**Research Article**

**Multidimensional Meteorological Variables for Wind Speed Forecasting in Qinghai Region of China: A Novel Approach**

He Jiang,1,2 Luo Shihua,1,2 and Yao Dong1,2

1School of Statistics, Jiangxi University of Finance and Economics, Nanchang 330013, China
2Applied Statistics Research Center, Jiangxi University of Finance and Economics, Nanchang 330013, China

Correspondence should be addressed to Luo Shihua; luoshihua@jxufe.edu.cn

Received 20 July 2019; Revised 5 February 2020; Accepted 12 February 2020; Published 6 May 2020

1. Introduction

The depletion of conventional fossil fuels and the deterioration of the ecological environment have been receiving increasing concern in recent years. It is inevitable that fossil fuels will be restricted or replaced by renewable energy sources, which include wind, solar, geothermal, and biomass ones [1, 2]. As a type of clean, inexhaustible, and environmentally preferable renewable energy, wind energy has been developed rapidly since the 1990s around the world and has received widespread attention [3]. Based on the annual report by the Global Wind Energy Council, the global market was 52.573 GW of the wind power installed capacity in 2017, thereby resulting in a cumulative installed capacity of 539.581 GW [4]. Figure 1 presents the top 10 countries in the world in terms of wind power installed capacity from January to December 2017. The wind power is a function of the cube of the wind speed; hence, small differences in the wind speed forecast cause large differences in the wind power forecast [5]. Moreover, the wind speed is an important variable that is related to the wind turbine output power, and the accurate forecast of wind speed facilitates the improvement of the efficiency of wind turbines and the enhancement of the stability of the power supply [6]. However, the wind speed exhibits the stochastic fluctuations and irregular intermittency due to the influence of several meteorological factors, and its distribution is not normal. Therefore, it is also critically challenging to forecast the wind speed accurately. Many scientific techniques have been developed for forecasting the wind speed. These algorithms can be divided into five types: physical-based methods, statistical-based methods, persistence methods, intelligent methods, and hybrid methods [7]. Physical-based methods, which are suitable for long-term wind speed forecasts, apply parametric models according to the physical changes in the atmosphere [8], such as the numerical weather prediction (NWP) model [9]. Statistical-based methods, such as statistical regression [10], the autoregressive moving average
ARMA models [11], Bayesian models [12], and Markov models [13], depend on the historical wind speed data and forecast the future pattern [6]. These models are suitable for forecasting short-term wind speed but have difficulties if the wind speed series exhibit high-frequency variations. Persistence methods, which are used in short-term wind speed forecasts, are based on the assumption that the wind speed at time $k+1$ will not change significantly from the wind speed at time $k$ [14]. The intelligent methods apply machine learning and artificial intelligence algorithms to forecast the wind speed and include Kalman filtering [15], artificial neural networks (ANNs) [16], support vector machine [17], extreme learning machine (ELM) [18], and high-dimensional feature selection [19]. The most popular intelligent methods for forecasting wind speed are ANNs. Zhang et al. proposed a new hybrid model that combines a shared-weight long short-term memory network (SWLSTM) and Gaussian process regression (GPR), and they applied SWLSTM-GPR to wind speed prediction cases [20]. Li et al. developed an innovative framework for multistep wind speed prediction that uses a recursive neural network that is based on the wind speed and the turbulence intensity. According to experimental results, the proposed model dramatically outperforms conventional machine learning models on multistep wind speed prediction [21].

Currently, the main focus has shifted from single methods to advanced hybrid methods, which combine techniques such as statistical and machine learning approaches and utilize the characteristics of individual methods. Sheela and Deepa used a hybrid computing model that integrates a multilayer perceptron network and self-organizing feature maps for wind speed prediction [22–24]. Shukur and Lee proposed a hybrid model that was based on ARIMA for forecasting the daily wind speed in Iraq and Malaysia; to improve the forecasting accuracy, an ANN and a Kalman filter (KF) could be utilized to overcome uncertainty and nonlinearity problems [25]. Doucoure et al. used wavelet decomposition and ANNs with predictability analysis to forecast wind speed time series [26]. Liu et al. developed a novel wind speed prediction model that was based on the WPD (wavelet packet decomposition), a convolutional neural network and a convolutional long short-term memory network, and they demonstrated that the proposed model can perform best in wind speed 1-step

| Country | MW  | Country  | MW  |
|---------|-----|----------|-----|
| China   | 19,500 | Brazil  | 2,022 |
| USA     | 7,017  | France   | 1,694 |
| Germany | 6,581  | Turkey   | 766  |
| UK      | 4,270  | Mexico   | 478  |
| India   | 4,148  | Belgium  | 467  |

Figure 1: Distribution of the global wind power installed capacity in 2017.
to 3-step predictions [27]. Zhu et al. proposed a wind speed forecasting model by integrating Gaussian process regression with a long short-term memory network, and the results demonstrated that the proposed method improves the forecast accuracy [28]. Cai et al. proposed a novel filtering strategy that integrates the statistical predictors and the NWP model outputs into a unified framework. Based on the proposed filtering strategy, a combined predictor, namely, SVR-SDA-UKF (support vector regression, stacked denoising autoencoder, and unscented Kalman filter), is proposed and validated [29].

In addition, wind speed time series exhibit nonlinearity and nonstationarity due to the weather and geographical factors. This may lead to incorrect conclusions regarding the activity of the wind speed when the techniques are used to forecast [15]. Therefore, several researchers have attempted to improve the forecasting precision via the addition of a preprocessing stage for time series decomposition and feature selection [30]. Time series decomposition approaches have been widely used to decompose the original complex-structured series into several simple series, such as the wavelet transform (WT) [31] and variational mode decomposition (VMD) [32]. Feature selection approaches facilitate the selection of suitable inputs for the forecasting models, which enhances the performances of the models substantially. Various feature selection approaches have yielded satisfactory results, such as Gram–Schmidt orthogonalization (GSO) [33], the least absolute shrinkage and selection operator (LASSO) [19], and forward regression [34].

The main contribution of this paper is the developed hybrid forecasting model, which uses random forest, a Gram–Schmidt-based feature selection method, and the whale optimization algorithm. The proposed method has inherited all the advantages of the above approaches, which are embedded together. Instead of combining them in parallel, random forest and the Gram–Schmidt-based feature selection method are embedded in the whale optimization algorithm; this embedding is a challenging task in terms of computation. The objective function value of the whale optimization algorithm is determined by these two methods. The details will be presented in Section 2, which describes the use of the complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN), random forest (RF), the Gram–Schmidt-based feature selection method (GSFS), and the whale optimization algorithm (WOA) to solve several chained problems.

(i) The first methodology (CEEMDAN in Section 2.1) addresses the problem that the signals that are detected from the measured objects are often mixed with various levels of noise. Wavelet analysis can filter out weak or useless signals and extract the information that users need. This paper introduces a data preprocessing strategy that is designed based on the theory of decomposition and reconstruction; it filters the noise from the original wind speed data efficiently and searches for the intrinsic values of the wind speed data. Finally, this preprocessing procedure utilizes wavelet packet decomposition for denoising and multiscale wavelet decomposition to complete data preprocessing, and it guarantees the accurate forecasting of the wind speed.

(ii) The second methodology (RF in Section 2.2) is a machine learning model that is utilized as a forecasting model after selecting all the important variables. This forecasting model is stable since it weakens the correlations between variables to address the multicollinearity issue. Furthermore, it is designed based on bootstrapping, which theoretically guarantees its stability and accurate forecasting results. In summary, compared with traditional programming methods, the utilized machine learning algorithm simplifies the code and improves the execution performance of the code. Second, machine learning technology can be used to solve problems that cannot be solved via traditional methods. Furthermore, machine learning can adapt to new data; namely, the system can be personalized according to the environment.

(iii) The third methodology (GSFS in Section 2.3) is an effective and efficient variable selection approach that is used to select the important variable after the preprocessing procedure. This approach is designed based on an orthogonalization method, which makes the selected variables orthogonal with each other. Furthermore, in contrast to other penalized variable selection methods, this novel approach does not introduce any bias into the model; hence, all nuisance variables can be eliminated effectively.

