Research Article

An Advanced Hybrid Forecasting System for Wind Speed Point Forecasting and Interval Forecasting

Haipeng Zhang\textsuperscript{1} and Hua Luo\textsuperscript{2}

\textsuperscript{1}School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China
\textsuperscript{2}School of Economics and Finance, Shanghai International Studies University, 1550 Wexiang Road, Shanghai 201620, China

Correspondence should be addressed to Hua Luo; loha_u@shisu.edu.cn

Received 15 June 2020; Revised 16 September 2020; Accepted 29 October 2020; Published 23 November 2020

Copyright © 2020 Haipeng Zhang and Hua Luo. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Ultra-short-term wind speed prediction can assist the operation and scheduling of wind turbines in the short term and further reduce the adverse effects of wind power integration. However, as wind is irregular, nonlinear, and nonstationary, to accurately predict wind speed is a difficult task. To this end, researchers have made many attempts; however, they often use only point forecasting or interval forecasting, resulting in imperfect prediction results. Therefore, in this paper, we developed a prediction system integrating an advanced data preprocessing strategy, a novel optimization model, and multiple prediction algorithms. This combined forecasting system can overcome the inherent disadvantages of the traditional forecasting methods and further improve the prediction performance. To test the effectiveness of the forecasting system, the 10-min and one-hour wind speed sequences from the Sotavento wind farm in Spain were applied for conducting comparison experiments. The results of both the interval forecasting and point forecasting indicated that, in terms of the forecasting capability and stability, the proposed system was better than the compared models. Therefore, because of the minimum prediction error and excellent generalization ability, we consider this forecasting system to be an effective tool to assist smart grid programming.

1. Introduction

Along with the rapid development of modern industry, the issues of environmental deterioration and fossil fuel depletion are becoming increasingly serious. Seeking clean and renewable energy is a hotspot topic in the world. Across the globe, the scale of wind power generation is growing rapidly and wind energy has attracted attention [1].

However, when wind power is connected to the grid, the intermittence and volatility of wind power will affect the balance of power demand and supply, the security of the power system, and the power quality. Therefore, in order to ensure the stability of the wind power system and promote the large-scale integration of wind power, it is necessary to predict the short-term wind power with high accuracy [2]. For obtaining more accurate prediction results, a great deal of efforts has been made and many studies have been performed. On the basis of applied data processing models, we summarized the previous research into four classes: (a) physical class, (b) statistical class, (c) spatial correlation class, and (d) artificial intelligence class [3].

Generally speaking, the physical strategy uses the atmospheric data and numerical weather simulation to predict the wind speed. Cassola and Burlando [4] used an improved numerical weather prediction (NWP) model to provide 2-year-long wind speed dataset of two wind farms in the eastern Liguria (Italy). Zhao et al. [5] put forward a novel multistep forecasting method according to the weather research and forecasting for predicting the one-day ahead wind speed and power. Dong et al. [6] proposed a new clustering analysis model for NWP information, which proved that the clustering analysis of NWP can be effectively used in the field of wind speed forecasting. Yang and Tsai [7] developed two linear regression models and a microgenetic algorithm (MGA) model to predict the gust wind speed during a typhoon. NWP needs to observe not only the
conventional data of temperature, humidity, wind pressure, and visibility but also the cloud top conditions, ozone, solar radiation, sea surface temperature, and so on, and it also requires the use of supercomputers. Therefore, it cannot deal with the short-term prediction problem.

Compared with the physical strategy, the statistical strategy has clear advantages in addressing short-term forecasting problems due to the full use of historical data. From the perspective of existing statistical methods, exponential smoothing (ES), ARMA, GARCH, and other typical models are universally employed and can achieve reliable prediction effects.

Kavasseri and Seetharaman [8] applied ARIMA for forecasting wind speeds on the day-ahead (24 h) and two-day-ahead (48 h) horizons, which resulted in a better forecasting accuracy. Qian and Wang [9] proposed an improved seasonal gray prediction model (GM (1, 1)) to forecast the wind generation of 2020 and 2021 in China for further development of the wind power industry. The spatial correlation algorithm takes the spatial relationship between different locations’ wind speeds into consideration. For instance, Chen et al. [10] examined the application of a multifactor spatial-temporal correlation model and achieved an improvement from other benchmark models. Nevertheless, these models had to measure the wind speed from multiple spatial correlation points. Because of the strict measurement requirements and time delay, the requirements are difficult to meet.

Apart from the abovementioned models, artificial intelligence arithmetic has attracted a great deal of attention since its development. Different from the previously mentioned strategies, artificial neural networks are completely dependent on historical data, which means that the models have good fault tolerance, strong robustness, and adaptability of wind speed prediction. In addition, artificial neural networks have a good ability in dealing with nonlinear data, self-adaptive adjustment, determination of prediction mode, fuzzy relationship judgments, and so on. Scholars have used artificial intelligence algorithms to forecast the speed of wind, using SVM [11, 12], ANN [13–15], methods of fuzzy logic (FL) [16, 17], and so on.

Some utilizations of artificial intelligence algorithms from previous works are shown below. Guo et al. [18] proposed a new method according to BPNN mixture with seasonal exponential regulation, and this greatly improved the wind speed forecasting accuracy. Liu et al. [19] designed a novel predicting method according to the neural network of Elman (ENN) and the quadratic decomposition method, which demonstrated excellent abilities in multistep wind speed predicting.

At present, artificial neural networks are always combined with other intelligent algorithms. Yang et al. [20] proposed a new forecasting system for electricity market management with multistep electricity price forecasting. A deep learning framework with two-stage feature selection is developed by Niu et al. [21] for multivariate financial time series prediction. Although the abovementioned models achieved satisfactory results, there are still some inherent defects in the neural network model, such as the high degree of fit, ease of falling into local optimization, and the slow convergence speed.

However, the majority of the methods only obtain the average point of forecasting results and ignore the reflecting forecasting volatility, which makes it difficult to meet the needs and the risk management of a wind system. Therefore, many researchers have studied wind speed interval forecasting models. Uncertainty in the forecast process is an accompanying factor, which has an important impact on the decision-making process of the energy system [22]. Therefore, some papers use the interval forecast to predict this uncertainty. Based on the existing literature, the interval forecasting model can be summed up in four types of models: point forecasting error models, bootstrap models, nonparametric theory models, and quantile regression models.

The quantile regression method focuses on quantum regression, and within this, each component must be modelled [23]. Bootstrap is a statistical model that applies data sampling and substitution to evaluate the robustness of the statistics, including the parameter of the confidence interval, the standard error, the coefficient correlation institution, and the regression coefficient [24]. In addition, an interval forecasting model, which applies both the ANNs and lower upper bound estimation (LUBE), was established to structure the predicting intervals for generating high-quality forecasting intervals in a short time [25].

By reviewing the previous literature, we concluded that there is no single model that can currently be regarded as the best model for wind speed prediction [26]. A more intuitive comparison of all models is drawn in Table 1.

From the above analysis, a conclusion can be drawn that each method has inherent disadvantages due to the particular characteristics. Therefore, in 1969, Bates and Granger [27] first proposed the combined forecasting theory, which significantly improved the prediction performance of a single model and has been widely used up to now. Stated thus, a novel combined system is put forward in this paper for conducting wind speed prediction. The system integrates an advanced data pretreatment theory, a fresh optimization technique, and several prediction methods, namely, ARIMA, GRNN, BPNN, and BiLSTM, respectively.

