Research on Combined Model Based on Multi-Objective Optimization and Application in Wind Speed Forecast

Shenghui Zhang 1, Yuewei Liu 1,* , Jianzhou Wang 2 and Chen Wang 3

1 School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China; zhangshh17@lzu.edu.cn
2 School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China; wangjz@dufe.edu.cn
3 School of Information Science and Engineering, Lanzhou University, Lanzhou 730000, China; chenwang15@lzu.edu.cn
* Correspondence: lyw@lzu.edu.cn; Tel.: +86-187-9311-7821

Received: 3 December 2018; Accepted: 17 January 2019; Published: 27 January 2019

Abstract: Wind power is an important part of a power system, and its use has been rapidly increasing as compared with fossil energy. However, due to the intermittence and randomness of wind speed, system operators and researchers urgently need to find more reliable wind-speed prediction methods. It was found that the time series of wind speed not only has linear characteristics, but also nonlinear. In addition, most methods only consider one criterion or rule (stability or accuracy), or one objective function, which can lead to poor forecasting results. So, wind-speed forecasting is still a difficult and challenging problem. The existing forecasting models based on combination-model theory can adapt to some time-series data and overcome the shortcomings of the single model, which achieves poor accuracy and instability. In this paper, a combined forecasting model based on data preprocessing, a nondominated sorting genetic algorithm (NSGA-III) with three objective functions and four models (two hybrid nonlinear models and two linear models) is proposed and was successfully applied to forecasting wind speed, which not only overcomes the issue of forecasting accuracy, but also solves the difficulties of forecasting stability. The experimental results show that the stability and accuracy of the proposed combined model are better than the single models, improving the mean absolute percentage error (MAPE) range from 0.007% to 2.31%, and the standard deviation mean absolute percentage error (STDMAPE) range from 0.0044 to 0.3497.

Keywords: multi-objective optimization; wind speed forecasting; combined model

1. Introduction

In recent years, wind energy has become the focus of managers and researchers in the energy field, due to the advantages of wind power, such as renewability and cleanliness.

In 2016, the global installed capacity of wind power exceeded 54 GW. This installed capacity is distributed between 90 countries, which have an installed capacity of more than 10 GW, with 29 countries having an installed capacity of 1 GW [1]. Cumulative installed capacity increased by 12.6%, and accumulated capacity reached 486.8 GW. In the United States, more than 53,000 wind turbines operate, generating more than 84.1 GW of electricity in 41 states [2]. On the other side of the globe, wind power may become China’s second largest power source by 2050 [3].

Accurate forecasting of wind speed is an important prerequisite for using wind power due to the following reasons: 1) it reduces rotating and operating costs of wind-farm equipment; 2) it helps dispatching departments to adjust their plans in time; 3) it reduces the impact on the entire power grid;
4) it effectively reduces or avoids the negative impact of wind farms on the power system, and 5) it improves the ability of wind power in power market competition.

However, because of the chaotic and random fluctuations of wind speed, it is a challenging and difficult task to obtain satisfactory wind-speed forecasting results [4–6]. To have satisfying results and reduce errors, and improve the accuracy and stability of the forecasting results, various wind-speed forecasting methods have been proposed and developed by former researchers. The methods can be classified in four categories: 1) physical models; 2) statistical models; 3) artificial-intelligence models; and 4) spatial-correlation models [7].

The common physical models include weather research and forecast, the consortium for small-scale modeling, and mesoscale model 5 (MM5) [8]. These methods are based on physics and the atmosphere, and forecast wind speed with additional background information [8]. Statistical models include the fuzzy methods [9], Autoregressive Moving Average (ARMA) model [10], the Autoregressive Integrate Moving Average (ARIMA) model [11,12], and the grey model [13], combined with neurofuzzy techniques [14] and Markov chains [15]. Statistical methods aim to mine the relationship between historical data and establish prediction models, including traditional statistical and machine-learning models, to describe the potential sampling of wind-speed forecasting [16]. In recent years, because of their powerful nonlinear character, many artificial-intelligence forecasting methods, such as artificial neural networks (ANN) [17–25] and support vector machines (SVM) [26,27], have been widely applied to wind-speed forecasting. The wind speed of the spatial relationship between different sites is considered in spatial-correlation models [28] to forecast wind speed.

However, these four categories of methods have disadvantages because of their character: 1) physical models need accurate numerical weather-prediction data and detailed information of wind-farm locality, input parameters, and data acquisition. Processing and calculation are complex [29,30]; 2) traditional statistical models can address forecasting with linear trends well, but times-series wind speed is always random, chaotic, and nonlinear, and the performance of these models does not satisfy researchers and managers; 3) although artificial-intelligence forecasting methods have some advantages, due to their nonlinear mapping capability compared with traditional statistical models, their main disadvantages and shortcomings are that they are easy to reach local optimal values, and they suffer from overfitting and slow convergence [31]; 4) spatial-correlation methods should consider many influential factors, such as the relationship of the time-series wind-speed data between different sites, original data from several stations, and if the time-series wind speed of the forecasting point and its neighbors is used for forecasting [28].

To overcome the problem of forecasting accuracy and stability and take advantage of single models, a combined model based on multi-objective optimization and combined theory is proposed by us to overcome their defects, including low accuracy and weak stability. Because wind-speed forecasting is a complex study, single-objective optimization is not enough to solve the wind-speed forecasting problem. Solving the single-objective optimization problem (SOP) is a direct task, but multiple-objective problems (like the wind-speed forecasting problem) are often complex, and objective functions are competing (or conflicting) with each other. Due to this reason, there is no single optimal solution, commonly known as a series of Pareto-optimal solutions, which simultaneously optimize all objective functions. A set of optimal solutions, called a Pareto-optimal Set (PS), of which mapping in the target space is named the Pareto front (PF), and not a single optimal solution, is the goal of solving the multiple-objective optimization problem (MOP). In MOPs, the favored solution from the PS is chosen by researchers, but not the optimal solution. Obviously, it is difficult to select one objective function from a number of objective functions and ensure that the selected objective function can achieve a lower mean absolute percentage error (MAPE) and stronger stability. Due to these reasons, multiple-objective optimization can solve the low accuracy and weak stability of the wind-speed forecasting problem through simultaneously searching the goal of multiple functions.

In this paper, a new combined model [32] that integrates nonpositive constraint theory [33], four branch models, including two nonlinear hybrid neural networks (Cuckoo Search-Back-Propagation
Neural Network (CS-BPNN) and DE-Online Extreme Learning Machine (DE-OSELM)), two linear models (ARIMA and HW), and a nondominated sorting genetic algorithm (NSGA-III) [34], which optimizes the weights of the branch models, is proposed in this paper. Ten-minute time-series wind-speed data from three wind-farm sites were applied to examine our proposed model. We also chose three combined models with two objective functions to test its performance.

The major contributions and innovations of this paper are as follows:

1. According to the observed time-series wind-speed data, the trajectory matrix was constructed, decomposed, and reconstructed to extract signals representing different components of the original time series, such as long-term trend signals, periodic signals, and noise signals, so that the structure of the time series could be analyzed and used for further forecasting.

2. Our proposed combined model is based on MOP theory, which can obtain both accuracy and stability. The wind-speed forecasting problem is a MOP, so the theory overcomes the difficulty of selecting one objective function from multiple functions to obtain higher precision and stronger stability.

3. Due to the wind-speed data having both linear and nonlinear characteristics, linear models (ARIMA and HW) and nonlinear models (BPNN and OSELM) were chosen to be the branch models of our proposed combined model. The combination of these two kinds of models can solve the wind-speed forecast problem with both linear and nonlinear characteristics.

4. Our novel developed combined model with three objective functions, which can be applied to forecast wind speed, is proposed in this paper. As compared with other models, the proposed model not only guarantees accuracy, but also has strong stability. The results of the experiments mean that the proposed integrated combined model is a more effective model for wind-speed forecasting and wind-farm management.

5. The forecasting performance of the combined model was evaluated scientifically and comprehensively. The evaluation system used five experiments and four performance metrics to effectively evaluate the forecasting accuracy and stability of the proposed combined model.

The remainder of the paper is arranged as follows. The strategy of the proposed combined model is shown in Section 2. Singular-spectrum analysis is presented in Section 3. Nonlinear back propagation, an extreme learning machine neural network, two linear models, autoregressive integrated moving average and Holt–Winters, heuristic algorithms, and the optimization procedure are introduced in Section 4. In Section 5, we show our proposed combined model. In Section 6, forecasting performance metrics, the forecasting results of individual models and of the proposed combined model, and comparisons are discussed, and the views and results of the entire paper are summarized. Finally, Section 7 concludes the study.

2. Strategy of Our Proposed Combined Model

(1) Preprocessing the time-series wind-speed data with denoising and reconstruction from China.

(2) Setting the train data and test data according to the character of the time-series wind-speed data; the interval of each two data items is ten minutes.

(3) Two kinds of models were selected as single models to build the combined model. The linear models include ARIMA and HW, and the nonlinear models include CS-BPNN and DE-OSELM, which were optimized by cuckoo search and differential evolution. The parameter values are shown in Table 1.

(4) Building a combined forecasting system model, which is based on nonpositive constraint theory, NSGA-III, which optimizes the weights of the branch models and four single models, which are mentioned before.

(5) Forecasting results were evaluated, and testing of the combined model is at the end.
Table 1. Four metrics.

| Metric | Definition | Equation |
|--------|------------|----------|
| AE | The average forecast error of n times forecast results | \( \text{AE} = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n) \) |
| MAE | The average absolute forecast error of n times forecast results | \( \text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |y_n - \hat{y}_n| \) |
| MSE | The average of the prediction error squares | \( \text{MSE} = \frac{1}{N} \sum_{n=1}^{N} (y_n - \hat{y}_n)^2 \) |
| MAPE | The average of absolute percentage error | \( \text{MAPE} = \frac{1}{N} \sum_{n=1}^{N} \left| \frac{y_n - \hat{y}_n}{y_n} \right| \times 100\% \) |

3. Results

The denoising process consists of four steps, embedding, singular-value decomposition, grouping, and diagonal averaging [35]. According to the observed time-series wind-speed data and these four steps, the trajectory matrix was constructed. The trajectory matrix was decomposed and reconstructed to extract signals representing different components of the original time series, such as long-term trend signals, periodic signals, and noise signals. In order to obtain satisfying results, the noise signals were abandoned. The detailed process is as follows:

1. Embedding:
   
   Form the trajectory matrix of the series \( X \), which is the \( L \times K \) matrix:
   
   \[
   X = [x_1, \ldots, x_N] = (x_{ij})_{i,j=1}^{L,K} = \begin{bmatrix}
   x_1 & x_2 & x_3 & \cdots & x_K \\
   x_2 & x_3 & x_4 & \cdots & x_{K+1} \\
   \vdots & \vdots & \vdots & \ddots & \vdots \\
   x_L & x_{L+1} & x_{L+2} & \cdots & x_N
   \end{bmatrix}
   \]

   where \( X = (x_1, \ldots, x_{1\L-1})^T, (1<i<K) \) are lagged vectors of size \( L \). Matrix \( X \) is a Hankel matrix that means that \( X \) has equal elements \( x_{ij} \) on the anti-diagonals \( i+j=\text{const} \).