(iv) The fourth methodology (WOA in Section 2.4) is a whale optimization algorithm, which is a new swarm intelligence optimization algorithm that simulates the predation behavior of humpback whales in the ocean. The algorithm has the advantages of simple structure, few parameters, strong search performance, and easy implementation. It is applied in the proposed model in an embedded way to determine the optimal weights for assignment to machine learning models. The fitness function in WOA is determined by the machine learning model and the variable selection method.

The hybrid forecasting model, namely, RF-GSFS-WOA, is proposed in Section 3. The experimental results and the corresponding analysis are presented in Section 4. Finally, Section 5 presents the conclusions of this study.

2. Materials and Methods

In this section, four techniques, namely, CEEMDAN, random forest, GSFS, and WOA, will be introduced. The proposed methodology is developed based on these important techniques.

2.1. Complementary Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN). The complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) originates from the ensemble empirical
mode decomposition (EEMD) algorithm. To solve the mode mixing problem, adaptive white noise that has a specified standard deviation can be added at every stage of the decomposition, and a unique residue can be calculated to generate each mode. Moreover, the decomposition is completed with a negligible error; hence, the CEEMDAN provides superior spectral separation of the modes and an accurate reconstruction of the original signal at a lower computational cost [35, 36].

The process of the CEEMEDAN is as follows.

Step 1: add realizations of a white Gaussian noise series and generate the corresponding signal:

\[ y'(n) = y(n) + \epsilon_n \omega'(n), \]  

where \( y(n) \) is the original signal, \( i(i = 1, 2, \ldots, I) \) is the number of the trials, \( \epsilon_n \) is the noise standard deviation, and \( \omega'(n) \) denotes the realizations of the white Gaussian noise series.

Step 2: decompose I realizations \( y'(n) \) via EMD to obtain their first intrinsic mode function (IMF):

\[ \text{IMF}_1(n) = \frac{1}{I} \sum_{i=1}^{I} \text{IMF}_{1i}(n), \]  

Step 3: calculate the first residue \( r_1(n) \) (at the first stage \( k = 1 \)):

\[ r_1(n) = y(n) - \text{IMF}_1(n). \]  

Step 4: decompose the noise-added residues until their first EMD mode and define the \((k+1)\)-th IMF of the CEEMDAN:

\[ \text{IMF}_{k+1}(n) = \frac{1}{I} \sum_{i=1}^{I} E_1(r_k(n) + \epsilon_n E_k(\omega'(n))), \]  

where the operator \( E_k(\cdot) \) produces the k-th mode of the EMD.

Step 5: calculate the k-th residue:

\[ r_k(n) = r_{k-1}(n) - \text{IMF}_k(n), \]  

where \( k = 2, \ldots, K, K \) is the total number of modes.

Step 6: repeat step 4 and step 5 for the next \( k \) until the obtained residue is not suitable for decomposition. The final residue is expressed as

\[ R(n) = y(n) - \sum_{i=1}^{K} \text{IMF}_i(n). \]  

Hence, the original signal can be expressed as

\[ y(n) = \sum_{i=1}^{K} \text{IMF}_i(n) + R(n). \]  

2.2. Random Forest. Random forest is an ensemble learning method that uses decision trees as base learners and measures the importance of each variable. Random forest is a variant of bagging (bootstrap aggregating), which considers all \( m \) attributes or variables when establishing a base learner. Decision trees tend to overfit the data, thereby leading to low bias but high variance. To overcome this drawback, random forest constructs decision trees and integrates the results of all individual trees to generate the final outcome [37]. In classification, the mode of the classes that are predicted by the trees is adopted as the output, while, in regression, the mean of the predictions of all individual trees is used.

Since bootstrapping is used to generate the sample from which an individual tree is grown, only approximately 63.2% of the observations are used to produce each base learner, and the remainder of the training set can be used as a validation set: suppose that there are \( m \) samples. The probability that not all the samples are chosen in the training set is \((1 - 1/m)^m\) which is \((1/e \approx 36.8\%)\) if \( m \to \infty \). For data \( x = (x, y) \), with \( x \) representing the observations and \( y \) representing the response variable, the out-of-bag (OOB) estimate, which is denoted by \( H_{\text{OOB}}(x) \), is produced by base learners [32, 33]. Given an ensemble of classifiers \( h_1(x), h_2(x), \ldots, h_T(x) \), it is defined as follows:

\[ H_{\text{OOB}}(x) = \arg\min_{y \in Y} \sum_{i=1}^{T} I(h_i(x) = y) \cdot I(x \notin D_i), \]  

where \( Y \) is the set that consists of all possible values of the response and \( D_i \) is the subsample that is applied to produce base learner \( h_i \). The OOB estimate of the generalization error is expressed as

\[ e_{\text{OOB}} = \frac{1}{|D|} \sum_{(x, y) \in D} I(H_{\text{OOB}}(x) \neq y), \]  

where \( D \) is the training set.

Both bagging and random forest measure the importance of each variable via the following procedure: (1) calculate the OOB error of each observation and average them to obtain the OOB error of the forest. (2) Permute the values of the \( j \)-th feature among the training set. (3) Compute the OOB error again on this perturbed data set, and the difference in the OOB error before and after the permutation can be used to measure the importance of the \( j \)-th variable, where a larger difference corresponds to higher importance of variable \( j \).

2.3. Gram–Schmidt-Based Feature Selection Method (GSFS). Feature selection substantially facilitates the reduction of the model complexity and the enhancement of the forecasting accuracy when establishing a forecasting model. Many feature selection methods are designed based on penalized regression, which shrinks the estimate; these methods yield inaccurate results. In this study, a subset selection method, namely, the Gram–Schmidt-based feature selection method (GSFS) [38], is used for feature selection. The objective of GSFS is to identify all the relevant features that contribute to the model forecasting. The details of GSFS are presented as follows.
GSFS iterations: let $G^{(0)}$ be the empty set and initialize $\gamma_0$ and $k$ to 0. Here, $\gamma_0$ denotes the initial forecasting values. Calculate the pseudoresponse $R^{(k)} = \gamma_k - \hat{y}_k$ and select the optimal variable $x_{(k)}$ by regressing $R^{(k)}$ on $x$:

$$\tilde{y}_{k+1} = \arg\min_{i=1}^{n} \left( R_i^{(k)} - \delta_i^{(k)} x_{ij}^2 \right)^2,$$

where $\delta_i^{(k)} = \sum_{t=1}^{n} R_i^{(k)} x_{ij} / \sum_{t=1}^{n} x_{ij}^2$. Then, $G^{(1)} = G^{(0)} \cup \{x_{(k)}\}$ and let $\tilde{y}_{k+1} = \tilde{y}_k + \delta_i^{(k)} x_{(k)}$. The objective in building $G^{(1)}$ is to add the optimal variables into the empty set.

In the next $M_n = \left( n / \log p_n \right)^{1/2}$ steps, we will conduct orthogonalization before selecting each new feature. If $k \geq 1$, orthogonalize $G^{(k)} = \{x_{(1)}, x_{(2)}, \ldots, x_{(k)}\}$ to $G_{\text{ord}}^{(k)} = \{x_{(1)}^2, x_{(2)}^2, \ldots, x_{(k)}^2\}$ via the Gram–Schmidt method. Similar to the first step, compute the pseudoresponse $G^{(k)} = y_i - \tilde{y}_k$ and consider

$$\tilde{y}_{k+1} = \arg\min_{i=1}^{n} \left( G_i^{(k)} - \eta_i^{(k)} x_{ij}^2 \right)^2,$$

where $\eta_i^{(k)} = \sum_{t=1}^{n} G_i^{(k)} x_{ij}^2 / \sum_{t=1}^{n} x_{ij}^2$. Update $\tilde{y}_{k+1} = \tilde{y}_k + \eta_i^{(k)} x_{(k)}$ and $G^{(k+1)} = G^{(k)} \cup \{x_{(k)}\}$. The objective of this procedure is to select variables that are orthogonal with each other sequentially and add them into an empty set.

The main advantage of GSFS is that it does not involve any penalized regression. It detects the relevant features by examining each feature sequentially. Although it may take more time, it will not miss any important data information. In the second step, GSFS computes the pseudoresponse as the target vector, which can be used to find the most strongly correlated features with the pseudoresponse. This pseudoresponse differs from the ordinary response that is applied in regression or classification problems. It is updated in every iteration of GSFS to eliminate the effects of selected features. An orthogonalization procedure that is based on the Gram–Schmidt method is applied to make all the features orthogonal with each other. This procedure removes the dependencies among features and shortens the computation time for inverting the projection matrix.