Specifically speaking, in contrast to the previous subjective judgment of the number of denoising layers, we performed a weighted reconstruction of each sequence obtained through decomposition to effectively control the interference of noise data. For capturing the linear and nonlinear characteristics hidden in collected wind speed sequence, BPNN, GRNN, BiLSTM, and ARIMA were selected as subsequences to forecast the pretreatment series. Then, an advanced weight-deciding technology, nondominated neighbor immune algorithm (NNIA), was effectively introduced to determine the output results.

This paper makes up for the deficiencies in the previous research in the following aspects:

(1) In contrast to the previous processing ideology, this paper proposes a more effective data preprocessing theory for controlling the interference of noise data. According to the disintegration and reconstruction
theory, the collected time series was disintegrated into a set of series with different frequencies and then these series were reconstructed with combination weights determined by NNIA. Through such preprocessing, the randomness and volatility of a wind speed sequence are successfully alleviated.

(2) For fully capturing all data characteristics and making up for the defects of individual methods, one linear algorithm and three neural networks were employed as submodels to build up a combination forecasting system. To expand, ARIMA performs well in seizing linear information, while BPNN, ELMNN, and BiLSTM are effective at grasping nonlinear characteristics. The combination system not only can integrate the advantages of these submodels but is also more suitable for wind speed time series with complex characteristics.

(3) To comprehensively compare the capability of different models, we established an efficient assessment system, mining the point prediction module and interval prediction module. This system contained eight assessment indicators, five of which were used to indicate point forecasting errors, while three were employed for the accuracy of interval prediction.

(4) The proposed combination system with high accuracy and stability is a powerful tool for the smart grid. On the basis of the dataset from the Sotavento wind farm in Spain, the simulation results demonstrated that relative to other benchmark models, the system was able to achieve reliable improvements.

The structure of this article is as follows. Section 2 provides the modelling structure and the related methodologies. In Section 3, we introduce the implementation process of the three experiments and the obtained results. And, Section 4 offers our main conclusions. The abbreviations that appear in the text are listed in Table 2.

### 2. Design of the Combined Forecasting System

This section provides the detailed introduction of the modelling structure and the related methodology.

#### 2.1. Modelling Structure

Figure 1 illustrates the whole process of building the combination system in this paper.

In the 1960s, Bates et al. found that by combining two or more prediction models, the prediction accuracy could be greatly improved. Since then, combinatorial theory has been widely used. We propose a novel combination wind speed prediction system, and the specific steps are expressed as follows.

**Step 1. Data preprocessing ideology:**

The collected wind speed sequence was first decomposed into a set of filters with the different frequencies.

| Table 1: Comparison on the advantages and disadvantages of existing wind speed forecasting methods. |
| --- | --- | --- | --- | --- |
| Category | Advantage | Disadvantage | Method sample | Sample advantage | Sample disadvantage |
| NWP models | Suitable for long-term forecast | Hard to deal with short-term horizons and costs a lot of computing time and resources | Computational fluid dynamics (FLUENT, CFD), Mesoscale numerical model (MM5, WRF), Physical schema (predictor) | High parallel efficiency, Accurate simulation and a large number of alternative parameterization schemes, Can predict spatial properties | High requirement of dataset length, limited simulation range, Suitable for weather forecast, not specifically optimized for wind prediction, Expensive calculation cost |
| Statistical strategies | Wide application, less running time | Assumption difficult to achieve | Time series (ARIMA, ARCH, GARCH, GM (1, 1)), Cluster | Simple model assumptions, good self-fitness, Strong fault tolerance, simple human-computer interaction | Poor extrapolation, limited prediction range, Low accuracy, lack of systematicness |
| Machine learning | Excellent robustness, little model risk, wide application | High complexity, high requirements of knowledge | Neural network (FNN, BPNN, GRNN, ELMAN), Artificial intelligence (PSO, GA, BA, DA), Data denoising (EMD, EEMD, CEEMD, SSA), Deep learning (CNN, RNN, LSTM) | Excellent fitting effect nonlinear property, Easy to be understood and combined with other methods, Eliminate the negative impact of noise data, Strong fault tolerance, simple human-computer interaction | The model is unstable, high degree of data dependence, Easy to fall into local extremum, The mode mixing problem in EMD and the residual noise in EEMD, Low accuracy when the amount of data is small, lack of systematicness |

Review of the previous literature shows that there is no single forecasting model that can be considered the best and applied in all cases for wind speed forecasting, as there are considerable differences among the wind speed time series.
Table 2: The abbreviations that appear in this article.

| Nomenclature          | Description                        |
|-----------------------|------------------------------------|
| NNIA                  | Nondominated neighbor immune algorithm |
| VMD                   | Variational mode decomposition      |
| Bi-LSTM               | Bidirectional long short-term memory |
| LSTM                  | Long short-term memory              |
| RNN                   | Recurrent neural network            |
| IMFs                  | Intrinsic mode functions            |
| GRNN                  | Generalized regression neural network |
| BPNN                  | Backpropagation neural network      |
| ARMA                  | Autoregressive moving average model |
| ES                    | Exponential smoothing               |
| EMD                   | Empirical mode decomposition        |
| CEEMD                 | Complementary ensemble empirical mode decomposition |
| SSA                   | Singular spectrum analysis          |
| ELM                   | Extreme learning machine neural network |
| SVM                   | Support vector machine              |
| NWP                   | Numerical weather prediction        |
| SSA                   | Generalized autoregressive conditional heteroskedasticity |
| GARCH                 | Generalized autoregressive conditional |
| LUBE                  | Lower upper bound estimation        |
| ANN                   | Artificial neural network           |
| PSO                   | Particle swarm optimization         |
| GA                    | The genetic algorithm               |
| MAE                   | The average of $N$ absolute percentage error |
| MAPE                  | The absolute error of $N$ forecasting results |
| RMSE                  | The square root of the average of the error squares |
| MSE                   | Mean squared error                  |
| DC                    | Directional change                  |
| FICP                  | Forecast interval coverage probability |
| FINAW                 | Forecast interval normalized average width |
| AWD                   | Accumulated width deviation of testing dataset |
| WACD                  | Weight adaptive combined denoising based models |

Advantages of the developed forecasting system from multiple perspectives, we constructed three different types of contrast models, which were single prediction models, nonoptimized models with different data preprocessing methods, and hybrid models with the same data preprocessing and internal structure optimized by the same optimization algorithm.

**Step 3. Point forecasting and interval forecasting:**
For point forecasting, the proposed combined system combined four different types of forecasting models to provide high-precision point-prediction results. The objective function is as follows:

$$\text{objective 1:} \min \text{ (MAE)},$$

$$\text{objective 2:} \min \text{ (MAPE)},$$

$$f(x) = \omega_1 \cdot y_{i1} + \omega_2 \cdot y_{i2} + \cdots + \omega_n \cdot y_{in},$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,$$

$$\text{MAPE} = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

where $\hat{y}_i$ is the $i$th forecasting data, $y_{ij}$ is the $i$th forecasting data of $j$th model, and $y_i$ is the $i$th actual data. The constraint for the weighting parameters $\omega$ is $[0, 1]$.

