Research Article

Research on Hybrid Wind Speed Prediction System Based on Artificial Intelligence and Double Prediction Scheme

Ying Nie,1 He Bo,2 Weiqun Zhang3, and Haipeng Zhang1

1School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
2Postdoctoral Research Station, Dongbei University of Finance and Economics, Dalian 116025, China
3School of Statistics, Xi’an University of Finance and Economics, Xi’an 710100, China

Correspondence should be addressed to Weiqun Zhang; 1304777607@qq.com

Received 27 November 2019; Revised 2 February 2020; Accepted 5 February 2020; Published 18 March 2020

1. Introduction

Recently, based on the exhaustion of fossil fuels and increasing demands for environmental protection, wind power and other new energy industries have developed rapidly [1]. Wind energy has the advantages of renewability and cleanliness, so the comprehensive development and utilization of wind energy have a wide range of social and economic benefits. Therefore, wind energy is a very promising resource around the world [2]. However, in practice, wind speed has the characteristics of inherent randomness and intermittence, meaning the effective and comprehensive development of wind systems is very limited, which poses a major challenge to the operation and management of power grids, particularly when considering wind power integration [3].

Generally, effective wind speed prediction can reduce the risks of wind power generation associated with uncertainty. Predicting wind speed accurately is a difficult task and major focus for wind farm decision-makers. It is very important to establish a suitable wind farm architecture and determine the nonlinear dynamic modes of wind speed precisely for the sake of efficient management and minimizing potential risk [4]. Wind speed prediction methods can be divided into four different types: ultrashort term (several seconds to 4 h), short term (4 to 24 h), medium term (1 to 7 d), and long term (more than 7 d). Different prediction horizons have different application values. For short-term and ultrashort-term wind speed prediction, the most critical impact is its role in power system operation [5]. For example, the output power of a wind farm in the United States can fluctuate by several hundred megawatts within an hour, which has a significant
impact on the safe and stable operation of the system. To avoid potential problems, short-term wind speed prediction is crucial for providing data support for the reasonable dispatching of power resources, improving the efficiency of optimal dispatching, and optimizing the use of various power generation methods [6].

In recent decades, many wind speed prediction methods have been proposed in three main categories: physical methods, statistical methods, and artificial intelligence methods [7–9]. Physical methods mainly use detailed data from the lower atmosphere for analysis, mining, and prediction [10]. Such models are based on the basic information of wind turbines provided by numerical weather forecasting systems and the parameterization of physical phenomena according to initial conditions and nonlinear partial differential equation systems, which can be used to obtain a series of different meteorological parameters [11]. However, such models do not analyze and mine historical data, so they ignore potentially useful information in historical data [12]. Additionally, based on the inclusion of different parametric models within a single larger model, there are some difficulties in applying such models to wind farms. Using such models alone for wind energy mining and prediction will produce relatively large system errors [13]. Statistical models use large amounts of historical data to perform prediction without considering the impact of many meteorological factors [14]. In early research on wind speed prediction, traditional statistical models largely relied on autoregression (AR) [15] and its extensions (e.g., ARIMA [14, 16], real ARIMA [17], and ARIMA-ARCH [18]). However, for wind speed series containing complex information, such models have difficulty in accurately mining patterns, particularly nonlinear patterns.

Since the initial development of artificial intelligence technology, intelligent prediction models have been designed and applied to wind energy forecasting [19], including artificial intelligence systems [20, 21], support vector machines [22], and fuzzy logic methods [23]. Additionally, based on the strong nonlinearity of wind energy data, only nonlinear models have reasonable prediction ability [24]. However, based on the inherent disadvantages of individual models, they cannot achieve the expected prediction results in all circumstances [25, 26]. Based on the increasing application of wind power generation in electronic systems, developing effective controls for prediction error is crucial [27]. To compensate for the shortcomings of individual models, some hybrid models have been designed for wind speed prediction to achieve better prediction performance [28]. Generally speaking, the superior ability of hybrid models makes it easier to achieve accurate wind speed predictions compared to the abilities of individual models. Therefore, many research reports on hybrid forecasting models are put forward every year. Such reports tend to focus on data preprocessing [29, 30], combining single models, and using heuristic algorithms, such as particle swarm optimization [31], genetic algorithms [32], and multiobjective algorithms [33], to optimize model parameters.

In recent years, hybrid models have been applied to both long-term and short-term wind speed prediction. For short-term wind speed forecasting, Ma et al. [34] developed a wind speed prediction model by using singular spectrum analysis (SSA) to derive a noise removal sequence corresponding to a real sequence to predict short-term wind speeds. Wang et al. [35] developed a hybrid wind speed prediction model using complete ensemble empirical mode decomposition (CEEMD), multiobjective whale optimization, and an Elman neural network. Meng et al. [36] proposed a short-term wind speed prediction hybrid model combining data preprocessing with an artificial neural network and various optimization methods. Assessments of the effectiveness of their model revealed that its prediction accuracy was significantly improved compared to several benchmark models. In [37], an effective short-term prediction framework for wind speed was proposed by combining a local linear fuzzy neural network, discrete wavelet transform, and singular spectrum analysis optimized by the seeker optimization algorithm. In [38], a hybrid model based on wind speed prediction was proposed to combine variational mode decomposition (VMD) with an extreme learning machine (ELM) optimized using the hybrid backtracking search optimization algorithm. This model achieved excellent performance in terms of describing nonlinear modes. These studies demonstrate not only that hybrid strategies are superior to individual models but also that such strategies can be used as an effective form of engineering application technology. Additionally, there have been a large number of studies on medium- and long-term wind speed prediction. For example, Wang et al. [39] combined support vector regression with seasonal index adjustment and an Elman recurrent neural network to construct hybrid models called PMERN and PAERN, which performed the mid-term prediction of wind speed effectively. Ulkut and Güney [40] proposed a method to determine wind speeds corresponding to specific positions without relying on previous wind speed data, which is effective for long-term wind speed prediction, by combining physical factors with an artificial neural network. The prediction results of hybrid models based on the mechanisms discussed above demonstrate the short- and long-term wind speed prediction effectiveness of hybrid models.

Another problem regarding wind energy prediction is that many studies focus on a single mode of prediction. Specifically, most previous studies have focused on either point prediction or interval prediction of wind speed alone, without considering how both models could be used together for predictive modeling and analysis. Therefore, existing models cannot meet the needs of engineering applications or guarantee the reliability of wind systems. Existing probability interval prediction methods can generate a large quantity of predictions that can help managers implement appropriate policies. However, the study of interval modeling and prediction is still insufficient. The main research direction for uncertainty quantification focuses on statistical methods, including quantile regression [41, 42], bootstrapping [43], and kernel density estimation [44]. Additionally, several interval prediction methods have been proposed based on artificial neural networks, lower bound estimation (LUBE) [45], and so forth.

Table 1 summarizes existing methods and models for wind speed point prediction and interval prediction, as well as
as the advantages and disadvantages of these methods. The main points in Table 1 can be summarized as follows.

In the area of point prediction, (1) although physical models provide good long-term prediction ability, their application is limited based on complicated meteorological conditions, difficult model initialization, and excessive computations. (2) Traditional statistical models, such as AR and ARIMA, have enhanced computational efficiency. However, the modeling of nonlinear time series, such as wind speed time series, is limited by the linear forms of such models. (3) An important issue related to artificial neural networks is that network iteration can easily fall into local optima, although such networks do provide good nonlinear time series modeling ability. (4) Although recent combined models successfully incorporate the advantages of individual models and improve prediction accuracy significantly, model combination technology in existing systems always revolves around linear combination. Based on the nonlinear characteristics of wind speed, this paper presents a method for combining individual prediction models in a nonlinear manner and optimizing model parameters using a multi-objective optimization algorithm to improve prediction effectiveness further.

Regarding interval prediction, (1) based on the unique advantages of quantile regression, most research has focused on this method. However, quantile regression is disadvantageous for developing prediction intervals because it must obtain a specific training dataset to establish prediction models. Additionally, every quantile must be considered, which increases computational complexity and the probability of discarding useful results during resampling [46]. (2) Bootstrapping methods are statistical methods that apply data resampling and replacement to evaluate the robustness of various statistics, including standard error, confidence interval parameters, correlation coefficients, and regression coefficients. Bootstrapping methods can compensate for the shortcomings of quantile regression methods but are only helpful for handling small sample sizes [47]. (3) Kernel density estimation can quickly calculate intervals based on point prediction results and a given statistical historical error distribution. However, such methods require the strict assumption of distributions [48]. (4) The LUBE method eliminates the shortcomings of traditional interval prediction methods and has high computational efficiency in terms of hypothesizing distributions, but its complex objective function cannot be obtained using conventional methods. In
summary, there is no unified interval prediction method and further research and investigation are required to obtain more effective results [49]. Therefore, we developed a novel interval prediction architecture that outperforms most individual interval prediction models based on assumed distributions. In the proposed interval prediction architecture, there are no hypotheses regarding distributions and models. Therefore, the established interval structure possesses powerful anti-interference ability in the presence of outliers in interval data.

Based on our review of the literature and methods described above, the major contribution of this article is the presentation of a hybrid double prediction system that is designed to combine the point prediction and probability interval prediction of wind speed, which compensates for the shortcomings of existing research. The proposed system is composed of a wind speed point prediction module and interval prediction module, which can provide numerous predictions for the managers of wind farms. Specifically, the proposed double prediction system includes a preprocessing module based on VMD, a prediction module based on a nonlinear combination model, an interval prediction module, and an evaluation module. As a relatively new signal processing technology, VMD decomposes wind speed sequences and then performs denoising and reconstruction to generate a time sequence with greater clarity. The nonlinear combination model proposed in this paper is an effective prediction. ELM [9], a generalized regression neural network (GRNN) [50], and ARIMA [51] are selected as the base models for combination. The prediction results from these three models are aggregated using backpropagation (BP) [52], which is a form of nonlinear combination. BP is very sensitive to the selection parameters, which directly determine the effectiveness of point prediction and interval prediction. Therefore, to identify the optimal parameters for the BP model, we adopted the multiobjective evolutionary algorithm based on decomposition (MOEA/D). Additionally, to verify the performance of the proposed prediction architecture, we selected ten indicators to judge the accuracy of prediction. We present a thorough discussion on the verification of the effectiveness of the prediction system in this paper.