The determination of $M_n$ plays an important role in guaranteeing the accuracy and the selection consistency. Let $p_n \rightarrow \infty$; namely, let the number of features increase with the sample size. We must impose the following regularity condition before proving the theorem.

(C1) $\log p_n = o(n)$. It follows that $\lim_{n \rightarrow \infty} (\log p_n)/n = 0$.

Assume that the noise term is independent of each feature and let $u$ and $u_0$ be constants; namely,

(C2) $E(\exp(u \cdot z_i^2)) < \infty$ for $u < u_0$.

Let $\sigma_i^2 = E(x_i^2)$, which represents the noise level and let $z_i = (x_i / \sigma_i)$ and $\tilde{z}_{ij} = (x_{ij} / \sigma_j)$. Assume that there exists $u_1$ such that

(C3) $\limsup_{n \rightarrow \infty} \max_{1 \leq j \leq p_n} E(\exp(u_1 \cdot z_i^2)) < \infty$.

(C4) $\sup_{i=1}^{n} \sum_{j=1}^{p_n} |\delta_i \sigma_j| < \infty$.

(C5) There exists $0 \leq \gamma \leq 1$ such that $n^\gamma = o \left( (n / \log p_n)^{1/2} \right)$ and $\liminf_{n \rightarrow \infty} n^\gamma \min_{1 \leq j \leq p_n; \delta_j \neq 0} \delta_j^\gamma > 0$.

Theorem 1. Suppose that regularity conditions (C1)–(C5) are satisfied. Let $M_n = O \left( (n / \log p_n)^{1/2} \right)$ and $M_n = O \left( (n / \log p_n)^{1/2} \right)$. Then, $\lim_{n \rightarrow \infty} P \left( N_n = S_{M_n} \right) = 1$, where $N_n = \{1 \leq j \leq p_n; \delta_j^\gamma \neq 0\}$ denotes the set of important variables.

Theorem 1 establishes the selection consistency of the GSFS method and provides a theoretical guarantee on the choice of $M_n$. Due to the selection consistency, the feature selection method can detect all the important features as the sample size approaches infinity. Here, $G_{\text{M}}$ contains the features that are selected in step $M_n$. To select the optimal model with all the important features, high-dimensional information criteria (HDIC) will be applied after GSFS. The proof of Theorem 1 is similar to [34] and is omitted here.

2.4. Whale Optimization Algorithm (WOA). In 2016, Mirjalili and Lewis [39] proposed the whale optimization algorithm (WOA), which is a novel metaheuristic that mimics the bubble-net foraging behavior of humpback whales in searching for their prey. It is illustrated in Figure 2(a). The humpback whales attack small fishes or schools of krill that are close to the surface by swimming in a shrinking circle and creating distinctive bubbles along a spiral-shaped path. The algorithm includes two phases: the first phase is encircling a prey and applying the spiral bubble-net attacking method, and the second phase is searching randomly for a prey. The details of each phase are presented as follows.

2.4.1. Encircling Prey. The humpback whales dive downwards and generate bubbles in a spiral shape, which makes the prey (typically krill) feel threatened and, hence, gather together. It is assumed that there are $N$ humpback whales in a $d$-dimensional search space. The position of each whale is denoted by $X_i = (X_{i1}, X_{i2}, \ldots, X_{id}) (i = 1, 2, \ldots, N)$. The position of the prey represents the globally optimal solution to the problem. The humpback whales can identify the prey’s position and encircle it. However, there is no priori information about the optimal solution. Therefore, in the WOA, the current best position is adopted as the position that corresponds to the optimal solution. Once this position has been identified, other humpback whales will gather around it and update their position as follows [36]:

$$X(t + 1) = X^* (t) - A \cdot D,$$

$$D = |C \cdot X^* (t) - X(t)|,$$
where $t$ is the current iteration, $X(t)$ is the current position, $X^*(t)$ is the current best position, and $A$ and $C$ are the coefficient vectors:

$$
A = 2a \cdot \text{rand}_1 - a,
C = 2 \cdot \text{rand}_2,
$$

(14)

where $\text{rand}_1$ and $\text{rand}_2$ are 2 random numbers that are in the interval $(0, 1)$ and $a$ is a convergence factor, which decreases linearly as the iteration proceeds. Define $t$ as the iteration number and $t_{\text{max}}$ as the maximum allowed number of iterations. Factor $a$ is defined as

$$
a = 2 \cdot \frac{2t}{t_{\text{max}}},
$$

(15)

Two mechanisms are designed to describe the hunting behavior of the humpback whales, namely, the shrinking encircling mechanism and the spiral position updating mechanism, as illustrated in Figures 2(b) and 2(c). Shrinking encircling is realized as the convergence factor $a$ decreases, while, in spiral position updating, the spiral movement of the humpback whales is simulated by the following:

$$
X(t + 1) = D' \cdot e^{ibl} \cdot \cos(2\pi l) + X^*(t),
$$

(16)

where $D' = |X^*(t) - X(t)|$ is the distance between the humpback and the prey, $b$ is the shape parameter of the spiral, and $l$ is a random number that is in the interval $(-1, 1)$. Since the humpback whales perform these two behaviors simultaneously, in each iteration, each of these mechanisms has a probability of 0.5 to be utilized. The corresponding model is as follows:

$$
X(t + 1) = \begin{cases} 
X^*(t) - A \cdot D, & f \ p < 0.5, \\
D' \cdot e^{ibl} \cdot \cos(2\pi l) + X^*(t), & f \ p \geq 0.5 
\end{cases}
$$

(17)

where $p$ is a random number in $(0, 1)$.

### 2.4.2. Search for Prey.

To enhance the exploration in WOA, instead of updating the positions of the humpback whales based on the best position so far, a random position is chosen to guide the search accordingly. A with random values that are greater than 1 or less than $-1$ is applied to force the position to move far away from the best position. The model of searching for prey is defined as [40]

$$
D'' = |C \cdot X_{\text{rand}} - X|,
$$

(18)

$$
X(t + 1) = X_{\text{rand}} - A \cdot D'',
$$

(19)

where $X_{\text{rand}}$ is a random position (or a random whale) that is chosen from the current population.

### 3. Proposed Hybrid Forecasting Model: RF-GSFS-WOA

In this paper, the proposed RF-GSFS-WOA is defined by the following procedure.
Step 1: the original wind speed time series \( y(n) \) is decomposed into \( I \) IMFs (let \( \text{IMF}_1 = R \ (n) \)) by following Step 1–Step 6 in Section 2.1.

\[
\text{IMF}_1 = a_1^{(1)} x_1 + a_2^{(1)} x_2 + \ldots + a_p^{(1)} x_p + a_1^{(1)} x_1^2 + a_2^{(1)} x_2^2 + \ldots + a_p^{(1)} x_p^2 + \epsilon_1, \\
\text{IMF}_2 = a_1^{(1)} x_1 + a_2^{(2)} x_2 + \ldots + a_p^{(2)} x_p + a_1^{(1)} x_1^2 + a_2^{(1)} x_2^2 + \ldots + a_p^{(1)} x_p^2 + \epsilon_2, \\
\vdots \\
\text{IMF}_I = a_1^{(1)} x_1 + a_2^{(I)} x_2 + \ldots + a_p^{(I)} x_p + a_1^{(I)} x_1^2 + a_2^{(I)} x_2^2 + \ldots + a_p^{(I)} x_p^2 + \epsilon_n, 
\]

where \( x_1, \ldots, x_p \) are variables (first order) in the original dataset; \( x_1^2, x_2^2, \ldots, x_p^2 \) are quadratic variables (second order) that are constructed by calculating the component-wise product of \( x_j \) and \( x_j \); \( a_1, a_2, \ldots, a_p, a_1, a_2, \ldots, a_p \) are the corresponding coefficients of the variables and quadratic variables; and \( \epsilon_1, \ldots, \epsilon_n \) are independent Gaussian noise errors that follow \( N(0, \sigma^2) \). I quadratic models will be obtained, which are denoted by \( \text{QM}_1, \ldots, \text{QM}_I \). As \( \sigma^2 + p \) variables are included in the QM, the computational cost is high even if the value of \( p \) is moderate. Therefore, a variable selection method is urgently needed to reduce the computational cost.

