
	19 h	20 h	21 h	22 h	23 h	24 h
SBL-CS	-2.00	-2.12	-2.07	-2.02	-2.02	-1.93
OSBL-CS	-1.67	-1.73	-1.71	-1.68	-1.67	-1.63

model. We can see that the OSBL-CS model gives good probabilistic forecasting performance in both farms, particularly in the horizon of 1–4 hours due to the high reliability and the small Sharpness Evaluation values. It should be noted that, in the horizon of 10–15 hours for WF-G, the SC values are a bit higher, which means that the performance is relatively worse. The wind characteristics and the geographical environment of WF-G may account for this phenomenon. During this period, the wind speed of WF-G may be high enough that the wind turbine would be close to the full state, and the temperature in this place is always high. Thus, the OSBL-CS model is not absolutely suitable for probabilistic forecasting in this horizon for WF-G.

6. Conclusion

Research on probabilistic short-term wind power forecasting is an important and challenging task. In this paper, a new supervised learning method SBL is introduced and optimized for the training of the forecast model. In order to improve the forecast accuracy, some methods were used: the wind speed data in NWP is firstly corrected before it is used to forecast; selectors are used to determine the inputs of

each model and for 1–4-hour-ahead forecasts, and historical data is added to modeling process; the SBL is optimized by adopting modified-Gaussian kernel function and Particle Swarm Optimization (PSO) to optimize the parameters of the kernel function. The simulation results in deterministic and probabilistic evaluation reveal the effectiveness of the proposed OSBL-CS model.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

The authors are thankful to the National Fund for Creative Groups of China (Grant no. 61121003) for their support of this study.

References

- [1] G. W. E. Council, "Global wind statistics 2013," Global Wind Report, 2013.
- [2] J. C. Smith, M. R. Milligan, E. A. DeMeo, and B. Parsons, "Utility wind integration and operating impact state of the art," *IEEE Transactions on Power Systems*, vol. 22, no. 3, pp. 900–908, 2007.
- [3] P. D. Fu, L. Hui, and L. Y. Fei, "A wind speed forecasting optimization model for wind farms based on time series analysis and Kalman filter algorithm," *Power System Technology*, vol. 32, pp. 82–86, 2008.
- [4] H. Liu, E. Erdem, and J. Shi, "Comprehensive evaluation of ARMA-GARCH(-M) approaches for modeling the mean and volatility of wind speed," *Applied Energy*, vol. 88, no. 3, pp. 724–732, 2011.
- [5] R. G. Kavasseri and K. Seetharaman, "Day-ahead wind speed forecasting using f-ARIMA models," *Renewable Energy*, vol. 34, no. 5, pp. 1388–1393, 2009.
- [6] Y.-Y. Hong, H.-L. Chang, and C.-S. Chiu, "Hour-ahead wind power and speed forecasting using simultaneous perturbation stochastic approximation (SPSA) algorithm and neural network with fuzzy inputs," *Energy*, vol. 35, no. 9, pp. 3870–3876, 2010.
- [7] J. P. S. Catalão, H. M. I. Pousinho, and V. M. F. Mendes, "Short-term wind power forecasting in Portugal by neural networks and wavelet transform," *Renewable Energy*, vol. 36, no. 4, pp. 1245–1251, 2011.
- [8] S. Salcedo-Sanz, E. G. Ortiz-García, Á. M. Pérez-Bellido, A. Portilla-Figueras, and L. Prieto, "Short term wind speed prediction based on evolutionary support vector regression algorithms," *Expert Systems with Applications*, vol. 38, no. 4, pp. 4052–4057, 2011.
- [9] J. W. Zeng and W. Qiao, "Short-term solar power prediction using a support vector machine," *Renewable Energy*, vol. 52, pp. 118–127, 2013.
- [10] X. Jiang, B. Dong, L. Xie et al., "Adaptive Gaussian process for short-term wind speed forecasting," in *Proceedings of the 19th European Conference on Artificial Intelligence (ECAI '10)*, vol. 215 of *Frontiers in Artificial Intelligence and Applications*, pp. 661–666, IOS Press, Lisbon, Portugal, August 2010.
- [11] S. Al-Yahyai, Y. Charabi, and A. Gastli, "Review of the use of numerical weather prediction (NWP) models for wind energy

