*Article*

# Forecasting Models for Wind Power Using Extreme-Point Symmetric Mode Decomposition and Artificial Neural Networks

**Jianguo Zhou, Xiaolei Xu \*, Xuejing Huo and Yushuo Li**

Department of Economics and Management, North China Electric Power University, 689 Huadian Road, Baoding 071000, China; dldxzjg@126.com (J.Z.); 2172218069@ncepu.edu.cn (X.H.); 2172218019@ncepu.edu.cn (Y.L.)

**\*** Correspondence: 2172218036@ncepu.edu.cn

check for updates

**Abstract:** The randomness and volatility of wind power poses a serious threat to the stability, continuity, and adjustability of the power system when it is connected to the grid. Accurate short-term wind power prediction methods have important practical value for achieving high-precision prediction of wind farm power generation and safety and economic dispatch. Therefore, this paper proposes a novel combined model to improve the accuracy of short-term wind power prediction, which involves grey correlation degree analysis, ESMD (extreme-point symmetric mode decomposition), sample entropy (SampEn) theory, and a hybrid prediction model based on three prediction algorithms. The meteorological data at different times and altitudes is firstly selected as the influencing factors of wind power. Then, the wind power sub-series obtained by the ESMD method is reconstructed into three wind power characteristic components, namely PHC (high frequency component of wind power), PMC (medium frequency component of wind power), and PLC (low frequency component of wind power). Similarly, the wind speed sub-series obtained by the ESMD method is reconstructed into three wind speed characteristic components, called SHC (high frequency component of wind speed), SMC (medium frequency component of wind speed), and SLC (low frequency component of wind speed). Subsequently, the Bat-BP model, Adaboost-ENN model, and ENN (Elman neural network), which have high forecasting accuracy, are selected to predict PHC, PMC, and PLC, respectively. Finally, the prediction results of three characteristic components are aggregated into the final prediction values of the original wind power series. To evaluate the prediction performance of the proposed combined model, 15-min wind power and meteorological data from the wind farm in China are adopted as case studies. The prediction results show that the combined model shows better performance in short-term wind power prediction compared with other models.

**Keywords:** short-term wind power prediction; extreme-point symmetric mode decomposition; sample entropy theory; combined model

## 1. Introduction

With energy shortage and environmental pollution's further deterioration, the development and utilization of renewable energy is receiving more and more attention from the whole world. Wind power, one of the most promising renewable energy sources, has experienced the most growth in the past several decades from a global perspective [1]. However, wind power generation is characterized by volatility and intermittency, which has a seriously negative influence on the electrical energy quality and the safe and stable operation of a power system [2]. Therefore, it is necessary to

predict wind power accurately to reduce the negative influence of the integration of wind turbines into power systems.

The increasing interest in the integration of wind power plants has heightened the need of accurate wind power prediction methods, which mainly includes physical methods, statistical methods, intelligent methods, and combined models. Physical methods, based on the meteorological and geographical information [3], have higher prediction accuracy when the environment of the wind power plant remains stable. For example, Wu et al. [4] developed a prediction system by combining statistical models and physical models to test the data of the Wattle Point wind farm in Australia and found that the system was effective for predicting the power output of wind farms. Cheng et al. [5] assimilated anemometer wind speed observations from wind farm turbines into a numerical weather forecast system, which can effectively improve the accuracy of wind power and wind speed prediction. The statistical method is to establish a mapping relationship between the input of the Numerical Weather Prediction (NWP) system and the wind power, mainly including the autoregressive (AR) model [6], auto-regressive moving average (ARMA) approach [7], auto-regressive integrated moving average (ARIMA) approach [8], and seasonal autoregressive integrated moving average (SARIMA) model [9], and so forth. Liu et al. [10] used the autoregressive moving average-generalized autoregressive conditional heteroscedasticity (ARMA-GARCH) method to simulate the mean and volatility of wind speed at the observation site in Colorado, USA. The results showed that the method can effectively capture the trend change of the mean and volatility of wind speed. Matallah et al. [11] developed a new wind speed forecasting model by combining the Hammerstein model and an autoregressive model, which reflects good prediction performance.

In addition, intelligent methods have good capabilities, such as a strong nonlinear fitting ability, simple learning rules, and high robustness, and they have been extensively applied by researchers for wind power forecasting. In [12], a least squares support vector machine (LSSVM) model optimized by PSOSA was used to predict wind power, and actual calculation examples demonstrated that the prediction method used in the article has high prediction precision. Wang et al. [13] used the support vector machine (SVM) based on the structure risk minimization principle to predict the short-term wind power. Zhang et al. [14] employed a radial basis function neural network (RBFNN) and a multi-objective optimization method to perform interval forecasting of wind speed and ultimately achieved a higher forecasting precision. The back propagation neural network (BPNN) has a long history in prediction, and it has made outstanding contributions to forecasting, especially in the state of uncovering nonlinearity between the inputs and outputs, even with a lack of sufficient information about the relationship between them [15]. For instance, Wang et al. [16] proposed a wind power range prediction model based on the multiple output property of BPNN. The simulation results of a practical example showed that the proposed wind power range prediction model can effectively forecast the output power interval. Sun et al. [17] developed a novel wind speed prediction model by combining fast ensemble empirical mode decomposition (FEEMD), phase space reconstruction, and improved BPNN, which ultimately obtained a higher forecasting accuracy. BPNN could approximate complex nonlinear functions, but it easily falls into local minima, and often exhibits over-fitting. To overcome the shortcomings of BPNN, some optimization algorithms are applied. Chao et al. [18] introduced a method called IS-PSO-BP that combines PSO-BP with comprehensive parameter selection to predict wind speed, and the experiment results clearly show that the proposed method achieves much a better forecasting performance than the BPNN and ARIMA model. Wang et al. [19] proposed a forecasting method based on improved empirical mode decomposition (EMD) and the GA-BP neural network to the prediction of wind speed on a wind farm in Inner Mongolia, China. The simulation with MATLAB shows that the proposed method can improve the forecasting accuracy and computational efficiency. A novel hybrid forecasting model called E-SA-BP, which combines ensemble empirical mode decomposition (EEMD), a simulated annealing (SA) algorithm, and BPNN, was developed to perform wind speed forecasting in [20]. Although the optimization algorithm mentioned above can improve the prediction performance of BPNN to a certain extent, they still have some defects of

a slow convergence speed, they easily fall into a local optimum, and have a long training process. Therefore, this paper uses the Bat algorithm to update the learning rule and network weights of BPNN. Compared with the optimization algorithm mentioned above, the Bat algorithm has the advantages of fewer parameters and better global optimization ability [21]. Additionally, the mathematical model involved in the Bat algorithm is relatively simple and computationally efficient.

As the recurrent neural network, ENN has been proven to be helpful for the prediction of discrete-time series, contributing to the advantage of modeling nonlinear dynamic systems and learning time-varying patterns [22], which has an abundant application in the prediction field. Zhang et al. [23] introduced a novel EMD–ENN approach to forecast wind speed, and the simulation results show that the proposed approach consistently has the minimum statistical error. Liu et al. [24] presented a novel hybrid model, with FEEMD and wavelet packet decomposition, and ENN, which had a desirable performance in multi-step ahead wind speed forecasting. Yu et al. [25] designed a new hybrid model combining improved wavelet transform (IWT) and ENN, which exhibited satisfactory performance in wind speed forecasting. Although ENN has made many contributions in the field of wind power and wind speed prediction, wind power has intermittent and volatility characteristics. Therefore, based on ENN, the Adaboost algorithm is used to further enhance the prediction ability of ENN for nonlinear, chaotic, and volatile data, which has quickly become a new research hotspot [26]. Liu and Tian et al. [27] applied the multilayer perceptron (MLP) neural networks optimized by the Adaboost algorithm to predict wind speed, and the prediction results show that the Adaboost algorithm has promoted the forecasting performance of MLP neural networks considerably and the Adaboost-MLP model is effective for wind speed predictions. Shao et al. [28] proposed a novel solution using the AdaBoost neural network in combination with wavelet decomposition to solve the defect of lower accuracy and enhance model robustness, and the experimental evaluation demonstrates that the proposed strategy can significantly enhance model robustness and effectively improve the prediction accuracy. Xiao et al. [29] developed a reliable combination model for wind speed forecasting based on an improved Adaboost algorithm named the time-vary-forecasting-effectiveness (TW-FE-Adaboost) algorithm to improve the overall forecasting accuracy. The increasing application of the Adaboost algorithm can be attributed to two aspects: One is that it can improve the prediction performance by combining multiple predictor models; another aspect is that the Adaboost algorithm has the advantages of simple calculation and small error. Therefore, ENN optimized by the Adaboost algorithm is applied to predict short-term wind power, which can maximize the merit of Adaboost algorithms and enhance prediction ability of ENN for nonlinear, chaotic, and volatile data.

Over the past few decades, numerous wind power forecasting approaches have been presented, which have enhanced the prediction accuracy of wind power series. However, since the relatively noisy and unstable characteristics of wind power data, wind power prediction by directly using original data would lead to substantial forecasting errors and poor performance [30]. Hence, the signal decomposition technique has been considered and applied for wind power forecasting to improve prediction performance, especially EMD [31] and EEMD [32]. Although these two techniques have improved the forecasting performance to a certain extent, they still have some disadvantages, such as the mode mixing problem in EMD and the residual noise in EEMD [33]. To overcome these defects of EMD and EEMD, a novel technique called ESMD, proposed by Wang et al. [34], is employed for reducing the noise and uncertainty of wind speed and wind power series. Moreover, compared with some classical time-frequency transform methods [35–37], the ESMD method proposes a "direct difference (DI) method" for data, broking through the traditional concept of using integral changes. Additionally, the ESMD method has been applied in many fields at present, such as climate change issues [38] and seismology [39].

A review of the previous literature indicates that the prediction methods discussed above have some inherent disadvantages. The shortcomings of these methods are summarized as follows:

(1) Physical methods are extremely weak in coping with short-term horizons; therefore, these methods do not have accurate and effective results in short-term forecasting [33]. Moreover,

physical methods are characterized by high cost and complexity. (2) Relative to physical methods, statistical methods based on single data are a relatively simple method, however, its ability to handle abrupt information is poor. Additionally, wind speed or wind power prediction is affected by meteorological data, but statistical methods are usually applied to single data. (3) Different from other methods, intelligent methods have been widely researched and applied to address complicated relationships and effectively perform forecasting, which could successfully capture hidden non-linear relationships among given historical data [40]. However, there are still many disadvantages and defects with intelligence methods, for example, easily getting into a local optimum, over-fitting, and exhibiting a relatively low convergence rate [41]. (4) The individual prediction model does not take into account the necessity and importance of data preprocessing techniques, so it cannot achieve high prediction accuracy and meet the requirements of time series prediction. Therefore, since each model has its inevitable shortcomings, the above prediction method cannot always capture the wind speed or wind power trend and cannot be applied in all cases. Consequently, a combined model has often been taken into consideration, which is deemed as an excellent method that utilizes the advantages of individual approaches to obtain a higher forecasting accuracy [42].