Step 3: to reduce the computational cost of QM in Step 2, GSFS is used to complete the variable selection task. With slight abuse of notation, \( \text{IMF}_i (1 \leq i \leq I) \) is denoted by \( \text{IMF}_0 \), where \( \text{IMF}_0 = 0, k = 0 \), and \( \text{IMF}_0 \) represents the initial forecasting value, \( \text{R}_i = \text{IMF}_0 - \text{IMF}_1 \), and the optimal variables are selected by regressing \( \text{R}_i \) on \( \text{X}^0: = \{ x_1, \ldots, x_p, x_1^2, x_2^2, \ldots, x_p^2 \} = \{ x_p \}_1 \leq j \leq p + p \), which is expressed as

\[
\hat{j}_{k+1} = \arg \min_n \sum_{i=1}^n \left( R_i^{(k)} - \delta^{(k)} x_i^{(j)} \right)^2, \tag{21}
\]

where \( \delta^{(k)} = \sum_{i=1}^n R_i^{(k)} x_i^{(j)} / \sum_{i=1}^n x_i^{(j)} \). Then, \( G^{(1)} = G^{(0)} \cup \{ x_i^{(j)} \}_1 \) and let \( \text{IMF}^{(k)}_k = \text{IMF}_k + \delta^{(k)} x_i^{(j)} \). In the next \( M_n = \left( (n \log p)^{1/2} / 2 \right) \) steps, the data will be orthogonalized. If \( k \geq 1 \), orthogonalize \( G^{(k)} = \{ x_i^{(j)} \} \) to \( G^{(k)} = \{ x_i^{(j)} \} \) via the Gram–Schmidt method. Similar to the first step, compute the pseudoresponse \( \text{O}^{(k)}_i = \text{IMF}_i - \text{IMF}^{(k)}_k \) and consider

\[
\hat{j}_{k+1} = \arg \min_n \sum_{i=1}^n \left( O_i^{(k)} - \gamma^{(k)} x_i^{(j)} \right)^2, \tag{22}
\]

where \( \gamma^{(k)} = \sum_{i=1}^n O_i^{(k)} x_i^{(j)} / \sum_{i=1}^n x_i^{(j)} \). Update \( \text{IMF}^{(k+1)} = \text{IMF}^{(k)} + \gamma^{(k)} x_i^{(j)} \) and \( G^{(k+1)} = G^{(k)} \cup \{ x_i^{(j)} \} \).

Step 4: random forest model RF is built over the variables that were selected in Step 3.

Step 5: assign a weight \( \omega_i \) to each RF, and the final forecast model is provided by the ensemble model (EM):

\[
\text{EM} = \sum_{i=1}^I \omega_i \text{RF}_i, \tag{23}
\]

with \( \sum_{i=1}^I \omega_i = 1 \), where \( \omega_i \) is obtained via the WOA algorithm.

The RF-GSFS-WOA algorithm is presented in detail as Algorithm 1. The proposed combination forecasting method, namely, RF-GSFS-WOA, has the following advantages. Instead of applying a single decision tree, RF grows many decision trees and, hence, can handle the correlation among the features. It bootstraps samples, which promotes the diversity of the base learners and, consequently, reduces the number of variables of the model. Furthermore, RF outperforms bagging as the number of base learners increases. It trains fast since it applies a fraction of the features instead of considering all possible features. The time that is required for training RF is often shorter than that for bagging. The overfitting problem of RF can be controlled well via the selection of a suitable number of features in each decision tree. GSFS is a greedy feature selection method, which is a type of forward selection method. In contrast to penalized feature selection methods such as LASSO and SCAD, it does not introduce any bias into the model. WOA is a meta-heuristic method that can obtain the globally optimal solutions of nonlinear optimization problems. A flowchart of the proposed RF-GSFS-WOA is presented in Figure 3. The RF-GSFS-WOA algorithm is presented in the Appendix.

4. Case Studies

This section describes the data collection process, establishes the forecasting assessment criteria, and presents, discusses, and analyzes the experiment results. The main objective of the experiments is to evaluate the performance of the proposed hybrid forecasting model, namely, RF-GSFS-WOA.
4.1. Data Collection. Qinghai province is located in the northeast of the Qinghai-Tibetan Plateau, and most of the area is rich in wind energy resources; hence, it is suitable for the installation of large-scale wind farm. Thus, it is meaningful to investigate sites in this region. The hourly data were collected between 0:00 and 24:00 from 1/1/2014 to 12/31/2014 at six sites in the Qinghai region. The dataset that is used in this study was collected by National Renewable Energy Laboratory (NREL) and is available at https://www.nrel.gov/gis/tools.html. To simplify the description, the six sites are denoted by Site 1, Site 2, Site 3, Site 4, Site 5, and Site 6. The longitudes and latitudes of these locations are specified in Figure 4. Five meteorological factors, namely, the temperature \(T\), pressure \(P\), relative humidity \(RH\), wind precipitable water \(PW\), and wind direction \(WD\), are considered in the forecasting of the wind speed for the

\[
(i) \text{ INPUTS: } D = \{X, y\}, \text{ the wind speed dataset.}
\]

\[
(ii) \text{ PARAMETERS:}
\]

- \(N_{\text{std}}\): noise standard deviation in CEEMDAN
- \(NR\): number of realizations in CEEMDAN
- \(\text{MaxIter}\): maximum number of sifting iterations that are allowed in CEEMDAN
- \(q\): variable that is selected in GSFS
- \(\text{MaxIterRF}\): maximum number of iterations in random forest
- \(\text{NTress}\): number of trees in random forest
- \(N\): number of whales in WOA

\[
\text{MaxIterWOA: maximum number of iterations in WOA}
\]

\[
(iii) \text{ OUTPUTS: the forecasting error}
\]

\[
(1) \text{ The original dataset } D \text{ is divided into training dataset } D_{\text{trn}} = (X_{\text{trn}}, y_{\text{trn}}) \text{ and test dataset } D_{\text{tst}} = (X_{\text{tst}}, y_{\text{tst}}).
\]

\[
(2) \text{ Decomposition using CEEMDAN, which results in } I \text{ IMFs: } \text{IMF}_1, \text{IMF}_2, \ldots, \text{IMF}_I.
\]

\[
(3) \text{ Establish a quadratic model of } X_{\text{trn}} \text{ and } X_{\text{tst}}, \text{which results in } X_{\text{trn}}^{o} \text{ and } X_{\text{tst}}^{o}.
\]

\[
(4) \text{ INITIALIZATION: set } G^{(0)} = \emptyset, \text{IMF}^{\wedge}_{0} = 0, k = 0.
\]

\[
(5) \text{ IF } k = 0 \text{ DO}
\]

\[
(6) \text{ Calculate the pseudoresponse}
\]

\[
(7) \text{ Perform regression on the data with a pseudoresponse and select the most strongly correlated variable}
\]

\[
(8) \text{ Update the response } \text{IMF}^{\wedge}_{k} \text{ and } G^{(k+1)} = G^{(k)} \cup x_{k+1}^{o}.
\]

\[
(9) \text{ END IF}
\]

\[
(10) \text{ FOR } k = 1 \text{ to } M_{n} \text{ DO}
\]

\[
(11) \text{ Orthogonalize } G^{(k)} \text{ via the Gram–Schmidt method.}
\]

\[
(12) \text{ Perform regression on the data with a pseudoresponse and select the most strongly correlated variable}
\]

\[
(13) \text{ Update the response } \text{IMF}^{\wedge}_{k} \text{ and } G^{(k+1)} = G^{(k)} \cup x_{k+1}^{o}.
\]

\[
(14) \text{ END FOR}
\]

\[
(15) \text{ Build the solution path } G = \{G^{(k)} : 1 \leq k \leq M_{n}\} \text{ and use HDIC to select the optimal solution, including } q \text{ important variables.}
\]

\[
(16) \text{ Grow a random forest with } NTress = 3000 \text{ trees using the selected variables with MaxIterRF iterations.}
\]

\[
(17) \text{ Initialize the whale population for weight matrix } W \in R^{I \times d} \text{ in which the columns are weights for established RF models.}
\]

\[
(18) \text{ Calculate the fitness function using } W \text{ and RF models}
\]

\[
(19) \text{ Let } W^{*} \text{ be the current best weights}
\]