On the basis of the point forecasting, we utilized the fuzzy c-means clustering (FCM) algorithm and optimization algorithm to build an appropriate forecast interval. As the value of the wind speed will affect the construction of the wind speed prediction interval, FCM was used to cluster the wind speed data into 10 categories and, then, to construct the prediction interval for each category. After that, NNIA was employed to search for a local optimal of $\omega$, which is the coefficient of the interval, and the same data have the same $\omega$. From the optimization point of view, a forecast interval coverage probability (FICP) higher than AWD and a lower FINRW are the two objectives for high-quality interval forecasting. Therefore, the primary problem for NNIA is modelled as a constrained single-objective problem:

$$\text{objective:} \min \text{ FINAW (}\omega\text{)},$$

$$\text{constraints:} \text{ FINAW (}\omega\text{)} > 0, \mu \text{ FICP (}\omega\text{)} \leq 100\%,$$

where $\mu$ is the nominal confidence level.

Finally, the complete prediction interval is constructed.

**Step 4. Establishment of developed forecasting system:**
The preprocessed wind speed sequence was divided into two parts: the training set and test set. The best
ratio, 4:1, was determined from many experiments; therefore, in the dataset with a total length of 1440, the first 1152 data were used to train the model’s structure and the remaining 288 data were input as the test set. After obtaining the fitting results of a single model, a new weight determination technique, NNIA, was employed to integrate these individual models and obtain the final prediction results.

2.2. The Related Methods. This part introduces the implementation process of several algorithms related to this paper.

2.2.1. Data Pretreatment Methodology. For each signal, the correlation analysis signal was calculated using the Hilbert transform to obtain a unilateral frequency spectrum. The Hilbert transform process can be expressed as the Cauchy principal value (denoted \(\text{p.v.}\)) of the convolution integral [28]:

\[
\text{Hilbertf}(t) = \frac{1}{\pi} \text{p.v.} \int_{-\infty}^{\infty} \frac{f(v)}{t-v} dv.
\]  

Next, the bandwidth was calculated according to the Gaussian smoothness of the demodulated signal, i.e., the square gradient. Then, the decomposition problem was transformed into a constrained variational problem:
\[
\begin{align*}
\min, & \left\{ \sum_k \left[ \delta(t) + \frac{j}{\pi t} \right] \ast \mu_k(t) \right\} e^{-j\kappa t} \right\}^2, \\
\mathrm{s.t.} & \sum_k \mu_k = f_0,
\end{align*}
\]

By alternately updating \( \mu_k, \varphi_k, \) and \( \lambda, \) the minimum point of the extended Lagrangian expression can be obtained. The equation of \( \mu_k, \varphi_k, \) and \( \lambda \) are as follows:

\[
\begin{align*}
\hat{\mu}^{n+1}_k(\varphi) & = f(\varphi) - \sum_{k} \hat{\mu}^{n+1}_k(\varphi) - \sum_{k} \hat{\mu}^{n+1}_k(\lambda^\theta(\varphi))/2, \\
\varphi^{n+1}_k & = \int_{\varphi}^{\infty} \frac{\hat{\mu}^{n+1}_k(\varphi)}{\varphi^{n+1}_k} d\varphi, \\
\tilde{\lambda}^{n+1}(\varphi) & = \tilde{\lambda}^n(\varphi) + \tau \left( \tilde{\lambda}(\varphi) - \sum_k \hat{\mu}^{n+1}_k(\varphi) \right).
\end{align*}
\]

Finally, based on the idea of sequence reconstruction, the processed sequence is expressed in the following form:

\[
\text{signal}_{\text{hit}} = \sum_k w_k \ast \mu_k = w_1 \ast \mu_1 + \cdots + w_k \ast \mu_k,
\]

where the weights, \( w_1, w_2, \ldots, w_k, \) are determined by a multiobjective optimization algorithm, NNIA.

2.2.2. Optimization Theory. To build up a complete combination system, we introduce a novel multiobjective optimization algorithm, nondominated neighbor immune algorithm (NNIA). The algorithm is used to optimize the weights in a combined system, including a data preprocessing module and different single forecasting models. The main steps of NNIA are as follows [30]:

**Step 1.** Initialize the parameters in the model: create one initial antibody population \( B_0 \) with a length of \( N_B. \) When \( t = 0, D_0 = \phi, A_0 = \phi, \) and \( C_0 = \phi. \)

**Step 2.** Update the dominated populations: determine the dominated antibodies in \( B_t. \) Copy the whole dominated antibodies to form the temporary dominated population (expressed as \( D_{T_t+1}. \)) When the size of \( D_{T_t+1} \) is smaller or equal to \( ND, \) set \( D_{t+1} = D_{T_t+1}. \) If not, after calculating all individuals’ crowding distances, according to these values sort the individuals in \( D_{T_t+1} \) in descending order, and then the first \( N_D \) individuals are selected to form \( D_{t+1}. \)

**Step 3.** Determine whether the termination conditions are met: if \( t \geq G_{\text{max}} \) (maximum number of generations), \( D_{t+1} \) is derived as the output and stops; otherwise, \( t = t + 1. \)

**Step 4.** The selection of the nondominant neighbor: if the size of \( D_t \) is smaller or equal to \( N_A \) (maximum size of active population), set \( A_t = D_t. \) If not, after calculating all individuals’ crowding distance, according to these values sort the individuals in \( D_t \) in descending order, and the first \( N_A \) individuals are chosen for forming \( A_t. \)

**Step 5.** Proportional cloning: obtain the clone population \( C_t \) through making proportional cloning to \( A_t. \) The proportional cloning \( C_t \) on the active population \( A = [a_1, a_2, \ldots, a_{|A|}] \) can be expressed as

\[
T_C(a_1 + a_2 + \cdots + a_{|A|}) = T_C(a_1) + T_C(a_2) + \cdots + T_C(a_{|A|})
\]

where \( a_i = a_i + a_i + \cdots + a_i \), \( a_i = \cdots, a_i, i = 1, 2, \ldots, |A|, j = 1, 2, \ldots, q_i, \) and \( q_i \) stands for a self-adaptive parameter. The symbol ‘+’ not only represents the arithmetical operator but also separates the antibodies here. \( q_i = 1 \) means no cloning on antibody \( a_i. \)

In this formula, \( T_C(a_i) = \{a_i^1 + a_i^2 + \cdots + a_i^q\}, \) where \( a_i^j = a_i, j = 1, 2, \ldots, |A|, j = 1, 2, \ldots, q_i, \) and \( q_i \) means no cloning on antibody \( a_i. \)

The greater the crowding-distance value of the individual, the larger of the \( q_i \) of the individual is. The calculation of \( q_i \) is as follows:

\[
q_i = \frac{\zeta(a_i, A)}{\sum_{j=1}^{\lfloor |A|/2 \rfloor} \zeta(a_i, A)}.
\]
where $\zeta (a_j, A)$ is the crowding-distance value of the active antibodies $a_j$ and $n_C$ stands for an expectant value of the size of the clone population.

**Step 6.** Hypermutation and recombination: carry out hypermutation and recombination on $C_i$ and let $C_i$ be the generated population.

**Step 7.** Combine $C_i$ and $D_i$ to obtain the antibody population $B_i$; return to repeat Step 2.

### 2.2.3. The Basic Theory about Submodels in a Combined System

The core forecasting algorithms involved in the combination system are ARIMA, BPNN, GRNN, and BiLSTM. In this section, their basic theories are described briefly [31].