Based on the analysis above, this study introduces a novel combined model to improve the accuracy of short-term wind power prediction. It combines meteorological data (humidity, pressure, temperature, wind direction, and wind speed), the signal decomposition technique, SampEn theory, and several forecasting algorithms, namely the Bat-BP model, Adaboost-ENN model, and ENN. It successfully exploits the advantages of each prediction model for further improvement. More specifically, considering that wind power is influenced by meteorological data at different times and altitudes (10 m, 30 m, 50 m, 70 m, and 100 m), we use grey correlation degree analysis to select meteorological data at different altitudes in the same location as part of the input variable to the prediction model. Then, to deal with the randomness and instability of wind speed and wind power time series, the ESMD method is employed to decompose the original wind speed and wind power time series into several sub-series, called Intrinsic Mode Functions (IMFs) and a residue R, respectively. Subsequently, SampEn theory is employed to calculate the complexity of each wind power sub-series obtained by the ESMD method, and reconstruct them into three characteristic components, namely PHC, PMC, and PLC. Similarly, wind speed sub-series are reconstructed into SHC, SMC, and SLC. Then, the training samples of PHC, PMC, and PLC are determined by PACF theory. Next, Bat-BP, Adaboost-ENN, and ENN, which have high prediction accuracy, are selected as forecasting models to predict PHC, PMC, and PLC, respectively. Finally, the prediction results of each characteristic component are aggregated into the final prediction values of the original wind power series.

On the whole, the novelty of this study can be described as follows:

(1) Based on the historical wind power data, the meteorological data at different times and altitudes are selected as the input variables of the wind power prediction to improve the forecasting accuracy. The grey correlation degree analysis is used to determine the influencing factors of wind power, which can more fully consider the influence of the external environment on wind power.

(2) The ESMD method, based on the direct difference (DI) method, is adopted to handle the complexity and volatility of wind speed and wind power series for the first time, which can smooth the data and extract the main characteristics of the data. The original wind power and wind speed series are decomposed and reconstructed into three characteristic components to decrease the instability of wind power and wind speed series.

(3) The proposed combined model based on three prediction algorithms improves the accuracy of wind power prediction to a certain extent. The proposed combined model utilizes the advantages of each individual model and overcomes the limitation of the low accuracy and instability of a single model.

The remainder of this paper is structured as follows. Section 2 describes the specific prediction model, including the Bat-BP model, Adaboost-ENN model, and ENN. The framework of this paper

can be seen in Section 3. Section 4 discusses the prediction results and prediction performance of the proposed combined model. Finally, Section 5 concludes the results of this study.

## 2. Methodology

This section aims to provide a brief introduction to the methods used in this study, including the ESMD method, Bat-BP model, and Adaboost-ENN model.

### 2.1. Extreme-Point Symmetric Mode Decomposition

ESMD, a data processing method, was proposed by Wang and Li et al., which is a new development of the Hilbert-Huang transform (HHT) method [34]. The ESMD method is similar to the EMD method, which can smooth the complex signals and obtain several IMFs and a residue R. The basic idea of ESMD draws on the EMD method, which changes a cubic spline interpolant of the upper and lower envelopes in EMD to the internal pole symmetric interpolation. At the same time, the ESMD method uses the least squares method to optimize R as the "adaptive global average" of the entire data to determine the optimal number of filters. The specific algorithm of ESMD is as follows:

Step 1: Find all the poles that exist in the original data, $Y_t$, and record them as $E_i(i = 1, 2, \cdots, n)$.

Step 2: Connect the adjacent poles, $E_i$, with line segments and mark the midpoint of each line segment as $F_i(i = 1, 2, \cdots, n - 1)$.

Step 3: Use the linear interpolation method to set midpoints of the left and right borders to $F_0$ and $F_n$, respectively.

Step 4: Construct $p$ interpolation curves, $L_1, L_2, \cdots, L_p(p \geq 1)$, using $n + 1$ midpoints, and an average curve can be obtained, $L^* = (L_1 + L_2 + \cdots + L_p)/p$.

Step 5: Calculate $Y_t - L^*$ by repeating step 1-step 4 until $|L^*| \leq \varepsilon$ is satisfied or the number of screenings reaches the preset maximum value, $K$, then the first $IMF_1$ is obtained. In general, the set of permitted error, $\varepsilon$, for $0.001\sigma_0$, and $\sigma_0$ is the standard deviation of the original data, $Y_t$.

Step 6: Repeat steps 1 to 5 for $Y_t - IMF_1$ to obtain $IMF_2, \cdots, IMF_n$ and a residual $R$ (The number of poles of $R$ is set as at least 4 poles).

Step 7: Make $K$ vary between $K_{\min}$ and $K_{\max}$ and repeat steps 1 to 6. Then, calculate the relative standard deviation, $\sigma$, of $Y_t - R$ when $K$ takes different values.

Step 8: Find the $K_0$ corresponding to the minimum variance ratio, $v = \sigma/\sigma_0$ (where $R$ is the best fit curve of the data), and the output decomposition results of steps 1 to 6 are repeated again.

It is worth noting that in step 4, ESMD mainly includes three forms according to the difference of the P number, which are called *ESMD_I*, *ESMD_II*, and *ESMD_III*. It has been proven by practice that *ESMD_II* shows better signal decomposition characteristics [43]. Therefore, this paper chooses *ESMD_II* as a signal decomposition method. In step 8, the minimum variance ratio, $v$, reflects the trend change of the trend term and the original signal. The value is smaller, the decomposition result is better.

### 2.2. Bat-BP Neural Networks

As a multi-layer forward network, BPNN is trained by the error back propagation algorithm, whose network topology mainly consists of an input layer, hidden layer, and output layer [44,45]. In the BPNN algorithm, the weight and threshold of the network are usually adjusted along the negative gradient direction of the network error change, and finally the network error reaches a minimum value. Although BPNN is widely used, it has the disadvantages of a poor global search ability, slow convergence rate, and it easily falls into local minimum values. In order to overcome the shortcomings of BPNN, the bat optimization algorithm is used to optimize the parameters of BPNN to further improve its prediction performance.

The bat algorithm is a new type of bionic algorithm. It is very suitable for the excellent selection of complex problems for simplicity and robustness, and it is widely used in various fields, such as optimization and classification. The main principle of the bat algorithm is to simulate the process

of bats searching for prey. The updated formula of frequency, velocity, and position is shown in Equations (1)–(3):

$$f_i = f_{\min} + (f_{\max} - f_{\min})\alpha. \tag{1}$$

$$V_i^{t+1} = V_i^t + (X_i^t - X^*)f_i. \tag{2}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1}. \tag{3}$$

where $\alpha \in [0,1]$ is a random number subjecting to uniform distribution; $f_i$ represents the update frequency of bat $i$ in the range $[f_{\min}, f_{\max}]$; $X^*$ indicates the current global best location; $V_i^{t+1}$ and $V_i^t$ are the velocity values of the bat $i$ at time step $t+1$ and $t$; $X_i^{t+1}$ and $X_i^t$ represent the specific location of the bat $i$ at time step $t+1$ and $t$, respectively.

The local search is proposed to improve the performance of the bat algorithm. Each bat gets a partial new solution by random walks when a solution is selected in the current best solutions:

$$X_i^t = X^* + \varepsilon A^t. \tag{4}$$

where $\varepsilon \in [0,1]$ is a random number, $A^t$ is the average loudness of all bats in the same time period.

Further, the loudness, $A_i$, is strong and the rate, $r_i$, pulse emission is low, contributing to expand the search space. After the prey is discovered, the loudness is gradually reduced and the rate of pulse emission is increased, which helps to accurately grasp the spatial position of the prey. This search feature is shown by Equations (5)–(6):

$$A_i^{t+1} = \alpha \bullet A_i^t. \tag{5}$$

$$r_i^{t+1} = r_i^0[1 - \exp(-\gamma t)]. \tag{6}$$

where, $A_i^{t+1}$ and $A_i^t$ are the loudness of the bat $i$ at time step $t+1$ and $t$; $\alpha \in [0,1]$ is the attenuation coefficient of loudness. Here, $r_i^{t+1}$ indicates the rate of the pulse emission of the bat $i$ at time step $t+1$; $r_i^0$ represents the maximum rate of pulse emission of the bat, $i$; $\gamma > 0$ is the increasing coefficient of the rate of the pulse emission.

This specific algorithm of Bat-BP is briefly described as follows:

Step 1: Construct the BPNN model and experiment repeatedly to determine its structure.

Step 2: Produce the number of bat, n, and form an initial population size of bats. Initialize bat position, $x_i$, velocity, $v_i$, loudness, $A_i$, and pulse rates, $r_i$.

Step 3: Contact the location of each individual bat to the fitness function $f(X)$.

Step 4: Determine whether the algorithm reaches the maximum number of iterations. If yes, the algorithm ends and the best solution is output; otherwise, go to step 5.

Step 5: Update the velocity and location of the bat by Equations (1)–(3).

Step 6: Generate a random number, *rand*1. If *rand*1 > $r_i$, a new solution will be obtained by Equation (4).

Step 7: Evaluate the quality of the solution.

Step 8: Generate a random number, *rand*2. If *rand*2 < $A_i$ and $f(X_i^t) < f(X^*)$, the bat individual will be updated. $A_i$ will be reduced by Equation (5) and $r_i$ will be enlarged by Equation (6), respectively.

Step 9: Output target value if the termination condition is met, and the termination program is executed. Otherwise, return to step 3.

Step 10: Use the weights and thresholds obtained by the bat algorithm to train the established BPNN model.

## 2.3. Adaboost-ENN Model

ENN is a recurrent neural network, with local memory units and feedback connections. Compared with the forward networks, ENN not only has an input layer, a hidden layer, and an output layer, but also has a context layer. The neurons in the input layer only play the role of signal transmission, and the output layer applying purelin function is linearly weighted. The hidden layer, a single-layer

network structure, adopts the tansig function, which can reduce the running time by improving the convergence speed of the ENN model. Compared with the forward networks, ENN adds a special layer, called the context layer. The context layer acts as a one-step delay operator to achieve the purpose of memory, so that the system has the ability to adapt to time-varying characteristics. Moreover, it can directly reflect the characteristics of the dynamic process system. Although ENN has been improved on the basis of the forward networks, there are still deficiencies due to its inherent characteristics. For example, ENN uses the gradient descent method to modify weights and thresholds between neurons in each layer, which reduces the convergence accuracy of the network [46]. For the inherent defects of ENN, the Adaboost algorithm is applied to optimize ENN to improve the prediction performance of the model.

The Adaboost algorithm was developed by Schapire for regression with time series data, which is a typical example of the boosting algorithm [47]. The Adaboost algorithm can generate different weak learners by repeatedly training the same data set, and then combines these weak learners into powerful learners to improve the prediction accuracy and generalization ability of the weak learners. The core idea of the Adaboost algorithm is to value the samples with large prediction error and the weak learners with good performance, that is, to improve the sample weights with a poor training effect and the weak learner weights with a strong learning ability. On the contrary, it reduces the sample weights with a good training effect and the weak learner weights with a weak learning ability.

According to Wang et al. [48], this detailed prediction process of the Adaboost-ENN model is as follows:

Step 1: Preprocess original data by data quantification and normalization.