Wind speed data from Penglai in Shandong province were selected as experiment datasets on which to test the performance of the proposed system. Shandong province is located on the east coast of China and is rich in wind energy resources. To meet the needs of social development, energy conservation, and environmental protection, Shandong has developed many wind power stations. As a coastal province, Shandong has one of the largest wind farms in China. By the end of 2018, the total installed capacity of wind power was 11.26 million kilowatts. In this study, Penglai city, which is located in northern Shandong province, was selected as a research area based on its huge energy potential and valuable wind energy resources.

The major innovations of the proposed predictive system can be summarized as follows.

1. In this paper, a novel double prediction system for wind speed is established based on certain point prediction and uncertain interval prediction. The goal of the proposed system is to enhance the accuracy of point prediction, enhance the construction efficiency of prediction intervals, and enhance the operation level of wind power systems. Numerical simulation results demonstrate that our model has satisfactory prediction abilities.

2. A nonlinear combination method based on a BP optimization method is proposed. To determine the optimal combination mode for each model and overcome the limitations of existing linear combination models for nonlinear wind speed data, a nonlinear aggregation mechanism based on ELM is used to combine different models to compensate for the inherent defects of individual models and linear combinations. The MOEA/D algorithm is applied to search for the best parameters for ELM to improve prediction accuracy further.

3. An interval prediction method based on fuzzy clustering is established. Compared to traditional parametric statistical models, one unique advantage of the proposed prediction model is its convenience because it does not need to know distribution shapes. This feature significantly reduces the complexity of the model and enhances the overall efficiency of the system.

The remainder of this article is organized as follows. Section 2 discusses the relevant methods applied in the proposed double prediction architecture. In Section 3, the double prediction model is established comprehensively. In Section 4, data are introduced and experimental results are analyzed. Further discussion is provided in Section 5. Finally, our conclusions are summarized in Section 6.

1.1. Knowledge and Tools for Model Preparation. When constructing our model, several methods were selected based on their unique advantages and combined to enhance the overall performance of the model. Here, we introduce the two main methods, which are variational mode decomposition [53] and multiobjective evolutionary algorithm based on decomposition.

First, we will discuss variational mode decomposition (VMD).

VMD is an effective data preprocessing method proposed by Dragomiretskiy and Zosso [54] in 2014. The goal of VMD is to decompose a real input signal sequence $f$ into a series of sub-signal sequences $y_k$ called modes, which have specific sparsity characteristics when reproducing the input. For completeness, signals $f$ and modes $y_k$ are required to be complete and square-integrable to the second derivative (i.e., $f, y_k \in L^1 \cap W^{2,2}$). Each mode $k$ maximally pulsates around a center $w_k$.

Step 1. Accessing the bandwidth of each mode.

For each mode $y_k$, the Hilbert transform is applied to calculate a correlation analysis signal and a unilateral frequency spectrum is obtained. Next, mixed exponents are adjusted to their estimated central frequencies to transfer the
Complexity

spectrum of the mode to the baseband. The bandwidth is estimated based on the $H^1$ Gauss smoothness of the demodulated signal (i.e., the square $L^2$-norm of the gradient). The constrained variational problem is defined as follows:

$$\min \left\{ \sum_{k=1}^{K} \left\| \delta_{\lambda}(t) \times y_k(t) \right\|_2^2 \right\}$$

subject to

$$\sum_{k=1}^{K} y_k = f(t),$$

where $\delta(t)$ represents the Dirac distribution and $k$ and $t$ represent the number of modes and time spaces, respectively. Furthermore, $\{y_k\}$ is the set of modes $\{y_1, y_2, \ldots, y_k\}$ and $\{w_k\}$ is the set of center pulsations $\{w_1, w_2, \ldots, w_k\}$.

Step 2. Defining the optimization problem.

Considering the penalty term and Lagrange multiplier $\lambda$, the constrained optimization problems above can be redefined as follows:

$$L(\{y_k\}, \{w_k\}, \lambda) = \alpha \sum_{k=1}^{K} \left\| \delta_{\lambda}(t) \times y_k(t) \right\|_2^2 + \left\| f(t) - \sum_{k=1}^{K} y_k(t) \right\|_2^2 + \left\| \lambda(t) - f(t) \right\|_2^2,$$

where $\alpha$ represents the equilibrium parameter of the data fidelity constraint.

Step 3. Solving for the modes $\{y_k\}$ and center pulsations $\{w_k\}$.

By adopting the multiplier alternating direction method, the process of solving for $y_k$ and $w_k$ can be defined as follows. For $y_k$, the minimization problem is defined as

$$y_k^{n+1} = \arg\min_{y_k \in X} \left\{ \alpha \left\| \delta_{\lambda}(t) \times y_k(t) \right\|_2^2 + \left\| f(t) - \sum_{i} y_i(t) \right\|_2^2 + \left\| \lambda(t) - f(t) \right\|_2^2 \right\}.$$

which can be rewritten as

$$\tilde{y}_k^{n+1} = \arg\min_{\tilde{y}_k \in X} \left\{ \alpha \left\| \delta_{\lambda}(t) \times \tilde{y}_k(t) \right\|_2^2 + \left\| f(t) - \sum_{i} \tilde{y}_i(t) + \frac{\lambda(t)}{2} \right\|_2^2 \right\}.$$
In this case, $\mathbf{V}$ is said to be dominated by $\mathbf{U}$, which can be denoted as $\mathbf{u} \succ \mathbf{v}$, where $\succ$ represents a set of dominant relationships.

If there is no point $x \in \Omega$ that makes $F(x)$ dominate $F(x')$, then the point $x^* \in \Omega$ is Pareto optimal. There is only one optimal set of compromise solutions called non-dominated solutions (i.e., not dominated by all other solutions). The values of the Pareto optimization solution in the determined space and target space are defined as the Pareto solution set (PS) and Pareto frontier, respectively [56].

MOEA/D has strong search ability for continuous optimization, combinatorial optimization, and PS complex problems. The main principle of this algorithm can be summarized as follows.

If a multiobjective optimal problem (e.g., equation (10)) and weight vector $\lambda = (\lambda_1, \ldots, \lambda_m)$ are given, and the given weight vector satisfies $\sum_{i=1}^{m} \lambda_i = 1, \lambda_i \geq 0, i = 1, 2, \ldots, m$, then the MOEA/D can be applied. MOEA/D based on Tchebycheff decomposition uses the weight vector to optimize a MOP into several subproblems based on the following methods:

$$\min_{x \in \Omega} g^c(x \mid \lambda^i, z^*) = \min_{x \in \Omega} \max_{1 \leq i \leq m} \left\{ \lambda_i \mid f_i(x) - z_i^* \right\}, \quad (12)$$

where $z^* = (z_1^*, z_2^*, \ldots, z_m^*)$ (i.e., $z_i^* \leq \min_{x \in \Omega} f_i(x)$) is the ideal point and $\lambda^i = (\lambda_1^i, \ldots, \lambda_m^i)^T$. By solving multiple subproblems with different weight vectors based on equation (12), a Pareto optimal solution set [57] with good diversity can be obtained.

It is known that $g^c$ is continuous in $\lambda$, so if $\lambda^i$ is close to $\lambda^j$, then the $g^c(x \mid \lambda^j, z^*)$ solution must be close to the $g^c(x \mid \lambda^j, z^*)$ solution. Therefore, a useful tool for $g^c(x \mid \lambda^i, z^*)$ optimization is information regarding $g^c$ with weight vectors near $\lambda^i$.

In the MOEA/D, the population is made up of the optimal solutions to the current subproblem. Each subproblem maintains a list of neighbors, and this list preserves subproblems with weight vectors similar to those of the current subproblem. Therefore, under the assumption of continuity, two neighboring subproblems should have similar optimal solutions. In each generation of MOEA/D, each subproblem is optimized using only the information from its neighboring subproblems.

For each generation $t$, MOEA/D using the Tchebycheff decomposition satisfies the following conditions.

1. A point group $x^1, \ldots, x^N \in \mathbb{N}$, where the $x^i$ is the current solution for the $i$-th subproblem.
2. $FV^1, \ldots, FV^N$, where $FV^i = F(x^i)$ for each $i = 1, \ldots, N$.
3. $z = (z_1, \ldots, z_m)^T$ and $z_i$ is the best value found now for objective $f_i$.
4. An external population is available to store the nondominated solutions found during the search.

The pseudocode of MOEA/D is described as Algorithm 1 below:

### 2. Construction of the Wind Speed Double Prediction System

This section discusses the proposed wind speed double prediction system architecture, including system establishment and evaluation.

#### 2.1. System Establishment

The prediction system proposed in this paper consists of two modules: a point prediction module and interval prediction module. The following subsections describe the system construction process and the system structure is illustrated in Figure 1.

#### 2.1.1. Point Prediction Module

In this section, we propose a novel type of nonlinear hybrid point forecasting model using ELM, GRNN, and ARIMA, as well as a BP network, the MOEA/D, and a nonlinear combination mechanism, to achieve stable high-precision wind speed point prediction results. Considering the excellent prediction performance of BP networks, the proposed method adopts a BP network for nonlinear combination.

The point prediction module in the designed system is composed of four stages. The details of each stage are discussed below.

(i) First stage: wind speed data preprocessing.

To remove the noise and extract helpful information from a wind speed sequence, we use VMD technology to disintegrate an original sequence and reconstruct a smooth time series. Specifically, an original sequence is decomposed into several intrinsic mode functions (IMFs). IMFs with higher frequencies are eliminated to filter the time series. Here, we remove IMF1, IMF2, and IMF3, and the remaining IMFs are reconstructed to derive the final series.

(ii) Second stage: single-model prediction.