\[
(20) \text{ WHILE } t < \text{MaxIterWOA} \text{ DO}
\]

\[
(21) \text{ FOR each weight DO}
\]

\[
(22) \text{ Update } a, A, C, l \text{ and } p
\]

\[
(23) \text{ IF } p < 0.5 \text{ THEN}
\]

\[
(24) \text{ IF } |A| < 1 \text{ THEN}
\]

\[
(25) \text{ Update the positions of the current weight via (12) and (13)}
\]

\[
(26) \text{ ELSE IF } |A| \geq 1 \text{ THEN}
\]

\[
(27) \text{ Randomly select a search agent}
\]

\[
(28) \text{ Update the position of the current search agent via (17)}
\]

\[
(29) \text{ END IF}
\]

\[
(30) \text{ ELSE IF } p \geq 0.5 \text{ THEN}
\]

\[
(31) \text{ Update the position of the current weight via (16)}
\]

\[
(32) \text{ END FOR}
\]

\[
(33) \text{ Check if any search agent goes beyond the search space}
\]

\[
(34) \text{ Calculate the fitness function of each weight}
\]

\[
(35) \text{ Update } W^{*} \text{ if there is a better solution}
\]

\[
(36) t = t + 1
\]

\[
(37) \text{ END WHILE}
\]

\[
(38) \text{ Obtain the optimal weights}
\]

\[
(39) \text{ Calculate the forecasting error using the optimal weights and the RF model that is generated by } X_{\text{tst}}^{o} \text{ and } y_{\text{tst}}.
\]

Algorithm 1: RF-GSFS-WOA algorithm.
Qinghai region, which facilitates the design of wind power plants to efficiently generate electricity for the local population [41]. There are four seasons in each year: spring (Mar., Apr., and May), summer (Jun., Jul., and Aug.), autumn (Sep., Oct., and Nov.), and winter (Dec., Jan., and Feb.). The data are divided into training data and forecasting data as follows: 20 days in each season are randomly chosen as the training data (20 days × 24 hours/day = 480) for building the model, and the forecasting data include 5 days (5 days × 24 hours/day = 120) that are randomly chosen from the remaining days in each season. The performances of various forecasting models are evaluated in the four seasons. The training sets and forecasting sets for the six sites are described in Table 1. The forecasting sets account for approximately 20% of the total data, which is reasonable.

4.2. Forecasting Assessment Criteria. To evaluate the results of the forecasting models quantitatively, four assessment measures, namely, the mean absolute error (MAE), root mean square error (RMSE), mean absolute percent error (MAPE), and Theil inequality coefficient (TIC), are applied. All are calculated between the actual data and the forecasted values, and the smaller their values, the better the forecasting results. These criteria are expressed as follows:

Figure 3: Flowchart of the developed model: RF-GSFS-WOA.
The hourly wind speed data collected from six sites in Qinghai province of China.

### Table 1: Training sets and forecasting sets for hourly wind speed data at six sites.

| Season         | Training set (size: 20 × 24 = 480)                          | Forecasting set (size: 5 × 24 = 120) | Relative meteorological factors |
|----------------|-------------------------------------------------------------|-------------------------------------|---------------------------------|
| Spring (Mar., Apr., May) | Mar. 9<sup>th</sup>, Mar. 12<sup>th</sup>, Mar. 19<sup>th</sup>, and Mar. 21<sup>st</sup> | May 16<sup>th</sup> and May 20<sup>th</sup> | Temperature, pressure, precipitable water, and wind direction |
|                | Mar. 26<sup>th</sup>, Mar. 29<sup>th</sup>, Apr. 5<sup>th</sup>, and Apr. 6<sup>th</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
|                | Apr. 13<sup>th</sup>, Apr. 15<sup>th</sup>, Apr. 16<sup>th</sup>, and Apr. 21<sup>st</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
|                | Apr. 22<sup>nd</sup>, Apr. 23<sup>rd</sup>, Apr. 24<sup>th</sup>, and Apr. 25<sup>th</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
|                | Apr. 28<sup>th</sup>, May 1<sup>st</sup>, May 8<sup>th</sup>, and May 9<sup>th</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
|                | Jun. 1<sup>st</sup>, Jun. 3<sup>rd</sup>, Jun. 4<sup>th</sup>, and Jun. 6<sup>th</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
|                | Jun. 13<sup>th</sup>, Jun. 16<sup>th</sup>, Jun. 22<sup>nd</sup>, and Jun. 26<sup>th</sup> | May 25<sup>th</sup>, May 28<sup>th</sup>, and May 29<sup>th</sup> |                                  |
| Summer (Jun., Jul., Aug.) | Jun. 30<sup>th</sup>, Jul. 2<sup>nd</sup>, Jul. 7<sup>th</sup>, and Jul. 13<sup>th</sup> | Aug. 12<sup>th</sup> and Aug. 20<sup>th</sup> |                                  |
|                | Jul. 15<sup>th</sup>, Jul. 18<sup>th</sup>, Jul. 22<sup>nd</sup>, and Jul. 28<sup>th</sup> | Aug. 12<sup>th</sup> and Aug. 20<sup>th</sup> |                                  |
|                | Jul. 31<sup>st</sup>, Aug. 1<sup>st</sup>, Aug. 3<sup>rd</sup>, and Aug. 11<sup>th</sup> | Aug. 22<sup>nd</sup>, Aug. 24<sup>th</sup>, and Aug. 25<sup>th</sup> |                                    |
|                | Sep. 3<sup>rd</sup>, Sep. 6<sup>th</sup>, Sep. 8<sup>th</sup>, and Sep. 11<sup>th</sup> | Aug. 22<sup>nd</sup>, Aug. 24<sup>th</sup>, and Aug. 25<sup>th</sup> |                                    |
|                | Sep. 14<sup>th</sup>, Sep. 15<sup>th</sup>, Sep. 21<sup>st</sup>, and Sep. 24<sup>th</sup> | Aug. 22<sup>nd</sup>, Aug. 24<sup>th</sup>, and Aug. 25<sup>th</sup> |                                    |
| Autumn (Sep., Oct., Nov.) | Sep. 30<sup>th</sup>, Oct. 4<sup>th</sup>, Oct. 6<sup>th</sup>, and Oct. 13<sup>th</sup> | Nov. 18<sup>th</sup> and Nov. 23<sup>rd</sup> |                                  |
|                | Oct. 18<sup>th</sup>, Oct. 31<sup>st</sup>, Nov. 2<sup>nd</sup>, and Nov. 3<sup>rd</sup> | Nov. 18<sup>th</sup> and Nov. 23<sup>rd</sup> |                                  |
|                | Nov. 4<sup>th</sup>, Nov. 8<sup>th</sup>, Nov. 12<sup>th</sup>, and Nov. 14<sup>th</sup> | Nov. 18<sup>th</sup> and Nov. 23<sup>rd</sup> |                                  |
|                | Dec. 2<sup>nd</sup>, Dec. 3<sup>rd</sup>, Dec. 4<sup>th</sup>, and Dec. 6<sup>th</sup> | Nov. 24<sup>th</sup>, Nov. 27<sup>th</sup>, and Nov. 29<sup>th</sup> |                                  |
| Winter (Dec., Jan., Feb.) | Dec. 9<sup>th</sup>, Dec. 13<sup>th</sup>, Dec. 14<sup>th</sup>, and Dec. 18<sup>th</sup> | Nov. 24<sup>th</sup>, Nov. 27<sup>th</sup>, and Nov. 29<sup>th</sup> |                                  |
|                | Dec. 20<sup>th</sup>, Dec. 25<sup>th</sup>, Dec. 26<sup>th</sup>, and Dec. 29<sup>th</sup> | Nov. 24<sup>th</sup>, Nov. 27<sup>th</sup>, and Nov. 29<sup>th</sup> |                                  |
|                | Dec. 30<sup>th</sup>, Jan. 5<sup>th</sup>, Jan. 8<sup>th</sup>, and Jan. 9<sup>th</sup> | Nov. 24<sup>th</sup>, Nov. 27<sup>th</sup>, and Nov. 29<sup>th</sup> |                                  |
|                | Jan. 16<sup>th</sup>, Jan. 18<sup>th</sup>, Jan. 28<sup>th</sup>, and Feb. 5<sup>th</sup> | Nov. 24<sup>th</sup>, Nov. 27<sup>th</sup>, and Nov. 29<sup>th</sup> |                                  |
\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i|,
\]
\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2},
\]
\[
\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i - \hat{y}_i}{y_i}\right) \times 100\%,
\]
\[
\text{TIC} = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \hat{y}_i^2} + \sqrt{\frac{1}{N} \sum_{i=1}^{N} y_i^2}} \times 100\%,
\]
where \(N\) is the size of the forecasting sample and \(y_i\) and \(\hat{y}_i\) are the actual value and the forecasted value, respectively, for time period \(i\).