Step 2: Assume training set, $X = (x_1, y_1), \cdots, (x_m, y_m), i = 1, 2, \cdots, m$, and the initial distribution weight of the sample on the training set is initialized: $D_1(i) = 1/m$. The neural network structure is determined by the input and output dimensions, and the weights and thresholds of ENN are initialized.

Step 3: Find the weak predictor, $h_j(j = 1, 2, \cdots, T)$. When the *jth* weak predictor is trained, the ENN is trained with the training set and the prediction results are output. Then, the sum of prediction error, $\varepsilon_j$, of the prediction series, $h(j)$, can be obtained, which can be expressed as follows:

$$\varepsilon_j = \sum_{i=1}^{m} D_t(x_i), i = 1, 2, \cdots, m(h_j(x_i) \neq y_i). \tag{7}$$

where $h_j(x_i)$ is the prediction results, and $y_i$ is the expected values.

Step 4: Update weight. According to the $\varepsilon_j$, the weight of the series is calculated as:

$$\alpha_j = 1/2 * \ln\left[(1 - \varepsilon_j)/\varepsilon_j\right]. \tag{8}$$

Then, the weight of the next training sample is adjusted, the adjustment formula is:

$$D_{j+1}(i) = \frac{D_j(i) \exp\left[-\alpha_j y_i h_j(x_i)\right]}{Z_j}, i = 1, 2, \cdots, m. \tag{9}$$

where $Z_j$ is the normalization factor, and $\sum_{i=1}^{N} D_{j+1}(x_i) = 1$.

Step 5: Obtain $T$ strong prediction functions, $h_j(x)(j = 1, 2, \cdots, T)$, through T-round training. A strong prediction function is formed:

$$H_{final}(x) = sign(\sum_{j=1}^{T} \alpha_j h_j(x)). \tag{10}$$

## 3. The Framework of the Combined Model

In this study, a novel combined model is successfully developed to improve the wind power prediction effectiveness. The main operational steps of the combined model proposed in this paper are as follows. In addition, we also plotted the corresponding flow chart, as shown in Figure 1.

(1) Use the grey correlation degree analysis to screen meteorological data at different times and altitudes to determine the influencing factors of the original wind power series.

(2) Adopt the ESMD method to decompose the wind speed and wind power series into several sub-series, called wind power sub-series and wind speed sub-series, respectively.

(3) Utilize sample entropy theory to calculate the complexity of each wind power sub-series obtained by the ESMD method, and reconstruct them into three wind power characteristic components of reduced order of complexities, namely PHC, PMC, and PLC. Meanwhile, the complexity of each wind speed sub-series is calculated by sample entropy theory, and they also are reconstructed into SHC, SHC, and SLC.

(4) Apply PACF theory to determine the input–output samples of PHC, PMC, and PLC to improve the accuracy of wind power prediction.

(5) Do a group of trial experiments to select the best training algorithm for PHC, PMC, and PLC, respectively.

(6) Employ the Bat-BP model, Adaboost-ENN model, and ENN to do one-step rolling prediction of PHC, PMC, and PLC, respectively, to forecast wind power in the next 24 h.

(7) Conduct aggregate calculation for the one-step rolling prediction results in PHC, PMC, and PLC to obtain the final prediction values of the original wind power series.

(8) To compare the prediction performance, the proposed combined prediction model will be compared with other models through four sets of comparative experiments, such as NMD-ESMD-Bat-BP model, ESMD-Adaboost-Elman model, EMD, and ENN, etc.
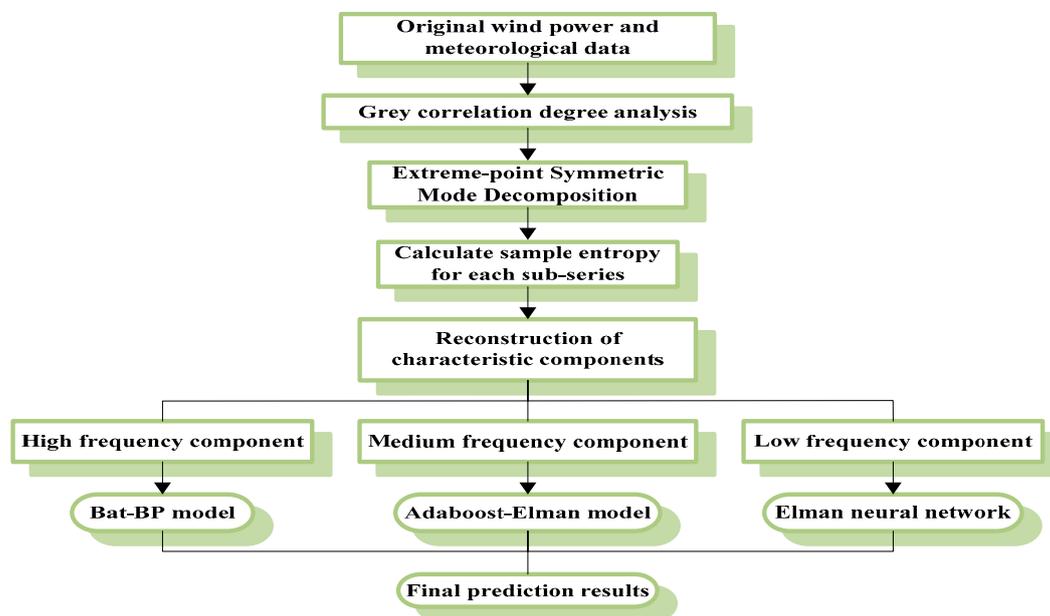


**Figure 1.** The overall framework for the combined prediction model.

## 4. Experimental Design, Results, and Discussion

### 4.1. Datasets

To validate the prediction performance of the proposed combined prediction model, 15-min wind power and meteorological data were collected from wind farm in China, which were generated from 09:00 24 April 2016 to 09:00 9 May 2016, a total of 15 days. Among them, meteorological data were

measured at heights of 10 m, 30 m, 50 m, 70 m, and 100 m from the ground. The rated installed capacity of the wind farm is 49.5 MW. Each dataset was divided into a training set and testing set, in which the data of the first 14 days was used as the training set to train the prediction model, and the data of the last day were used as the testing set to estimate the prediction performance of the model [49]. The data statistical description of the wind power and meteorological data is shown in Table A1. Meanwhile, the description of the training set and testing set is presented in Table A2. Although the wind power and meteorological data for the 15 days are continuous without missing data points, there are still a small number of outliers. Therefore, we used fractal interpolation to process the outliers that were rejected [50].

### 4.2. Evaluation Criteria

In this paper, three error metrics were used to evaluate the prediction performance of the proposed model, namely MAPE (mean absolute percent error), NMAE (normalized mean absolute error), and NRMSE (normalized root mean square error). The specific formula of each error metric is as follows:

$$MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{\hat{y}_t - y_t}{y_t} \right| * 100\%. \tag{11}$$

$$NMAE = \frac{1}{P_{inst} N} \sum_{t=1}^{N} |\hat{y}_t - y_t| * 100\%. \tag{12}$$

$$NRMSE = \frac{1}{P_{inst}} \sqrt{\frac{1}{N} \sum_{t=1}^{N} [\hat{y}_t - y_t]^2} * 100\%. \tag{13}$$

where $y_t$ is the real wind power, $\hat{y}_t$ represents the predicted wind power, $P_{inst}$ is the rated installed capacity of wind farm, and $N$ indicates the test samples number for the prediction model.

Moreover, we used three percentage error indices to compare the prediction performance of the built model, called $P_{MAPE}$ (promoting percentages of mean absolute percentage error), $P_{NMAE}$ (promoting percentages of normalized mean absolute error), and $P_{NRMSE}$ (promoting percentages of normalized root mean square error). Their detailed formulas are as follows.

$$P_{MAPE} = \left| \frac{MAPE_1 - MAPE_2}{MAPE_1} \right|. \tag{14}$$

$$P_{NMAE} = \left| \frac{NMAE_1 - NMAE_2}{NMAE_1} \right|. \tag{15}$$

$$P_{NRMSE} = \left| \frac{NRMSE_1 - NRMSE_2}{NRMSE_1} \right|. \tag{16}$$

### 4.3. Simulation

#### 4.3.1. Grey Correlation Degree Analysis

To determine the influencing factors of wind power, we used grey correlation degree analysis to measure the relatedness between the wind power series and meteorological data at different times and altitudes. As a multi-factor statistical analysis method, the grey correlation degree analysis is used to describe the relationship between variables. The gray relational degree changes in the range of 0 to 1, generally believed that greater than 0.5 has a strong correlation. As illustrated in Table 1, the grey relational degree between the meteorological data and wind power series exceeds 0.6 at different times and altitudes, which indicates that the correlation between the meteorological data and wind power is significant. Among the five types of meteorological data, the correlation between wind speed and wind power is the strongest, especially the wind speed (50 m). Additionally, humidity

(30 m) and temperature (100 m) are also strongly correlated with wind power. In addition, pressure at different times and altitudes are parallel and direction series at different times and altitudes are the same, so the correlation between the two types of meteorological data and wind power at different times and altitudes is 0.60583 and 0.63686, respectively. In view of the above analysis, we selected humidity (30 m), pressure (50 m), temperature (100 m), wind direction (50 m), and wind speed (50 m), respectively, as the influencing factors of wind power.

**Table 1.** The grey correlation degree between wind power series and meteorological data at different times and altitudes.

| | Meteorological Data | | | | |
|---|---|---|---|---|---|
| | **Humidity** | **Pressure** | **Temperature** | **Wind Direction** | **Wind Speed** |
| 10 m | 0.62561 | 0.60583 | 0.66530 | 0.63686 | 0.70173 |
| 30 m | 0.62564 | 0.60583 | 0.66538 | 0.63686 | 0.70175 |
| 50 m | 0.62562 | 0.60583 | 0.66545 | 0.63686 | 0.70181 |
| 70 m | 0.62559 | 0.60583 | 0.66550 | 0.63686 | 0.70152 |
| 100 m | 0.62561 | 0.60583 | 0.66562 | 0.63686 | 0.70163 |

### 4.3.2. ESMD Decomposition

Through the analysis of the grey correlation degree, it can be seen that wind power is mainly affected by wind speed, which has the characteristics of instability and intermittentency. Therefore, the ESMD method was employed to decompose original wind power and wind speed series (50 m). Before decomposing the original time series, we determined the number of screening times corresponding to the minimum variance ratio, $v$, by repeatedly adjusting the number of residual component extreme points and the number of iterations. Therefore, the number of residual component extreme points of wind power is 20 and the number of iterations is 56 after repeated tests and comparisons. Similarly, the number of residual component extreme points of wind speed is 20 and the number of iterations is 73. The number of iterations corresponding to the variance ratio of wind speed and wind power is presented in Figure 2.
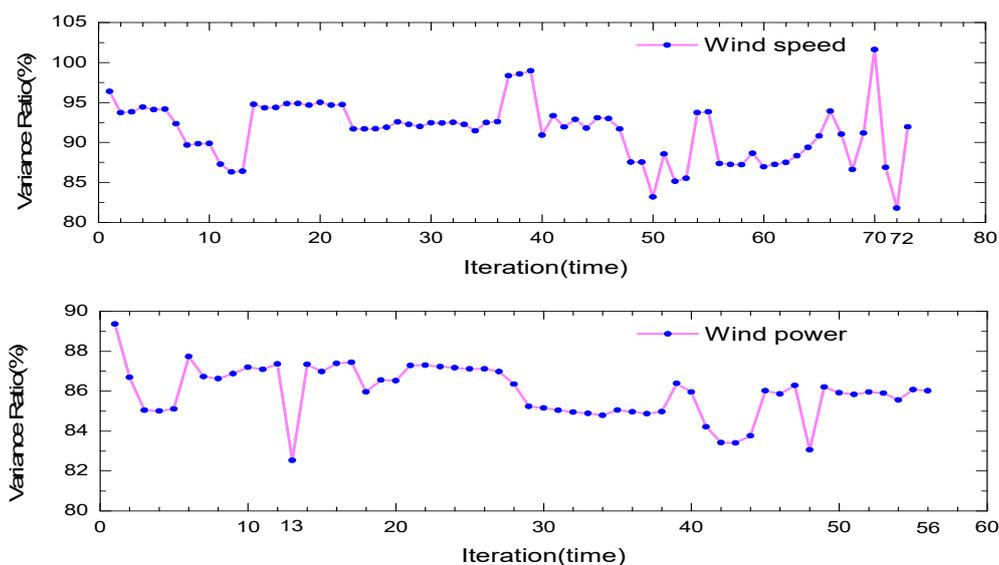


**Figure 2.** The number of iterations corresponding to the variance ratio of wind speed and wind power.

It can be seen from Figure 2 that the variance ratio is the smallest when the number of iterations of the wind speed is the 72nd time, so the best screening times for wind speed is 72. Similarly, the best screening times for wind power is 13, which corresponds to the minimum variance ratio.