- assessment,” *Renewable & Sustainable Energy Reviews*, vol. 14, no. 9, pp. 3192–3198, 2010.
- [12] M. G. de Giorgi, A. Ficarella, and M. Tarantino, “Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods,” *Energy*, vol. 36, no. 7, pp. 3968–3978, 2011.
- [13] S. Salcedo-Sanz, Á. M. Pérez-Bellido, E. G. Ortiz-García, A. Portilla-Figuera, L. Prieto, and D. Paredes, “Hybridizing the fifth generation mesoscale model with artificial neural networks for short-term wind speed prediction,” *Renewable Energy*, vol. 34, no. 6, pp. 1451–1457, 2009.
- [14] S. Salcedo-Sanz, Á. M. Pérez-Bellido, E. G. Ortiz-García, A. Portilla-Figuera, L. Prieto, and F. Correoso, “Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks,” *Neurocomputing*, vol. 72, no. 4–6, pp. 1336–1341, 2009.
- [15] J. Juban, N. Siebert, and G. N. Kariniotakis, “Probabilistic short-term wind power forecasting for the optimal management of wind generation,” in *Proceedings of the IEEE Lausanne Power Tech*, pp. 683–688, IEEE, Lausanne, Switzerland, July 2007.
- [16] J. B. Bremnes, “Probabilistic wind power forecasts using local quantile regression,” *Wind Energy*, vol. 7, no. 1, pp. 47–54, 2004.
- [17] H. Liu, J. Shi, and X. Qu, “Empirical investigation on using wind speed volatility to estimate the operation probability and power output of wind turbines,” *Energy Conversion and Management*, vol. 67, pp. 8–17, 2013.
- [18] P. Kou, F. Gao, and X. Guan, “Sparse online warped Gaussian process for wind power probabilistic forecasting,” *Applied Energy*, vol. 108, pp. 410–428, 2013.
- [19] Z. Zhang, Y. Sun, G. Li, L. Cheng, and J. Lin, “A solution of economic dispatch problem considering wind power uncertainty,” *Automation of Electric Power Systems*, vol. 35, no. 22, pp. 125–130, 2011.
- [20] A. Fabbri, T. G. S. Román, J. R. Abbad, and V. H. M. Quezada, “Assessment of the cost associated with wind generation prediction errors in a liberalized electricity market,” *IEEE Transactions on Power Systems*, vol. 20, no. 3, pp. 1440–1446, 2005.
- [21] Z.-S. Zhang, Y.-Z. Sun, D. W. Gao, J. Lin, and L. Cheng, “A versatile probability distribution model for wind power forecast errors and its application in economic dispatch,” *IEEE Transactions on Power Systems*, vol. 28, no. 3, pp. 3114–3125, 2013.
- [22] B.-M. Hodge and M. Milligan, “Wind power forecasting error distributions over multiple timescales,” in *Proceedings of the IEEE PES General Meeting: The Electrification of Transportation and the Grid of the Future*, Detroit, Mich, USA, July 2011.
- [23] D. Hosansky, *Weather forecast accuracy gets boost with new computer model*, University Corporation for Atmospheric Research (UCAR), 2006.
- [24] G. Li and J. Shi, “Application of Bayesian model averaging in modeling long-term wind speed distributions,” *Renewable Energy*, vol. 35, no. 6, pp. 1192–1202, 2010.
- [25] M. E. Tipping, “Sparse Bayesian learning and the relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [26] S. R. Yang and H. Y. Shen, “Research and application of machine learning algorithm based on Relevance Vector Machine,” *Computing Technology and Automation*, vol. 29, pp. 43–47, 2010.
- [27] B. Demir and S. Ertürk, “Hyperspectral image classification using relevance vector machines,” *IEEE Geoscience and Remote Sensing Letters*, vol. 4, no. 4, pp. 586–590, 2007.
- [28] N. Chen, Z. Qian, I. T. Nabney, and X. Meng, “Wind power forecasts using gaussian processes and numerical weather prediction,” *IEEE Transactions on Power Systems*, vol. 29, no. 2, pp. 656–665, 2014.
- [29] J. Yan, Y. Liu, S. Han, and M. Qiu, “Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine,” *Renewable & Sustainable Energy Reviews*, vol. 27, pp. 613–621, 2013.
- [30] J. Kennedy and R. Eberhart, “Particle swarm optimization,” in *Proceedings of the 4th IEEE International Conference on Neural Networks*, pp. 1942–1948, 1995.
- [31] P. Pinson and G. Kariniotakis, “Conditional prediction intervals of wind power generation,” *IEEE Transactions on Power Systems*, vol. 25, no. 4, pp. 1845–1856, 2010.
- [32] H. Bludszweit, J. A. Domínguez-Navarro, and A. Llombart, “Statistical analysis of wind power forecast error,” *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 983–991, 2008.
- [33] M. Xu, Y. Qiao, and Z. Lu, “A comprehensive error evaluation method for short-term wind power prediction,” *Automation of Electric Power Systems*, vol. 35, no. 12, pp. 20–26, 2011.
- [34] P. Pinson, H. A. Nielsen, J. K. Møller, H. Madsen, and G. N. Kariniotakis, “Non-parametric probabilistic forecasts of wind power: required properties and evaluation,” *Wind Energy*, vol. 10, no. 6, pp. 497–516, 2007.
- [35] Y. Zhang, J. X. Wang, and X. F. Wang, “Review on probabilistic forecasting of wind power generation,” *Renewable & Sustainable Energy Reviews*, vol. 32, pp. 255–270, 2014.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