1. **ARIMA.** The model is also known as ARIMA $(p, d, q)$. The ARIMA model has three parameters: $p$, $d$, and $q$.
   - $p$ is the lag number of the data of ARIMA
   - $d$ is the orders that the data are differential $d$ times in order to make the data stable
   - $q$ is the lag number of forecasting error
   ARIMA can be expressed as follows:
   $$y_t = \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j y_{t-j} + \epsilon_t,$$
   where $\epsilon_t$ is the error term at time $t$, $\varphi_i$ is the $j$-th coefficient of AR, and $\theta_j$ is the $i$-th coefficient of MA.

2. **BPNN.** BPNN simulates the activation and transmission of human neurons. Taking a three-layer neural network as an example, the BP neural network consists of three layers: input, hidden, and output layers. The information of each layer of neurons is transmitted to the next layer of neurons through the activation function:
   $$f \text{ (input) } = \frac{1}{1 - e^{-\text{input}}}. \quad (14)$$
   The BP algorithm is used to optimize the weights of the network to keep the mapping between input and output consistent. Then, BP minimizes the objective function by changing the weight of each layer:
   $$F_{\text{objective}} = \frac{1}{2 \times N_{\text{training}}} \sum_{i=1}^{n} \sum_{j=1}^{d_{\text{input}}} (y_{ij} - \tilde{y}_{ij})^2, \quad (15)$$
   where $N$ is the number of training samples, $\tilde{y}_{ij}$ is the forecasting value, and $y_{ij}$ is the actual value.
   Sigmoid function is used as neuron transfer function. No matter how complex the network structure is, each parameter can be evaluated by calculating the gradient. This is the basic idea of multilayer perceptron backpropagation algorithm. The sigmoid function is as follows:
   $$(3) \text{ GRNN.} \text{ The learning speed and nonlinear fitting capability of GRNN are strong. Even if the sample size is small, the forecasting effect of GRNN is accurate. Furthermore, the instibility data can be forecasted by GRNN. Generally, the GRNN can be established by linear neurons and radial basis function neurons.}$$
   GRNN consists of four layers: output layer, mode layer, sum layer, and input layer. The activation function of GRNN can be summarized as
   $$p_i = \exp \left[ -\frac{(\text{input} - \mu_i)^T (\text{input} - \mu_i)}{2\sigma^2} \right]. \quad (16)$$

4. **BiLSTM.** The LSTM was firstly proposed by Hochreiter and Schmidhuber [32], and they have been refined and promoted by many others. BiLSTM is a kind of special RNN network, which can learn the dependence between long sequences. BiLSTM has been widely used in function fitting. The BiLSTM model has two LSTM cells, a hidden layer, and an output layer. The forward and backward LSTM cells can learn the依赖 between long sequences. BiLSTM has been widely used in function fitting. The BiLSTM model has two LSTM cells, a hidden layer, and an output layer. The forward and backward LSTM cells can learn the dependence between long sequences. BiLSTM has been widely used in function fitting. The BiLSTM model has two LSTM cells, a hidden layer, and an output layer. The forward and backward LSTM cells can learn the dependence between long sequences.

### 3. Experiments and Analysis

We adopted three experiments and the associated analyses to verify the forecasting capacity of the proposed model. The data sources, evaluation metrics, and the experiments are also introduced in this section.

#### 3.1. Study Area Description and Datasets

Original wind speed sequences were collected from Sotavento, which is located in the southwest of Europe, in Galicia, Spain. Sotavento is highlighted as an example of good practice by UNESCO in 2014. The time intervals were 10 minutes and one hour. The preprocessed wind speed sequence was divided into two parts: the training set and test set. The best ratio, $4:1$, was determined from many experiments; therefore, in the dataset with a total length of 1440, the first 1152 data were used to train the model’s structure and the remaining 288 data were input as the test set. Dataset 1 and dataset 2 contained 10-minute intervals; dataset 3 contained one-hour intervals. The data structure of the three sites is listed in Table 3.

#### 3.2. Evaluation Metrics

After obtaining the prediction results, there must be a clear and unified standard to evaluate the prediction ability of the model. Therefore, we established a comprehensive evaluation system composed of seven indexes and evaluated the prediction ability from two aspects of point prediction performance and interval precision. In this paper, MAE, MAPE, MSE, RMSE, and DC were selected
to reflect the prediction error. The FICP, FINAW, and AWD represent the interval prediction capability. Table 4 shows the specific information regarding these indexes.

$$AWD_i = \frac{(\text{forecast} - U_i)}{(U_i - L_i)}$$

3.3. Experiment Design. According to the obtained wind speed time series, three experiments were conducted to make clear the comparison between the proposed forecasting system and other models. The input was the wind speed value of the first four days, and the output was the wind speed value of the fifth day.

3.3.1. The Related Parameters of the MATLAB Operation of the Proposed Forecasting Model. In this part, the relevant parameters of each model used in this paper are discussed to achieve the best prediction effect.

In the VMD method, there are six parameters that must be set to achieve a better denoising effect. Alpha is the balancing parameter of the data fidelity constraint, and we set it to 2000; we chose 0 for the noise slack (tau); and the number of modes K to be recovered was 10 in this paper. We did not require the DC part; therefore, the DC was 0. All omegas started at 0; therefore, init = 0. The tolerance of the convergence criterion (tol) was $10^{-7}$.

As for NNIA, the population size, analog binary cross parameters, polynomial variation parameter, frequency minimum, frequency maximum, and iterations were the parameters that were required to be set. A small range of changes in one of the parameters has little effect on the accuracy; however, the program run time is quite different. Thus, considering the two factors, the default parameters were set up as follows: the population size was 10, analog binary cross parameter was 0.25, polynomial variation parameter was 0.5, frequency minimum was 0, frequency maximum was 2, and iterations were 100. The weights of IMFs of dataset 1 calculated by NNIA of the Sotavento wind farm were 1.00, 1.00, 1.00, 1.00, 1.00, 0.79, 0.53, 0.08, and 0.00.

The parameters of BiLSTM were as follows: the learning method was deep learning; the maximum number of training epochs was 300; the minimum batch size was 120; the count of hidden layers in BiLSTM unit one was 250; the number of hidden layers in the feedforward network was 20–50; and the execution environment was CPU.

3.3.2. Experiment I: Comparison with Other Weight Adaptive Combined Denoising (WACD)-Based Models. In this experiment, a group of hybrid models that were also preprocessed by weight adaptive combined denoising (WACD) was selected as benchmark models. The experimental results are exhibited in Figure 2 and Table 5. The following results can be drawn.

In both one-step and two-step forecasting for dataset 1, the developed system obtained a remarkable performance. For example, the developed system achieved the smallest MAE values at 0.1648 and 0.2435, which were better than the other hybrid models. As for the DC, the correct directional change was 92.0139%, which proved that the forecasting of trend of the proposed model was very accurate. For interval prediction, FICP and FINAW cannot be analyzed separately. If the FICP is large but the interval normalized average width is bigger at the same time, the model is bad. According to Table 5, when the FINAW is similar to other models, the FICP of the proposed model is bigger than the WACD-ARIMA, WACD-BPNN, WACD-GRNN, and WACD-BILSTM. In addition, the accumulated width deviation of the proposed model was 0.0193, which is smaller than the other models. Dataset 2 was similar to dataset 1. Therefore, we concluded that the forecasting effect of the developed model was better than other models in both point forecasting and interval forecasting for 10-minute wind speed data.