After determining the best number of screening times, the wind speed and wind power series were decomposed into six IMFs and a residue, respectively. The specific decomposition results of the wind power and wind speed series are shown in Figure 3.
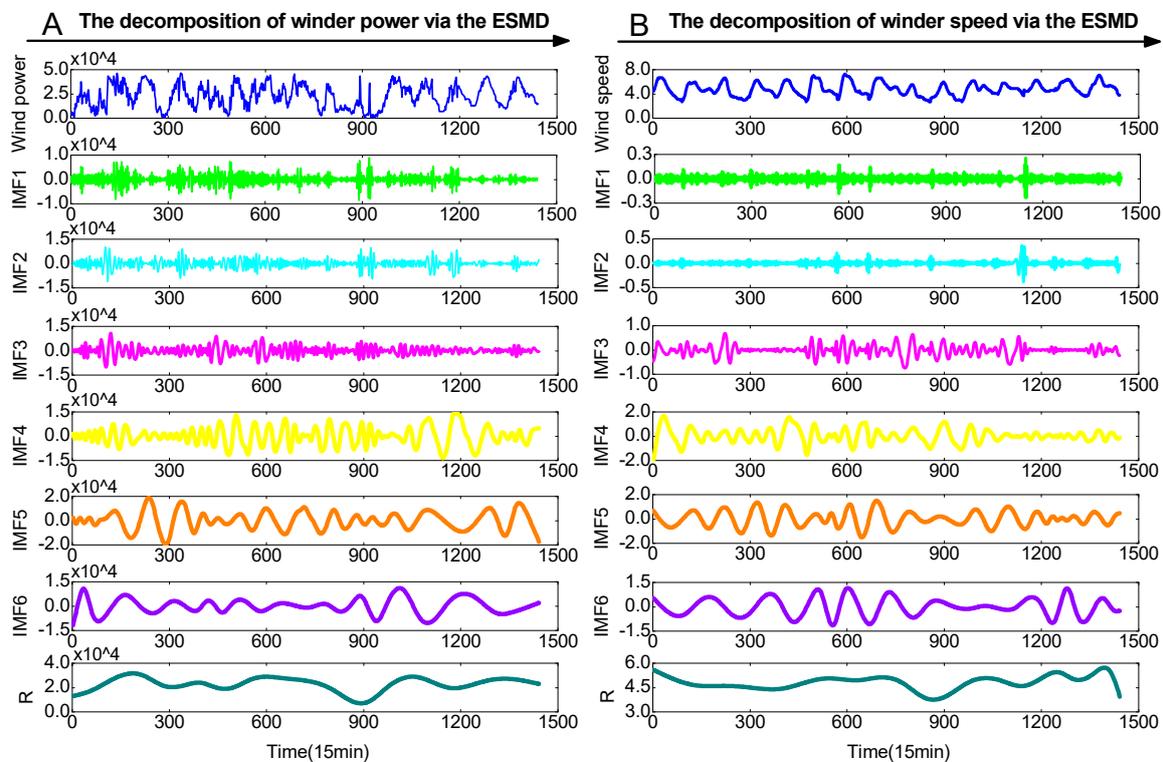


**Figure 3.** The decomposed results of wind power and wind speed series using the ESMD method: (**A**) Wind power series is decomposed by the ESMD method into six IMFs and a residue R (from top to bottom); (**B**) wind speed series is decomposed by the ESMD method into six IMFs and a residue R (from top to bottom).

### 4.3.3. Sample Entropy Theory

As a method for measuring the complexity of time series, sample entropy was proposed by Richman in 2000, which has been widely used in many fields [51]. Specifically, the sample entropy has an anti-noise ability and anti-interference ability, and is suitable for mixed signals composed of random components [52]. However, the wind power series are characterized by instability and volatility. Therefore, we used the sample entropy method to calculate the sample entropy values of each wind power sub-series decomposed by the ESMD method, and grouped them into PHC, PMC, and PLC. According to [53], the specific grouping rules are as follows:

Step 1: Calculate the sample entropy values of the original wind power series and each wind power sub-series, denoted as $SampEn_{data}$ and $SampEn_{sub-series}$.

Step 2: Record the $SampEn_{data}$ as the initial threshold. If $SampEn_{data} > SampEn_{sub-series}$, $SampEn_{sub-series}$ is grouped into a characteristic component; otherwise, $SampEn_{sub-series}$ is grouped into another characteristic component.

Step 3: Set a special threshold, $u$, and an upper bound line ($SampEn_{sub-series} - SampEn_{data} = u$) and a lower bound line ($SampEn_{data} - SampEn_{sub-series} = u$) will form.

Step 4: According to the results of Step 2 and Step 3, the wind power sub-series with similar sample entropy values can be grouped into PHC, PMC, and PLC.

This paper set the $u$ value as 0.3, and the detailed grouping process is shown in Figure 4. At the same time, the composition of the IMF number in each characteristic component is given in Table 2. After the sample entropy values of each wind power sub-series were calculated, the sub-series were

reconstructed by aggregating the sub-series in each characteristic component, respectively, as shown in Figure 5. It can be seen that the original wind power series was reconstructed into three new wind power components, which contributes to improving the prediction performance. Similarly, we used the sample entropy method to calculate the complexity of each wind speed sub-series obtained by the ESMD method, and reconstructed them into SHC, SMC, and SLC, as shown in Table 2 and Figure 5.
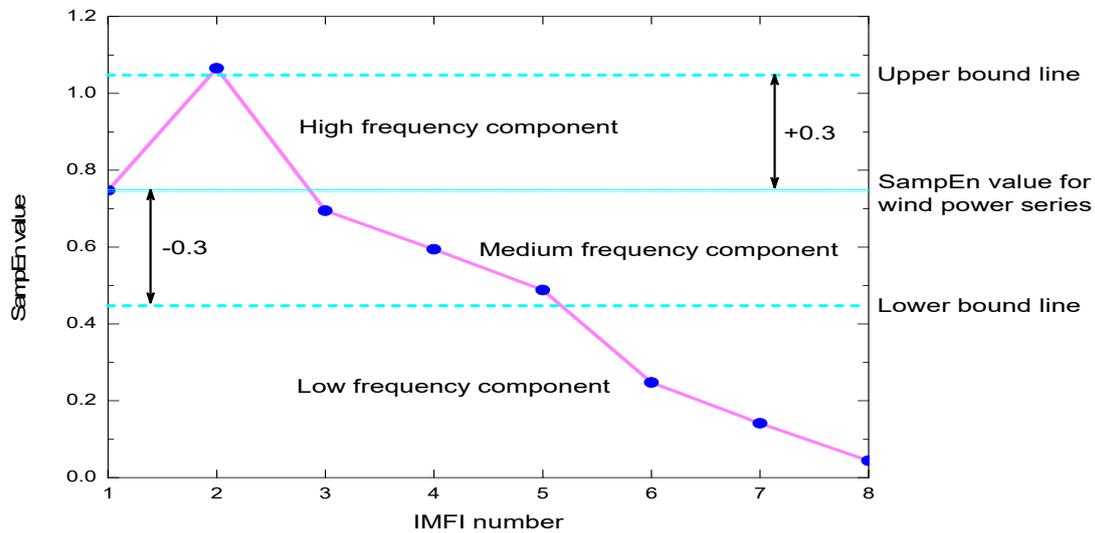


**Figure 4.** The SampEn value of each sub-series for the wind power series obtained by ESMD.

**Table 2.** The composition of the sub-series number in different characteristic components.

|  | HC | MC | LC |
|---|---|---|---|
| Wind power sub-series | IMF1 | IMFs 2–4 [1] | IMFs 5–6 & R [2] |
| Wind speed sub-series | IMF1 | IMFs 2–4 | IMFs 5–6 & R |

[1] IMFs 2–4 indicates IMF2, IMF3 and IMF4; [2] IMFs 5–6 & R indicates IMF5, IMF6 and R.
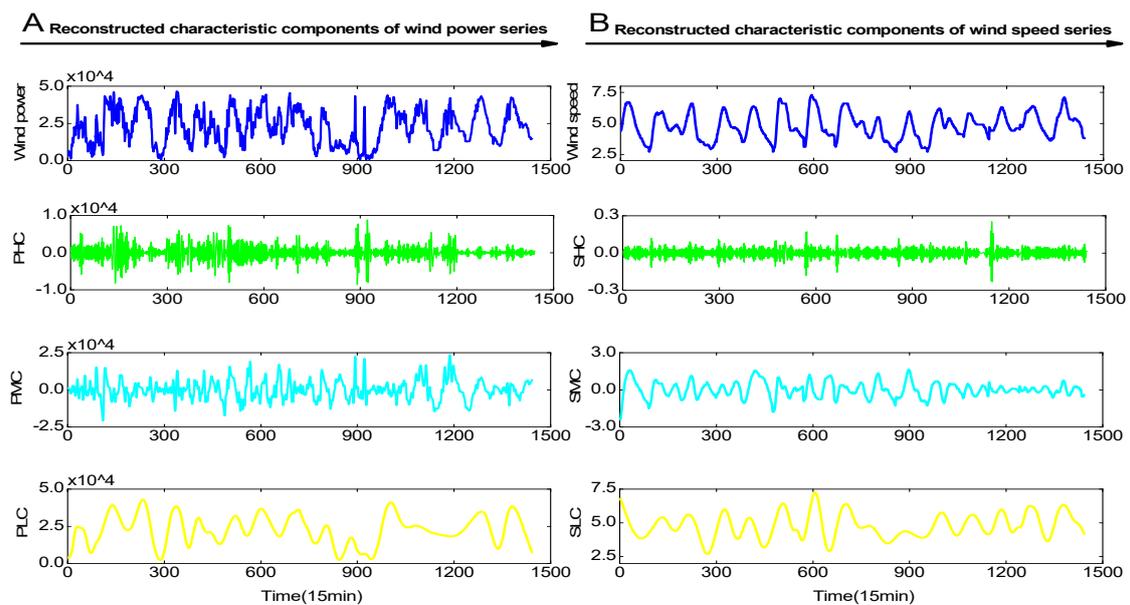


**Figure 5.** Reconstructed characteristic components of wind power and wind speed series: (**A**) wind power series is reconstructed into PHC, PMC and PLC; (**B**) wind power series is reconstructed into SHC, SMC and SLC.