2. Singular-Value Decomposition:
   
   Perform the singular value decomposition (SVD) of trajectory matrix \( X \). Set \( S = XX^T \) and denote by \( \lambda_1, \ldots, \lambda_L \) the eigenvalues of \( S \) taken in the decreasing order of magnitude \( \lambda_1 \geq \cdots \geq \lambda_L \geq 0 \), and by \( U_1, \ldots, U_L \) the orthonormal system of the eigenvectors of matrix \( S \) corresponding to these eigenvalues.

   Set \( d = \text{rank} X = \max\{i, \text{ such that } \lambda_i > 0\} \) (note that \( d = L \) for a typical real-life series) and \( V_i = X^T U_i / \sqrt{\lambda_i} (i = 1, \ldots, d) \). In this notation, the SVD of trajectory matrix \( X \) can be written as \( X = X_1 + \cdots + X_N \).

   Where \( X_i = \sqrt{\lambda_i} U_i V_i^T \) are matrices having rank 1; these are called elementary matrices. Collection \( \{\sqrt{\lambda_i}, U_i, V_i^T\} \) is the \( i \)th eigentriple (ET) of the SVD. Vectors \( U_i \) are the left singular vectors of matrix \( X \), and numbers \( \sqrt{\lambda_i} \) are the singular values and provide the singular spectrum of \( X \); this gives the name to SSA. Vectors \( V_i \sqrt{\lambda_i} = X_i U_i \) are the vectors of principal components (PCs).

3. Eigentriple grouping:
   
   Partition set of indices \( \{1, \ldots, d\} \) into \( m \) disjoint subsets \( I_1, \ldots, I_m \).

   Let \( I = \{i_1, \ldots, i_p\} \). Then, resultant matrix \( X_I \) corresponding to group \( I \) is defined as \( X \) is a Hankel matrix transformed into a new series of length \( N \) using the one-to-one correspondence between Hankel matrices and the time series. Diagonal averaging applied to a resultant matrix \( X_{I_i} \) produces a
reconstructed series \( \tilde{X}^{(k)} = (\tilde{X}_1^{(k)}, \ldots, \tilde{X}_N^{(k)}) \). In this way, the initial series \( x_1, \ldots, x_N \) is decomposed into a sum of \( m \) reconstructed subseries:

\[
x_n = \sum_{k=1}^{m} \tilde{x}_n^{(k)}, \quad (n = 1, 2, \ldots, N)
\]

This decomposition is the main result of the SSA algorithm. The decomposition is meaningful if each reconstructed subseries could be forecast as a part of either a trend, some periodic component, or noise.

The pseudocode for denoising is provided in Appendix A, Algorithm 1.

4. Methods and Heuristic Algorithm

The linear and nonlinear models selected to build the combined system are shown in this section. The heuristic algorithm, cuckoo search, and differential evolution to optimize BPNN and OSELM are also introduced in this section.

4.1. Nonlinear Models

Because of the character of wind-speed data, nonlinear models can obtain excellent performance in forecasting. In this paper, two models, BPNN and ELM, were selected from nonlinear models and the parameter was optimized by the algorithm to obtain better performance.

4.1.1. Back-Propagation Neural Network (BPNN) Model

BPNN is a type of multilayer feed-forward neural network with a wide variety of applications. It is based on a gradient-descent method that minimizes the sum of the squared errors between actual and desired output values. The transfer function is of the neuron type. The output function is between 0 and 1, and can transform input to output for continuous nonlinear mapping [36].

The topology of the BPNN is as follows:

\[
X' = \{X'i\} = 2 \times \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} - 1, \quad (i = 1, 2, \ldots, n), \quad X' \subset [-1, 1]
\]

where \( X_{\text{min}} \) and \( X_{\text{max}} \) are the minimum and maximum value of the input array or output vectors, and \( X'i \) denotes the real value of each vector.

Step 1. Calculate outputs of all hidden layer nodes:

\[
y_j = f(\sum_i w_{ij}x_i + b_j) = f(\text{net}_j) \quad (i = 1, \ldots, n ; j = 1, \ldots, 2n)
\]

\[
\text{net}_j = \sum_i w_{ij}x_i + b_j \quad (j = 1, \ldots, 2n)
\]

where the activation value of node \( j \) is \( \text{net}_j \), \( w_{ij} \) representing the connection weight from input node \( i \) to hidden node \( j \), \( b_j \) represents the bias of neuron \( j \), \( y_j \) represents the output of hidden layer node \( j \), and \( f \) is the activation function of a node, which is usually a sigmoid function.

Step 2. Calculate the output data of the neural network:

\[
O_1 = f_0(\sum_j w_{0j}y_j + b_0), \quad (i = 1, \ldots, 2n)
\]

where \( w_{0j} \) represents the connection threshold from hidden node \( j \) to the output node, \( b_0 \) represents the bias of the neuron, \( O_1 \) represents the output data of the network, and \( f_0 \) is the activation function of the output layer node.
Step 3. Minimize the global error via the training algorithm:

\[
\text{Mean Square Error} = \frac{1}{m} \sum (O_1 - Z)^2
\]  

(7)

where \(Z\) represents the real data vector of the output, \(m\) represents the number of output.

4.1.2. CS Algorithm

CS, which was proposed by Yang and Deb [37] in 2009, is derived from the action of cuckoos laying their eggs in other birds’ nests to let those birds hatch the eggs for them. However, once the host birds discover the cuckoo eggs, the host birds throw them away or abandon their nests and rebuild a new nest elsewhere. The CS algorithm is constructed based on three assumptions: a) only one egg is randomly laid by each cuckoo in a selected nest; b) the following generations would begin in the best nest; and c) it is a constant of the number of available host nests, and the probability value of the host bird discovering the egg laid by a cuckoo is \(p\), which has the range of 0 to 1. In CS, every nest stands for a solution. The steeps of BPNN optimized by CS are shown in Figure 1.

![Figure 1. The flow chart of hybrid Back-Propagation Neural Network (BPNN).](image)

4.1.3. OSELM

An Extreme Learning Machine (ELM) is a simple and effective single hidden-layer feed-forward neural network proposed by Huang et al. [38]. In many practical applications, however, learning is a continuous process. For this reason, Liang et al. proposed an online learning algorithm, the OSELM [39].

For different N training sample \(Z\), \(Z = \{(x_i, t_i) | i = 1, \ldots, N\}\), where \(x_i = [x_{i1}, x_{i2}, \ldots, x_{in}] \in R^n\), \(t_i = [t_{i1}, t_{i2}, \ldots, t_{im}] \in R^m\), ELM with \(L\) hidden layer nodes, and \(g(x)\) activation function can approximate arbitrary N samples with zero errors by arbitrarily specifying \(a_j\) and \(b_j\), which can be expressed as follows:

\[
O_i = \sum_{j=1}^{L} \beta_j g(a_j, b_j, x_i) = t_i
\]  

(8)
where \(a_j\) is the input weight, \(b_j\) is the threshold of hidden layer nodes, \(x_i\) is the input vector, \(O_i\) is the output vector, and \(\beta_i\) is the output weight \(\beta_{L \times m} = [\beta_1, \beta_2, \ldots, \beta_L]^T\).

Formula (8) can be simplified as:

\[
H\beta = T
\]  
(9)

where, \(\beta_{L \times m} = [\beta_1, \beta_2, \ldots, \beta_L]^T\), \(T_{N \times m} = [T_1, T_2, \ldots, T_N]^T\), and \(H\) is the hidden layer output matrix.

\[
H(a_1, \ldots, a_L; b_1, \ldots, b_L; x_1, \ldots, x_N) = \begin{bmatrix}
g(a_1, b_1, x_1) & \cdots & g(a_L, b_L, x_1) \\
\vdots & \ddots & \vdots \\
g(a_1, b_1, x_N) & \cdots & g(a_L, b_L, x_N)
\end{bmatrix}_{N \times L}
\]

represents the output of the \(j\) hidden layer node for input \(x_1, x_2, \ldots, x_N\).

After the pseudoinverse matrix is incorporated, the least-squares solution of the above linear system is:

\[
\hat{\beta} = (H^TH)^{-1}H^TT \\
(10)
\]

Based on a recursive least-squares algorithm, the algorithm flow of OSELM can be described as follows:

1. **Initialization Stage**
   - Given activation function \(g\), the number of hidden layer nodes is \(L\). A small training set is given to initialize the network and obtain the initial value. The output weight value is \(\beta(0)\). Take \(k = 0\), where \(k\) is the number of data segments sent to the network, which is represented.

2. **Online Learning Stage**
   - Given the \(k + 1\) data segment, calculate output weight beta \(\beta^{(k+1)}\), take \(k = k + 1\), then return to the online learning stage, and constantly update the calculated output weight until data learning is completed.

4.1.4. Differential Evolution

The differential-evolution algorithm is a parallel, direct, and random-search algorithm based on group evolution [40]. Firstly, the population can be randomly initialized in the feasible solution space of the problem. The population can be represented by \(NP\) (population size) \(D\) (number of decision variables), dimension parameter \(x_{hl}\), where \(h = 1, 2, \ldots, NP\), \(l = 1, 2, \ldots, D\).

Two different individual vectors are randomly selected and subtracted to generate difference vectors. The difference vectors are weighted and added to the third random selected individual vectors to generate variation vectors. This operation is called variation. Utilizing Formula (11) to implement a mutation operation on each individual \(x_h^l\) at \(t\)-time, the corresponding mutation individual \(v_h^{l+1}\) is obtained, that is:

\[
v_h^{l+1} = x_h^l + K(x_{r_2}^l - x_{r_3}^l) \\
(11)
\]

where, \(r_1, r_2, r_3 \in \{1, 2, \ldots, NP\}\) are different, and different to \(h\).