For the 1-hour wind speed data (dataset 3), the WACD-NNIA-EBGB system still provided more distinguished prediction performance, not only in one step but also in two steps. For instance, for two-step prediction, our proposed system with the smallest errors was the WACD-BPNN, WACD-GRNN, and WACD-BILSTM. In addition, the accumulated width deviation of the proposed model was 0.0193, which is smaller than the other models. Dataset 2 was similar to dataset 1. Therefore, we concluded that the forecasting effect of the developed model was better than other models in both point forecasting and interval forecasting for 10-minute wind speed data.

Remark 1. According to Figure 2, compared with the mixed model, there is clear evidence that the developed prediction system with the lowest statistical error values in both point and interval prediction had the optimal prediction performance. In addition, other assessment indicator values were also the most ideal, which fully proves the feasibility of this combination system in short-term wind speed prediction.

3.3.3. Experiment II: Comparison with Models with Different Data Preprocessing Technologies. The aim of experiment II was to make clear the comparison situation between several common data preprocessing technologies, namely, EMD, CEEMD, SSA, and VMD and the developed pretreatment strategy. The assessment indicator values are listed in Table 6 and Figure 3. The prediction results are shown as Figure 4.

For Set 1, the WACD-based combination system was more successful in processing noise data, as observed from all indicator values. In addition, this strategy obtained higher
Table 4: The indexes for evaluating the prediction ability.

| Metric   | Definition                                      | Equation                                                                 |
|----------|-------------------------------------------------|--------------------------------------------------------------------------|
| MAPE     | The average of N absolute percentage error       | $MSE = \frac{\sum_{i=1}^{N} |(A_i - F_i)|/A_i \times 100%}{N}$               |
| MAE      | The mean absolute error of N forecasting results | $MSE = \frac{\sum_{i=1}^{N} |(A_i - F_i)|}{N}$                                |
| RMSE     | The square root of the average of the error squares | $RMSE = \sqrt{\frac{\sum_{i=1}^{N} ((A_i - F_i)^2)}{N}}$                  |
| MSE      | Mean squared error                               | $MSE = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N}$                       |
| DC       | Directional change                               | $DC = \frac{(100/N - 1)\sum_{i=1}^{N-1} a_i}{N}$, $a_t = \begin{cases} 0, & \text{otherwise}, \\ 1, & \text{if } ((y(t+1) - y(t)) (\hat{y}(t+1) - y(t))) > 0, \end{cases}$ |
| FICP     | Forecast interval coverage probability           | $FICP = \frac{\sum_{i=1}^{N} c_i \times 100\%}{N}$                       |
| FINAW    | Forecast interval normalized average width       | $FINAW = \frac{\sum_{i=1}^{N} (U_i - L_i) \times 100\%}{N \times [\max (\text{data}) - \min (\text{data})]}$ |
| AWD      | Accumulated width deviation of testing dataset   | $AWD = \frac{\sum_{i=1}^{N} AWD_i}{N}$                                       |

In the formulas, $L_i$ and $U_i$ represent the lower and the upper limit of the production interval, respectively. And, $c_i$ is the number of the truth value contained in constructed interval. $N$ is the number of the testing set [16]. Data are the actual data. Besides, $AWD_i$ is the width deviation of the construction interval of each sample.

Figure 2: Experiment I: comparison with other WACD-based models.
| Dataset | Model           | One step |   |   |   |   |   |   | Two steps |   |   |   |   |   |   |
|---------|----------------|----------|---|---|---|---|---|---|-----------|---|---|---|---|---|---|
|         | MAE            | MAPE     | RMSE | MSE | DC | FICP | FINAW | AWD | MAE       | MAPE | RMSE | MSE | DC | FICP | FINAW | AWD |
| Set 1   | WACD-ARIMA     | 0.1662   | 3.9433 | 0.2296 | 0.0527 | 90.7850 | 89.4558 | 0.5469 | 12.9943 | 0.7321 | 0.5359 | 48.6301 | 53.5836 | 0.1074 | 0.3382 |
|         | WACD-BPNN      | 0.1959   | 4.7382 | 0.2716 | 0.0737 | 85.6655 | 87.0748 | 0.4073 | 9.8769 | 0.5509 | 0.3035 | 56.1644 | 61.7747 | 0.1049 | 0.2207 |
|         | WACD-GRNN      | 0.3930   | 9.5245 | 0.5161 | 0.2663 | 63.4812 | 55.1020 | 0.5288 | 12.8015 | 0.6925 | 0.4796 | 59.2466 | 42.6621 | 0.0836 | 0.5223 |
|         | WACD-BILSTM    | 0.3165   | 7.0425 | 0.4150 | 0.1722 | 69.9659 | 65.9864 | 0.3312 | 7.9995 | 0.4088 | 0.1671 | 60.0683 | 66.6667 | 0.1060 | 0.0923 |
|         | Proposed model | 0.1648   | 3.8354 | 0.2284 | 0.0527 | 90.7850 | 89.4558 | 0.5469 | 12.9943 | 0.7321 | 0.5359 | 48.6301 | 53.5836 | 0.1074 | 0.3382 |
| Set 2   | WACD-ARIMA     | 0.2273   | 7.3885 | 0.3126 | 0.0977 | 88.3959 | 90.1361 | 0.7532 | 24.1722 | 1.0796 | 1.1655 | 46.9178 | 51.1945 | 0.1373 | 0.2955 |
|         | WACD-BPNN      | 0.3149   | 12.3639 | 0.4125 | 0.1702 | 80.5461 | 80.9524 | 0.6529 | 26.4495 | 0.8918 | 0.7954 | 56.1644 | 52.9010 | 0.1255 | 0.3144 |
|         | WACD-GRNN      | 0.5112   | 17.3599 | 0.5687 | 0.5623 | 69.7279 | 75.2455 | 0.7939 | 32.0372 | 1.0743 | 1.1542 | 54.4521 | 51.7406 | 0.0849 | 1.4376 |
|         | WACD-BILSTM    | 0.4272   | 12.8502 | 0.5800 | 0.3363 | 70.6485 | 72.1088 | 0.4335 | 14.7300 | 0.5962 | 0.3555 | 65.8703 | 72.1088 | 0.1282 | 0.1073 |
|         | Proposed model | 0.2159   | 6.94987 | 0.3065 | 0.0937 | 90.2778 | 94.2177 | 0.3974 | 11.6754 | 0.5477 | 0.2999 | 80.4795 | 81.5700 | 0.1987 | 0.0533 |
| Set 3   | WACD-ARIMA     | 0.3021   | 8.7753 | 0.3828 | 0.1465 | 87.7133 | 84.6939 | 0.8111 | 25.1829 | 1.1431 | 1.3067 | 49.4916 | 54.9128 | 0.2216 | 0.2982 |
|         | WACD-BPNN      | 0.3111   | 9.1970 | 0.3962 | 0.1570 | 87.7133 | 84.6939 | 0.5495 | 18.4000 | 0.7313 | 0.5347 | 75.6849 | 70.3072 | 0.2183 | 0.1445 |
|         | WACD-GRNN      | 0.7059   | 18.9352 | 0.8934 | 0.7981 | 62.7986 | 46.9388 | 0.9215 | 25.7201 | 1.1458 | 1.3129 | 57.5342 | 33.1058 | 0.1515 | 0.5364 |
|         | WACD-BILSTM    | 0.4087   | 10.8551 | 0.5195 | 0.2698 | 82.2526 | 74.4898 | 0.4353 | 11.8373 | 0.5721 | 0.3272 | 76.7918 | 73.1292 | 0.1872 | 0.0900 |
|         | Proposed model | 0.2946   | 8.2657 | 0.3794 | 0.1456 | 89.9305 | 94.5578 | 0.3655 | 9.6888 | 0.4736 | 0.2243 | 85.6164 | 88.7372 | 0.2349 | 0.0188 |

The most notable results are bolded. **WACD** is weight adaptive combined denoising. **MAPE** is the average of absolute percentage error. **MAE** is the mean absolute error of forecasting results. **RMSE** is the square root of the average of the errors squares. **MSE** is mean squared error. **DC** is directional change. **FICP** is forecast interval coverage probability. **FINAW** is forecast interval normalized average width. **AWD** is accumulated width deviation of testing dataset.