### 4.3.4. PACF Theory

The input–output samples in each wind power component were determined by PACF (Partial autocorrelation function) theory, which is a useful tool for analyzing the correlation between candidate variables and historical datasets [54]. Therefore, PACF theory was used to determine the input vector for each wind power characteristic component. After normalizing the training samples, the PACF values of each wind power characteristic component were calculated and shown in Figure 6. It can be seen from Figure 6 that the dimension values of each wind power characteristic component are seven, five, and eight, respectively.

Once the input–output samples of each wind power component were determined, the corresponding data format of PHC can be defined in Figure 7. Moreover, the corresponding data format of PMC and PLC were omitted due to a restriction of space. It is worth noting that the training set and testing set both use the same rolling prediction mechanism. Therefore, the rolling prediction mechanism was adopted for wind power prediction of one day based on one-step prediction results.
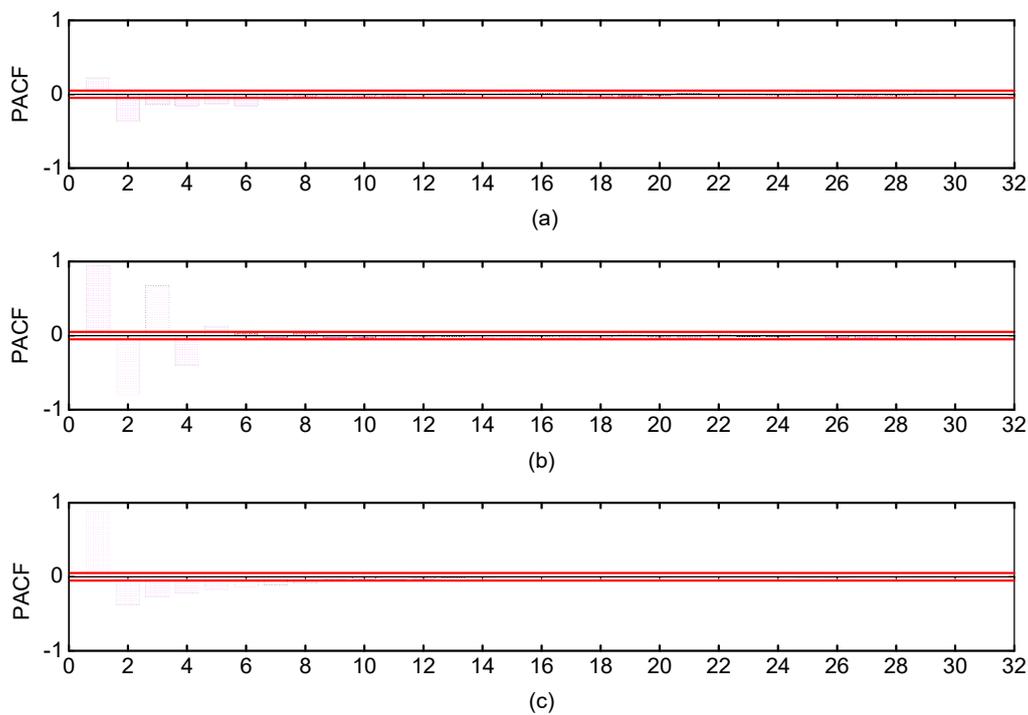


**Figure 6.** The PACF values of three wind power characteristic components: (**a**) The PACF values of PHC; (**b**) the PACF values of PMC; (**c**) the PACF values of PLC.
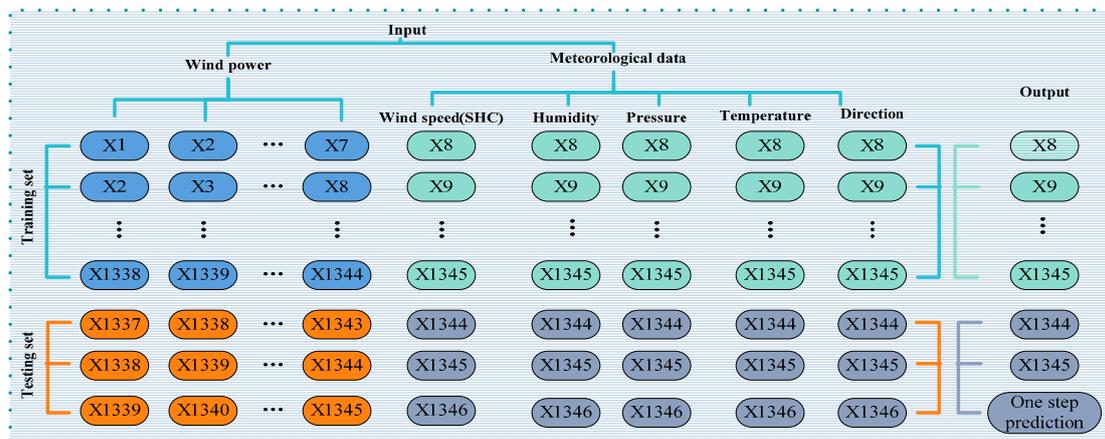


**Figure 7.** The data format of inputting and outputting for PHC.

4.3.5. The Parameter Setting of the Prediction Models

In this paper, the main procedures were performed in a Windows 7 PC with a 32-bit, 2.60 GHz Intel Core i5-3230M CPU, and 4 GB of memory. The wind power prediction experiments were employed in MATLAB R2014a. The experimental parameters are shown in Table 3.

**Table 3.** Setting the experimental parameters.

|  | Experimental Parameters | Default Value |
|---|---|---|
| Bat-BP | initial population size | 10–20 |
| | loudness | 0.25 |
| | pulse rate | 0.25 |
| | minimum frequency | 0 |
| | maximum frequency | 2 |
| | max- iteration number | 200 |
| | maximum number of convergences | 100 |
| | learning rate | 0.0001 |
| Adaboost-ENN | weak predictors | 8–20 |
| | hidden layer | 10–30 |
| | maximum number of convergences | 530 |
| | learning rate | 0.00001 |
| **ENN** | neuron number of the hidden layer | 10–30 |
| | maximum number of convergences | 1500 |
| | display interval | 2 |

*4.4. Analysis of Forecast Results and Comparisons of Different Models*

In this study, wind power sub-series obtained by the ESMD method were reconstructed into PHC, PMC, and PLC, and the data characteristics of each characteristic component are different. Among them, PHC is a nonlinear system whose fluctuations are nonlinear and abrupt; the fluctuation of PMC is nonlinear and chaotic and its volatility is relatively moderate; and the fluctuation of PLC is relatively flat and nonlinear [55]. However, Bat-BP, Adaboost-ENN, and ENN are all suitable for dealing with nonlinear, fluctuating, abrupt, and chaotic data, and have made great contributions to wind power prediction. Therefore, to select the best prediction model for each wind power characteristic component, we performed a set of comparative experiments by using Bat-BP, Adaboost-ENN, and ENN in each characteristic component, with the prediction results shown in Table 4. It can be seen from Table 4 that the Bat-BP model is the best for PHC prediction. Similarly, the Adaboost-ENN model and ENN model are suitable for PMC and PLC prediction, respectively.

To estimate the prediction accuracy of the proposed combined model, we set up four different experiments. The comparisons of MAPE for different prediction models and specific results of four groups of experiments are shown in Figure 8, part A and B. It can be seen from Figure 8 that wind power prediction results of the proposed model are superior to others in wind power forecasting. Additionally, the prediction error is lower than others in terms of MAPE. The specific purpose of each experiment and corresponding analysis are as follows.

The first group of experiments was built to validate that meteorological data at different times and altitudes has an influence on the accuracy of wind power prediction, which was compared to other prediction models that did not consider meteorological data (NMD), which were the NMD combined model, NMD-ESMD-Bat-BP, NMD-ESMD-Adaboost-ENN, and NMD-ESDM-ENN. From Table 5, in all prediction results, the proposed model is superior to other models that do not consider meteorological data, with MAPE values of 5.60%, 6.76%, 8.29%, and 8.84%, respectively. In addition, the NMD combined model has a better prediction performance among the remaining four methods, with NMAE values of 2.96%. The prediction results obtained by the NMD combined model can further illustrate that although the meteorological data has not been considered, the combined model based on three prediction models has significantly improved the accuracy of wind power prediction. However,

NMD-ESDM-ENN achieves the worst wind power prediction accuracy with a maximum NRMSE of 6.23%, which shows that individual model has lower prediction performance.

The second group of experiments was conducted to evaluate the accuracy of the proposed combined model compared with the other three ESMD-based models, namely ESMD-Bat-BP, ESMD-Adaboost-ENN, and ESDM-ENN. It can be seen from Table 5 that the proposed model has the best MAPE, NMAE, and NEMSE at 3.73%, 1.91%, and 2.51%, respectively. In addition, the models from highest to lowest based on the prediction accuracy are ESMD-Adaboost-ENN, ESMD-Bat-BP, and ESMD-ENN, with MAPE values of 6.64%, 6.76%, and 7.21%, respectively. Moreover, although ESMD-Adaboost-ENN has a lower MAPE value than ESMD-Bat-BP, its NMAE and NRMSE are higher than ESMD-Bat-BP, indicating that the prediction performance of ESMD-Bat-BP is better than ESMD-Adaboost-ENN.

The third group of experiments illustrates the prediction performance of the wind power series by comparing the combined model with models using different signal preprocessing methods, called EMD and EEMD. The comparison results are presented in Table 5. According to the evaluation criteria, the performance metric values calculated from the proposed combined model are evidently lower than those calculated from the models using other data preprocessing methods. In addition, models using three prediction methods after signal preprocessing are always superior relative to the models using the individual prediction method (including ENN, Adaboost-ENN, and Bat-BP), with NMAE values of 2.68% and 2.18%, respectively.

The fourth group of experiments presents a comparison between the proposed combined model and some classic individual models without signal preprocessing methods, namely ENN, Adaboost-ENN, and Bat-BP. The detailed prediction performances are shown in Table 5. Compared with three other individual models, the developed combined model achieves the best prediction performance with an NMAE value of 1.91%. However, the other individual models have fairly lower NMAE values of 5.19%, 4.59%, and 4.41%. Based on the MAPE values in Table 4, among the remaining three models, Bat-BP is ranked as the second most effective model in wind power prediction, with lower MAPE values at 8.08%. Meanwhile, the ENN has the highest MAPE values at 9.51%, which illuminates its poor prediction accuracy.