In the proposed method, we first use individual models to predict points. Specifically, we use ELM, GRNN, and ARIMA as individual prediction models to construct a combined model. ELM and GRNN are adopted to handle the nonlinear characteristics of wind speed and ARIMA is adept at discerning the linear characteristics of wind speed data. In this study, we divided 4464 pieces of wind speed data into a training set $\text{train}_1$ and testing set $\text{test}_1$, where $\text{train}_1$ contained 3964 pieces of data and $\text{test}_1$ contained 500 pieces of data. In general, there are no clear regulations regarding the ratio of training sets and testing sets for neural networks. It is common practice to use approximately 2/3 to 4/5 of the sample data for training and the remaining samples for testing. When the quantity of data is large, the data proportion in the training set can be increased appropriately [58]. As the proportion of training data increases, the neural network can achieve better prediction accuracy [59, 60].
Input:
(i) MOP—multiobjective optimization problem
(ii) N— the number of the MOEA/D subproblems
(iii) $\lambda_1, \ldots, \lambda_N$—a uniform distribution of $N$ weight vectors
(iv) $T$—the number of the weight vectors in the neighborhood of each weight vector
(v) max_gen—the maximum number of generations

Output:
(i) EP—external population

Setup:
(i) Set EP = $\emptyset$
(ii) gen = 0

Step 1: Initialization
(i) /*Initialize an primary internal population uniformly randomly.*/
\[ P_{i} = \{x_1, \ldots, x_N\} \text{ and } FV_{i} = F(x) \]
(ii) /*Initialize $z = (z_1, \ldots, z_n)^T$ by a specific problem method.*/
(iii) /* Calculate the Euclidean distance between any two weight vectors, and then calculate the closest $T$ weight vectors to each weight vector.*/
(iv) $\forall i = 1, \ldots, N$, set $B(i) = \{i_1, \ldots, i_T\}$ $\lambda_{i_1}, \ldots, \lambda_{i_T}$ represent the $T$ closest weight vectors to $\lambda_i$

Step 2: Updating
(i) /* WHILE ($t < \text{max} \_\text{gen}$) DO */
(ii) /* FOR EACH $i = 1, \ldots, N$ DO */
\[ F_{V_{i}} = F(x) \]
(iii) /* FOR EACH $j = 1, \ldots, n$ DO */
\[ g_{o_{e}}(y | \lambda_{j}, z) \leq g_{o_{e}}(x_{j} | \lambda_{j}, z^x) \]
if $z_j < f_j(y)$, then set $z_j = f_j(y)$
END FOR
(iv) /* FOR EACH index $j \in B(i)$ DO */
\[ F_{V_{j}} = F(y) \]
if $z_j < f_j(y)$, then set $x_j = y$ and $F_{V_{j}} = F(y)$.
END FOR
(v) /* Update of EP. */
/* Remove from EP all the vectors dominated by $F(y)$. Add $F(y)$ to EP if no vector in EP dominate $F(y)$.*
END FOR
(vi) $t = t + 1$
END WHILE
(vii) RETURN EP

Algorithm 1: MOEA/D.

Therefore, this division of the model is sufficient to construct a model and verify its accuracy.

The input and output structures of $\text{train}_1$ are described in equations (13) and (14), respectively. By using ELM and GRNN models trained on $\text{train}_1$ to predict the wind speeds in the test set, the prediction sequences $\text{predict}_1$ and $\text{predict}_2$ are obtained, respectively. Similarly, ARIMA is used to obtain the prediction sequence $\text{predict}_3$.

\[
\text{train}_1 = \begin{bmatrix}
  x(1) & x(2) & \cdots & x(l) \\
  x(2) & x(3) & \cdots & x(l+1) \\
  \vdots & \vdots & \ddots & \vdots \\
  x(n-1) & x(n-l+1) & \cdots & x(n-1) \\
\end{bmatrix}, \quad (13)
\]
\[
\text{test}_1 = \begin{bmatrix}
  x(l+1) & x(l+2) & \cdots & x(n) \\
\end{bmatrix}, \quad (14)
\]

where $n$ is the number of samples in $\text{train}_1$, $l$ is the look-back time lag, and $x(k)$ is the wind speed value at time $k$. For example, consider $x(1)$ to $x(l)$ as inputs and $x(l+1)$ as an output. Here, we set $l = 5$.

(iii) Third stage: nonlinear combination model construction.

To obtain an effective combination of each model, a nonlinear decision-making method based on an optimized BP neural network is proposed to obtain optimal results. Specifically, we divide $\text{test}_1$ into a training set $\text{train}_2 (356)$ and testing set $\text{test}_2 (144)$ and then use $\text{predict}_1$, $\text{predict}_2$, and $\text{predict}_3$ as BP network inputs and the 356 data in $\text{train}_2$ as outputs. It is worth noting that it is difficult to determine the weights and thresholds for the neurons in each layer of the BP network, so MOEA/D is adopted to search
for the best weights and thresholds for the neural network. The input and output structure for BP network training are defined in equations (15) and (16), respectively.

\[
\text{input: } [\hat{x}_1(1) \hat{x}_2(1) \hat{x}_3(1)]\]
\[
= \begin{bmatrix}
\hat{x}_1(1) & \hat{x}_2(1) & \hat{x}_3(1) \\
\hat{x}_1(2) & \hat{x}_2(2) & \hat{x}_3(2) \\
\vdots & \vdots & \vdots \\
\hat{x}_1(N) & \hat{x}_2(N) & \hat{x}_3(N)
\end{bmatrix},
\]

(15)

\[
\text{output: } \text{test}_2 = [ x(1) \ x(2) \ \cdots \ x(N) ],
\]

(16)

where \( N \) represents the number of samples in \( \text{test}_2 \) and \( \hat{x}_1(k), \hat{x}_2(k), \hat{x}_3(k) \) represent the \( k \)th wind speed values predicted by ELM, GRNN, and ARIMA, respectively. By using the \textit{input} and \textit{output}, an optimized BP neural network can be trained.

(iv) Fourth stage: wind speed point prediction.

According to the established nonlinear prediction model, a rolling prediction method is used for multistep prediction and final prediction results are obtained. Evaluation indexes are calculated using the prediction results and \( \text{test}_2 \), and the performance of the model is evaluated.

In particular, multistep forecasting means forecasting multiple load values in the future. A time index \( t \) is the forecast origin and a positive integer \( l \) is the forecast horizon. It can be assumed that the time index \( t \) is exactly the time point that we are in, and our target is to obtain the forecasting value \( \hat{y}_{test}(l \geq 1), l = 1, 2, 3 \) corresponds to 1 step, 2 steps, and 3 steps, respectively.

2.1.2. Interval Prediction Module. The interval prediction method in the proposed system was developed using point prediction results based on a fuzzy system. The three main steps in this module are summarized below.

(i) First stage: data classification.
In this step, the training set \textit{train}, of wind speed data is clustered into several classes using fuzzy c-means clustering. We assume that the data in each category follows the same normal distribution. Therefore, we can derive a set of interval classes $F_1, F_2, \ldots, F_k$.

Here, we consider \textit{site 1} as an example. One can see that the data is divided into ten categories, where the scope of each category is defined in (17).

\begin{align}
1: & \ (-\infty, 5.59] & 2: & \ (5.59, 6.04] & 3: & \ (6.04, 6.31] & 4: & \ (6.31, 6.50] & 5: & \ (6.50, 6.71] \\
6: & \ (6.71, 6.93] & 7: & \ (6.93, 7.18] & 8: & \ (7.18, 7.75] & 9: & \ (7.75, 8.42] & 10: & \ (8.42, \infty+].
\end{align}

(ii) Second stage: wind speed interval estimation.

The confidence degree of each category interval is 95%. According to the mean and variance of each category of data, a corresponding confidence interval is constructed. Different categories have different widths of unified prediction intervals. This process of constructing different adaptive intervals according to different data characteristics is one of the main innovations of our model. According to the testing set $test_t$ of point prediction results from the point prediction module, we identify the category $F$ to which each prediction value belongs. Then, according to the constructed confidence interval for each category, the prediction interval for each prediction value is calculated as follows:

$$\left[x_i - z_{1-(\alpha/2)} \frac{s_j}{\sqrt{n_j}}, x_i + z_{1-(\alpha/2)} \frac{s_j}{\sqrt{n_j}}\right],$$

where $x_i$ is a point prediction value, $j$ is the category number of $x_i$, $s_j$ is the standard deviation of category $j$, and $n_j$ is the number of data samples in category $j$.

(iii) Third stage: sorting prediction results.

According to the prediction intervals derived above, final interval estimations for wind speed can be obtained.

### 2.2. System Evaluation

The evaluation indexes for the designed double prediction system are introduced in this section, including four indexes for point prediction and six indexes for interval prediction.

#### 2.2.1. Point Prediction Evaluation

Generally speaking, evaluation criteria are not unique to a given prediction system. This paper uses four common evaluation standards to evaluate the ability of the developed model and other comparative models, namely, mean absolute error (MAE) [61], mean squared error (MSE) [62], mean absolute percentage error (MAPE) [63], and direction change (DC). The smaller the values of MAPE, MSE, and MAE, the better the prediction performance. If the DC value is relatively large, the predicted direction of motion is considered to be consistent with the real value. Table 2 provides additional details regarding these four indexes.

#### 2.2.2. Interval Prediction Evaluation

For interval prediction, we selected six evaluation indexes: prediction interval coverage probability (PICP) [64], prediction interval normalized average width (PINAW) [65], average coverage error (ACE), average width of the constructed PIs (MPI), Winkler score (WS), and accumulated width deviation (AWD) [66]. Table 3 provides specific definitions for these six indexes.

In this paper, PICP specifically refers to the PICP of the testing dataset, which is the main evaluation index for interval prediction. It indicates the coverage effect of the obtained confidence intervals relative to the target value. Given a confidence level, if the PICP is greater than or equal to (1-alpha), then the constructed interval is valid. Otherwise, the constructed interval is invalid. PINAW refers to the NAW of the prediction interval of the testing dataset. The cost of reducing the width diminishes the probability of achieving the expected target coverage. Increasing coverage requires increasing the width of the interval, so PICP and PINAW are essentially contradictory [48]. ACE represents the difference between the coverage and confidence of a prediction interval. MPI represents the average width of an obtained interval [6]. Similarly, the quality of an interval can also be assessed by its Winkler score. A high-quality interval has a smaller Winkler absolute value for an assigned nominal confidence level [66]. AWD refers to the AWD of the testing dataset, which can be obtained by calculating its relative deviation degree. The cumulative sum of AWD, represents the relative deviation degree [67]. Table 3 provides specific descriptions of these formulas.

Among the formula, $y_i$ and $\bar{y}_i$ represent the true value and predicted value of wind speed, respectively. $N$ represents the testing set number.