### 4.3. Experiment Results and Corresponding Analysis.

The experimental settings are described in detail in Table 2. The parameters in the complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) are set as follows: the noise standard deviation \((N\text{std})\) \(\epsilon_0\) is set as 0.2, the number of realizations \((Nr)\) is set as 500, and the maximum allowed number of sifting iterations \((\text{MaxIter})\) is set as 500. The regularization parameters in LASSO and

| Nstd | Nr  | MaxIter | \(\lambda\) | MaxIterNN | Func   | \(h\) | \(q\) | MaxIterRF |
|------|-----|---------|-------------|-----------|--------|------|------|-----------|
| 0.2  | 500 | 500     | -           | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | 0.875       | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | 0.625       | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | -           | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | 0.875       | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | 0.625       | 2000      | Tansig | 10   | -    | -         |
| 0.2  | 500 | 500     | -           | 2000      | Tansig | -30  | 30   | 500       |

### Table 3: Forecasting performances of the proposed model and other models in four seasons at Site 1 and Site 2.

| Site | Seasons | Criteria | Forecasting models |
|------|---------|----------|--------------------|
|      |         | MAE (m/s) | Elman Elman-LASSO Elman-SCAD BP BP-LASSO BP-SCAD RF-GSFS-WOA |
|      | Spring  | 0.38  | 0.35 | 0.31 | 0.47 | 0.40 | 0.35 | 0.35 | 0.08 |
|      |         | RMSE (m/s) | 0.49 | 0.43 | 0.37 | 0.59 | 0.49 | 0.43 | 0.43 | 0.10 |
|      |         | MAPE (%) | 17.27 | 13.32 | 12.27 | 20.71 | 17.15 | 13.78 | 3.03 |
|      |         | TIC (%)  | 6.35 | 5.51 | 4.79 | 7.67 | 6.38 | 5.60 | 1.29 |
|      |         | MAE (m/s) | 0.38 | 0.33 | 0.29 | 0.49 | 0.39 | 0.36 | 0.36 | 0.08 |
|      |         | RMSE (m/s) | 0.49 | 0.42 | 0.36 | 0.59 | 0.49 | 0.44 | 0.44 | 0.10 |
|      |         | MAPE (%) | 12.76 | 10.30 | 8.99 | 15.38 | 11.86 | 11.85 | 2.44 |
|      |         | TIC (%)  | 6.11 | 5.24 | 4.99 | 7.25 | 6.15 | 5.52 | 1.24 |
|      | Summer  | MAE (m/s) | 0.39 | 0.28 | 0.22 | 0.44 | 0.40 | 0.34 | 0.34 | 0.08 |
|      |         | RMSE (m/s) | 0.48 | 0.35 | 0.27 | 0.56 | 0.49 | 0.43 | 0.43 | 0.10 |
|      |         | MAPE (%) | 26.46 | 19.85 | 13.76 | 30.44 | 27.60 | 21.55 | 5.90 |
|      |         | TIC (%)  | 8.57 | 6.31 | 4.90 | 10.17 | 8.75 | 7.72 | 1.81 |
|      |         | MAE (m/s) | 0.38 | 0.33 | 0.29 | 0.41 | 0.40 | 0.31 | 0.31 | 0.08 |
|      |         | RMSE (m/s) | 0.49 | 0.42 | 0.36 | 0.51 | 0.36 | 0.51 | 0.40 | 0.10 |
|      |         | MAPE (%) | 18.39 | 17.22 | 15.68 | 22.11 | 18.24 | 16.64 | 4.78 |
|      |         | TIC (%)  | 7.15 | 6.13 | 5.27 | 7.43 | 7.38 | 5.99 | 1.46 |

for Site 2.
SCAD are chosen via 10-fold cross-validation. In the experiments on the neural networks, the maximum number of iterations (MaxIterNN) is set as 2000 for both BP and Elman. Each activation function from the hidden layer to the output layer is set as the Tansig function (Func). The number of hidden neurons is crucial since it mainly determines the network complexity. To determine the optimal number of hidden neurons, values are selected from the following set: \{5, 8, 10, 15, 20, 25\}. Furthermore, Elman is trained using backpropagation through time, and the corresponding weights are selected based on a uniform distribution. Based on the experimental results, the gradient descent with momentum is set as 0.82, and the adaptive learning rate is set as 0.02. In GSFS, we set \( q \) as 30; hence, 30 variables will be chosen after the GSFS procedure, and, thus, there are 1200 variables in the quadratic model. The number of trees in RF is set as 3000, and the maximum number of iterations (MaxIterRF) is 500. RF is implemented using MATLAB function TreeBagger. The fitness function of WOA is selected as “F1,” which is the square error loss function, and the range is (0, 1). The weights are determined after normalization such that the sum of all the weights is 1.

To evaluate the forecasting performance of the proposed hybrid model for forecasting hourly wind speed, experiments for six sites and four seasons are designed in this paper. Four types of assessment criteria are used in all experiments to evaluate the performance. The proposed hybrid forecasting model is RF-GSFS-WOA, and six forecasting models, namely, an Elman neural network, Elman-LASSO, Elman-SCAD, a BP neural network (BP), BP-LASSO, and BP-SCAD, are applied as comparison models to evaluate the performance of the developed hybrid model. Wind speed data exhibit volatility or instability due to changes in the weather; hence, this study conducts a preprocessing step before building all the models; namely, the complementary ensemble empirical mode decomposition with adaptive noise (CEEMDAN) is used to eliminate the noise and smooth the wind speed timeseries data before the developed hybrid model and six comparison models are established for hourly wind speed forecasting in four seasons at six sites.

In each season, 20 days of hourly wind speed time series are selected randomly as the training phase for the construction of the models, and 5 days of wind speed time series are selected randomly as the forecasting phase.
the data will be used as training data; the remaining data (20%) will be used as test data. The reason for this splitting scheme is that the model can be trained sufficiently with more observations (80% of the data) and overfitting can be avoided. However, the number of observations in the test data is also important: the test accuracy cannot be evaluated accurately if the number of observations is too small (less than 20%).

The multiple output strategy (5 days are used as outputs), which is one of the multistep forecasting techniques, is applied in the experiments.

The wind speed is influenced by several meteorological variables. Consequently, seven forecasting models consider multidimensional meteorological factors, such as the temperature, pressure, relative humidity, precipitable water, and wind direction, as dependent variables in multivariate models. The use of multivariable inputs increases the difficulty of determining the model parameters.

Four performance criteria, namely, MAPE, RMSE, MSE, and TIC, for each forecasting model in the four seasons at the six sites, are listed in Tables 3–5. Figures 5–7 present the forecasting results of the seven models for observation sites 1–6, which successfully illustrate that the proposed hybrid model RF-GSFS-WOA yields closer results to the real wind speed data than any other forecasting model in all instances. Furthermore, the developing hybrid framework realizes the best MAPE, RMSE, MSE, and TIC values. For hourly wind speed forecasting with multidimensional meteorological variables, the forecasting performance is described in detail in Tables 3–5 and Figures 5–8, from which the following findings are obtained.

(I) From Table 3 and Figure 5, it is found that Elman-based models, namely, Elman, Elman-LASSO, and Elman-SCAD, outperform the BP-based models, namely, BP, BP-LASSO, and BP-SCAD, and that Elman and BP yield the worst forecasting results. For instance, at Site 1, the developed hybrid model, namely, RF-GSFS-WOA, outperforms the others six forecasting models. The RF-GSFS-WOA model realizes the smallest MAPE of 2.44% in summer, and similar results are obtained in terms of the MAE, RMSE, and TIC criteria.