The variation vector is mixed with the target vector to generate test vectors. This process is called a crossover. Formula (12) is used to cross-operate \(x_h^l\) with variant individual \(v_h^{l+1}\) generated by Formula (11) to generate experimental individual \(u_h^{l+1}\), that is:

\[
u_h^{l+1} = \begin{cases} 
  v_h^{l+1}, & \text{if } (rand(l) \leq CR) \text{ or } l = rnbr(h); \\
  x_h^{l+l'}, & \text{otherwise.}
\end{cases}
(12)
\]

where, \(rand(l)\) is a uniformly distributed random number in the range of [0,1]; \(CR\) is a cross probability in the range of [0,1]; and \(rnbr(h)\) is a random variable in the range of \([1, 2, \ldots, D]\).

If the fitness of the test vector is better than that of the target vector, the next generation is formed by replacing the target vector with the test vector. This operation is called selection. Fitness functions \(f\)
of $u^{t+1}_h$ and $x^t_i$ were compared by Formula (6). For the minimization problem, the individual with a low fitness function value was selected as the individual $x^{t+1}_h$ of the new population, that is:

$$x^{t+1}_h = \begin{cases} u^{t+1}_h, & \text{if } f(u^{t+1}_h) < f(x^t_h); \\ x^t_h, & \text{otherwise.} \end{cases}$$

(13)

where $J$ is the fitness function. The Figure 2 shows the flow chart of hybrid OSELM.

![Figure 2. The flow chart of hybrid Online Extreme Learning Machine (OSELM).](image)

### 4.2. Linear Models

Although wind-speed data are usually nonlinear, according to our test in Section 6.1, the time-series wind-speed data also have a linear character. We can say that the linear models used to build the combined system are correct and appropriate. In this section, two linear models are briefly introduced.

#### 4.2.1. ARIMA

The ARIMA model is one of the most popular forecasting models [41]. The ARIMA model can be expressed as follows:

$$y_t = \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \cdots + \varphi_p y_{t-p} + \epsilon - \theta_q \epsilon_{t-q}$$

(14)

where $y_i (i = 1, 2, \ldots, t)$ is the actual value, $\epsilon_j (i = 1, 2, \ldots, t)$ is the random error at time $t$, $\varphi_i$ and $\theta_i$ represent the coefficients, and $p$ and $q$ are internumbers that are often referred to as autoregressive and moving polynomial factors, respectively.

#### 4.2.2. Holt–Winters (HW)

The output of the HW method is written as $F_{t+m}$, an estimate of the value of $x$ at time $t+m$, $m > 0$ based on the raw data up to time $t$; suppose we have a sequence of observations $\{x_t\}$ beginning at time $t = 0$ with a cycle of seasonal change of length $L$. The formula and recursive updating equations are the following:

$$F_{t+m} = s_t + mb_t + c_{t-L+1+(m-1)}$$

(15)

where $s_t = \alpha(x_t - c_{t-L}) + (1 - \alpha)(s_{t-1} + b_{t-1})$, $b_t = \beta(s_t - s_{t-1}) + (1 - \beta)b_{t-1}$, $c_t = \gamma(x_t - s_{t-L} - b_{t-1}) + (1 - \gamma)c_{t-L}$. $\alpha$ is the data-smoothing factor, $\beta$ is the trend-smoothing factor, $\gamma$ is the seasonal change-smoothing factor, and these three parameters are between 0 to 1. $|s_1|$ and $|b_1|$ represent...
the smoothed value of the constant part for time \( t \) and the sequence of best estimates of the linear trend that are superimposed on the seasonal changes, respectively. \( \{c_t\} \) is the sequence of seasonal-correction factors [42].

5. Our Proposed Combined Model

The combined model theory, MOPs, our proposed combined system with three objective functions, and the compared combined model with two objective functions are presented in this section. The flowchart is shown in Figure 3.

Flow chart of the wind speed forecasting propose model

![Flow chart of the wind speed forecasting propose model](image)

**Figure 3.** The flow chart of our proposed combined model.

5.1. Combined-Model Theory

The forecasting model based on combination, which was initiated by Bates and Granger [32], has long been considered as representing an improvement over individual models and also as an efficient and simple way to perfect forecasting accuracy and stability. A new combined system that consolidates...
several neural networks, NSGA-III, and nonpositive constraint theory [34] was successfully proposed in this paper.

**Definition 1.** Let $\hat{x}_{j,t}$ denote the unbiased out-of-sample forecast for $x_t$, which is obtained by the $j$th individual model. Then, the combined output at time $t$ of the combining methods has the following weighted average form [43,44]:

$$\hat{x}_{c,t} = \sum_{j=1}^{m} w_j \hat{x}_{j,t}, \quad t = 1, 2, \ldots$$

where $\hat{x}_{c,t}$ is the combined output, $m$ is the number of the component models, and $w_j$ is the weight on the $j$th component model. These weights have no limitation in the range of $[0,1]$. The experiment results show that the combination model can obtain desirable results when the weight vector has a value in the range of $[-2,2]$ [33].

5.2. Multiobjective Optimization Problem

Generally, MOPs can be classified into two groups: constrained and unconstrained problems. Constrained problems with $J$ inequality and $K$ equality constraints can be formulated as:

**Definition 2.**

Minimize $F(x) = (f_1(x), f_2(x), \ldots, f_M(x))^T$

s.t. $g_j(x) \geq 0$, $j = 1, 2, \ldots, J$,

$h_k(x) = 0$, $k = 1, 2, \ldots, K$,

$x \in \Omega$. (18)

where $M$ is the number of objectives, and $M \geq 4$, and $x = (x_1, x_2, \ldots, x_n)^T$ is the decision vector, where $n$ is the number of decision variables. In MOP (1), $\Omega = \prod_{i=1}^{n} [x_i^L, x_i^U] \subseteq R^n$ is called the decision space, where $x_i^L$ and $x_i^U$ are the lower and upper bounds of the decision variable $x_i$, respectively.

When the inequality and equality constraints in MOP (1) were omitted, then the unconstrained (with box constraints only) problems were obtained, which can be stated as follows:

**Definition 3.**

Minimize $F(x) = (f_1(x), f_2(x), \ldots, f_M(x))^T$

s.t. $x \in \Omega$. (20)

In order to solve MOP, the NSGA-III is presented in the next section.

5.3. Introduction of Objective Functions and NSGA-III

Since wind-speed prediction is not a single-objective problem, it is necessary to consider multiple objectives to obtain both accuracy and stability. In this part, we show a multi-objective optimization algorithm that optimizes the weights of four models and three proposed objective functions to comprehensively consider this problem.

5.3.1. Objective Functions

In this paper, we chose three functions to be the objective functions.

(1) The Theil Inequality Coefficient (TIC) can be indicated as follows:

$$\text{TIC}(\hat{Y}, Y) = \frac{\sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} \hat{y}_i^2} + \sqrt{\frac{1}{N} \sum_{i=1}^{N} y_i^2}}$$

(21)
TIC is always between 0 to 1; a lower numerical value is equal to better performance.

(2) Root mean squared error (RMSE). A smaller absolute value of RMSE ($\hat{Y}, Y$) indicates a more accurate forecasting performance of the system. It can be defined as follows:

$$\text{RMSE}(\hat{Y}, Y) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - y_i)^2}$$  \hspace{1cm} (22)

(3) MAPE can be represented as:

$$\text{MAPE}(\hat{Y}, Y) = \frac{1}{N} \sum_{i=1}^{N} \frac{|\hat{y}_i - y_i|}{y_i}$$  \hspace{1cm} (23)

The lower MAPE ($\hat{Y}, Y$) is, the more accurate the system.

Thus, in this combined system, the fitness function for the accuracy and stability objectives can be defined as follows:

$$\text{Minimize} = \begin{cases} f_1 = \text{TIC}(\hat{Y}, Y) \\ f_2 = \text{RMSE}(\hat{Y}, Y) \\ f_3 = \text{MAPE}(\hat{Y}, Y) \end{cases}$$  \hspace{1cm} (24)

5.3.2. NSGA-III

A random population of size $N$ and a set of widely distributed prespecified $M$-dimensional reference points $H$ on a unit hyperplane, which is placed in a manner so that it intersects each objective axis at one, having a normal vector of ones covering the entire $R^M_+$ region, were initialized by NSGA-III [45]. The more detailed description of NSGA-III is shown in the Appendix A.

5.4. Compared Combined Models with Two Objective Functions

There are three two-object-function models that were chosen to compare with our combined system:

(1) Combined model with Object Function (21) and (22);

$$\text{Minimize} = \begin{cases} f_1 = \text{TIC}(\hat{Y}, Y) \\ f_2 = \text{RMSE}(\hat{Y}, Y) \end{cases}$$  \hspace{1cm} (25)

(2) Combined model with Object Function (21) and (23);

$$\text{Minimize} = \begin{cases} f_1 = \text{TIC}(\hat{Y}, Y) \\ f_2 = \text{MAPE}(\hat{Y}, Y) \end{cases}$$  \hspace{1cm} (26)

(3) Combined model with Object Function (22) and (23).

$$\text{Minimize} = \begin{cases} f_1 = \text{RMSE}(\hat{Y}, Y) \\ f_2 = \text{MAPE}(\hat{Y}, Y) \end{cases}$$  \hspace{1cm} (27)

6. Experiments

Some performance metrics must be described to comprehensively understand the model characteristics. Diebold-Mariano and forecasting effectiveness were used in this study. Four experiments are given in this section to text the data and our proposed combined system.
6.1. The Performance Metric

Some performance metrics must be described to comprehensively understand the model characteristics. Four metrics, that is, AE, MAE, MSE, and MAPE, are shown in Table 1.

6.1.1. Diebold Mariano test

A comparison test, the Diebold Mariano (DM) test, was proposed by Diebold, F.X., and Mariano [46], which focused on the predictive accuracy and evaluating the forecasting performance between two or more-time series models.

Actual values:
\[ \{ y_t; t = 1, \ldots, n + m \} \] (28)

Two forecasts:
\[ \{ \hat{y}_t^{(1)}; t = 1, \ldots, n + m \} \] (29)
\[ \{ \hat{y}_t^{(2)}; t = 1, \ldots, n + m \} \] (30)

The forecast errors from the two models are:
\[ \{ \epsilon_{n+h}^{(1)} = y_{n+h} - \hat{y}_{n+h}^{(1)}, h = 1, 2, \ldots, m \} \] (31)
\[ \{ \epsilon_{n+h}^{(2)} = y_{n+h} - \hat{y}_{n+h}^{(2)}, h = 1, 2, \ldots, m \} \] (32)

A suitable loss function, \( L(\epsilon_{n+h}^{(i)}) \) \( i = 1, 2, \ldots \), was applied to measure the accuracy of each forecast.