Besides, $a_i$ is the directional factor and is calculated as

$$a_i = \begin{cases} 
0, & \text{otherwise}, \\
1, & \text{if } ((y(t+1) - y(t))((\bar{y}(t+1) - y(t))) > 0.
\end{cases}$$

$$\text{(19)}$$

### Table 3

| Index | Formula | Description |
|-------|---------|-------------|
| PICP  | $\text{PICP} = \frac{\text{Number of Samples in Constructed Interval}}{\text{Number of Samples in Test Set}}$ | Prediction interval coverage probability |
| PINAW | $\text{PINAW} = \frac{\text{Average Width of Constructed PIs}}{\text{Average Width of Target Interval}}$ | Prediction interval normalized average width |
| ACE   | $\text{ACE} = \frac{\text{Average Coverage Error}}{\text{Average Coverage of Target Interval}}$ | Average coverage error |
| MPI   | $\text{MPI} = \frac{\text{Average Width of Constructed PIs}}{\text{Number of Samples}}$ | Average width of the constructed PIs |
| WS    | $\text{WS} = \frac{\text{Winkler Score}}{\text{Number of Samples}}$ | Winkler score |
| AWD   | $\text{AWD} = \frac{\text{Accumulated Width Deviation}}{\text{Number of Samples}}$ | Accumulated width deviation |

\[\text{(17)}\]
Table 2: Four point prediction evaluation metrics.

| Indicator | Definition | Equation |
|-----------|------------|----------|
| MAE       | Mean absolute error | $MAE = (1/N) \sum_{i=1}^{N} |y_i - \hat{y}_i|$ |
| MAPE      | Mean absolute percentage error | $MAPE = (1/N) \sum_{i=1}^{N} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\%$ |
| MSE       | Mean squared error | $MSE = (1/N) \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$ |
| DC        | Directional change | $DC = (100N - 1) \sum_{i=1}^{N} \alpha_i$ |

Table 3: Six interval prediction evaluation metrics.

| Indicator | Definition | Equation |
|-----------|------------|----------|
| PICP      | Prediction interval coverage probability | $PICP = (1/N) \sum_{i=1}^{N} c_i \times 100\%$ |
| PINAW     | Prediction interval normalized average width | $PINAW = (1/N) \sum_{i=1}^{N} \left( U_i - L_i / (y_{max} - y_{min}) \right) \times 100\%$ |
| ACE       | Average coverage error | $ACE = PICP - \mu$ |
| MPI       | Average width of the constructed PIs | $MPI = (1/N) \sum_{i=1}^{N} |U_i - L_i|$ |
| WS        | Winkler score | $WS = (1/N) \sum_{i=1}^{N} S_i$ |
| AWD       | Accumulated width deviation | $AWD = (1/N) \sum_{i=1}^{N} AWD_i$ |

3. Experiments and Analysis

This section discusses the application of the double prediction model and several comparative models. The comparisons are divided into three experimental demonstrations. The operating environment of the experiments was a PC with a 2.40 GHz CPU, 4.00 GB of RAM, Windows 7 operating system, and MATLAB R2016A platform. Considering random factors, to guarantee the reliability of final results, 20 trials were conducted for each experiment and the average values were recorded.

3.1. Dataset Description. The wind speed data for three sites at Penglai in Shandong Province are chosen as experimental datasets on which to test the performance of the established double prediction system. Basic information regarding this wind speed data is provided in Table 4. Descriptive statistical analysis uses four statistical indicators, namely, the maximum, minimum, and average values, as well as standard deviation (Std.). The basic information and original data for the selected sites are presented in Figure 2.

For the sake of estimating the prediction effects of the models, 10 min wind speed data blocks from the Penglai wind farm from January 1, 2011, to January 31, 2011, were selected as experimental data. This wind farm consists of three different sites. Each dataset contained 4464 data points, which were divided into a training set train1 and testing set test1. The training set train1 contained 3964 data points and the testing set test1 contained 500 data points. For nonlinear aggregation, the testing set test1 was subdivided into a training set train2 and testing set test2 containing 356 and 144 points, respectively. For both the training and testing sets, we used a rolling forecasting mechanism to predict wind speed and produce one-step and two-step prediction results. The data structure details of the double prediction model are presented in Figure 2.

Wind speed of Penglai, Shandong province (37.48N, 120.45E), from January 7 to January 17, 2011.

3.2. Diebold-Mariano Test. To determine if the designed hybrid model provides better forecasting results than the comparative models, we adopted an effective verification method called the DM test, which was proposed by Diebold and Mariano RS [46]. The theory behind the DM test is summarized below.

Considering a significance level $\alpha$, the zero hypothesis $H_0$ indicates that the predictive effectiveness levels of the proposed model and a comparative model are not significantly different. The meaning of $H_1$ is opposite to that of $H_0$.

The relevant formulas are defined as follows:

$$H_0: E[L(\text{err}_1^{\text{r}})] = E[L(\text{err}_2^{\text{r}})]$$
$$H_1: E[L(\text{err}_1^{\text{r}})] \neq E[L(\text{err}_2^{\text{r}})]$$

where $L$ represents the loss function for prediction error and $\text{err}_1^{\text{r}}$ and $\text{err}_2^{\text{r}}$ are the error sequences predicted by the selected models.

Additionally, the statistics of the DM test can be defined as follows:

$$DM = \frac{\sum_{i=1}^{n} (L(\text{err}_1^{\text{r}}) - L(\text{err}_2^{\text{r}}))^2}{\sqrt{S^2/n}}$$

where $S^2$ is the estimate of the variance of $d_i = L(\text{err}_1^{\text{r}}) - L(\text{err}_2^{\text{r}})$. Assuming a certain significance level $\alpha$, the obtained DM value is compared to $z(\alpha/2)$. Once the DM statistics exceed the interval $[-z(\alpha/2), z(\alpha/2)]$, $H_0$ can be rejected. This indicates that the predictive performances of the target model and a comparative model are significantly different, meaning $H_1$ will be accepted.
3.3. Results and Analysis of Point Prediction. For the sake of verifying the applicability of the proposed point prediction module, two experiments are presented in this section, which are denoted as experiment I and experiment II, respectively. The main purpose of experiment I was to prove the superiority of the nonlinear combination model in the point prediction module compared to a single model, thereby reasonably proving the validity of hybrid modeling. Additionally, the results of experiment I demonstrate the necessity of data preprocessing. Similarly, to demonstrate the rational and superior ability of the VMD technology adopted in our system, it was compared to other common data preprocessing methods in experiment II. Detailed analysis of each experiment is provided below.

3.3.1. Experiment I: Comparison to Individual Models. In this experiment, all experimental datasets were considered to assess the effectiveness of the point prediction module based on three comparisons. In the first comparison, the proposed model was compared to three preprocessed data models, X1, X2, X3, X4, X5, X6. The 10-minute wind speed data of Penglai (37.48N,120.45E) wind farm from January 1, 2011 to January 31, 2011 is selected as the experimental data, which contains three datasets.

Table 4: Data structure of three selected datasets in Penglai, Shandong Province.

| Site  | Samples | Numbers | Max | Min  | Mean | Std.  |
|-------|---------|---------|-----|------|------|-------|
| Site 1| All samples | 4464 | 17.8 | 0.9 | 7.4970 | 2.7793 |
|       | Training | 3964 | 17.8 | 1.9 | 7.6431 | 2.8474 |
|       | Testing  | 500  | 11.0 | 0.9 | 6.3386 | 1.7834 |
| Site 2| All samples | 4464 | 18.0 | 0.9 | 8.6648 | 2.9352 |
|       | Training | 3964 | 18.0 | 1.5 | 8.8022 | 2.9923 |
|       | Testing  | 500  | 13.5 | 0.9 | 7.5711 | 2.1417 |
| Site 3| All samples | 4464 | 18.4 | 0.7 | 7.9148 | 3.0421 |
|       | Training | 3964 | 18.4 | 1.5 | 8.1144 | 3.0827 |
|       | Testing  | 500  | 11.8 | 0.7 | 6.3326 | 2.1147 |

Figure 2: Description of observations in three datasets. On the left is the geographic location obtained by data points. The right part of the graph shows the dividing method of training set and testing set.

The 10-minute wind speed data of Penglai (37.48N,120.45E) wind farm from January 1, 2011 to January 31, 2011 is selected as the experimental data, which contains three datasets.

Table 4: Data structure of three selected datasets in Penglai, Shandong Province.

| Site  | Samples | Numbers | Max | Min  | Mean | Std.  |
|-------|---------|---------|-----|------|------|-------|
| Site 1| All samples | 4464 | 17.8 | 0.9 | 7.4970 | 2.7793 |
|       | Training | 3964 | 17.8 | 1.9 | 7.6431 | 2.8474 |
|       | Testing  | 500  | 11.0 | 0.9 | 6.3386 | 1.7834 |
| Site 2| All samples | 4464 | 18.0 | 0.9 | 8.6648 | 2.9352 |
|       | Training | 3964 | 18.0 | 1.5 | 8.8022 | 2.9923 |
|       | Testing  | 500  | 13.5 | 0.9 | 7.5711 | 2.1417 |
| Site 3| All samples | 4464 | 18.4 | 0.7 | 7.9148 | 3.0421 |
|       | Training | 3964 | 18.4 | 1.5 | 8.1144 | 3.0827 |
|       | Testing  | 500  | 11.8 | 0.7 | 6.3326 | 2.1147 |
namely, VMD-ARIMA, VMD-GRNN, and VMD-ELM, to analyze the advantages of the combination model and nonlinear combination method. In the second comparison, the three VMD-based models were compared to ARIMA, GRNN, and ELM, respectively. In the third comparison, the effectiveness of the designed prediction model was evaluated further by using the traditional wave.nn model and BP as comparative methods. The predicted results are presented in Table 5 and Figure 3, and the comparison results are summarized below.

(1) Regarding the first comparison, the hybrid nonlinear model yielded the best results for one-step and two-step wind speed prediction on all three datasets according to the error indexes. For example, for one-step prediction, the MAPE value of the developed model is approximately 2% to 3%, while the best accuracy values of the VMD-based models are more than 1% lower than that of the developed model. Two-step prediction yields similar results.

(2) Regarding the second comparison, when comparing VMD-ARIMA, VMD-GRNN, and VMD-ELM to ARIMA, GRNN, and ELM, respectively, without data preprocessing, one can see that data preprocessing is very important for enhancing wind speed prediction. For site 1, the MAPE values of ARIMA and ELM are 4.2190% and 6.7442% higher than those of VMD-ARIMA and VMD-ELM, respectively, for one-step prediction and 4.2927% and 6.4793% higher, respectively, for two-step prediction. The accuracy of VMD-GRNN is also slightly improved. For sites 2 and 3, the results are very similar.