---

**Table 5:** Prediction performance comparison table of the proposed model and hybrid models using the same data pretreatment technique.
| Dataset | Model           | One step | Two steps |
|---------|----------------|----------|-----------|
|         | MAE | MAPE | RMSE | MSE | DC | FICP | FINAW | AWD | MAE | MAPE | RMSE | MSE | DC | FICP | FINAW | AWD |
| Set 1   | EMD-based model | 0.2388 | 5.4728 | 0.3147 | 0.0991 | 80.2048 | 80.9524 | 0.1164 | 0.0538 | 0.3351 | 7.6773 | 0.4486 | 0.2012 | 79.7945 | 69.9659 | 0.1170 | 0.1303 |
|         | CEEMD-based model | 0.2396 | 5.7345 | 0.3134 | 0.0982 | 80.8874 | 81.6327 | 0.1120 | 0.0472 | 0.4077 | 9.6214 | 0.5311 | 0.2821 | 75.3425 | 60.7509 | 0.1133 | 0.1796 |
|         | SSA-based model | 0.3933 | 9.1076 | 0.5015 | 0.2515 | 60.7509 | 60.5442 | 0.1311 | 0.1425 | 0.3958 | 9.1705 | 0.5024 | 0.2524 | 60.9589 | 60.4096 | 0.1136 | 0.1449 |
|         | VMD-based model | 0.2379 | 5.5063 | 0.3225 | 0.0400 | 83.2765 | 82.6531 | 0.1217 | 0.0548 | 0.5132 | 12.0669 | 0.6844 | 0.4684 | 68.1507 | 52.9010 | 0.1342 | 0.2736 |
|         | Proposed model  | 0.1648 | 3.8354 | 0.2284 | 0.0498 | 92.0139 | 93.4028 | 0.0984 | 0.0193 | 0.2435 | 5.6371 | 0.3329 | 0.1109 | 85.9589 | 83.2765 | 0.1284 | 0.0494 |
| Set 2   | EMD-based model | 0.3594 | 10.3580 | 0.5560 | 0.3091 | 76.1092 | 83.6735 | 0.1977 | 0.0486 | 0.5209 | 15.2106 | 0.7310 | 0.5343 | 71.5753 | 69.9659 | 0.2342 | 0.0990 |
|         | CEEMD-based model | 0.3295 | 10.4210 | 0.4697 | 0.2206 | 78.8396 | 88.0952 | 0.1810 | 0.0335 | 0.5969 | 18.6063 | 0.8420 | 0.7090 | 68.4932 | 67.5768 | 0.1875 | 0.1524 |
|         | SSA-based model | 0.5030 | 14.5762 | 0.7126 | 0.5078 | 60.4096 | 70.4082 | 0.2208 | 0.0983 | 0.5120 | 14.9608 | 0.7470 | 0.5580 | 71.2329 | 69.2833 | 0.2216 | 0.1105 |
|         | VMD-based model | 0.3715 | 11.0545 | 0.5239 | 0.2745 | 79.5222 | 84.0136 | 0.1806 | 0.0550 | 0.7715 | 23.4133 | 1.0700 | 1.1450 | 62.6712 | 51.8771 | 0.2005 | 0.2378 |
|         | Proposed model  | 0.2159 | 6.94987 | 0.3065 | 0.0937 | 90.2778 | 94.2177 | 0.1260 | 0.0128 | 0.3974 | 11.6754 | 0.5477 | 0.2999 | 80.4795 | 81.5700 | 0.1987 | 0.0533 |
| Set 3   | EMD-based model | 0.4462 | 12.3784 | 0.5637 | 0.3177 | 76.4505 | 77.5510 | 0.2071 | 0.0427 | 0.5967 | 15.9707 | 0.7418 | 0.5503 | 74.3151 | 68.9420 | 0.2025 | 0.0716 |
|         | CEEMD-based model | 0.4004 | 10.9472 | 0.5195 | 0.2699 | 81.5700 | 83.3333 | 0.2014 | 0.0297 | 0.6858 | 18.8034 | 0.8687 | 0.7546 | 76.3699 | 60.7509 | 0.2084 | 0.1217 |
|         | SSA-based model | 1.0028 | 26.9696 | 1.2404 | 1.5387 | 56.3140 | 44.2177 | 0.3372 | 0.2572 | 1.0115 | 26.8555 | 1.2534 | 1.5711 | 56.8493 | 44.3686 | 0.3550 | 0.2524 |
|         | VMD-based model | 0.3265 | 8.8089 | 0.4150 | 0.1722 | 87.7133 | 91.5651 | 0.2108 | 0.0141 | 0.9349 | 25.2413 | 1.1851 | 1.4044 | 65.0685 | 48.1229 | 0.2513 | 0.2248 |
|         | Proposed model  | 0.2946 | 8.2657 | 0.3794 | 0.1456 | 89.9305 | 94.5578 | 0.1860 | 0.0072 | 0.3655 | 9.6888 | 0.4736 | 0.2243 | 85.6164 | 88.7372 | 0.2349 | 0.0188 |

The most notable results are bolded. MAPE is the average of absolute percentage error. MAE is the mean absolute error of forecasting results. RMSE is the square root of the average of the error squares. MSE is mean squared error. DC is directional change. FICP is forecast interval coverage probability. FINAW is forecast interval normalized average width. AWD is accumulated width deviation of testing dataset.
values, meaning that after removing the interference, the forecasting models better fit the changes of wind speed. The denoising effect of EMD technology followed the developed system, with MAPE values of 5.4278%, and 7.6773%, respectively. Moreover, when the number of prediction steps varied, the prediction ability of CEEMD-based and VMD-based combined models undulated greatly. When the FINAW of all models was almost the same, the forecast interval coverage probability of the proposed model was approximately 10% more than the other models in one-step forecasting. The AWD of the proposed model was 0.0345 smaller than the EMD-based model.

For Set 2, the results showed that the prediction errors obtained by the WACD-based combined system under 7% were smaller than the other data that were pretreatment-based. Taking one-step prediction as example, the developed system’s RMSE was 0.3065, improving 30.40% and 54.12%, respectively, relative to the values of the CEEMD-based (0.4697) and SSA-based prediction systems (0.7126).

In Set 3, for both one-step and multistep prediction, any evaluation index of the prediction system put forward in this paper was at the optimal level. More specifically, under the one-step forecasting situation, the WACD-based system had the best MAE, MAPE, RMSE, MSE, and DC at 0.2946, 8.2657%, 0.3794, 0.1456, and 89.9305, respectively. Nevertheless, the prediction performance of the SSA-based system was the worst with the biggest MAPE at 26.9696%.