Square error loss:
\[ L(\epsilon_{n+h}^{(i)}) = (\epsilon_{n+h}^{(i)})^2 \] (33)

Absolute deviation loss:
\[ L(\epsilon_{n+h}^{(i)}) = |\epsilon_{n+h}^{(i)}| \] (34)

The DM test statistics estimate the forecasts according to arbitrary loss function \( L(g) \):
\[ DM = \frac{\sum_{i=1}^{m} (L(\epsilon_{n+h}^{(1)}) - L(\epsilon_{n+h}^{(2)}))/m}{\sqrt{S^2/m}} \] (35)

where \( S^2 \) is an estimator of the variance of \( d_h = L(\epsilon_{n+h}^{(1)}) - L(\epsilon_{n+h}^{(2)}) \).

The null hypothesis is:
\[ H_0: E(d_h) = 0 \ \forall \ t \] (36)

versus the alternative hypothesis, which is:
\[ H_1: E(d_h) \neq 0 \] (37)

The null hypothesis means that the two forecasts have the same accuracy. The alternative hypothesis means that the two forecasts have different levels of accuracy. Under the null hypothesis, test statistic \( DM \) is asymptotically \( N(0, 1) \) distributed. The null hypothesis of no difference is rejected if the computed \( DM \) statistic falls outside the range from \(-z_{\alpha/2}\) to \( z_{\alpha/2}\), that is, if
\[ |DM| > z_{\alpha/2} \] (38)

where \( z_{\alpha/2} \) is the upper (or positive) z-value from the standard normal table, corresponding to half of the desired \( \alpha \) level of the test.
6.1.2. Forecasting Effectiveness

Both the square sum of forecasting error and the mean and mean squared deviation were applied to measure the forecasting accuracy by forecasting effectiveness. In some practical cases, it is necessary to further consider the kurtosis and skewness of the forecasting-accuracy distribution. On this basis, the general discrete form of forecasting effectiveness is given in this section [47].

**Definition 4.** Let

\[ \varepsilon_n = \begin{cases} -1, & (y_n - \hat{y}_n)/y_n < -1 \\ (y_n - \hat{y}_n)/y_n, & -1 < (y_n - \hat{y}_n) < 1 \\ 1, & (y_n - \hat{y}_n)/y_n > 1 \end{cases} \]  

(39)

**Definition 5.** \( A_n = 1 - |\varepsilon_n| \) is called the forecasting accuracy at time \( n \).

**Definition 6.** \( m^k = \sum_{n=1}^{N} Q_n A_n^k \) is the forecasting-accuracy effectiveness unit, \( k \) is a positive integer, \( \{Q_n, n = 1, 2, \ldots, N\} \) is the discrete-probability distribution at time \( n \), and \( \sum_{n=1}^{N} Q_n = 1, Q_n > 0 \). Especially if the a priori information of the discrete-probability distribution is unknown, we define \( Q_n = 1/N, n = 1, 2, \ldots, N \).

**Definition 7.** \( m^k \) is the \( k \)-order forecasting-effectiveness unit, and \( H \) is a continuous function of a certain \( k \) unit. \( H(m^1, m^2, \ldots, m^k) \) is the \( k \)-order forecasting effectiveness.

**Definition 8.** When \( H(x) = x \) is a continuous function of one variable, \( H(m^1) = m^1 \) is the first-order forecasting effectiveness. When \( H(x, y) = x(1 - \sqrt{y - x^2}) \), it is a continuous function of two variables.

6.2. Experiment I: Use the Linear and Nonlinear Functions to Test the Feature of a Wind-Speed Series.

Only by better understanding the characteristics of the research data can we better select the model for future work. To achieve better results, we must consider the characteristics of the data.

In general, the linear model fits the linear data better, just as the non-linear model fits the non-linear data better. Only by understanding the characteristics of the data can we achieve good results in future forecasting work. For data, it is not only linear or non-linear, but also may both. Therefore, it is necessary to judge the linear nonlinearity of the data used in this paper. Therefore, we carried out the following experiment. In order to verify the linear or nonlinear character of wind speed, three functions were structured: (1) linear function \( f_1(x) = \pi + \sum_{i=1}^{4} b_i x_i \); (2) nonlinear function \( f_2(x) = 1/(\pi + \exp(\sum_{i=1}^{4} b_i x_i)) + b_5 \); and (3) nonlinear function \( f_3(x) = 2/(\pi + \exp(\sum_{i=1}^{4} b_i x_i + b_5)) - 1 \).

From the results of Tables 2 and 3, the wind speed data are both linear and nonlinear by hypothesis test. Therefore, the linear models and nonlinear models considered in our proposed forecasting model are correct and necessary.
Table 2. Testing wind speed data by adjusting to linear functions or nonlinear functions.

|               | Number of Observations | Error Degrees of Freedom | R-Squared | Adjusted R-Squared | F-Statistic vs. Constant Model | p-Value |
|---------------|------------------------|--------------------------|-----------|-------------------|-------------------------------|---------|
| $y \sim f_1(x)$ | Site 1 2008            | 2008                     | 1996      | 0.954             | 0.954                         | 4680    | 0       |
|               | Site 2 2008            | 1996                     | 0.939     | 0.939             | 4300                          | 0       |
|               | Site 3 2008            | 1996                     | 0.938     | 0.938             | 4460                          | 0       |
| $y \sim f_2(x)$ | Site 1 2008            | 2008                     | 1996      | 0.935             | 0.935                         | 9840    | 0       |
|               | Site 2 2008            | 1996                     | 0.933     | 0.932             | 9620                          | 0       |
|               | Site 3 2008            | 1996                     | 0.936     | 0.936             | 10710                         | 0       |
| $y \sim f_3(x)$ | Site 1 2008            | 2008                     | 1996      | 0.923             | 0.923                         | 8510    | 0       |
|               | Site 2 2008            | 1996                     | 0.925     | 0.924             | 8350                          | 0       |
|               | Site 3 2008            | 1996                     | 0.911     | 0.910             | 8460                          | 0       |

Table 3. The explanations of the test parameters.

| Number of Observations | Number of Rows without Any NaN Values. |
|------------------------|----------------------------------------|
| Error degrees of freedom | $n - p$, where $n$ is the number of observations, and $p$ is the number of coefficients in the model, including the intercept. |
| R-squared and Adjusted R-squared | Coefficient of determination and adjusted coefficient of determination, respectively. |
| F-statistic vs. constant model | Test statistic for the F-test on the regression model. It tests for a significant regression relationship between the response variable and the predictor variables. |
| p-Value | $p$-value for the F statistic of the hypotheses test that the corresponding coefficient is equal to zero or not. |

6.3. Experiment II: Models Tested with Wind-Speed Data From Site 1

In order to evaluate the accuracy and stability of the forecasting system, we selected the average results of 100 trials, which were divided into two parts: accuracy and stability. In the accuracy section, we compare the AE, MAE, MSE, and MAPE values for a single model and combined model (as shown in Table 4).

Table 4. The forecasting results from Site 1.

|                | Mon       | Tue       | Wed       | Thu       | Fri       | Sat       | Sun       |
|----------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|
| AE CSBPNN      | −0.0466   | −0.0547   | −0.0269   | −0.0265   | −0.0194   | −0.0261   | −0.0386   |
| AE DE-OSELM    | −0.0120   | −0.0164   | −0.0108   | −0.0238   | −0.0444   | −0.0108   | −0.0297   |
| AE ARIMA       | −0.2191   | −0.2508   | −0.2658   | −0.2054   | −0.3563   | −0.3744   | −0.3101   |
| AE HW          | 0.0184    | 0.0178    | −0.0187   | 0.0043    | −0.0083   | −0.2792   | −0.0068   |
| AE Combined Model | −0.0013   | −0.0091   | −0.0158   | −0.0156   | −0.0104   | −0.0119   | 0.0047    |
| MAE CSBPNN     | 0.2810    | 0.2520    | 0.2129    | 0.2289    | 0.4479    | 0.4294    | 0.3179    |
| MAE DE-OSELM   | 0.2759    | 0.2531    | 0.2144    | 0.2234    | 0.4472    | 0.4282    | 0.3182    |
| MAE ARIMA      | 0.2770    | 0.2638    | 0.2717    | 0.2363    | 0.4559    | 0.4312    | 0.3255    |
| MAE HW         | 0.2942    | 0.2652    | 0.2366    | 0.2324    | 0.4252    | 0.4153    | 0.3412    |
| MAE Combined Model | 0.2420    | 0.2520    | 0.2141    | 0.2207    | 0.4445    | 0.4196    | 0.3161    |
| MSE CSBPNN     | 0.1916    | 0.1074    | 0.0769    | 0.0780    | 0.4441    | 0.3709    | 0.1713    |
| MSE DE-OSELM   | 0.1876    | 0.1075    | 0.0754    | 0.0812    | 0.4351    | 0.3705    | 0.1695    |
| MSE ARIMA      | 0.1818    | 0.1591    | 0.0873    | 0.0821    | 0.3966    | 0.3914    | 0.1287    |
| MSE HW         | 0.1999    | 0.1107    | 0.0926    | 0.0882    | 0.3486    | 0.2850    | 0.1901    |
| MSE Combined Model | 0.1812    | 0.1053    | 0.0752    | 0.0771    | 0.3322    | 0.3580    | 0.1269    |
| MAPE CSBPNN    | 5.908%    | 3.726%    | 4.081%    | 5.173%    | 5.729%    | 6.628%    | 5.103%    |
| MAPE DE-OSELM  | 5.768%    | 3.728%    | 4.097%    | 5.445%    | 5.748%    | 6.547%    | 5.093%    |
| MAPE ARIMA     | 6.227%    | 4.299%    | 5.513%    | 5.610%    | 5.967%    | 6.469%    | 5.522%    |
| MAPE HW        | 6.493%    | 3.999%    | 4.517%    | 5.509%    | 5.721%    | 6.457%    | 5.446%    |
| MAPE Combined Model | 5.691%    | 3.698%    | 4.078%    | 5.166%    | 5.692%    | 6.443%    | 5.082%    |
(a) The results of Table 4 show the following:

1. CS-BPNN reached the best results in Tuesday, Wednesday, and Thursday compared to other branch models, with MAPE values being 3.726%, 4.081%, and 5.173%, respectively.