(3) Regarding the third comparison, based on the four indexes of MAPE, MAE, MSE, and DC, one can see that the developed model is more accurate than other individual models, such as wave.nn and the BP neural network. Additionally, the individual models with the highest prediction accuracy are ARIMA, BP, and GRNN. Therefore, we selected BP as a model for nonlinear combination. Because ARIMA is a linear model, it can determine if wind speed data has certain linear characteristics, so it is intuitive to consider ARIMA in the proposed model. The other three models, namely, ARIMA, GRNN, and ELM, are submodels of the combined model.

### Table 5: Prediction abilities of the established model and single models.

| Site | Model   | MAPE (%) | MAE | MSE | DC (%) |
|------|---------|----------|-----|-----|--------|
|      |         | 1-step   | 2-step | 1-step | 2-step | 1-step | 2-step | 1-step | 2-step |
| 1    | ARIMA   | 6.836    | 13.2649 | 0.5193 | 0.7687 | 0.5268 | 1.0810 | 63.1263 | 62.6506 |
|      | GRNN    | 8.906    | 11.9739 | 0.3106 | 0.6815 | 0.5109 | 0.8447 | 61.5230 | 62.6506 |
|      | ELMNN   | 11.130    | 12.2825 | 0.561    | 0.6833 | 0.6137 | 0.8455 | 59.7194 | 61.6466 |
|      | wave.nn | 9.2901   | 18.2638 | 0.5509   | 1.1013 | 0.5654 | 1.8153 | 57.1142 | 56.6251 |
| Site 2 | BPNN   | 8.7207   | 11.6458 | 0.5212   | 0.6883 | 0.5265 | 0.8178 | 60.1202 | 63.6546 |
|      | VMD-ARIMA | 4.4646 | 8.9722    | 0.2719  | 0.5342 | 0.1421 | 0.5217 | 86.7735 | 65.4618 |
|      | VMD-GRNN | 8.6501   | 11.3655 | 0.4955   | 0.6390 | 0.4397 | 0.7379 | 68.1363 | 63.2530 |
|      | VMD-ELMNN | 4.4188 | 5.8032    | 0.2684  | 0.3512 | 0.1378 | 0.2338 | 86.9739 | 83.3333 |
|      | Proposed model | 2.6816 | 3.8291    | 0.1768  | 0.2556 | 0.0611 | 0.1135 | 91.6084 | 83.9161 |
| 3    | ARIMA   | 7.5150   | 12.3326 | 0.5283   | 0.8315 | 0.5458 | 1.2054 | 57.3146 | 56.2249 |
|      | GRNN    | 7.5967   | 10.6678 | 0.5212   | 0.7039 | 0.5404 | 0.8955 | 58.3166 | 59.0361 |
|      | ELMNN   | 10.5427   | 12.3717 | 0.5770   | 0.7465 | 0.6631 | 0.9663 | 54.9098 | 55.8233 |
|      | wave.nn | 17.1573   | 20.9437 | 0.9954   | 1.1856 | 1.6001 | 2.6076 | 55.3106 | 57.0281 |
| Site 2 | BPNN   | 7.7633   | 10.3159 | 0.5284   | 0.6886 | 0.5490 | 0.8433 | 58.5170 | 56.6265 |
|      | VMD-ARIMA | 3.8247 | 7.8950    | 0.2727  | 0.5568 | 0.1509 | 0.5524 | 86.9739 | 58.0321 |
|      | VMD-GRNN | 8.1833   | 10.5210 | 0.5168   | 0.6788 | 0.4827 | 0.8068 | 63.9279 | 57.8313 |
|      | VMD-ELMNN | 3.8669 | 5.0213    | 0.2756  | 0.3588 | 0.1525 | 0.2434 | 87.1743 | 81.3253 |
|      | Proposed model | 2.5660 | 3.1463    | 0.1981  | 0.2510 | 0.0711 | 0.1054 | 88.8112 | 83.9161 |
| 3    | ARIMA   | 8.6656   | 13.9963 | 0.5136   | 0.8113 | 0.4833 | 1.1452 | 59.1182 | 58.3239 |
|      | GRNN    | 8.7151   | 12.1387 | 0.5074   | 0.6773 | 0.4645 | 0.8202 | 56.7134 | 57.0281 |
|      | ELMNN   | 12.0519   | 12.6477 | 0.5735   | 0.6862 | 0.6019 | 0.8388 | 54.7094 | 59.3894 |
|      | wave.nn | 14.0062   | 18.2186 | 0.7417   | 0.9660 | 0.9241 | 1.5119 | 56.5130 | 56.6265 |
| Site 3 | BPNN   | 9.0109   | 11.9842 | 0.5051   | 0.6571 | 0.4539 | 0.7718 | 61.1222 | 58.6345 |
|      | VMD-ARIMA | 4.2864 | 9.1965    | 0.2596  | 0.5341 | 0.1219 | 0.4998 | 85.9719 | 60.8434 |
|      | VMD-GRNN | 9.3575   | 12.0019 | 0.5136   | 0.6673 | 0.4574 | 0.7730 | 65.9319 | 60.2410 |
|      | VMD-ELMNN | 4.2089 | 5.8116    | 0.2527  | 0.3445 | 0.1143 | 0.2091 | 86.9739 | 80.7229 |
|      | Proposed model | 3.0342 | 4.3360    | 0.1874  | 0.2629 | 0.0614 | 0.1179 | 86.7133 | 81.8182 |

In particular, ARIMA is single linear prediction model; GRNN, ELMNN, wave.nn, and BPNN are single nonlinear prediction models. VMD-ARIMA, VMD-GRNN, and VMD-ELM are single prediction models after data preprocessing.
different data preprocessing methods are the EMD-based model, EEMD-based model, CEEMD-based model, and SSA-based model. These models only use different decomposition methods during the data preprocessing stage. In this experiment, we tested whether or not the proposed prediction model is reasonable and identified the best method for removing noise to improve prediction effectiveness.

The results obtained by models using different data preprocessing methods are listed in Table 6. Figure 4 presents a clearer and more intuitive comparison. In Table 6 and Figure 4, one can see that the model based on VMD technology has superior performance compared to the other decomposition-based prediction models. The MAPE value of the VMD-based proposed model is 0.3 to 4 percentage points higher than those of the EMD-based model, EEMD-based model, CEEMD-based model, and SSA-based model. Of all the benchmark models, the SSA-based model performs the worst. Compared to the other models, the MAE, MSE, and DC values for one-step and two-step prediction by the proposed model are improved to some extent, which demonstrates the superiority of the data preprocessing method adopted in our hybrid model.

(1) Remark Regarding the Point Prediction Module. Experiments I and II focused on proving the advantages of the proposed point prediction module and verifying it from the perspective of single prediction models, combination models, and data preprocessing. The results show that, in both cases, the proposed point prediction model is superior to all the comparative models. This proves that the combination of data preprocessing technology, optimization algorithms, and nonlinear combined methods can successfully resolve the issues of wind energy prediction based on the selection of appropriate prediction methods. Based on the superior effectiveness of the designed point prediction model, it has very promising application potential.

In particular, EMD, EEMD, and CEEMD are a series of processes of the same principle; the changing process of $\text{EMD} \rightarrow \text{EEMD} \rightarrow \text{CEEMD}$ can be summarized as follows:

\[
\text{The signal formula of EMD is } s_i(t) = x(t) .
\]

\[
\text{The signal formula of EEMD is } s_i(t) = x(t) + n_i(t) .
\]

\[
\text{The signal formula of CEEMD is } m_i^+(t) = x(t) + n_i^+(t), \quad m_i^-(t) = x(t) + n_i^-(t),
\]

where $x(t)$ is the original signal, $n_i(t)$ is the noise sequence, and $n_i^+(t)$ and $n_i^-(t)$ are positive noise and negative noise sequence. On the basis of EMD, noise sequence is added to form EEMD. CEEMD further decomposes the noise sequence into positive noise sequence and negative noise sequence.
3.4. Results and Analysis of Interval Prediction (Experiment III). Based on wind speed point prediction results, probability interval prediction can derive additional wind speed information. In this study, we developed a method based on fuzzy clustering which performs interval prediction based on point prediction results. Three datasets were considered in this experiment. For the sake of verifying the effectiveness of the designed interval prediction module, we used all of the comparative models for point prediction and performed multistep prediction to verify the interval prediction results. The results of the proposed interval prediction model and other models are listed in Table 7. Based on space limitations, Table 7 only lists the results for site 3. We set the confidence interval to 90% to assess the effectiveness of the interval prediction results. Table 6: Performances of models with different data preprocessing techniques.

| Site | Model          | MAPE (%) | MAE   | MSE   | DC (%) |
|------|----------------|----------|-------|-------|--------|
|      | 1-step | 2-step | 1-step | 2-step | 1-step | 2-step | 1-step | 2-step |
| Site 1 | EMD-based model | 3.4387  | 3.8291 | 0.2261 | 0.2556 | 0.0832 | 0.1135 | 79.7203 | 83.9161 |
|       | EEMD-based model | 2.9010  | 3.9826 | 0.1948 | 0.2987 | 0.0697 | 0.1184 | 88.8112 | 79.7203 |
|       | CEEMD-based model | 3.2030  | 6.4977 | 0.2085 | 0.4314 | 0.0781 | 0.3212 | 74.8252 | 71.3278 |
|       | SSA-based model  | 6.4600  | 6.6521 | 0.4223 | 0.4323 | 0.3121 | 0.3212 | 74.8252 | 71.3278 |
|       | Proposed model   | 2.6816  | 3.8291 | 0.1768 | 0.2556 | 0.0611 | 0.1135 | 91.6084 | 83.9161 |
| Site 3 | EMD-based model | 3.3054  | 4.2136 | 0.2657 | 0.3369 | 0.1230 | 0.1919 | 86.0140 | 74.8252 |
|       | EEMD-based model | 2.8489  | 3.8400 | 0.2108 | 0.2751 | 0.0763 | 0.1315 | 83.2168 | 81.8182 |
|       | CEEMD-based model | 3.1061  | 5.5675 | 0.2498 | 0.4394 | 0.1147 | 0.3940 | 81.1189 | 61.5385 |
|       | SSA-based model  | 5.8965  | 5.7573 | 0.4578 | 0.4484 | 0.3726 | 0.3786 | 67.1329 | 63.6364 |
|       | Proposed model   | 2.5660  | 3.1463 | 0.1981 | 0.2510 | 0.0711 | 0.1054 | 88.8112 | 83.9161 |
| Site 4 | EMD-based model | 3.5904  | 5.6482 | 0.2231 | 0.3375 | 0.0817 | 0.2023 | 84.6154 | 72.7273 |
|       | EEMD-based model | 3.3608  | 4.7642 | 0.1965 | 0.2930 | 0.0667 | 0.1254 | 86.0140 | 79.7203 |
|       | CEEMD-based model | 4.5358  | 7.5806 | 0.2734 | 0.4468 | 0.1134 | 0.3048 | 81.8182 | 64.3357 |
|       | SSA-based model  | 7.6309  | 7.3411 | 0.4589 | 0.4457 | 0.3247 | 0.3206 | 63.6364 | 65.7343 |
|       | Proposed model   | 3.0342  | 4.3360 | 0.1874 | 0.2629 | 0.0614 | 0.1179 | 86.7133 | 81.8182 |

EMD-based model, EEMD-based model, CEEMD-based model, and SSA-based model are prediction models after data preprocessing. These models are the same in combination and optimization methods, and the only difference is that they have different data preprocessing methods.
complexity. From Table 7, one can draw the following conclusions.