**Remark 2.** Different from the recombination idea of the denoising layer number based on subjective judgment, each subsequence obtained though decomposition was weighted and reconstructed to eliminate the adverse influence of high frequency noise. The results of experiment II show that the combination system based on WACD was always better than the other preprocessing technologies, which indicates that the idea of adaptive combination reconstruction of weight has great potential in the future.

### 3.3.4. Experiment III: Comparison with Traditional Forecasting Models

This experiment highlighted the contrast...
that exists between the prediction capability of the developed system and several traditional prediction methods, including ARIMA, BPNN, ELMNN, GRNN, and ELMAN. The detailed numerical values are shown in Table 7 and Figure 4.

For Set 1, although the assessment index values were changed with the number of prediction steps, the point prediction and interval prediction performance of the proposed combination system was always optimal among all models. This phenomenon reveals that the combined system put forward in this study not only had the smallest prediction error but also achieved the highest stability.

For Set 2, the RMSE values of ARIMA, BPNN, ELMAN, GRNN, and ELM were 1.1928, 1.0728, 1.1324, 1.0922, and 1.0630, respectively, in two-step forecasting. Apart from the worst ELMAN model, there was no significant difference among the prediction results obtained from the other single models; however, they were significantly weaker than the proposed composite system, whose RMSE was 0.5477, even smaller than 50% of the above model.

In Set 3, when comparing the interval forecasting capability, the developed system was still the best relative to the other comparison methods in terms of the values of the predictive performance indexes. The values of FICP and FINAW should be studied together. For instance, with the two-step prediction, the developed system had the smallest FINAW and the biggest FICP, which were almost two-times greater than the single models.

**Remark 3.** Compared with the single models, the advantages of the combined strategy were outstanding. This conclusion has nothing to do with the selected evaluation index and the number of prediction steps. Therefore, these traditional individual prediction models are no longer suitable for current wind speed prediction.
Table 7: Prediction performance comparison table of the proposed model and classic individual models.

| Dataset | Model   | One step | Two steps |
|---------|---------|----------|-----------|
|         |         | MAE      | MAPE      | RMSE      | MSE | DC       | FICP | FINAW | AWD | MAE      | MAPE      | RMSE      | MSE | DC       | FICP | FINAW | AWD |
| Set 1   | ARIMA   | 0.4011   | 9.5513    | 0.5344    | 0.2856    | 53.9249 | 63.9456 | 0.1042 | 0.1969 | 0.6462 | 15.7336   | 0.8202    | 0.6727    | 53.0822 | 41.9795 | 0.1042 | 0.4227 |
|         | BPNN    | 0.4065   | 9.7395    | 0.5467    | 0.2989    | 53.9249 | 56.1224 | 0.1042 | 0.2433 | 0.5395 | 13.4091   | 0.7108    | 0.5052    | 55.1370 | 47.0990 | 0.0978 | 0.4435 |
|         | ELMAN   | 0.5378   | 12.9708   | 0.4662    | 0.2682    | 51.5358 | 44.2177 | 0.1002 | 0.4116 | 0.6125 | 14.8276   | 0.7740    | 0.5990    | 51.7123 | 41.9795 | 0.1052 | 0.4647 |
|         | GRNN    | 0.4559   | 11.4420   | 0.5919    | 0.3504    | 55.2901 | 46.9388 | 0.0941 | 0.3536 | 0.5698 | 14.4080   | 0.7385    | 0.5454    | 56.8493 | 38.2253 | 0.0887 | 0.6553 |
|         | ELMNN   | 0.3976   | 9.6203    | 0.5314    | 0.2824    | 59.3857 | 59.5238 | 0.1015 | 0.2135 | 0.5391 | 13.3873   | 0.6952    | 0.4833    | 51.0274 | 44.3686 | 0.0967 | 0.4563 |
|         | Proposed | 0.1648   | 3.8354    | 0.2284    | 0.0498    | 92.0139 | 93.4028 | 0.0984 | 0.0193 | 0.2435 | 5.6317    | 0.3329    | 0.1109    | 85.9589 | 83.2765 | 0.1284 | 0.0494 |
| Set 2   | ARIMA   | 0.5104   | 16.0071   | 0.7449    | 0.5549    | 53.9249 | 71.7687 | 0.1348 | 0.1530 | 0.8363 | 27.6248   | 1.1928    | 1.4228    | 52.0548 | 44.0273 | 0.1169 | 0.4541 |
|         | BPNN    | 0.6360   | 26.6598   | 0.8468    | 0.7171    | 54.6075 | 48.6395 | 0.1221 | 0.3209 | 0.7450 | 27.5824   | 1.0728    | 1.1509    | 53.0822 | 47.4403 | 0.1265 | 0.4329 |
|         | ELMAN   | 0.6979   | 22.7232   | 0.9921    | 0.9843    | 53.5836 | 50.0000 | 0.1175 | 0.3213 | 0.7994 | 26.1353   | 1.1322    | 1.2819    | 53.4247 | 46.0751 | 0.1178 | 0.4054 |
|         | GRNN    | 0.6720   | 27.8385   | 0.9280    | 0.8612    | 54.6075 | 40.8163 | 0.0896 | 0.6960 | 0.7864 | 30.3558   | 1.0922    | 1.1928    | 55.1370 | 34.1297 | 0.0912 | 0.9722 |
|         | ELMNN   | 0.4948   | 16.0898   | 0.7033    | 0.4946    | 68.6007 | 61.2245 | 0.1069 | 0.2164 | 0.7516 | 28.8463   | 1.0630    | 1.1299    | 54.7945 | 42.6621 | 0.1207 | 0.4272 |
|         | Proposed | 0.2159   | 6.94987   | 0.3065    | 0.0937    | 90.2778 | 94.2177 | 0.1260 | 0.0128 | 0.3974 | 11.6754   | 0.5477    | 0.2999    | 80.4795 | 81.5700 | 0.1987 | 0.0533 |
| Set 3   | ARIMA   | 0.7076   | 20.4992   | 0.9131    | 0.8388    | 59.3857 | 59.5238 | 0.2155 | 0.2166 | 1.1537 | 33.8070   | 1.4614    | 2.1538    | 53.7671 | 37.8840 | 0.1978 | 0.4784 |
|         | BPNN    | 0.7149   | 20.3467   | 0.9188    | 0.8442    | 56.3140 | 57.6830 | 0.2029 | 0.2235 | 0.9870 | 28.4973   | 1.2463    | 1.5532    | 53.7671 | 39.2491 | 0.1831 | 0.4745 |
|         | ELMAN   | 0.9611   | 27.7013   | 1.2250    | 1.5005    | 56.6563 | 42.8571 | 0.1853 | 0.4171 | 1.3019 | 37.0247   | 1.6473    | 2.7137    | 57.1918 | 40.6143 | 0.2449 | 0.4282 |
|         | GRNN    | 0.7658   | 23.3317   | 0.9823    | 0.9684    | 58.7081 | 48.9796 | 0.1773 | 0.3137 | 1.0044 | 31.0281   | 1.2558    | 1.5771    | 57.1918 | 33.7884 | 0.1629 | 0.6429 |
|         | ELMNN   | 0.6912   | 20.5330   | 0.9017    | 0.8131    | 61.0922 | 60.2041 | 0.1983 | 0.2272 | 0.9731 | 29.5352   | 1.2352    | 1.5257    | 54.1096 | 40.2730 | 0.1863 | 0.4719 |
|         | Proposed | 0.2946   | 8.2657    | 0.3794    | 0.1456    | 89.9305 | 94.5578 | 0.1860 | 0.0072 | 0.3655 | 9.6888    | 0.4736    | 0.2243    | 85.6164 | 88.7372 | 0.2349 | 0.0188 |

The most notable results are bolded. MAPE is the average of absolute percentage error. MAE is the mean absolute error of forecasting results. RMSE is the square root of the average of the error squares. MSE is mean squared error. DC is directional change. FICP is forecast interval coverage probability. FINAW is forecast interval normalized average width. AWD is accumulated width deviation of testing dataset.
4. Conclusions

Wind energy, as one of the fastest growing natural resources, is gradually occupying the leading position in energy structures. Wind speed forecasting, as the basis of wind energy utilization, has become a research focus [21].