2. On Monday and Sunday, DE-OSELM obtained the lowest MAPE compared with other branch models.

3. HW achieved the most accurate forecasting value of all branch models on Friday and Saturday.

4. Although ARIMA could achieve higher forecasting precision, their forecasting performance was worse than nonlinear models and HW.

5. Our combined model had a significant improvement in forecasting accuracy with a lower MAPE compared with all branch models. The MAPE values from Monday to Sunday are 5.691%, 3.698%, 4.078%, 5.166%, 5.692%, 6.443%, and 5.082%, respectively.

(b) The results of Figure 4 show the following:

1. Part A shows the MAE, MSE, and MAPE of five models, although our combined model did not achieve the lowest MAE every day; the lowest MSE and MAPE were obtained by our proposed model.

2. Part B shows the forecasting results of CS-BPNN, DE-OSELM, ARIMA, HW, and the combined model.

3. Part C also shows the 95% confidence intervals (CIs) obtained by each model; the figure indicates that both the upper and lower CIs were close between four branch models but, for linear models, there were more points in the confidence interval. As Part C shows, the errors of the combined model were very small, and our combined model also achieved a small CI.

Remark 1. From Table 4 and Figure 4, the results indicate that our proposed model showed better performance than the other branch models. In brief, it can be explained that SSA, which could denoise the time-series
wind-speed data as a preprocessing method, and the combined model, which took advantage of the branch models, could improve forecasting accuracy.

6.4. Experiment III: The Performance of Branch Models at Each Time Point.

In this experiment, four models, two nonlinear hybrid models (CS-BPNN and DE-OSELM) and two linear models (ARIMA and HW), were tested for performance at each time points. In this part the, wind-speed data of Tuesday from Site 2 were used in our experiments.

(a) All Tuesday results in Figure 5 show the following:

(1) Part A shows the MAE, MSE, and MAPE of five models, although our combined model did not achieve the lowest MAE every day; the lowest MSE and MAPE were obtained by our proposed model.

(2) Part B shows the forecasting results of CS-BPNN, DE-OSELM, ARIMA, HW, and the combined model.

(3) Part C also shows the 95% confidence intervals (CIs) obtained by each model; the figure indicates that both the upper and lower CIs were close between four branch models but, for linear models, there were more points in the confidence interval. As Part C shows, the errors of the combined model were very small, and our combined model also achieved a small CI.

Remark: From Table 4 and Figure 4, the results indicate that our proposed model showed better performance than the other branch models. In brief, it can be explained that SSA, which could denoise the time-series wind-speed data as a preprocessing method, and the combined model, which took advantage of the branch models, could improve forecasting accuracy.

(b) Every-hours’ Tuesday results in Table 5 show the following:
Table 5. The Tuesday forecasting results of each time point from Site 2.

|      | CS-BPNN   | MAE   | MSE   | MAPE  | DE-OSELMNN | MAE   | MSE   | MAPE  | ARIMA    | MAE   | MSE   | MAPE  | HW      | MAE   | MSE   | MAPE  |
|------|-----------|-------|-------|-------|------------|-------|-------|-------|----------|-------|-------|-------|---------|-------|-------|-------|
| 0:00 | 0.2906    | 0.0844| 2.91% |       | 0.2845     | 0.0809| 2.85% |       | 0.4337   | 0.1881| 4.34% |       | 1.1450  | 1.3110| 11.45%|       |
| 1:00 | 0.2043    | 0.0418| 1.87% |       | 0.1720     | 0.0296| 1.58% |       | 0.2185   | 0.0477| 2.00% |       | 0.3371  | 0.1137| 3.09% |       |
| 2:00 | 0.2459    | 0.0605| 2.20% |       | 0.2792     | 0.0780| 2.49% |       | 0.0642   | 0.0041| 0.57% |       | 0.3456  | 0.1195| 3.09% |       |
| 3:00 | 0.2195    | 0.0482| 1.91% |       | 0.2612     | 0.0682| 2.27% |       | 0.0811   | 0.0066| 0.71% |       | 0.2612  | 0.0682| 2.27% |       |
| 4:00 | 0.6579    | 0.4328| 7.00% |       | 0.6607     | 0.4365| 7.03% |       | 0.4615   | 0.2130| 4.91% |       | 0.4275  | 0.1827| 4.55% |       |
| 5:00 | 0.1567    | 0.0246| 1.65% |       | 0.1472     | 0.0217| 1.55% |       | 0.2550   | 0.0650| 2.68% |       | 0.2491  | 0.0620| 2.62% |       |
| 6:00 | 0.2231    | 0.0498| 2.59% |       | 0.2047     | 0.0419| 2.38% |       | 0.4538   | 0.2060| 5.28% |       | 0.0373  | 0.0014| 0.43% |       |
| 7:00 | 0.3123    | 0.0975| 3.90% |       | 0.3237     | 0.1048| 4.05% |       | 0.5266   | 0.2773| 6.58% |       | 0.3450  | 0.1190| 4.31% |       |
| 8:00 | 0.1321    | 0.0174| 1.63% |       | 0.1151     | 0.0133| 1.42% |       | 0.2286   | 0.0522| 2.82% |       | 0.1605  | 0.0258| 1.98% |       |
| 9:00 | 0.1588    | 0.0252| 1.87% |       | 0.1588     | 0.0252| 1.87% |       | 0.1485   | 0.0220| 1.75% |       | 0.1168  | 0.0136| 1.37% |       |
| 10:00| 0.5914    | 0.3497| 5.97% |       | 0.5923     | 0.3509| 5.98% |       | 0.0805   | 0.0065| 0.81% |       | 0.2916  | 0.0850| 2.95% |       |
| 11:00| 0.6408    | 0.4106| 6.47% |       | 0.6467     | 0.4182| 6.53% |       | 0.0611   | 0.0037| 0.62% |       | 0.5104  | 0.2605| 5.16% |       |
| 12:00| 0.3194    | 0.1020| 3.67% |       | 0.3130     | 0.0980| 3.60% |       | 0.2415   | 0.0583| 2.78% |       | 0.2863  | 0.0820| 3.29% |       |
| 13:00| 0.7038    | 0.4953| 9.78% |       | 0.6736     | 0.4537| 9.36% |       | 0.4774   | 0.2279| 6.63% |       | 0.0910  | 0.0083| 1.26% |       |
| 14:00| 0.0005    | 0.0001| 0.07% |       | 0.0098     | 0.0002| 0.13% |       | 0.2089   | 0.0436| 2.68% |       | 0.0553  | 0.0031| 0.71% |       |
| 15:00| 0.6799    | 0.4622| 10.79%|       | 0.6923     | 0.4793| 10.99%|       | 0.5366   | 0.2879| 8.52% |       | 0.5679  | 0.3225| 9.01% |       |
| 16:00| 0.2851    | 0.0813| 4.60% |       | 0.3025     | 0.0915| 4.88% |       | 0.2584   | 0.0667| 4.17% |       | 0.5191  | 0.2694| 8.37% |       |
| 17:00| 0.1840    | 0.0338| 4.09% |       | 0.2230     | 0.0497| 4.95% |       | 0.6237   | 0.3890| 13.86%|       | 0.1046  | 0.0109| 2.32% |       |
| 18:00| 0.2602    | 0.0677| 5.91% |       | 0.2561     | 0.0656| 5.82% |       | 0.3537   | 0.1251| 8.04% |       | 0.5095  | 0.2596| 11.58%|       |
| 19:00| 0.0176    | 0.0003| 0.46% |       | 0.0734     | 0.0054| 1.93% |       | 0.3991   | 0.1593| 10.50%|       | 0.0206  | 0.0004| 0.54% |       |
| 20:00| 0.1342    | 0.0180| 2.74% |       | 0.1254     | 0.0157| 2.56% |       | 0.2098   | 0.0440| 4.28% |       | 0.0451  | 0.0020| 0.92% |       |
| 21:00| 0.3694    | 0.1364| 5.96% |       | 0.4101     | 0.1682| 6.61% |       | 0.0774   | 0.0060| 1.25% |       | 0.4237  | 0.1795| 6.83% |       |
| 22:00| 0.3561    | 0.1268| 5.93% |       | 0.3627     | 0.1316| 6.05% |       | 0.4876   | 0.2476| 8.29% |       | 0.7679  | 0.5897| 12.80%|       |
| 23:00| 0.1832    | 0.0336| 2.78% |       | 0.0927     | 0.0086| 1.40% |       | 0.1492   | 0.0223| 2.26% |       | 0.5378  | 0.2892| 8.15% |       |
| Average| 0.3188 | 0.1559 | 4.11% |       | 0.3254 | 0.1657 | 4.23% |       | 0.2927 | 0.1093 | 4.36% |       | 0.2821 | 0.1340 | 3.78% |       |
(1) CS-BPNN obtained the lowest MAPE values of all branch models at 1:00, 7:00, 14:00, and 22:00, and the values are 1.87%, 3.90%, 0.07%, and 5.93%, respectively.
(2) At 1:00, 5:00, 8:00, 18:00, and 22:00, DE-OSELM reached the most accurate forecasting value.
(3) With the MAPE values 0.57%, 0.71%, 0.81%, 0.62%, 2.78%, 8.52%, 4.17%, and 1.25% at eight time points, ARIMA obtained the maximum time-point accuracy results on Tuesday from Site 2.
(4) HW achieved the lowest MAPE values of all branch models at 4:00, 6:00, 9:00, 13:00, 17:00, 19:00, and 20:00, and HW is also the model that had the best performance than the other branch models at all hours.
(5) The results reveal that there is no model that can reach the best results at every time point.

**Remark 2.** There is no one model that can reach the best results at every time point; each model has advantages and disadvantages. The combined models can add up the forecasting models to overcome these dilemmas. This experiment provides a reason to apply combined-model theory to forecast wind speed.

### 6.5. Experiment IV: Stability Comparison with Branch Models

In Experiments II and III, we tested the time-point performance and accuracy of the branch models. In this experiment, we evaluated the stability of the proposed combined model by comparing the standard deviation of the MAPE (STD-MAPE) values (as shown in Table 6). Because there is no randomness in the mathematical models, we just tested CS-BPNN and DE-OSELMNN of branch models in terms of stability. Here, we show the wind-speed data from Sites 1 and 2, and the results of MAPE and STD-MAPE from 100 time experiments.