More analyses are shown as follows. First, in general, individual models do not achieve good performance compared to all remaining models. Specifically, the prediction performance of the three classic individual models in the fourth group of experiments is relatively poor, especially the ENN. Then, the classic individual model was used to predict the short-term wind power based on the signal preprocessing. A comparison between the model considering meteorological data with the model not considering meteorological data shows that the wind power time series is affected by the external environment, as the model considering the meteorological data can effectively identify external factors, thus obtaining a more precise prediction. The exception is that the NMD-ESMD-Adaboost-ENN model's three error metrics are lower than the ESMD-ENN model, indicating that the Adaboost algorithm effectively integrates single ENN to form a strong predictor. Finally, comparisons among the combined model, EMD, and EEMD models reveal that the combined models based on meteorological data are very effective in enhancing the forecasting accuracy, and the proposed combined model adequately makes use of the advantages of the external influence factors, decomposition method, sample entropy method, and hybrid prediction algorithm, and integrates them well.

Furthermore, to compare performance differences between the proposed model and the other models considering meteorological data, we used three indexes (including PMAPE, PNMAE, and PNRMSE) to measure the difference between these models, as shown in Table 6. When comparing the single ENN model with the ESMD-ENN model, the latter has improved the performance of the former considerably. For instance, the MAPE promoting percentages by the ESMD is 31.90%; the NMAE promoting percentages by the ESMD is 35.86%; the NRMSE promoting percentages by the ESMD is 42.76%. When comparing the proposed model with the ESMD-ENN model, the hybrid prediction model has improved the performance of the ESMD-ENN model significantly. Similarly,

the prediction performance of the ESMD-Adaboost-ENN model (or ESMD-Bat-BP model) is better than the Adaboost-ENN model (or Bat-BP model), and the prediction performance of the proposed model is further improved. Compared with the EMD model, the EEMD method effectively improves the prediction performance. However, the prediction accuracy and stability of the proposed model is better than that of the EEMD model. For instance, the MAPE promoting percentages by the ESMD method is 17.69%; the NMAE promoting percentages by the ESMD method is 14.14%; the NRMSE promoting percentages by the ESMD is 11.95%.

**Table 4.** The results of the prediction model experiment.

| Model | PHC | | | PMC | | | PLC | | |
|---|---|---|---|---|---|---|---|---|---|
| | MAPE (%) | NMAE (%) | NRMSE (%) | MAPE (%) | NMAE (%) | NRMSE (%) | MAPE (%) | NMAE (%) | NRMSE (%) |
| Bat-BP | 154.12 | 1.09 | 1.41 | 70.98 | 1.70 | 2.21 | 1.91 | 0.71 | 0.96 |
| Adaboost-ENN | 160.07 | 1.12 | 1.45 | 59.63 | 1.32 | 1.84 | 1.57 | 0.62 | 0.74 |
| ENN | 168.23 | 1.18 | 1.53 | 74.76 | 1.76 | 2.28 | 1.31 | 0.51 | 0.69 |

**Table 5.** The prediction performances for three different experiments.

| Model | MAPE (%) | NMAE (%) | NRMSE (%) |
|---|---|---|---|
| ENN | 9.51 | 5.19 | 6.21 |
| Adaboost-ENN | 8.71 | 4.59 | 5.53 |
| Bat-BP | 8.08 | 4.41 | 5.65 |
| NMD-ESDM-ENN | 8.29 | 4.65 | 6.23 |
| NMD-ESMD-Adaboost-ENN | 6.96 | 3.42 | 4.34 |
| NMD-ESMD-Bat-BP | 8.84 | 4.33 | 5.06 |
| ESMD-ENN | 7.21 | 3.82 | 4.35 |
| ESMD-Adaboost-ENN | 6.64 | 3.40 | 4.06 |
| ESMD-Bat-BP | 6.76 | 3.16 | 3.64 |
| NMD combined model | 5.60 | 2.96 | 3.87 |
| EMD | 5.11 | 2.68 | 3.30 |
| EEMD | 4.39 | 2.18 | 2.81 |
| Proposed model | 3.73 | 1.91 | 2.51 |

**Table 6.** The promoting percentages for three different experiments.

| | $P_{MAPE}$ (%) | $P_{NMAE}$ (%) | $P_{NRMSE}$ (%) |
|---|---|---|---|
| ESMD-ENN vs. ENN | 31.90 | 35.86 | 42.76 |
| Proposed Model vs. ESMD-ENN | 93.30 | 98.53 | 73.31 |
| ESMD-Adaboost-ENN vs. Adaboost-ENN | 31.17 | 35.00 | 36.21 |
| Proposed Model vs. ESMD-Adaboost-ENN | 78.02 | 78.01 | 61.75 |
| ESMD-Bat-BP vs. Bat-BP | 19.53 | 39.56 | 55.22 |
| Proposed Model vs. ESMD-Bat-BP | 81.23 | 65.45 | 45.02 |
| EEMD vs. EMD | 16.40 | 22.94 | 17.44 |
| Proposed Model vs. EEMD | 17.69 | 14.14 | 11.95 |
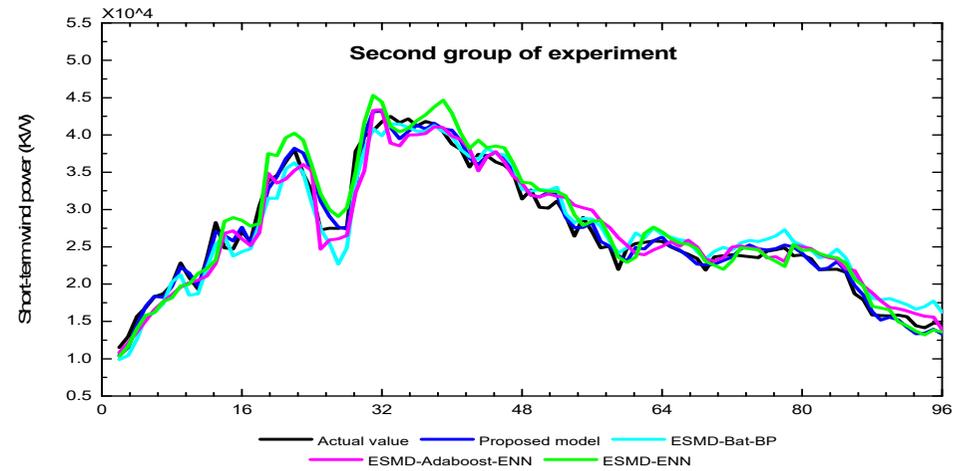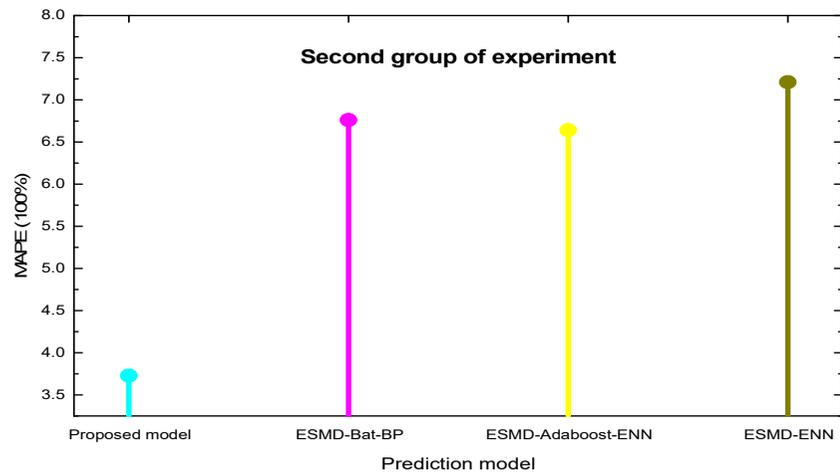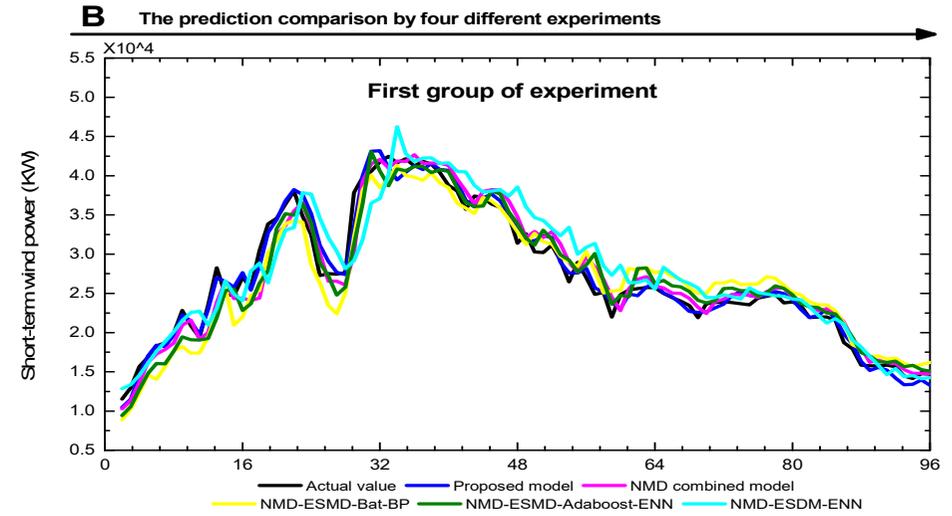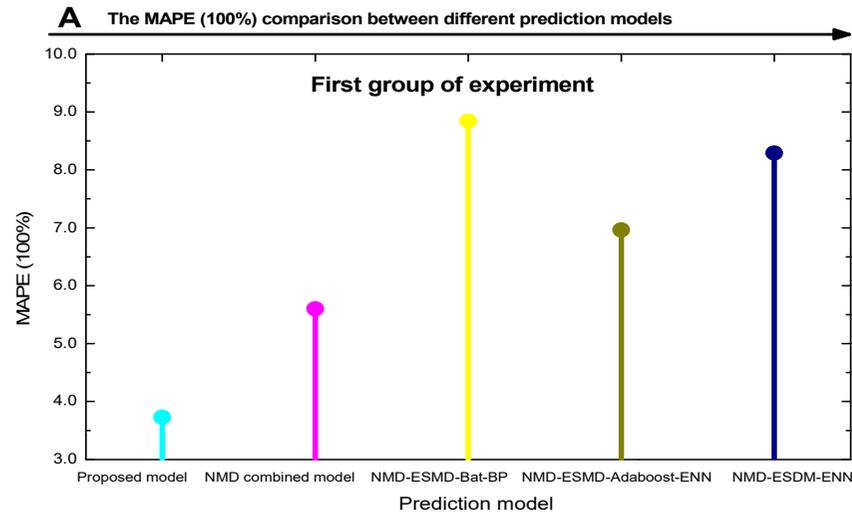
**Figure 8.** *Cont.*
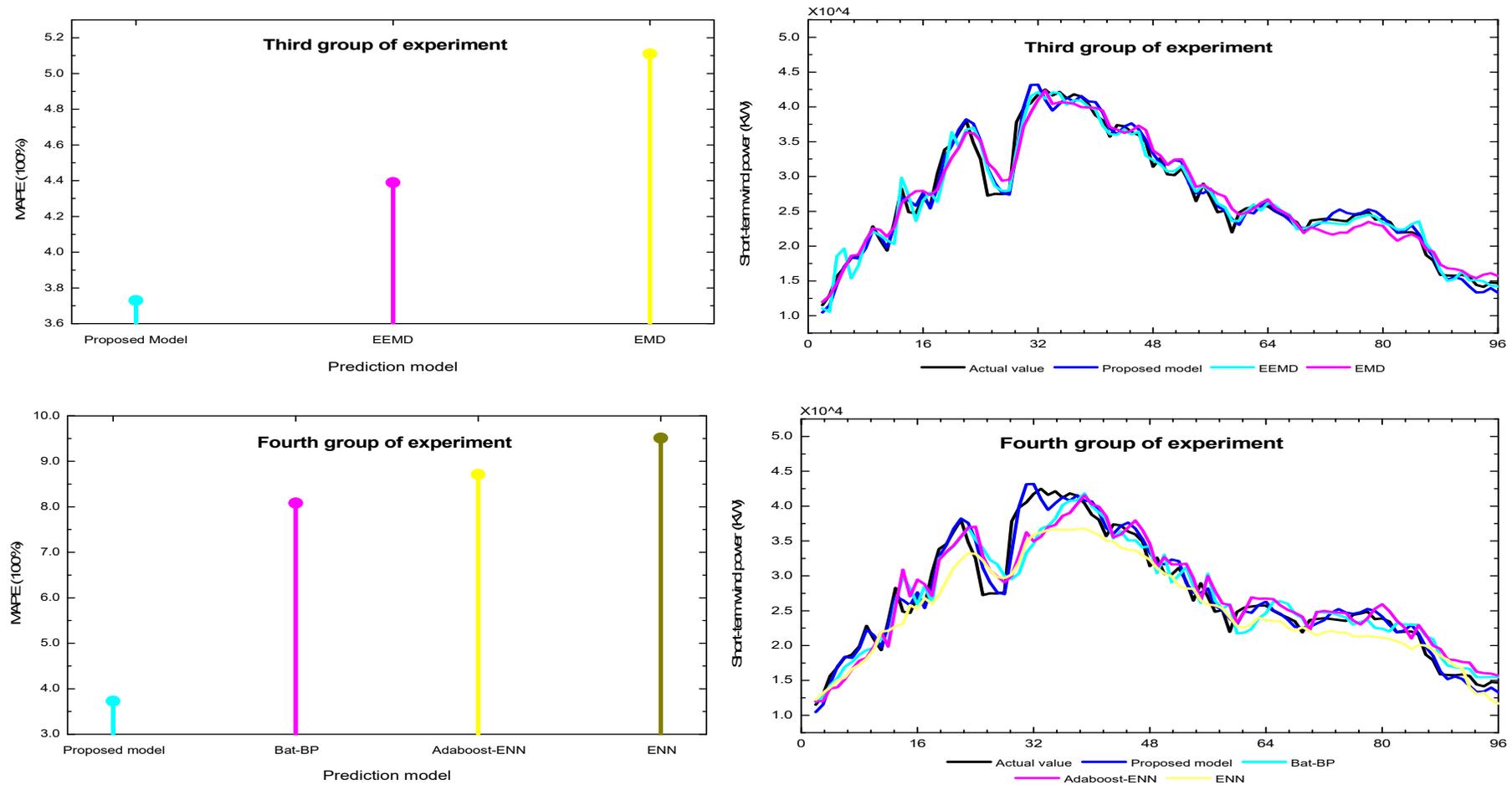
**Figure 8.** Forecast results and comparison of MAPE between different models: (**A**) Comparison of MAPE (100%) in different models from four different experiments (from top to bottom in part A; first group of experiment, second group of experiment, third group of experiment, fourth group of experiment); (**B**) comparison of forecasts by four different experiments (from top to bottom in part B; first group of experiment, second group of experiment, third group of experiment, fourth group of experiment).