The best values for all indexes among all models are obtained by the proposed prediction model. The coverage probability of the prediction interval is 96.5278% in one-step PICP and 90.1944% in two-step PICP. The average width of the interval is 1.3399 for one-step prediction and 1.2158 for two-step prediction according to MPI. In terms of the absolute value of wind speed, the interval width is relatively accurate. The AWD is 0.0066 for one-step prediction and 0.0264 for two-step prediction, indicating that the deviation degree of the constructed interval is small. All indexes indicate that the predicted interval is qualified. In contrast, for the PICPs of individual prediction models, none of the one-step predictions reach more than 90% and all the two-step predictions are below 80%. By combining PINAW with PICP, for the proposed model, when the PICP value is very high, PINAW is relatively small, which demonstrates the superiority of the proposed model. For one-step and two-step prediction, the AWD values of most other benchmark models are more than ten times that of the proposed model. ACE is the difference between the coverage and confidence of the prediction interval. The ACE values of all models except for the proposed model are negative, indicating that the coverage of the developed model is much better than that of the other models. The absolute value of the WS index of the proposed model is the smallest, indicating its reliability. These six indexes fully reflect the superior prediction performance of the proposed model.

To present the comparison results intuitively, the results of the designed module and comparative methods are visualized in Figure 5. These conclusions are consistent with the results in Table 7, providing intuitive evidence that verifies the superior abilities of the proposed system for wind speed interval prediction. As shown in Figure 5, compared to the other methods, the proposed model yields superior interval prediction results. The prediction range not only covers most of the wind speed values but also is the smoothest range among all models. This demonstrates that the proposed model is more stable than the other models. Therefore, our model is more advantageous for the three experimental datasets.

Table 7: Performances of the designed model compared with all other models for site 4.

| Model              | PICP (%) | ACE  | MPI   | AWD   | PINAW (%) | WS   |
|--------------------|----------|------|-------|-------|-----------|------|
|                    | 1-step   | 2-step| 1-step| 2-step| 1-step    | 2-step| 1-step| 2-step| 1-step| 2-step| 1-step| 2-step|
| ARIMA              | 62.6000  | 44.8898 | -0.2740 | -0.4511 | 1.1787 | 1.1389 | 0.1886 | 0.4532 | 0.1062 | 0.1026 | (0.8087) | (1.7072) |
| wavenn             | 37.6000  | 28.0561 | -0.5240 | -0.6194 | 0.9494 | 0.9042 | 0.4624 | 0.8502 | 0.0855 | 0.0815 | (1.6103) | (2.4889) |
| BPNN               | 61.2000  | 53.3066 | -0.2880 | -0.3669 | 1.0902 | 1.1526 | 0.1797 | 0.3192 | 0.0982 | 0.1038 | (0.7670) | (1.2361) |
| GRNN               | 61.4000  | 49.2986 | -0.2860 | -0.4070 | 1.1414 | 1.0516 | 0.1874 | 0.3558 | 0.1028 | 0.0947 | (0.7896) | (1.3146) |
| ELM                | 56.8000  | 47.4950 | -0.3320 | -0.4251 | 1.0636 | 1.0469 | 0.2347 | 0.3782 | 0.0958 | 0.0943 | (0.9800) | (1.3606) |
| VMD-ARIMA          | 87.4000  | 61.5230 | -0.0260 | -0.2848 | 1.1412 | 1.1894 | 0.0274 | 0.1868 | 0.1028 | 0.1072 | (0.2198) | (0.8190) |
| VMD-GRNN           | 58.0000  | 45.2906 | -0.3200 | -0.4471 | 0.9662 | 0.9779 | 0.2167 | 0.3511 | 0.0870 | 0.0881 | (0.8388) | (1.3117) |
| VMD-ELM            | 87.2000  | 76.9539 | -0.0280 | -0.1305 | 1.1262 | 1.1545 | 0.0250 | 0.0585 | 0.1015 | 0.1040 | (0.2093) | (0.3419) |
| EMD-based model    | 95.1389  | 81.2500 | 0.0514 | -0.0875 | 1.2890 | 1.3389 | 0.0083 | 0.0568 | 0.1953 | 0.2029 | (0.1536) | (0.3205) |
| EEMD-based model   | 92.3611  | 84.7222 | 0.0236 | -0.0528 | 1.1643 | 1.1788 | 0.0089 | 0.0218 | 0.1764 | 0.1786 | (0.1380) | (0.1855) |
| CEEMD-based model  | 90.2778  | 70.8333 | 0.0028 | -0.1917 | 1.2326 | 1.2098 | 0.0193 | 0.1202 | 0.1868 | 0.1827 | (0.1818) | (0.4801) |
| SSA-based model    | 63.8889  | 63.8889 | -0.2611 | -0.2611 | 1.0555 | 1.0545 | 0.1159 | 0.1026 | 0.1675 | 0.1598 | (0.4876) | (0.4974) |
| Proposed model     | 96.5278  | 90.1944 | 0.0653 | 0.0019 | 1.3399 | 1.2158 | 0.0066 | 0.0264 | 0.2030 | 0.1842 | (0.1538) | (0.2033) |

(1) Remark Regarding the Interval Prediction Module. Similar to the comparison model used for point forecasting, 12 different models based on three datasets and multistep forecasting were compared. The results demonstrated that the designed interval model is superior to all the comparative models. Based on the excellent results of the designed interval prediction module using fuzzy clustering, it is a very promising interval prediction method for wind speed.

All the above comparison models are the comparison models of experiment 1 and experiment 2. We still use them to compare the performance of interval prediction, so as to prove the interval prediction performance of the developed model. In particular, the WS value in the table is bracketed to indicate its absolute value.

4. Discussion

For the sake of discussing our experimental conclusions in detail and reducing the error of wind speed forecasting, the validity of the established model, combination mechanism of the combined model, and its practical application to wind power systems are discussed in this section.

4.1. DM Test. First, the validity of the proposed model was verified via DM testing in which all of the other models were compared to the proposed double prediction model. Based on the DM testing theory, the zero hypothesis is that the forecasting results of two models contain no significant differences. The alternative hypothesis is opposite to the zero hypothesis. We chose two scales with alpha values of 0.1 and 0.05 as the criteria for judging the significance of results with $Z_{0.05}/2 = 1.96$ and $Z_{0.1}/2 = 1.645$, respectively. Table 8 lists the DM statistics and averages for the three test sites.

Table 8 reveals that most of the DM test values calculated by the developed model and comparative models are greater than the upper limit of a 5% significance level. However, for the VMD-ARIMA-, VMD-ELM-, and EEMD-based models, the results do not reveal significant differences compared to the proposed model. Therefore, we can reject the zero hypothesis at a threshold of 10% significance. For example, the DM test statistic for the
VMD-ELM model at site 1 is 1.7554, which is not significantly different from that of the developed model at a 5% significance level, but is significantly different from the developed model at a 10% significance level. At a 10% significance level, all distinctions between the designed model and benchmark models are significant. Therefore, it can be concluded that the designed hybrid double prediction model is preferable to the other models.

4.2. Combination Mechanism of the Combined Model. For the sake of verifying the effectiveness of the designed nonlinear combination mechanism (MOEA/D-BP), a simple averaging strategy and linear combination mechanism were selected as comparative methods in this study. The simple averaging strategy computes the mean value of the prediction results of each model, while the linear combination mechanism uses the MOEA/D as a weight determination method to derive the final prediction results. Comparative results for the developed model and the other two methods are listed in Table 9.

The effects of each combination mechanism are compared based on four point prediction error measurement rules and six interval error prediction measurement rules. One can see that the prediction effectiveness of the nonlinear combination model is greater than that of the simple averaging strategy and linear combination mechanism, regardless of the location and prediction steps. The linear combination mechanism is often more effective than the simple averaging strategy. In other words, the simple averaging strategy performs the worst. Therefore, the developed MOEA/D-BP mechanism successfully improves forecasting effectiveness for wind speed.

The simple average method is to use the simple average formula under statistical sense to calculate the final predicted value. The method formula is briefly introduced as follows:

\[ p = \frac{p_1 + p_2 + p_3}{3}, \]  

where \( p_i \) is the prediction results of the corresponding model. The linear combination of the models is the weighted combination of the results of the three single models, and a final

Figure 5: The final interval results of some models in multistep prediction in experiment III for dataset 3. The figure shows the interval forecasting results of the proposed model, VMD-ELM model, CEEMD-based model, and BP single model, respectively. Besides, the figure shows the relationship between the interval prediction band and the real value.
To analyze the performance of the MOEA/D, different population size numbers were selected to test its abilities using four test functions. Three multiobjective optimization algorithms, namely, MOGWO, MOALO, and MODA, were used as comparative models. Table 12 contains the details of

4.3. Parameter Settings. An artificial intelligence algorithm called BP was used to combine wind speed results. In a BP neural network, the weights and thresholds of input, hidden, and output layers play crucial roles in terms of network performance. To determine the appropriate connection weights and node thresholds efficiently, we adopted the MOEA/D for parameter optimization. The parameters for BP and the MOEA/D are listed in Tables 10 and 11, respectively.