**Table 6.** The stability and accuracy from three model.

| Site 1 | Site 2 |
|-------|-------|
| **MAPE** | **STD-MAPE** |
| CS-BPNN | DE-OSELM | Combined Model | CS-BPNN | DE-OSELM | Combined Model |
| Mon | 5.91% | 5.77% | 5.69% | 4.55% | 4.80% | 4.20% |
| Tue | 3.73% | 3.73% | 3.70% | 4.11% | 4.23% | 3.74% |
| Wed | 4.08% | 4.10% | 4.08% | 4.07% | 4.19% | 3.67% |
| Thu | 5.17% | 5.44% | 5.17% | 5.29% | 5.50% | 4.81% |
| Fri | 5.73% | 5.75% | 5.69% | 6.15% | 6.10% | 5.38% |
| Sat | 6.63% | 6.55% | 6.44% | 5.70% | 5.79% | 5.04% |
| Sun | 5.10% | 5.09% | 5.08% | 4.38% | 4.40% | 3.93% |

| Mon | 0.4910 | 0.5220 | 0.4617 | 0.3704 | 0.3492 | 0.3448 |
| Tue | 0.2555 | 0.2676 | 0.1871 | 0.2668 | 0.2434 | 0.2001 |
| Wed | 0.2389 | 0.2752 | 0.2176 | 0.2051 | 0.2430 | 0.1846 |
| Thu | 0.7592 | 0.8261 | 0.6506 | 1.1536 | 0.9245 | 0.8039 |
| Fri | 0.3601 | 0.3754 | 0.3243 | 0.4756 | 0.4489 | 0.3702 |
| Sat | 0.3714 | 0.4134 | 0.3636 | 0.3705 | 0.3519 | 0.3470 |
| Sun | 0.3162 | 0.3562 | 0.2797 | 0.2827 | 0.2759 | 0.2048 |

The results in Table 6 show the following:
(1) CS-BPNN performs better than DE-OSELM in accuracy on most test days.
(2) In terms of stability, CS-BPNN obtained lower STD-MAPE than DE-OSELM in Site 1, but in Site 2, DE-OSELM was better than CS-BPNN.
(3) Our proposed combined model reached the best accuracy and stability values compared to the two other models.
(4) The lowest STD-MAPE value of CS-BPNN was achieved on Wednesday on Site 2; at the same time, the STD-MAPE of DE-OSELM and our combined model were 0.2430 and 1846, respectively.
Remark 3. From the results shown in Table 6, our combined model obtained the lowest MAPE and STD-MAPE, which means that our combined-model functions could achieve high accuracy and strong stability compared with the single models.

6.6. Experiment V: Comparison with Three Combined Models (Two Objective Functions)

We proposed three other combined models with two objective functions in Section 5.2. In this experiment, we tested the accuracy and stability between our proposed system and the other three combined models. Here, wind-speed data from Site 3 were used in our experiments, and all of the models in this part were run 100 times.

(a) The results in Tables 7 and 8 show the following:

(1) Table 7 shows the average AE, MAE, and MSE. It can be seen that the forecast performance and the results of our proposed combined model were better in accuracy.

(2) The average results of the three two-objective-function models were close to the combined system (three objective functions). Especially on Friday, the difference between the combined model (three objective functions) and the combined models (two objective functions) was only 0.65%.

(3) As shown in Table 8, after running the models for 100 times, the standard-deviation MAPE and the MAPE range of the two-objective-function models was much higher and larger than the combined model. This shows that the stability of our combined model’s performance was much better than the combined models with the two objective functions.

(b). Figure 6 shows the following:

(1) Part A shows that, with regard to the minimum, maximum, and average results of MAPE from 100 experiments, our proposed combined model with three objective functions reached the lowest values compared with the other combined models on each day.

(2) Part C presents that the MAE and MSE of our proposed combined model with three objective functions were also lower than the other models.

Remark 4. The results show that the performance of the combined model with two objective functions was close to the combined model with three objective functions when we chose the result average. However, the single results of the combined model with three objective functions could be given a high degree of trust, as accuracy and stability were successfully enhanced by our combined model based on NSGA-III in the forecasting problems.
Table 7. The results for the combined model compared with combined model with two objective functions.

| Combined Model * | Combined Model ** | Combined Model *** | Combined Model |
|------------------|-------------------|--------------------|----------------|
|                  | AE    | MAE  | MSE  | AE    | MAE  | MSE  | AE    | MAE  | MSE  |
| Monday           | 0.01815 | 0.33185 | 0.00233 | 0.01568 | 0.32888 | 0.00225 | 0.07684 | 0.34506 | 0.00590 |
| Tuesday          | −0.03139 | 0.36454 | 0.00099 | −0.03672 | 0.36132 | 0.00135 | 0.04987 | 0.35975 | 0.00249 |
| Wednesday        | −0.00596 | 0.28940 | 0.00204 | −0.01435 | 0.28707 | 0.00221 | 0.04960 | 0.29631 | 0.00246 |
| Thursday         | 0.03749 | 0.32268 | 0.00341 | 0.03271 | 0.32185 | 0.00307 | 0.09834 | 0.33558 | 0.00967 |
| Friday           | −0.05878 | 0.55131 | 0.00345 | −0.06042 | 0.54689 | 0.00365 | 0.02861 | 0.55014 | 0.00082 |
| Saturday         | −0.06960 | 0.49825 | 0.00484 | −0.07203 | 0.49099 | 0.00519 | 0.01423 | 0.50229 | 0.00020 |
| Sunday           | −0.05840 | 0.38680 | 0.00341 | −0.06404 | 0.38348 | 0.00410 | 0.02042 | 0.38329 | 0.00042 |

* Combined Model with the objective functions: TIC (\(\hat{Y}, Y\)) and RMSE (\(\hat{Y}, Y\)). ** Combined Model with the objective functions: TIC (\(\hat{Y}, Y\)) and MAPE (\(\hat{Y}, Y\)). *** Combined Model with the objective functions: RMSE (\(\hat{Y}, Y\)) and MAPE (\(\hat{Y}, Y\)).

Table 8. Mean absolute percentage error (MAPE) of the results for the combined model compared with combined model with two objective functions.

| Combined Model * | Combined Model ** | Combined Model *** | Combined Model |
|------------------|-------------------|--------------------|----------------|
|                  | MAX   | MIN   | AVE  | STD  | MAX   | MIN   | AVE  | STD  | MAX   | MIN   | AVE  | STD  | MAX   | MIN   | AVE  | STD  |
| Monday           | 11.47% | 6.55% | 7.60% | 0.7837 | 11.08% | 5.94% | 7.43% | 0.7540 | 11.12% | 6.30% | 7.82% | 0.7138 | 8.83% | 5.01% | 5.86% | 0.6240 |
| Tuesday          | 5.92% | 4.38% | 4.86% | 0.2528 | 5.83% | 3.96% | 4.81% | 0.2772 | 5.99% | 3.93% | 4.86% | 0.2263 | 4.80% | 3.64% | 4.05% | 0.2023 |
| Wednesday        | 6.27% | 4.65% | 5.25% | 0.3012 | 6.34% | 4.37% | 5.21% | 0.3227 | 6.62% | 4.32% | 5.42% | 0.2811 | 5.69% | 3.46% | 3.93% | 0.2775 |
| Thursday         | 18.21% | 6.73% | 8.60% | 1.6167 | 16.66% | 6.19% | 8.41% | 1.6433 | 14.05% | 7.03% | 8.95% | 1.6301 | 13.98% | 4.91% | 6.64% | 1.5877 |
| Friday           | 7.98% | 5.75% | 6.60% | 0.4075 | 8.07% | 5.41% | 6.55% | 0.4319 | 8.33% | 5.30% | 6.60% | 0.3719 | 6.67% | 5.16% | 5.90% | 0.3199 |
| Saturday         | 8.03% | 5.67% | 6.62% | 0.3858 | 7.75% | 5.51% | 6.55% | 0.3903 | 8.32% | 5.36% | 6.69% | 0.3555 | 6.33% | 4.94% | 5.63% | 0.2820 |
| Sunday           | 6.08% | 4.37% | 4.94% | 0.3127 | 6.16% | 4.10% | 4.88% | 0.3043 | 6.38% | 3.94% | 4.90% | 0.3148 | 4.70% | 3.37% | 3.82% | 0.2345 |

* Combined Model with the objective functions: TIC (\(\hat{Y}, Y\)) and RMSE (\(\hat{Y}, Y\)). ** Combined Model with the objective functions: TIC (\(\hat{Y}, Y\)) and MAPE (\(\hat{Y}, Y\)). *** Combined Model with the objective functions: RMSE (\(\hat{Y}, Y\)) and MAPE (\(\hat{Y}, Y\)).
As shown in Table 8, after running the models for 100 times, the standard-deviation MAPE and the MAPE range of the two-objective-function models was much higher and larger than the combined model. This shows that the stability of our combined model’s performance was much better than the combined models with the two objective functions.

Part B shows the following:

1. Part A shows that, with regard to the minimum, maximum, and average results of MAPE from 100 experiments, our proposed combined model with three objective functions reached the lowest values compared with the other combined models on each day.

2. Part C presents that the MAE and MSE of our proposed combined model with three objective functions were also lower than the other models.

Remark: The results show that the performance of the combined model with two objective functions was close to the combined model with three objective functions when we chose the result average. However, the single results of the combined model with three objective functions could be given a high degree of trust, as accuracy and stability were successfully enhanced by our combined model based on NSGA-III in the forecasting problems.

6.7. Experiment VI: Forecasting Results Test

To evaluate the forecasting results of these models, two important evaluation metrics, the DM test and forecasting availability, were applied in this part. We discuss the results from Site 3.

1. The results of the DM test are shown in Table 9, which indicate that the combined system was different from the other models at a significant difference level in different datasets.