(II) According to Table 4 and Figure 6, the proposed RF-GSFS-WOA still outperforms other forecasting models in all seasons at Site 5 and Site 6.
models by realizing lower values of MAE, RMSE, MAPE, and TIC. The forecasting models that were established based on Elman substantially outperform the models that were constructed based on BP. Elman-SCAD outperforms BP-SCAD. For example, the MAPE of Elman-SCAD is 7.61, which is critically lower than that of BP-SCAD, which is 10.37 in spring at Site 3. In autumn at Site 4, the RMSE of Elman-SCAD is 0.27, which is approximately 34.15% lower than that of BP-SCAD. Furthermore, similar to the results from Table 3 and Figure 5, the SCAD-related forecasting models outperform the LASSO-related forecasting models. This is mainly because SCAD can better handle the collinearity of variables than LASSO. Similar phenomena are observed in Table 5 and Figure 7. RF-GSFS-WOA realizes remarkable forecasting performances.

(III) According to Figure 8, which presents the average MAPE and RMSE values of the compared forecasting methods in the four seasons, RF-GSFS-WOA realizes the lowest MAPE values over all seasons, which are 4.04%, 3.73%, 3.02%, 3.08%, 2.81%, and 4.13% at the six sites. Elman-SCAD still outperforms Elman-LASSO and BP-SCAD. The worst value is obtained by BP, which is approximately 20%. BP-LASSO and BP-SCAD slightly outperform BP, which does not perform as well as Elman.

Therefore, from the discussions above, the proposed hybrid forecasting model, namely, RF-GSFS-WOA, performs well in the wind speed forecasting task by combining the advantages of machine learning module RF, feature selection approach GSFS, and metaheuristic method WOA.

4.4. Statistical Test. To further evaluate the performance of the proposed forecasting model, a nonparametric statistical test, namely, the Friedman test, is conducted. The forecasting performances of the compared methods differ among the six sites [42]. The following hypothesis, namely, the null
hypothesis $H_0$, along with the alternative hypothesis $H_1$, is considered in the Friedman test:

$$H_0 : L_1 = L_2 = \ldots = L_k,$$

where $L_j = (1/B)\sum_i a_{ij}^j$, with $a_{ij}^j$ representing the $j$-th rank of $k$ algorithms on a dataset of size $B$. The Friedman statistic is defined as follows [43]:

$$\chi^2_F = \frac{12B}{k(k+1)} \sum_j L_j^2 - \frac{k(k+1)^2}{4}.$$  \hspace{1cm} (25)

This statistic follows chi-square distribution $\chi^2_k (k-1)$. An $F$-distribution statistic, namely, $F(k-1, (k-1)(B-1))$, is calculated as follows:

$$F_F = \frac{B(k-1)\chi^2_F}{B(k-1) - \chi^2_F}.$$  \hspace{1cm} (26)

Suppose that there are significant differences among the comparing forecasting methods. Then, the null hypothesis $H_0$ will be rejected. If this occurs, critical differences (CDs) will be applied for further comparisons, which is known as a post hoc test. A Bonferroni-Dunn test is used with $CD$ values $q_{0.1} = \sqrt{k(k+1)/6B}$, where $q_{0.1}$ is determined as 2.291 in [44].

The Friedman test is described in detail in Table 6. All the $F_F$ values are much larger than the critical value $F(6, 30) = 2.42$. This suggests that the null hypothesis $H_0$ should be rejected and a post hoc test is needed. Using the above-described method, the $CD$ value is calculated as $q_{0.1} = \sqrt{6B} = 2.291 \times \sqrt{7/12} = 2.8574$, which is used as a benchmark for comparison. The proposed RF-GSFS-WOA significantly outperforms BP-LASSO, BP, and Elman since the differences between the average ranks of these methods exceed 2.85. For instance, the difference between RF-GSFS-WOA and BP-LASSO is 4.67 for MAE, which is larger than 2.85. Compared to other methods, the proposed method does not show strong advantages in forecasting. However, based on the above analysis, the proposed method delivers remarkable forecasting results.
Figure 7: Forecasting results of seven models at Site 5 and Site 6.

Figure 8: Continued.
Figure 8: Average values of MAPE and RMSE for four seasons at six sites.

Table 6: Results of Friedman test for all forecasting methods.

| Models        | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 | Site 6 | Average | $\chi^2_F$ | $F_F$  |
|---------------|--------|--------|--------|--------|--------|--------|---------|-----------|--------|
| MAE           |        |        |        |        |        |        |         |           |        |
| Elman         | 5      | 5      | 5      | 6      | 5      | 6      | 5.33    |           |        |
| Elman-LASSO   | 3      | 3      | 3      | 4      | 3      | 3      | 3.17    |           |        |
| Elman-SCAD    | 2      | 2      | 2      | 2      | 2      | 2      | 2.00    |           |        |
| BP            | 7      | 7      | 7      | 7      | 7      | 7      | 7.00    | 35.07    | 188.84 |
| BP-LASSO      | 6      | 6      | 6      | 5      | 6      | 5      | 5.67    |           |        |
| BP-SCAD       | 4      | 4      | 4      | 3      | 4      | 4      | 3.83    |           |        |
| RF-GSFS-WOA   | 1      | 1      | 1      | 1      | 1      | 1      | 1.00    |           |        |
| RMSE          |        |        |        |        |        |        |         |           |        |
| Elman         | 5      | 6      | 5      | 6      | 6      | 6      | 5.67    |           |        |
| Elman-LASSO   | 3      | 3      | 3      | 4      | 3      | 3      | 3.17    |           |        |
| Elman-SCAD    | 2      | 2      | 2      | 2      | 2      | 2      | 2.00    |           |        |
| BP            | 7      | 7      | 7      | 7      | 7      | 7      | 7.00    | 35.07    | 188.84 |
| BP-LASSO      | 6      | 5      | 6      | 5      | 5      | 5      | 5.33    |           |        |
| BP-SCAD       | 4      | 4      | 4      | 3      | 4      | 4      | 3.83    |           |        |
| RF-GSFS-WOA   | 1      | 1      | 1      | 1      | 1      | 1      | 1.00    |           |        |
| MAPE          |        |        |        |        |        |        |         |           |        |
| Elman         | 5      | 6      | 5      | 6      | 6      | 6      | 5.50    |           |        |
| Elman-LASSO   | 3      | 3      | 3      | 4      | 3      | 3      | 3.17    |           |        |
| Elman-SCAD    | 2      | 2      | 2      | 2      | 2      | 2      | 2.00    |           |        |
| BP            | 7      | 7      | 7      | 7      | 7      | 7      | 7.00    | 35       | 175    |
| BP-LASSO      | 6      | 5      | 6      | 5      | 5      | 5      | 5.50    |           |        |
| BP-SCAD       | 4      | 4      | 4      | 3      | 4      | 4      | 3.83    |           |        |
| RF-GSFS-WOA   | 1      | 1      | 1      | 1      | 1      | 1      | 1.00    |           |        |
| TIC           |        |        |        |        |        |        |         |           |        |
| Elman         | 5      | 6      | 5      | 6      | 6      | 6      | 5.67    |           |        |
| Elman-LASSO   | 3      | 3      | 3      | 4      | 3      | 3      | 3.17    |           |        |
| Elman-SCAD    | 2      | 2      | 2      | 2      | 2      | 2      | 2.00    |           |        |
| BP            | 7      | 7      | 7      | 7      | 7      | 7      | 7.00    | 35.07    | 188.84 |
| BP-LASSO      | 6      | 5      | 6      | 5      | 5      | 5      | 5.33    |           |        |
| BP-SCAD       | 4      | 4      | 4      | 3      | 4      | 4      | 3.83    |           |        |
| RF-GSFS-WOA   | 1      | 1      | 1      | 1      | 1      | 1      | 1.00    |           |        |
5. Conclusions
Accurate wind speed forecasting has attracted the attention of researchers of wind energy for decades. This paper proposed and investigated a novel hybrid wind speed forecasting model that combines the advantages of CEEMDAN, GSFS, RF, and WOA. The proposed model has been applied to data from six sites in Qinghai province to forecast the wind speed. The forecasting performances are evaluated using four statistical measures: MAE, RMSE, MAPE, and TIC. The forecasting results demonstrate that the proposed hybrid model outperforms other compared forecasting models, namely, Elman, Elman-LASSO, Elman-SCAD, BP, BP-LASSO, and BP-SCAD, which are famous wind speed forecasting models and variable selection methods. Moreover, the Elman-based forecasting models deliver better results than the BP-based models, and the SCAD-based models outperform the LASSO-based models. Furthermore, by using variable selection methods such as GSFS, LASSO, and SCAD, the model performances are boosted substantially. Thus, variable selection is necessary since it can be used to select the meteorological factors that make the most important contributions to the forecasting model while discarding useless factors. To further evaluate the performance of the proposed model, a nonparametric test, namely, the Friedman test, is conducted to evaluate the forecasting results. The test results are consistent with the findings of the analysis of the results. Consequently, according to the experimental results, the proposed hybrid model forecasts the wind speed accurately and efficiently [38, 45].