4.3.1. Parameter Settings. This section first introduces the parameter settings for the BP network and MOEA/D and then presents convergence testing results for metaheuristic algorithms.
the four test functions. By comparing different optimization methods, it was proven that the prediction ability of the MOEA/D is superior to that of other multiobjective algorithms. Twenty experiments were conducted for each case and average values were obtained. The calculation results for each index are listed in Table 13.

We selected two performance indexes as the criteria for evaluating the optimization algorithms, namely, the IGD and SP index. Additionally, the running times of different algorithms were compared. IGD is an indicator of the convergence conditions of an algorithm and it can be used to judge the robustness and stability of algorithms. The smaller the IGD value, the better the performance of an algorithm. In a Pareto set, SP is typically used to evaluate the distribution of solutions. If SP is equal to zero, then all nondominant solutions are equidistant.

The final simulated results are listed in Table 13. For all of the algorithms, as the population size increases from 100 to 150, 200, and 300, convergence is enhanced. The MOEA/D yields the best performance for ZDT1, ZDT2, ZDT3, and ZDT6. The IDG of the MOEA/D is far less than that of the other algorithms, indicating that the MOEA/D provides the best convergence performance. MODA is the second-best algorithm. The convergence effect of the MOGWO algorithm is much weaker than that of the other algorithms. For SP, the MOEA/D yields the best allocation performance. The running time of the MOEA/D is significantly lower than the running times of the other three algorithms, which demonstrates that the MOEA/D is the fastest and most efficient algorithm.

IGD and SP are two important performance evaluation indexes of multiobjective algorithm solution set, of which the calculation formulas are as follows [60]:

\[
\text{IGD} = \frac{\sum_{v \in P} d(v, Q)}{|P|},
\]

(28)

where \(P\) is the set of points uniformly distributed on the real Pareto surface and \(|P|\) is the number of individuals of the set of points distributed on the real Pareto surface. \(Q\) is the optimal Pareto solution set obtained by the algorithm. \(d(v, Q)\) is the minimum Euclidean distance between individual \(v\) and population \(Q\) in \(P\). Therefore, IGD is to evaluate the comprehensive performance of the algorithm by calculating the average value of the minimum distance from the point set on the real Pareto surface to the obtained population.

\[
\text{SP} = \frac{1}{n-1} \sum_{i=1}^{n} (d_i - \overline{d})^2,
\]

(29)

where \(d_i\) represents the minimum distance from the \(i\)-th solution to other solutions in the solution set and \(\overline{d}\) represents the mean value of all \(d_i\). \(n\) is the number of solution set individuals. SP measures the standard deviation of the minimum distance from each solution to other solutions. The smaller the SP value, the more uniform the solution set.

4.4. Practical Application to a Power System. Wind power forecasting systems are of great importance for large-capacity wind power systems. Effective wind speed forecasting can be helpful in many areas, such as timely maintenance scheduling and electronic grid safety management. The contributions of an accurate wind speed forecasting model to a power system can be summarized as follows [50]:

1. To guarantee the best wind energy output quality, it is very important to assess the quantity of wind power. Wind power has a direct power relationship with wind speed, so the evaluation of wind power can be accomplished based on wind speed prediction. Therefore, precise wind speed forecasting can enhance decision-making for wind farms and is conducive to smart grid planning.

2. Accurate wind speed forecasting can provide essential guidance for the dispatching and control of wind turbines. Based on predicted wind speeds,
(3) Wind speed prediction effectiveness has an important impact on electronic grid dispatching and supervising. Power output fluctuates significantly and intermittently, which makes power system operation very challenging. Therefore, accurate prediction models can assist decision-makers in making timely decisions to avoid the problems discussed above.

5. Conclusions

Based on the depletion of traditional energy sources, wind energy is considered to be a promising alternative energy source because of its sustainability and cleanliness. However, based on its inherent intermittence and randomness, the administrators can control wind turbines immediately to ensure the best wind energy output quality. If the wind speed value is greater than the fan capacity, the fan should be closed to avoid damage and reduce operating costs.
extraction of wind energy is very limited, which can endanger the dispatching and management of wind power systems.

To analyze the uncertainty characteristics of wind speed more comprehensively, a double prediction system was successfully developed in this study. The proposed system compensates for the shortcomings of previous methods. The proposed system consists of two main parts: a point prediction module based on nonlinear combination and interval prediction module based on fuzzy clustering. It is of great significance to explore the predictability and modeling of wind speed comprehensively. Unlike previous works, we implemented a BP neural network using MOEA/D optimization as a novel nonlinear combination mechanism to derive final prediction results, which enhances the accuracy of point prediction and improves final prediction accuracy. To improve the accuracy of point prediction, wind speed data was divided into different categories based on fuzzy clustering and different intervals were constructed according to the prediction data in different categories. This method of constructing different intervals according to different data characteristics has been proven to be an effective interval prediction method. Finally, a large number of experiments were conducted using quantitative indexes, which demonstrated the effectiveness and superiority of the proposed system. Additionally, because the proposed system provides reliable performance, it can also be applied to load prediction, wind power forecasting, economic forecasting, and other fields.

Appendix

The model proposed in this paper is based on wind speed data from the Penglai wind farm which was collected in January of the year 2011. This dataset was randomly selected to train and test the proposed model. To explore the impact of different seasons on the proposed model, the authors also selected data from three other months in different seasons (April, July, and October) for comparison with the data from January. The results are listed in Table 14. One can see that there are no significant differences between the prediction accuracies of the models constructed from data from different months for one-step point prediction, two-step point prediction, or interval prediction. This demonstrates that the construction of the proposed model is not affected by seasonal changes. Similarly, data from different years can also be used to construct models for wind speed prediction. The proposed model can be used to predict general wind speeds to study time trends and seasonal characteristics.

Nomenclature

| Abbreviation | Description |
|--------------|-------------|
| ARIMA | Autoregressive integrated moving average |
| GRNN | Generalized regression neural network |
| ELM | Extreme learning machine neural network |
| Wavenn | Wavelet neural network |
| BP | Backpropagation neural network |
| EMD | Empirical mode decomposition |
| EEMD | Ensemble empirical mode decomposition |
| CEEMD | Complementary ensemble empirical mode decomposition |
| SSA | Singular spectrum analysis |
| VMD | Variational mode decomposition |
| IMFs | Intrinsic mode functions |
| ZDT | Test functions for multiobjective algorithm |
| AR | Autoregressive model |
| ARIMA-AR | Autoregressive conditional heteroscedasticity model |
| MOP | Multiobjective optimization problem |
| MAE | Mean absolute error |
| MAPE | Mean absolute percentage error |
| MSE | Mean squared error |
| DC | Directional change |
| PICP | Prediction interval coverage probability |
| PINAW | Prediction interval normalized average width |
| ACE | Average coverage error |
| MPI | Average width of the constructed PIs |
| WS | Winkler score |
| AWD | Accumulated width deviation |
| MOGWO | Multiobjective grey wolf optimization |
| MOALO | Multiobjective ant colony optimization |
| MODA | Multiobjective Dragonfly algorithm |
| MOEA/D | Multiobjective evolutionary algorithm based on decomposition |
| IGD | Inverted generational distance |
| SP | Spread performance |

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request. The 10-minute wind speed data of Penglai, Shandong province, which has one of the largest wind farms in China, are selected and three datasets of data are collected. The data are true and reliable. The authors will provide the data if necessary to assist the experimental proof.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] P. Pourbeik, V. Akhmatov, Y. Akiyama et al., Modelling and Dynamic Behavior of Wind Generation as It Relates to Power System Control and Dynamic Performance, International Council Large Electrical System (CIGRE), Paris, France, 2007.
[2] J. Yu, F. Ji, L. Zhang, and Y. Chen, “An over painted oriental arts: evaluation of the development of the Chinese renewable energy market using the wind power market as a model,” Energy Policy, vol. 37, no. 12, pp. 5221–5225, 2009.
[3] Z. Liu, P. Jiang, L. Zhang, and X. Niu, “A combined forecasting model for time series: Application to short-term wind speed forecasting,” Applied Energy, vol. 259, Article ID 114137, 2019.
[4] R. Li and Y. Jin, “A wind speed interval prediction system based on multi-objective optimization for machine learning method,” Applied Energy, vol. 228, pp. 2207–2220, 2018.

[5] L. Zhang, Y. Dong, and J. Wang, “Wind speed forecasting using a two-stage forecasting system with an error correcting and nonlinear ensemble strategy,” IEEE Access, vol. 7, pp. 176000–176023, 2019.

[6] T. Niu, J. Wang, H. Lu, and P. Du, “Uncertainty modeling for chaotic time series based on optimal multi-input multi-output architecture: application to offshore wind speed,” Energy Conversion and Management, vol. 156, pp. 597–617, 2018.

[7] W. Zhang, Z. Qu, K. Zhang, W. Mao, Y. Ma, and X. Fan, “A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting,” Energy Conversion and Management, vol. 136, pp. 439–451, 2017.

[8] P. Jiang, H. Yang, and J. Heng, “A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting,” Applied Energy, vol. 235, pp. 786–801, 2019.

[9] Y. Hao and C. Tian, “A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting,” Applied Energy, vol. 238, pp. 368–383, 2019.

[10] X. Ma, Y. Wang, and C. Wang, “Low-carbon development of the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms,” Energy Conversion and Management, vol. 73, no. 7, pp. 567–574, 2014.

[11] E. Erdem and J. Shi, “ARMA based approaches for forecasting the tuple of wind speed and direction,” Applied Energy, vol. 88, no. 4, pp. 1405–1411, 2011.

[12] P. Poggi, M. Muselli, G. Notton, C. Cristofari, and A. Louche, “Forecasting and simulating wind speed in Corsica by using an autoregressive model,” Energy Conversion and Management, vol. 44, no. 20, pp. 3177–3196, 2003.

[13] J. L. Torres, A. Garcia, 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.

[14] 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.

[15] M.-D. Wang, Q. R. Qiu, and B. W. Cui, “Short-term wind speed forecasting combined time series method and arch model,” in Proceedings of the International Conference on Machine Learning and Cybernetics, vol. 3, pp. 924–927, Lanzhou, China, 2012.