2. As shown in Table 10, the first-order and second-order forecasting availabilities of the proposed combined system performed better than the other models for seven datasets from three regions in electricity-load forecasting. For example, in the Monday dataset, the first-order forecasting availabilities offered by each forecasting models were 0.9110, 0.9143, 0.8432, 0.8851, 0.9284, 0.9291, 0.9249, and 0.9456, respectively.
Table 9. Results for the Diebold Mariano test.

|       | CS-BPNN | DE-OSELM | ARIMA | HW | CM *  | CM ** | CM *** |
|-------|---------|----------|-------|----|-------|-------|--------|
| Mon   | 3.2420  | 2.9318   | 4.6351| 3.2245| 2.6696| 2.6639| 3.0670 |
| Tue   | 6.0215  | 5.9697   | 6.5906| 5.1321| 3.2101| 2.8908| 3.3343 |
| Wed   | 7.5394  | 7.4828   | 9.4464| 5.4080| 5.6746| 5.5454| 6.7144 |
| Thu   | 7.5371  | 7.2595   | 9.5547| 5.0369| 6.4416| 6.3672| 6.4491 |
| Fri   | 5.7344  | 5.3455   | 2.9753| 0.8264| 1.7598| 1.4657| 1.6069 |
| Sat   | 5.8464  | 5.7280   | 3.9762| 1.3795| 1.8977| 1.5754| 1.6169 |
| Sun   | 5.3999  | 5.3771   | 9.1815| 4.1056| 4.7345| 4.6653| 5.2127 |

* Combined Model with the objective functions: TIC ($\hat{Y}$, $Y$) and RMSE ($\hat{Y}$, $Y$). ** Combined Model with the objective functions: TIC ($\hat{Y}$, $Y$) and MAPE ($\hat{Y}$, $Y$). *** Combined Model with the objective functions: RMSE ($\hat{Y}$, $Y$) and MAPE ($\hat{Y}$, $Y$).

Table 10. Results for the forecasting effectiveness.

|       | CS-BPNN | DE-OSELM | ARIMA | HW | CM *  | CM ** | CM *** | CM   |
|-------|---------|----------|-------|----|-------|-------|--------|------|
| Mon   | 1-order | 0.9110   | 0.9143| 0.8432| 0.8851| 0.9284| 0.9291 | 0.9249| 0.9456|
|       | 2-order | 0.8311   | 0.8343| 0.7715| 0.8121| 0.8611| 0.8621 | 0.8562| 0.8973|
| Tue   | 1-order | 0.9453   | 0.9456| 0.9116| 0.9302| 0.9520| 0.9525 | 0.9521| 0.9605|
|       | 2-order | 0.9023   | 0.9030| 0.8709| 0.8825| 0.9162| 0.9146 | 0.9141| 0.9259|
| Wed   | 1-order | 0.9390   | 0.9398| 0.9135| 0.9399| 0.9481| 0.9486 | 0.9464| 0.9616|
|       | 2-order | 0.8973   | 0.8987| 0.8817| 0.9018| 0.9112| 0.9117 | 0.9098| 0.9313|
| Thu   | 1-order | 0.8985   | 0.9018| 0.8831| 0.9276| 0.9230| 0.9228 | 0.9166| 0.9471|
|       | 2-order | 0.7773   | 0.7776| 0.7836| 0.8693| 0.8281| 0.8281 | 0.8134| 0.8816|
| FRI   | 1-order | 0.9269   | 0.9279| 0.8812| 0.9212| 0.9351| 0.9356 | 0.9351| 0.9423|
|       | 2-order | 0.8623   | 0.8642| 0.8325| 0.8614| 0.8784| 0.8791 | 0.8775| 0.8862|
| Sat   | 1-order | 0.9262   | 0.9262| 0.8750| 0.9174| 0.9348| 0.9357 | 0.9342| 0.9446|
|       | 2-order | 0.8657   | 0.8655| 0.8269| 0.8675| 0.8817| 0.8826 | 0.8812| 0.8935|
| Sun   | 1-order | 0.9468   | 0.9466| 0.9110| 0.9473| 0.9513| 0.9518 | 0.9515| 0.9624|
|       | 2-order | 0.9041   | 0.9036| 0.8831| 0.9106| 0.9127| 0.9135 | 0.9131| 0.9317|

* Combined Model with the objective functions: TIC ($\hat{Y}$, $Y$) and RMSE ($\hat{Y}$, $Y$). ** Combined Model with the objective functions: TIC ($\hat{Y}$, $Y$) and MAPE ($\hat{Y}$, $Y$). *** Combined Model with the objective functions: RMSE ($\hat{Y}$, $Y$) and MAPE ($\hat{Y}$, $Y$).

7. Conclusions

Effective, accurate, and reliable forecast of wind speed is a crucial component of wind-farm management, and also a significant part of the economic development of a nation. However, previous studies only focused on one type of accuracy or stability. Due to this reason, former models cannot achieve satisfying results that obtain high accuracy and strong stability. Wind-speed forecast is also a multicriteria problem; considering only one criterion (accuracy or stability) cannot achieve satisfying results. Thus, it is difficult to select one objective function from multiple-objective functions, and arduous to ensure that the selected function can obtain low MAPE and strong stability. In this paper, a combined forecasting model based on data preprocessing and MOP theory, which could simultaneously improve the accuracy and stability of wind-speed forecasting and four models (two hybrid nonlinear models and two linear models), is proposed and was successfully applied to wind-speed forecasting, not only overcoming the issue of forecasting accuracy, but also solving the difficulty of forecasting stability. Then, as the results show, our proposed combined model with three objective functions achieved lower MAPE and STMAPE than the other models. From the results of our research, our proposed combined model improved the MAPE range from 0.007% to 2.31%, and the STDMAPE range from 0.0044 to 0.3497. Moreover, according to our study, the combined model could be used in large wind farms to forecast wind speed, evaluate wind-energy resources, and save operation costs and wind energy.
Author Contributions: S.Z. carried on programming and writing of the whole manuscript; Y.L. carried on the validation and visualization of experiment results; J.W. and C.W. provided the overall guide of conceptualization and methodology. 

Funding: The research was funded by National Natural Science Foundation of China (Grants No.71671029).

Acknowledgments: This work was supported by National Natural Science Foundation of China (Grants No.71671029).

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

Appendix A

List of abbreviations:

- MM5: Mesoscale Model 5
- HW: Holt-Winters
- CS: Cuckoo Search algorithm
- SVM: Support Vector Machine
- AE: Absolute Error
- MAE: Mean Absolute Error
- MSE: Mean Square Error
- PS: Pareto-optimal Set
- PF: Pareto Front
- DE: Differential Evolution
- MOP: Multiple Objective Optimization Problem
- ARIMA: Auto-Regressive Integrate Moving Average
- ARMA: Autoregressive Moving Average
- BPNN: Back-Propagation Neural Network
- NSGA-III: Non-dominated Sorting Genetic Algorithm
- SOP: Single Objective Optimization Problem
- OSELM: Online Extreme Learning Machine
- MAPE: Mean Absolute Percentage Error
- RMSE: Root Mean Squared Error
- ANN: Artificial Neural Networks

The Introduction of NSGA-III

NSGA-III performs the following operations, performed at a generation $t$. Primarily, different nondomination levels are classified from whole population $P_t$ in the same way in NSGA-II, following the principle of nondominated sorting. Then, $P_t$ creates an offspring population $Q_t$ by way of the usual recombination and mutation operators. For every reference point, only one individual of the population is anticipated to be found, so there is no necessity for any selection operation in NSGA-III. Then, there is a new population $R_t$ combined with $P_t$ and $Q_t$ ($R_t = P_t \cup Q_t$). Subsequently, points starting from the first nondominated front are selected for $P_{t+1}$ one at a time, until all solutions from a complete front cannot be included. We can denote the final front as $F_L$. In general, only a few solutions from $F_L$ need to be selected for $P_{t+1}$ using a niche-preserving operator, which could be described in the next one. Primarily, each population member of $P_{t+1}$ and $F_L$ is normalized by using the current population spread so that all objective vectors and reference points have commensurate values. Subsequently, each member of $P_{t+1}$ and $F_L$ is associated with a specific reference point by using the shortest perpendicular distance ($d()$)of each population member, with a reference line created by joining the origin with a supplied reference point. Then, a careful niching strategy is employed to choose those $F_L$ members that are associated with the least-represented reference points in $P_{t+1}$. The niching strategy puts an emphasis on selecting a population member for as many supplied reference points as possible. A population member associated with an under-represented or unrepresented reference point is immediately preferred. With continuous stress for emphasizing nondominated...
individuals, the whole process is then expected to find one population member corresponding to each supplied reference point close to the Pareto-optimal front, provided that the genetic variation operators (recombination and mutation) are capable of producing respective solutions. The use of well-spread reference points ensures a well-distributed set of trade-off points at the end.

The pseudocode for the NSGA-III is provided in Algorithm 2.

**Algorithm 1: Denoising**

**Input:**
Y = \{y_1, y_2, \ldots, y_N\} a sequence of original time series

**Output:**
\tilde{Y} = \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N\} — a sequence of decomposed time series

**Parameters:**
N — the length of a time series.
X — the trajectory matrix.
L — the windows length.
K — the number of lagged vectors.
K' — the maximum between L and K.
M — the number of repetitions of each trial.

1 /*Obtain the trajectory matrix.*/
2 FOR EACH i = 1:L DO
3 | X(i,:) = (y_i, \ldots, y_{K+i-1})
4 END FOR
5 X = \begin{bmatrix}
    y_1 & y_2 & \cdots & y_K \\
    y_2 & y_3 & \cdots & y_{K+1} \\
    \vdots & \vdots & \ddots & \vdots \\
    y_L & y_{L+1} & \cdots & y_N
\end{bmatrix}
6 S = XX^T /*Singular value decomposition.*/
7 /*Obtain \lambda_1, \ldots, \lambda_L the eigenvalue of S. (\lambda_1 \geq \cdots \geq \lambda_L \geq 0)*/
8 /*Obtain U_1, \ldots, U_L the orthonormal system of the eigenvectors of the matrix S corresponding to these eigenvalues.*/
9 FOR EACH i = 1:d DO
10 | V_i = X^T U_i \sqrt{\lambda_i}; X_i = \sqrt{\lambda_i} U_i V_i^T;
11 END FOR
12 X = X_1 + \cdots + X_d; Let I = [i_1, \ldots, i_p]; X_I = X_{i_1} + \cdots + X_{i_p};
13 /*The results of the grouping procedure partition the matrices X_i into several disjoint subsets.*/
14 /*Transform each matrix of the grouped decomposition into a new series of length N.*/
15 /*Diagonal averaging*/
16 \tilde{Y} = \{\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N\}
Algorithm 2: NSGA-III

**Input:**
Y = \{y_1, y_2, \cdots, y_N\}—a sequence of time series wind speed data

**Output:**
\hat{Y} = \{\hat{y}_1, \hat{y}_2, \cdots, \hat{y}_N\}—a sequence of time series wind speed forecasting data

**Fitness function**

\[
\text{Minimize} = \begin{cases} 
  f_1 = \text{TIC}(\hat{Y}, Y) \\
  f_2 = \text{RMSE}(\hat{Y}, Y) \\
  f_3 = \text{MAPE}(\hat{Y}, Y)
\end{cases}
\]