## 5. Conclusions

In this paper, a novel combined model was proposed, which is based on grey correlation degree analysis, the ESMD method, sample entropy theory, and hybrid prediction models combining Bat-BP, Adaboost-ENN, and ENN. The grey correlation degree analysis was used to screen meteorological data at different times and altitudes to determine the influencing factors of original wind power series. The ESMD method was employed to decompose the original wind power and wind speed series (50 m) into several sub-series, including IMFs and a residue R, respectively. On this basis, the sample entropy theory was adopted to calculate the complexity of each wind power sub-series, and to reconstruct them into PHC, PMC, and PLC, respectively. Similarly, wind speed sub-series were reconstructed into SHC, SMC, and SLC. Subsequently, the input–output samples in PHC, PMC, and PLC were determined by PACF theory. Next, the hybrid prediction model was utilized to predict PHC, PMC, and PLC respectively, and the prediction results were aggregated to obtain the final prediction results of the original wind power series. To evaluate the prediction performance of the proposed model, the proposed model was compared with other models by the wind power and meteorological data collected from a wind farm of China.

The experimental results demonstrate that: (1) Wind power has volatile and intermittent characteristics and is easily affected by the external environment. However, a large number of studies on wind power prediction do not take into account meteorological data or simply consider the wind speed time series, which may cause the wind power predictions to deviate from the true values. Therefore, this article performed a set of comparative experiments and found that the prediction performance of wind power considering meteorological data is better than that of wind power not considering meteorological data. (2) To improve prediction accuracy, many signal decomposition techniques were applied to wind speed prediction or wind power prediction, especially EMD and EEMD. These two methods can decompose the time series into several IMFs and R. Among these, a set of experiments was performed and the prediction accuracy of the ESMD-type was higher than that of the EMD-type and EEMD-type. (3) According to different data characteristics, different prediction models were constructed by combining multidisciplinary knowledge and cross domain research. The prediction performance of the developed combined model was evidently superior compared to all the other models.

Although the proposed model achieved good prediction performance in short-term wind power prediction, there are still some problems that need to be solved. The prediction of short-term wind power in this paper involves not only the wind power series, but also the corresponding meteorological data. Therefore, the requirements for data are relatively high, which also means the study only uses data from a wind farm. In future studies, the data from other wind farms can be used to validate this model if the data is available. Moreover, this study only made one-step prediction, and multi-step prediction can be tried in the next work.

**Author Contributions:** J.Z. and X.X. conceived and designed the experiments; X.X. performed the experiments; X.H. and Y.L. analyzed the data; X.X. contributed analysis tools; X.X. wrote the paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Appendix A

**Table A1.** The statistical descriptions of wind power and meteorological data.

| Data | | Size | Mean | Std | Max | Min |
|---|---|---|---|---|---|---|
| Wind power | | 1441 | 22887.3 | 11520.2 | 46566.7 | 588.9 |
| Humidity | 10 m | 1441 | 57.7 | 18.5 | 98.3 | 23.0 |
| | 30 m | 1441 | 57.1 | 18.3 | 97.3 | 22.8 |
| | 50 m | 1441 | 56.6 | 18.1 | 96.4 | 22.6 |
| | 70 m | 1441 | 56.0 | 17.9 | 95.4 | 22.3 |
| | 100 m | 1441 | 55.2 | 17.7 | 94.0 | 22.0 |
| Pressure | 10 m | 1441 | 782.1 | 1.6 | 784.8 | 776.8 |
| | 30 m | 1441 | 779.7 | 1.6 | 782.4 | 774.4 |
| | 50 m | 1441 | 777.3 | 1.6 | 780.0 | 772.0 |
| | 70 m | 1441 | 774.9 | 1.6 | 777.6 | 769.6 |
| | 100 m | 1441 | 771.3 | 1.6 | 774.0 | 766.0 |
| Temperature | 10 m | 1441 | 18.1 | 4.0 | 26.2 | 11.1 |
| | 30 m | 1441 | 18.0 | 4.0 | 26.1 | 11.0 |
| | 50 m | 1441 | 17.9 | 4.0 | 26.0 | 10.9 |
| | 70 m | 1441 | 17.8 | 4.0 | 25.9 | 10.8 |
| | 100 m | 1441 | 17.6 | 4.0 | 25.7 | 10.6 |
| Wind direction | 10 m | 1441 | 235.6 | 12.2 | 265.0 | 185.0 |
| | 30 m | 1441 | 235.6 | 12.2 | 265.0 | 185.0 |
| | 50 m | 1441 | 235.6 | 12.2 | 265.0 | 185.0 |
| | 70 m | 1441 | 235.6 | 12.2 | 265.0 | 185.0 |
| | 100 m | 1441 | 235.6 | 12.2 | 265.0 | 185.0 |
| Wind speed | 10 m | 1441 | 4.8 | 1.1 | 7.3 | 2.7 |
| | 30 m | 1441 | 5.6 | 1.2 | 8.5 | 3.1 |
| | 50 m | 1441 | 6.0 | 1.3 | 9.0 | 3.3 |
| | 70 m | 1441 | 6.3 | 1.4 | 9.4 | 3.5 |
| | 100 m | 1441 | 6.5 | 1.4 | 9.8 | 3.6 |

**Table A2.** The statistical descriptions of the training set and testing set.

| Label | | Size | Mean | Std | Max | Min |
|---|---|---|---|---|---|---|
| Training set | wind power | 1345 | 22605.7 | 11673.8 | 46566.7 | 588.9 |
| | humidity (30 m) | 1345 | 57.5 | 18.5 | 97.3 | 22.8 |
| | pressure (50 m) | 1345 | 777.2 | 1.6 | 780.0 | 772.0 |
| | temperature (100 m) | 1345 | 17.4 | 4.0 | 25.6 | 10.6 |
| | wind direction (50 m) | 1345 | 235.2 | 12.4 | 265.0 | 185.0 |
| | wind speed (50 m) | 1345 | 6.0 | 1.3 | 9.0 | 3.3 |
| Testing set | wind power | 96 | 26833.1 | 8178.1 | 42477.8 | 11522.2 |
| | humidity (30 m) | 96 | 52.1 | 14.9 | 73.6 | 31.1 |
| | pressure (50 m) | 96 | 778.1 | 1.3 | 780.0 | 776.0 |
| | temperature (100 m) | 96 | 19.7 | 4.0 | 25.7 | 14.6 |
| | wind direction (50 m) | 96 | 241.2 | 6.4 | 253.0 | 231.0 |
| | wind speed (50 m) | 96 | 6.8 | 1.0 | 8.8 | 4.8 |

## References

1. Xiao, L.; Wang, J.; Dong, Y.; Wu, J.J.R.; Reviews, S.E. Combined forecasting models for wind energy forecasting: A case study in China. *Renew. Sustain. Energy Rev.* **2015**, *44*, 271–288. [CrossRef]
2. Liu, K.; Zhang, Y.; Liang, Q. A novel combined forecasting model for short-term wind power based on ensemble empirical mode decomposition and optimal virtual prediction. *J. Renew. Sustain. Energy* **2016**, *8*, 013104. [CrossRef]

3.  Zhao, J.; Guo, Z.H.; Su, Z.Y.; Zhao, Z.Y.; Xiao, X.; Liu, F. An improved multi-step forecasting model based on WRF ensembles and creative fuzzy systems for wind speed. *Appl. Energy* **2016**, *162*, 808–826. [CrossRef]

4.  Wu, B.; Song, M.; Kai, C.; He, Z.; Xing, Z. Wind power prediction system for wind farm based on auto regressive statistical model and physical model. *J. Renew. Sustain. Energy* **2014**, *6*, 3192–3198. [CrossRef]