[16] P. Jiang and X. Ma, “A hybrid forecasting approach applied in the electrical power system based on data preprocessing, optimization and artificial intelligence algorithms,” Applied Mathematical Modelling, vol. 40, no. 23-24, pp. 10631–10649, 2016.

[17] J. Wang, W. Yang, P. Du, and Y. Li, “Research and application of a hybrid forecasting framework based on multi-objective optimization for electrical power system,” Energy, vol. 148, pp. 59–78, 2018.

[18] C. Tian, Y. Hao, and J. Hu, “A novel wind speed forecasting system based on hybrid data preprocessing and multi-objective optimization,” Applied Energy, vol. 231, pp. 301–319, 2018.

[19] J. Hu, J. Wang, and K. Ma, “A hybrid technique for short-term wind speed prediction,” Energy, vol. 81, pp. 563–574, 2015.

[20] 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.

[21] L. Ma, S. Luan, C. Jiang, H. Liu, and Y. Zhang, “A review on the forecasting of wind speed and generated power,” Renewable and Sustainable Energy Reviews, vol. 13, no. 4, pp. 915–920, 2009.

[22] P. Du, J. Wang, W. Yang, and T. Niu, “Multi-step ahead forecasting in electrical power system using a hybrid forecasting system,” Renewable Energy, vol. 122, pp. 533–550, 2018.

[23] X. Ma, Y. Wang, and C. Wang, “Low-carbon development of China’s thermal power industry based on an international comparison: review, analysis and forecast,” Renewable and Sustainable Energy Reviews, vol. 80, pp. 942–970, 2017.

[24] H. Bludszuweit, J. A. Dominguez-Navarro, and A. Llombart, “Statistical analysis of wind power forecast error,” IEEE Transactions on Power Systems, vol. 23, no. 3, pp. 983–991, 2008.

[25] C. Tian and Y. Hao, “A novel nonlinear combined forecasting system for short-term load forecasting,” Energies, vol. 11, no. 4, p. 712, 2018.

[26] Z. Guo, W. Zhao, H. Lu, and J. Wang, “Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model,” Renewable Energy, vol. 37, pp. 241–249, 2012.

[27] H. Liu, H.-q. Tian, X.-f. Liang, and Y.-f. Li, “Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks,” Applied Energy, vol. 157, pp. 183–194, 2015.

[28] H. Liu, H.-q. Tian, C. Chen, and Y.-f. Li, “An experimental investigation of two wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization,” International Journal of Electrical Power & Energy Systems, vol. 52, pp. 161–173, 2013.

[29] S. Wang, N. Zhang, L. Wu, and Y. Wang, “Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method,” Renewable Energy, vol. 94, pp. 629–636, 2016.

[30] P. Jiang and Z. Liu, “Variable weights combined model based on multi-objective optimization for short-term wind speed forecasting,” Applied Soft Computing, vol. 82, p. 105587, 2019.

[31] X. Ma, Y. Jin, and Q. Dong, “A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting,” Applied Soft Computing, vol. 54, pp. 296–312, 2017b.

[32] J. Wang, P. Du, T. Niu, and W. Yang, “A novel hybrid system based on a new proposed algorithm-multi-objective whale optimization algorithm for wind speed forecasting,” Applied Energy, vol. 208, pp. 344–360, 2017.

[33] A. Meng, J. Ge, H. Yin, and S. Chen, “Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm,” Energy Conversion and Management, vol. 114, pp. 75–88, 2016.
[37] Q. Dong, Y. Sun, and P. Li, “A novel forecasting model based on a hybrid processing strategy and an optimized local linear fuzzy neural network to make wind power forecasting: a case study of wind farms in China,” Renewable Energy, vol. 102, pp. 241–257, 2017.

[38] C. Zhang, J. Zhou, C. Li, W. Fu, and T. Peng, “A compound structure of ELM based on feature selection and parameter optimization using hybrid backtracking search algorithm for wind speed forecasting,” Energy Conversion and Management, vol. 143, pp. 360–376, 2017.

[39] J. Wang, S. Qin, Q. Zhou, and H. Jiang, “Medium-term wind speeds forecasting utilizing hybrid models for three different sites in Xinjiang, China,” Renewable Energy, vol. 76, pp. 91–101, 2015.

[40] D. Ulkay and M. E. Günday, “Prediction of mean monthly wind speed and optimization of wind power by artificial neural networks using geographical and atmospheric variables: case of Aegean Region of Turkey,” Neural Computing and Applications, vol. 30, no. 10, pp. 3037–3048, 2017.

[41] H. Z. Wang, G. B. Wang, G. Q. Li, J. C. Peng, and Y. T. Liu, “Deep belief network based deterministic and probabilistic wind speed forecasting approach,” Applied Energy, vol. 182, pp. 80–93, 2016.

[42] J. B. Bremnes, “Probabilistic wind power forecasts using local quantile regression,” Wind Energy, vol. 7, no. 1, pp. 47–54, 2004.

[43] R. Errouissi, J. Cardenas-Barrera, J. Meng, E. Castillo-Guerra, X. Gong, and L. Chang, “Bootstrap prediction interval estimation for wind speed forecasting,” in Proceedings of the 2015 IEEE Energy Conversion Congress and Exposition (ECCE), pp. 1919–1924, Montreal, Canada, 2015.

[44] 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 2007, pp. 683–688, Piscataway, NJ, USA, 2007.

[45] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, “Lower upper bound estimation method for construction of neural network-based prediction intervals,” IEEE Transactions on Neural Networks, vol. 22, no. 3, pp. 337–346, 2011.

[46] J. Yan, Y. Liu, S. Han, Y. Wang, and S. Feng, “Reviews on uncertainty analysis of wind power forecasting,” Renewable and Sustainable Energy Reviews, vol. 52, pp. 1322–1330, 2015.

[47] Bootstrapping and Resampling, http://www.ncl.ucar.edu/Applications/bootstrap.shtml.

[48] W. Yang, J. Wang, H. Lu et al., “Hybrid wind energy forecasting and analysis system based on divide and conquer scheme: a case study in China,” Journal of Cleaner Production, vol. 22, no. 3, pp. 337–346, 2019.

[49] I. Moghram and S. Rahman, “Analysis and evaluation of five short-term load forecasting techniques,” IEEE Transactions on Power Systems, vol. 4, no. 4, pp. 1484–1491, 1989.

[50] J. Song, J. Wang, and H. Lu, “A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting,” Applied Energy, vol. 215, pp. 643–658, 2018.

[51] R. Li, Y. Dong, Z. Zhu, C. Li, and H. Yang, “A dynamic evaluation framework for ambient air pollution monitoring,” Applied Mathematical Modelling, vol. 65, pp. 52–71, 2019.

[52] X. Zhang, J. Wang, and K. Zhang, “Short-term electric load forecasting based on singular spectrum analysis and support vector machine optimized by Cuckoo search algorithm,” Electric Power Systems Research, vol. 146, no. 2, pp. 270–285, 2017.

[53] Y. Hao and C. Tian, “The study and application of a novel hybrid system for air quality early-warning,” Applied Soft Computing, vol. 74, pp. 729–746, 2019.

[54] K. Dragomiretskiy and D. Zosso, “Variational mode decomposition,” IEEE Transactions on Signal Processing, vol. 62, no. 3, pp. 531–544, 2014.

[55] Q. Zhang and H. Li, “MOEA/D: a multiobjective evolutionary algorithm based on decomposition,” IEEE Transactions on Evolutionary Computation, vol. 11, no. 6, pp. 712–731, 2007.

[56] K. Deb and D. Kalyanmoy, Multi-Objective Optimization Using Evolutionary Algorithms, Vol. 2, Wiley & Sons, Hoboken, NJ, USA, 2001.

[57] H. Ishibuchi, Y. Sakane, N. Tsukamoto et al., “Simultaneous use of different scalarizing functions in MOEA/D,” in Proceedings of the 12th Annual Conference on Genetic and Evolutionary Computation, pp. 519–526, Kyoto, Japan, 2010.

[58] L. Liu, Q. Wang, J. Wang, and M. Liu, “A rolling grey model optimized by particle swarm optimization in economic prediction,” Computational Intelligence, vol. 32, no. 3, pp. 391–419, 2014.

[59] J. Wang, T. Niu, H. Lu et al., “A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting,” IEEE Transactions on Sustainable Energy, vol. 11, no. 1, pp. 337–349, 2019.

[60] W. Yang, J. Wang, T. Niu, and P. Du, “A hybrid forecasting system based on a dual decomposition strategy and multi-objective optimization for electricity price forecasting,” Applied Energy, vol. 235, pp. 1205–1225, 2019.

[61] J. Wang, P. Du, H. Lu, W. Yang, and T. Niu, “An improved grey model optimized by multi-objective ant lion optimization algorithm for annual electricity consumption forecasting,” Applied Soft Computing, vol. 72, pp. 321–337, 2018.

[62] H. Li, J. Wang, R. Li, and H. Lu, “Novel analysis–forecast system based on multi-objective optimization for air quality index,” Journal of Cleaner Production, vol. 208, pp. 1365–1383, 2019.

[63] X. Zhang, J. Wang, and Y. Gao, “A hybrid short-term electricity price forecasting framework: cuckoo search-based feature selection with singular spectrum analysis and SVM,” Energy Economics, vol. 81, pp. 899–913, 2019.

[64] Y. Jiang, R. Li, and H. Li, “Multi-objective algorithm for the design of prediction intervals for wind power forecasting model,” Applied Mathematical Modelling, vol. 67, pp. 101–122, 2019.

[65] J. Wang, N. Tong, H. Lu, W. Yang, and P. Du, “A novel framework of reservoir computing for deterministic and probabilistic wind power forecasting,” IEEE Transactions on Sustainable Energy, vol. 11, no. 1, pp. 337–349, 2020.

[66] A. Khosravi, S. Nahavandi, and D. Creighton, “Prediction interval construction and optimization for adaptive neuro-fuzzy inference systems,” IEEE Transactions on Fuzzy Systems, vol. 19, no. 5, pp. 983–988, 2011.

[67] C. Zhang, H. Wei, L. Xie, Y. Shen, and K. Zhang, “Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework,” Neurocomputing, vol. 205, pp. 53–63, 2016.