**Parameters:**
P_0—the initial population.
N_{Pop}—the size of population.
t— the number of iterations.
W_i—the weight of ith single model.
It_{max}—the number of maximum iteration.

```plaintext
1  /* Input Y and Initial P_0 and t=0*/
2  WHILE t < It_{max} DO
3     /*Create Offspring Q_t*/; Mutation on Q_t; Set R_t = P_t \cup Q_t
4     /*Apply non-dominated with fitness function sorting on R_t and find F_1, F_2,...*/
5     /*S_t=\{\} \forall i=1*/
6     WHILE
7         S_t = S_t \cup F_t; i=i+1;
8     END WHILE
9     IF |S_t| = N_{pop}
10    P_{t+1} = S_t
11   ELSE
12    P_{t+1} = \bigcup_{i=1}^{N_{pop}} F_i
13    /*Normalize S_t using min and intercept points of each objectives*/
14    /*Associate each member of S_t to a reference point*/
15    /*Choose N_{pop} \cdot |P_{t+1}| members from F_i by niche-preserving operator*/
16   END IF
17  t=t+1
18  END WHILE
19  return W_i /*Choose suitable W_i from P_t*/
20  Return \hat{Y}
```

**References**
1. Available online: http://www.sohu.com/a/192783727_472920 (accessed on 18 September 2018).
2. Wiser, R.; Bolinger, M. Wind Technologies Market Report; Tech. Rep.; Lawrence Berkeley National Laboratory: Berkeley, CA, USA, 2016.
3. Yu, J.; Ji, F.; Zhang, L.; Chen, Y. An over painted oriental arts: Evaluation of the development of the Chinese renewable energy market using the wind power market as a model. *Energy Policy* 2009, 37, 5221–5225. [CrossRef]
4. Soman, S.; Zareipour, H.; Malik, O.; Mandal, P. A review of wind power and wind speed forecasting methods with different time horizons. In Proceedings of the North American Power Symposium (NAPS), Arlington, TX, USA, 26–28 September 2010; pp. 1–8.

5. De Giorgi, M.G.; Ficarella, A.; Tarantino, M. Assessment of the benefits of numerical weather predictions in wind power forecasting based on statistical methods. Energy 2011, 36, 3968–3978. [CrossRef]

6. Cassola, F.; Burlando, M. Wind speed and wind energy forecast through kalman filtering of numerical weather prediction model output. Appl. Energy 2012, 99, 154–166. [CrossRef]

7. Wang, J.; Heng, J.; Xiao, L.; Wang, C. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. Energy 2017, 125, 591–613. [CrossRef]

8. 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]

9. Yang, H.; Jiang, Z.; Lu, H. A Hybrid Wind Speed Forecasting System Based on a ‘Decomposition and Ensemble’ Strategy and Fuzzy Time Series. Energies 2017, 10, 1422. [CrossRef]

10. Milligan, M.; Schwartz, M.; Wan, Y.H. Statistical Wind Power Forecasting for U.S. In Proceedings of the Wind Farms: Preprint Conference on Probability & Statistics in the Atmospheric Sciences/American Meteorological Society Meeting, Seattle, WA, USA, 11–15 January 2004.

11. Flores, J.J.; Loaeza, R.; Rodriguez, H.; Cadenas, E. Wind Speed Forecasting Using a Hybrid Neural-Evolutive Approach. In Mexican International Conference on Artificial Intelligence; Springer: Berlin/Heidelberg, Germany, 2009.

12. Flores, J.J.; Graff, M.; Rodriguez, H. Evolutive design of ARMA and ANN models for time series forecasting. Renew. Energy 2012, 44, 225–230. [CrossRef]

13. El-Fouly, T.H.M.; El-Saadany, E.F.; Salama, M.M.A. Grey predictor for wind energy conversion systems output power prediction. IEEE Trans. Power Syst. 2006, 21, 1450–1452. [CrossRef]

14. Atsalakis, G.; Nezis, D.; Zopounidis, C. Neuro-fuzzy versus traditional models for forecasting wind energy production. In Genetic Programming and Evolvable Machines, Advances in Data Analysis; Springer: Boston, MA, USA, 2010; pp. 275–287.

15. Tanga, J.; Brousteb, A.; Tsuia, K.L. Some improvements of wind speed markov chain modeling. Renew. Energy 2015, 81, 52–56. [CrossRef]

16. Esen, H.; Inalli, M.; Sengur, A.; Esen, M. Performance prediction of a ground-coupled heat pump system using artificial neural networks. Expert Syst. Appl. 2008, 35, 1940–1948. [CrossRef]

17. Zhao, X.; Wang, C.; Su, J.; Wang, J. Research and application based on the swarm intelligence algorithm and artificial intelligence for wind farm decision system. Renew. Energy 2019, 134, 681–697. [CrossRef]

18. Fu, T.; Wang, C. A Hybrid Wind Speed Forecasting Method and Wind Energy Resource Analysis Based on a Swarm Intelligence Optimization Algorithm and an Artificial Intelligence Model. Sustainability 2018, 10, 3913. [CrossRef]

19. Yao, Z.; Wang, C. A Hybrid Model Based on A Modified Optimization Algorithm and an Artificial Intelligence Algorithm for Short-Term Wind Speed Multi-Step Ahead Forecasting. Sustainability 2018, 10, 1443. [CrossRef]

20. Wang, Z.; Wang, C.; Wu, J. Wind Energy Potential Assessment and Forecasting Research Based on the Data Pre-Processing Technique and Swarm Intelligent Optimization Algorithms. Sustainability 2016, 8, 1191. [CrossRef]

21. Heng, J.; Wang, C.; Zhao, X.; Xiao, L. Research and Application Based on Adaptive Boosting Strategy and Modified CGFPA Algorithm: A Case Study for Wind Speed Forecasting. Sustainability 2016, 8, 235. [CrossRef]

22. Du, P.; Wang, J.; Guo, Z.; Yang, W. Research and application of a novel hybrid forecasting system based on multi-objective optimization for wind speed forecasting. Energy Convers. Manag. 2017, 150, 90–107. [CrossRef]

23. Wang, J.; Du, P.; Niu, T.; Yang, W. A novel hybrid system based on a new proposed algorithm—Multi-Objective Whale Optimization Algorithm for wind speed forecasting. Appl. Energy 2017, 208, 344–360. [CrossRef]

24. Jiang, P.; Yang, H.; Heng, J. A hybrid forecasting system based on fuzzy time series and multi-objective optimization for wind speed forecasting. Appl. Energy 2019, 235, 786–801. [CrossRef]

25. Dalto, M.; Matsuško, J.; Vašak, M. Deep neural networks for ultra-short-term wind forecasting. In Proceedings of the 2015 IEEE International Conference on Industrial Technology (ICIT), Seville, Spain, 17–19 March 2015.
26. Liu, D.; Niu, D.; Wang, H.; Fan, L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew. Energy* **2014**, *62*, 592–597. [CrossRef]

27. Hu, J.; Wang, J.; Ma, K. A hybrid technique for short-term wind speed prediction. *Energy* **2015**, *81*, 563–574. [CrossRef]

28. Barbounis, T.G.; Theocharis, J.B. A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation. *Neurocomputing* **2007**, *70*, 1525–1542. [CrossRef]

29. Focken, U.; Lange, M.; Waldl, H. Previento—A wind power prediction system with an innovative upscaling algorithm. In Proceedings of the European Wind Energy Conference, Copenhagen, Denmark, 2–6 July 2001; Volume 276.

30. Landberg, L. Short-term prediction of the power production from wind farms. *J. Wind Eng. Aerodyn.* **1999**, *80*, 207–220. [CrossRef]

31. 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.* **2015**, *32*, 981–990. [CrossRef]

32. Bates, J.M.; Granger, C.W.J. The combination of forecasts. In *Essays in Econometrics*; Cambridge University Press: Cambridge, UK, 2001; pp. 451–468.

33. 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]

34. Jain, H.; Deb, K. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach. *IEEE Trans. Evol. Comput.* **2014**, *18*, 602–622. [CrossRef]

35. Abdollahzade, M.; Miranian, A.; Hassani, H.; Iranmanesh, H. A new hybrid enhanced local linear neuro-fuzzy model based on the optimized singular spectrum analysis and its application for nonlinear and chaotic time series forecasting. *Inf. Sci.* **2015**, *295*, 107–125. [CrossRef]

36. Guo, Z.H.; Wu, J.; Lu, H.Y.; Wang, J.Z. A case study on a hybrid wind speed forecasting method using BP neural network. *Knowl.-Based Syst.* **2011**, *24*, 1048–1056. [CrossRef]

37. Yang, X.S.; Deb, S. Cuckoo search via Lévy flights. In Proceedings of the World Congress on Nature & Biologically Inspired Computing, Coimbatore, India, 9–11 December 2009; pp. 210–214.

38. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: A new learning scheme of feedforward neural networks. *Neurocomputing* **2004**, *2*, 985–990.

39. Liang, N.Y.; Huang, G.B.; Saratchandran, P.; Sundararajan, N. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans. Neural Netw.* **2006**, *17*, 1411–1423. [CrossRef]

40. Storn, R.; Price, K. *Differential Evolution—A Simple and Efficient Adaptive Scheme for Global Optimization over Continuous Spaces*; Tech. Rep.TR-95-012; ICSI: Berkeley, CA, USA, 1995.

41. 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]

42. Grubb, H.; Mason, A. Long lead-time forecasting of UK air passengers by Holt–Winters methods with damped trend. *Int. J. Forecast.* **2001**, *17*, 71–82. [CrossRef]

43. Stock, J.H.; Watson, M.W. Combination forecasts of output growth in a seven-country data set. *J. Forecast.* **2004**, *23*, 405–430. [CrossRef]

44. Geweke, J.; Amisano, G. Optimal prediction pools. *J. Econom.* **2011**, *164*, 130–141. [CrossRef]

45. Das, I.; Dennis, J.E. Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems. *SIAM J. Optim.* **1998**, *8*, 631–657. [CrossRef]

46. Diebold, F.X.; Mariano, R. Comparing predictive accuracy. *J. Bus. Econ. Stat.* **1995**, *13*, 253.

47. Chen, H.Y.; Hou, D.P. Research on superior combination forecasting models based on forecasting effective measures. *J. Univ. Sci. Technol. China* **2002**, *2*, 172–180.

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).