5.  Cheng, W.Y.Y.; Liu, Y.; Bourgeois, A.J.; Wu, Y.; Haupt, S.E. Short-term wind forecast of a data assimilation/weather forecasting system with wind turbine anemometer measurement assimilation. *Renew. Energy* **2017**, *107*, 340–351. [CrossRef]

6.  Lydia, M.; Kumar, S.S.; Selvakumar, A.I.; Kumar, G.E.P. Linear and non-linear autoregressive models for short-term wind speed forecasting. *Energy Convers. Manag.* **2016**, *112*, 115–124. [CrossRef]

7.  Erdem, E.; Shi, J. ARMA based approaches for forecasting the tuple of wind speed and direction. *Appl. Energy* **2011**, *88*, 1405–1414. [CrossRef]

8.  Wang, Y.; Wang, J.; Zhao, G.; Dong, Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy* **2012**, *48*, 284–294. [CrossRef]

9.  Fang, T. Evaluation of a multiple linear regression model and SARIMA model in forecasting heat demand for district heating system. *Appl. Energy* **2016**, *179*, 544–552. [CrossRef]

10. Liu, H.; Erdem, E.; Jing, S. Comprehensive evaluation of ARMA–GARCH(-M) approaches for modeling the mean and volatility of wind speed. *Appl. Energy* **2011**, *88*, 724–732. [CrossRef]

11. Maatallah, O.A.; Achuthan, A.; Janoyan, K.; Marzocca, P. Recursive wind speed forecasting based on Hammerstein Auto-Regressive model. *Appl. Energy* **2015**, *145*, 191–197. [CrossRef]

12. Wang, J.Z.; Wang, Y.; Jiang, P. The study and application of a novel hybrid forecasting model—A case study of wind speed forecasting in China. *Appl. Energy* **2015**, *143*, 472–488. [CrossRef]

13. Wang, J.; Sun, J.; Zhang, H. Short-term wind power forecasting based on support vector machine. In Proceedings of the 2013 5th International Conference on Power Electronics Systems and Applications (PESA), Hong Kong, China, 11–13 December 2013; pp. 1–5.

14. Chi, Z.; Wei, H.; Xie, L.; Yu, S.; Zhang, K. Direct interval forecasting of wind speed using radial basis function neural networks in a multi-objective optimization framework. *Neurocomputing* **2016**, *205*, 53–63.

15. Xu, Z.; Song, X.; Zhang, X.; Bai, J. A monte-carlo-based network method for source positioning in bioluminescence tomography. *J. Biomed. Imaging* **2014**, *2007*, 10. [CrossRef] [PubMed]

16. Wang, J.; Fang, K.; Pang, W.; Sun, J. Wind Power Interval Prediction Based on Improved PSO and BP Neural Network. *J. Electr. Eng. Technol.* **2017**, *12*, 989–995. [CrossRef]

17. Wei, S.; Wang, Y. Short-term wind speed forecasting based on fast ensemble empirical mode decomposition, phase space reconstruction, sample entropy and improved back-propagation neural network. *Energy Convers. Manag.* **2018**, *157*, 1–12.

18. Ren, C.; An, N.; Wang, J.; Li, L.; Hu, B.; Shang, D. Optimal parameters selection for BP neural network based on particle swarm optimization: A case study of wind speed forecasting. *Knowl.-Based Syst.* **2014**, *56*, 226–239. [CrossRef]

19. Wang, S.; Na, Z.; Lei, W.; Wang, Y. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew. Energy* **2016**, *94*, 629–636. [CrossRef]

20. Jiang, P.; Ge, Y.; Wang, C. Research and application of a hybrid forecasting model based on simulated annealing algorithm: A case study of wind speed forecasting. *J. Renew. Sustain. Energy* **2016**, *8*, 226–239. [CrossRef]

21. Yuan, T.; Zhu, N.; Shi, Y.; Chang, C.; Yang, K.; Ding, Y. Sample data selection method for improving the prediction accuracy of the heating energy consumption. *Energy Build.* **2018**, *158*, 234–243. [CrossRef]

22. Ardalani-Farsa, M.; Zolfaghari, S. Chaotic Time Series Prediction with Residual Analysis Method Using Hybrid Elman—NARX Neural Networks. *Neurocomputing* **2010**, *73*, 2540–2553. [CrossRef]

23. Wang, J.; Zhang, W.; Li, Y.; Wang, J.; Dang, Z. Forecasting wind speed using empirical mode decomposition and Elman neural network. *Appl. Soft Comput.* **2014**, *23*, 452–459. [CrossRef]

24. Liu, H.; Tian, H.Q.; Liang, X.F.; Li, Y.F. Wind speed forecasting approach using secondary decomposition algorithm and Elman neural networks. *Appl. Energy* **2015**, *157*, 183–194. [CrossRef]

25. Yu, C.; Li, Y.; Zhang, M. An improved Wavelet Transform using Singular Spectrum Analysis for wind speed forecasting based on Elman Neural Network. *Energy Convers. Manag.* **2017**, *148*, 895–904. [CrossRef]

26. Kai, L.I.; Huang, H.K. An Approach to Improving Diversity of Neural Network Ensemble. *Acta Electron. Sin.* **2005**, *33*, 1387–1390.

27. Liu, H.; Tian, H.Q.; Li, Y.F.; Zhang, L.J.E.C. Management, Comparison of four Adaboost algorithm based artificial neural networks in wind speed predictions. *Energy Convers. Manag.* **2015**, *92*, 67–81. [CrossRef]

28. Shao, H.; Xing, D.; Fang, C. Short-term Wind Speed Forecasting using the Wavelet Decomposition and AdaBoost Technique in Wind Farm of East China. *IET Gener. Transm. Distrib.* **2016**, *10*, 2585–2592. [CrossRef]

29. Ling, X.; Dong, Y.; Yao, D. An improved combination approach based on Adaboost algorithm for wind speed time series forecasting. *Energy Convers. Manag.* **2018**, *160*, 273–288.

30. Zhang, W.; Qu, Z.; Zhang, K.; Mao, W.; Ma, Y.; Fan, X. A combined model based on CEEMDAN and modified flower pollination algorithm for wind speed forecasting. *Energy Convers. Manag.* **2017**, *136*, 439–451. [CrossRef]

31. An, X.; Jiang, D.; Zhao, M.; Chao, L. Short-term prediction of wind power using EMD and chaotic theory. *Commun. Nonlinear Sci. Numer. Simul.* **2012**, *17*, 1036–1042. [CrossRef]

32. Wu, Q.; Peng, C. Wind Power Generation Forecasting Using Least Squares Support Vector Machine Combined with Ensemble Empirical Mode Decomposition, Principal Component Analysis and a Bat Algorithm. *Energies* **2016**, *9*, 261. [CrossRef]

33. Song, J.; Wang, J.; Lu, H. A novel combined model based on advanced optimization algorithm for short-term wind speed forecasting. *Appl. Energy* **2018**, *215*, 643–658. [CrossRef]

34. Wang, J.; Li, Z. The ESMD Method for Climate Data Analysis. *Clim. Chang. Res. Lett.* **2014**, *3*, 4.

35. Zheng, Q.; Rehman, S.; Alam, M.M.; Alhems, L.M.; Lashin, A. Decomposition of wind speed fluctuations at different time scales. *J. Earth Syst. Sci.* **2017**, *126*, 36. [CrossRef]

36. Hu, J.; Wang, J. Short-term wind speed prediction using empirical wavelet transform and Gaussian process regression. *Energy* **2015**, *93*, 1456–1466. [CrossRef]

37. Wang, J. A Hybrid Wavelet Transform Based Short-Term Wind Speed Forecasting Approach. *Sci. World J.* **2014**, *2014*, 914127. [CrossRef] [PubMed]

38. Wang B, L.X.D. Multi-scale fluctuation of European temperature revealed by EEMD analysis. *Acta Sci. Nat. Univ. Pekin.* **2011**, *47*, 627–635.

39. Zhang, Y.; Zhao, X.; Shen, Q.; Fang, Y. A New Method for Fault Line Selection of Small Current System Based on ESMD. *Electromech. Inf.* **2015**, *18*, 126–127.

40. Feng, C.; Cui, M.; Hodge, B.-M.; Zhang, J. A data-driven multi-model methodology with deep feature selection for short-term wind forecasting. *Appl. Energy* **2017**, *190*, 1245–1257. [CrossRef]

41. Iversen, E.B.; Morales, J.M.; Møller, J.K.; Madsen, H. Short-term probabilistic forecasting of wind speed using stochastic differential equations. *Int. J. Forecast.* **2016**, *32*, 981–990. [CrossRef]

42. Xiao, L.; Wang, J.; Hou, R.; Wu, J. A combined model based on data pre-analysis and weight coefficients optimization for electrical load forecasting. *Energy* **2015**, *82*, 524–549. [CrossRef]

43. Wang, J.L.Z. *Pole Symmetry Mode Decomposition Method*; Higher Education Press: Beijing, China, 2015.

44. Shao, J. Improving Nowcasts of Road Surface Temperature by a Backpropagation Neural Network. *Weather Forecast.* **1997**, *13*, 164–171. [CrossRef]

45. Demuth, H.B.; Beale, M.H.; Jess, O.D.; Hagan, M.T. *Neural Network Design*; Oklahoma State University: Stillwater, OK, USA, 2002.

46. Liu, C.P.; Chun-Ming, Y.E. Bat Algorithm with Chaotic Search Strategy and Analysis of Its Property. *J. Syst. Simul.* **2013**, *25*, 1183–1195.

47. Chen, W.H.; Yan, X.H.; Hua, F.U. On the innovated Elman neural network for forecasting the mining gas emission. *J. Saf. Environ.* **2015**, *15*, 19–24.

48. Wang, G.G.; Guo, L.H.; Duan, H.; Liu, L.; Wang, H.Q. The Model and Algorithm for the Target Threat Assessment Based on ElmanAdaBoost Strong Predictor. *Acta Electron. Sin.* **2012**, *40*, 91–96.

49. Niu, D.; Pu, D.; Dai, S. Ultra-short-term wind-power forecasting based on the weighted random forest optimized by the niche immune lion algorithm. *Energies* **2018**, *11*, 1098. [CrossRef]

50. Jiang, P.; Liu, F.; Wang, J.; Song, Y. Cuckoo search-designated fractal interpolation functions with winner combination for estimating missing values in time series. *Appl. Math. Model.* **2016**, *40*, 9692–9718. [CrossRef]

51. Yang, F. A detection-theoretic generalization of on-line learning and application to boosting. *J. Comput. Syst. Sci.* **2005**, *13*, 663–671.

52. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039. [CrossRef]

53. Yang, Z. *Study of Emotional Change Based on the Analysis of the Sample Entropy*; Dalian Jiaotong University: Dalian, China, 2017.

54. Zhang, G.; Wu, Y.; Liu, Y. An advanced wind speed multi-step ahead forecasting approach with characteristic component analysis. *J. Renew. Sustain. Energy* **2014**, *6*, 1663–1672. [CrossRef]

55. Xianzi, Z.C.Y. Forecasting of China's regional carbon market price based on multi-frequency combined model. *Syst. Eng.-Theory Pract.* **2016**, *36*, 3017–3025.