In this paper, an advanced combination system, which integrated the fortes of data preprocessing techniques and multiobjective optimization theory, was successfully proposed. The system achieved great improvements regarding both the prediction accuracy and stationarity of the previous models. To expand, the proposed forecasting system overcame the defects exposed in previous studies in the following aspects. (1) In contrast to the recombination idea of the denoising layer number based on subjective judgment, each subsequence obtained though decomposition was weighted and reconstructed to eliminate the adverse influence of high frequency noise. (2) One widespread applied statistical method, two traditional neural networks, and one deep learning model were used as base models for fully capturing all important characteristics in wind speed series. (3) An advanced multi-objective optimization algorithm, NNIA, was successfully applied in the forecasting system to decide the weights in the data preprocessing link and combined forecasting link.

Empirical studies on the basis of realistic wind speed data indicated that the developed WACD-NNIA-ABGB system was clearly superior to other benchmark models. (1) Experiment I made a comparison between the proposed system and hybrid models with the same data pretreatment technique, and the results showed that the developed system obtained the lowest prediction error, with the MAPE mean values at 4.7405%, 9.2661%, and 8.9389%, which were reduced by more than 50%, compared with other hybrid models. (2) When contrasted with the models with diverse data preprocessing techniques, similar results indicated that the forecasting ability assessment values were still better than the other comparative models, revealing that the proposed decomposition and weighted reconstruction strategy surpassed the previous strategies and was a more effective tool. (3) In experiment III, the proposed forecasting system was compared with several traditional algorithms. According to the experiment results, we observed that regardless of the sites, the prediction performance of our system always ranked first, indicating that the combination theory was effective and successful. (4) Further discussions indicated that the developed system realized higher precision and stability relative to the other models. In conclusion, it is evident that the prediction system proposed in this paper shows great application potential and will be a reliable tool in smart grids.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

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

Acknowledgments

This study was supported by the Natural Science Foundation of Liaoning Province (No. 2019-MS-109) and HSSF of Chinese Ministry of Education (No. 20YJ A790049).

References

[1] 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, 2020.
[2] Z. Cheng and J. Wang, “A new combined model based on multi-objective salp swarm optimization for wind speed forecasting,” Applied Soft Computing, vol. 92, Article ID 106294, 2020.
[3] X. Niu and J. Wang, “A combined model based on data preprocessing strategy and multi-objective optimization algorithm for short-term wind speed forecasting,” Applied Energy, vol. 241, pp. 519–539, 2019.
[4] 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.
[5] J. Zhao, Y. Guo, X. Xiao, J. Wang, D. Chi, and Z. Guo, “Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method,” Applied Energy, vol. 197, pp. 183–202, 2017.
[6] L. Dong, L. Wang, S. F. Kahlro, S. Gao, and X. Liao, “Wind power day-ahead prediction with cluster analysis of NWP,” Renewable and Sustainable Energy Reviews, vol. 60, pp. 1206–1212, 2016.
[7] T.-H. Yang and C.-C. Tsai, “Using numerical weather model outputs to forecast wind gusts during typhoons,” Journal of Wind Engineering and Industrial Aerodynamics, vol. 188, pp. 247–259, 2019.
[8] 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.
[9] W. Qian and J. Wang, “An improved seasonal GM (1, 1) model based on the hp filter for forecasting wind power generation in china,” Energy, vol. 209, Article ID 118499, 2020.
[10] Y. Chen, S. Zhang, W. Zhang, J. Peng, and Y. Cai, “Multifactor spatio-temporal correlation model based on a combination of convolutional neural network and long short-term memory neural network for wind speed forecasting,” Energy Conversion and Management, vol. 185, pp. 783–799, 2019.
[11] C.-H. Cheng and L.-Y. Wei, “A novel time-series model based on empirical mode decomposition for forecasting TAIEX,” Economic Modelling, vol. 36, pp. 136–141, 2014.
[12] N. Ramesh Babu and B. Jagan Mohan, “Fault classification in power systems using EMD and SVM,” Ain Shams Engineering Journal, vol. 8, no. 2, p. 103, 2017.
[13] H. Liu, H.-Q. Tian, D.-F. Pan, and Y.-F. Li, “Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks,” Applied Energy, vol. 107, pp. 191–208, 2013.
[14] 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.
[15] R. Blonbou, “Very short-term wind power forecasting with neural networks and adaptive Bayesian learning,” Renewable Energy, vol. 36, no. 3, pp. 1118–1124, 2011.
[16] 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.

[17] S. Chenthur Pandian, K. Duraiswamy, C. Christober Asir Rajan, and N. Kanagaraj, "Fuzzy approach for short term load forecasting," Electric Power Systems Research, vol. 76, no. 6-7, pp. 541–548, 2006.

[18] Z.-H. Guo, J. Wu, H.-Y. Lu, and J.-Z. Wang, "A case study on a hybrid wind speed forecasting method using BP neural network," Knowledge-Based Systems, vol. 24, no. 7, pp. 1048–1056, 2011.

[19] 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.

[20] W. Yang, J. Wang, T. Niu, and P. Du, "A novel system for multi-step electricity price forecasting for electricity market management," Applied Soft Computing, vol. 88, Article ID 106029, 2020.

[21] T. Niu, J. Wang, H. Lu, W. Yang, and P. Du, "Developing a deep learning framework with two-stage feature selection for multivariate financial time series forecasting," Expert Systems with Applications, vol. 148, Article ID 113237, 2020.

[22] A. Khosravi, S. Nahavandi, and D. Creighton, "A neural network-GARCH-based method for construction of Prediction Intervals," Electric Power Systems Research, vol. 96, pp. 185–193, 2013.

[23] 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.

[24] Bootstrapping and Resampling. <http://www.ncl.ucar.edu/Applications/bootstrap.shtml>.

[25] 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.

[26] 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.

[27] J. M. Bates and C. W. J. Granger, "The combination of forecasts," Journal of the Operational Research Society, vol. 20, no. 4, pp. 451–468, 1969.

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

[29] Y. Zhang, Y. Zhao, C. Kong, and B. Chen, "A new prediction method based on VMD-PRBF-ARMA-E model considering wind speed characteristic," Energy Conversion and Management, vol. 203, Article ID 112254, 2020.

[30] M. Gong, L. Jiao, H. Du, and L. Bo, "Multiobjective Immune algorithm with nondominated neighbor-based selection," Evolutionary Computation, vol. 16, no. 2, pp. 225–255, 2008.

[31] P. Jiang and Y. Nie, "A hybrid double forecasting system of short term power load based on swarm intelligence and nonlinear integration mechanism," Applied Sciences, vol. 10, no. 4, p. 1550, 2020.

[32] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.