Data Availability
The data that are used to support the findings of this study are included within the article.

Conflicts of Interest
The authors declare that there are no conflicts of interest regarding the publication of this article.

Acknowledgments
This research was supported by the National Natural Science Foundation of China (Grant nos. 71901109, 71761016, 61973145, 71861012, and 18ATJ001), Natural Science Foundation of Jiangxi, China (no. 2018BAB211020), Jiangxi Double Thousand Plan, Postdoctoral Foundation of Jiangxi Province (no. 2018KY08), Scientific Research Fund of Jiangxi Provincial Education Department (Grant nos. GJJ180287, GJJ180247, and GJJ190264), and Human and Social Science Foundation of Jiangxi Province (no. T19202).

References
[1] N. A. Shrivastava, K. Lohia, and B. K. Panigrahi, “A multi-objective framework for wind speed prediction interval forecasts,” Renewable Energy, vol. 87, pp. 903–910, 2016.
[2] E. T. Renani, M. F. M. Elias, and N. A. Rahim, “Using data-driven approach for wind power prediction: a comparative study,” Energy Conversion and Management, vol. 118, pp. 193–203, 2016.
[3] I. Okumus and A. Dinler, “Current status of wind energy forecasting and a hybrid method for hourly predictions,” Energy Conversion and Management, vol. 123, pp. 365–371, 2016.
[4] Global Wind Energy Council, “Global wind statistics,” Global Wind Energy Council, Brussels, Belgium, 2017, http://gwec.net/global-figures/.
[5] Y. Noorollahi, M. A. Jokar, and A. Kalhor, “Using artificial neural networks for temporal and spatial wind speed forecasting in Iran,” Energy Conversion and Management, vol. 115, pp. 17–25, 2016.
[6] A. B. Ashgar and X. Liu, “Adaptive neuro-fuzzy algorithm to estimate effective wind speed and optimal rotor speed for variable-speed wind turbine,” Neurocomputing, vol. 272, pp. 495–504, 2018.
[7] J. Shi, J. Guo, and S. Zheng, “Evaluation of hybrid forecasting approaches for wind speed and power generation time series,” Renewable and Sustainable Energy Reviews, vol. 16, no. 5, pp. 3471–3480, 2012.
[8] J. Zhao, Y. Guo, X. Xiao, J. Wang, D. Chi, and Z. Guo, “Multistep wind speed and power forecasts based on a WRF simulation and an optimized association method,” Applied Energy, vol. 197, pp. 183–202, 2017.
[9] F. Cassola and M. Burlando, “Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output,” Applied Energy, vol. 99, pp. 154–166, 2012.
[10] O. Ait Maatallah, A. Achuthan, K. Janoyan, and P. Marzocca, “Recursive wind speed forecasting based on Hammerstein autoregressive model,” Applied Energy, vol. 145, pp. 191–197, 2015.
[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. 1405–1414, 2011.
[12] S. Baran, “Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components,” Computational Statistics & Data Analysis, vol. 75, no. 6, pp. 227–238, 2014.
[13] Z. Song, Y. Jiang, and Z. Zhang, “Short-term wind speed forecasting with Markov-switching model,” Applied Energy, vol. 130, no. 5, pp. 103–112, 2014.
[14] C. W. Potter and M. Negnevitsky, “Very short-term wind forecasting for Tasmanian power generation,” IEEE Transactions on Power Systems, vol. 21, no. 2, pp. 965–972, 2006.
[15] C. D. Zuluaga, M. A. Álvarez, and E. Giraldo, “Short-term wind speed prediction based on robust Kalman filtering: an experimental comparison,” Applied Energy, vol. 156, pp. 321–330, 2015.
[16] G. Li and J. Shi, “On comparing three artificial neural networks for wind speed forecasting,” Applied Energy, vol. 87, no. 7, pp. 2313–2320, 2010.
[17] X. Kong, X. Liu, R. Shi, and K. Y. Lee, “Wind speed prediction using reduced support vector machines with feature selection,” Neurocomputing, vol. 169, pp. 449–456, 2015.
[18] L. Wang, X. Li, and Y. Bai, “Short-term wind speed prediction using an extreme learning machine model with error correction,” Energy Conversion and Management, vol. 162, pp. 239–250, 2018.
[19] D. Ambach and W. Schmid, “A new high-dimensional time series approach for wind speed, wind direction and air pressure forecasting,” Energy, vol. 135, pp. 833–850, 2017.
Z. Zhang, L. Ye, H. Qin et al., “Wind speed prediction method using shared weight long short-term memory network and Gaussian process regression,” Applied Energy, vol. 247, pp. 270–284, 2019.

F. Li, G. Ren, and J. Lee, “Multi-step wind speed prediction based on turbulence intensity and hybrid deep neural networks,” Energy Conversion and Management, vol. 186, pp. 306–322, 2019.

K. G. Sheela and S. N. Deepa, “Neural network based hybrid computing model for wind speed prediction,” Neurocomputing, vol. 122, pp. 425–429, 2013.

J. Wang, W. Yang, P. Du, and T. Niu, “A novel hybrid forecasting system of wind speed based on a newly developed multi-objective sine cosine algorithm,” Energy Conversion and Management, vol. 163, pp. 134–150, 2018.

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.

O. B. Shukur and M. H. Lee, “Daily wind speed forecasting through hybrid ARIMA,” Renewable Energy, vol. 76, pp. 637–647, 2015.

B. Doucoure, K. Agbossou, and A. Cardenas, “Time series prediction using artificial wavelet neural network and multi-resolution analysis: application to wind speed data,” Renewable Energy, vol. 92, pp. 202–211, 2016.

H. Liu, X. Mi, and Y. Li, “Smart deep learning based wind speed prediction model using wavelet packet decomposition, convolutional neural network and convolutional long short term memory network,” Energy Conversion and Management, vol. 166, pp. 120–131, 2018.

S. Zhu, X. Yuan, Z. Xu, X. Luo, and H. Zhang, “Gaussian mixture model coupled recurrent neural networks for wind speed interval forecast,” Energy Conversion and Management, vol. 198, p. 111772, 2019.

H. Cai, X. Jia, J. Feng et al., “A combined filtering strategy for short term and long term wind speed prediction with improved accuracy,” Renewable Energy, vol. 136, pp. 1082–1090, 2019.

C. Li, Z. Xiao, X. Xia, W. Zou, and C. Zhang, “A hybrid model based on synchronous optimisation for multi-step short-term wind speed forecasting,” Applied Energy, vol. 215, pp. 131–144, 2018.

A. Tascikaraoglu, B. M. Sanandaji, K. Poolla, and P. Varaiya, “Exploiting sparsity of interconnections in spatio-temporal wind speed forecasting using Wavelet Transform,” Applied Energy, vol. 165, pp. 735–747, 2016.

W. Sun and M. Liu, “Wind speed forecasting using FEEMD echo state networks with RELM in Hebei, China,” Energy Conversion and Management, vol. 114, pp. 197–208, 2016.

A. A. Abdoos, “A new intelligent method based on combination of VMD and ELM for short term wind power forecasting,” Neurocomputing, vol. 203, pp. 111–120, 2016.

H. Jiang and Y. Dong, “Global horizontal radiation forecast using forward regression on a quadratic kernel support vector machine: case study of the Tibet Autonomous Region in China,” Energy, vol. 133, pp. 270–283, 2017.

M. E. Torres, M. A. Colominas, G. Schlotthauer, and P. Flandrin, “A complete ensemble empirical mode decomposition with adaptive noise,” in Proceedings of the IEEE international conference on acoustics, speech and signal processing, ICASSP, pp. 4144–4147, Prague, Czech Republic, May 2011.