


Article

# A New Hybrid Approach for Wind Speed Forecasting Applying Support Vector Machine with Ensemble Empirical Mode Decomposition and Cuckoo Search Algorithm

Tongxiang Liu <sup>1</sup>, Shenzhong Liu <sup>2,\*</sup>, Jiani Heng <sup>2</sup>  and Yuyang Gao <sup>2</sup>

<sup>1</sup> Faculty of Professions, University of Adelaide, Adelaide 5000, Australia; tongxiang.liu@student.adelaide.edu.au

<sup>2</sup> School of Statistics, Dongbei University of Finance and Economics, Dalian 116025, China; hengjin13@lzu.edu.cn (J.H.); gaoyuyang0315@gmail.com (Y.G.)

\* Correspondence: liusz@dufe.edu.cn; Tel.: +86-13804258916

Received: 12 July 2018; Accepted: 20 September 2018; Published: 28 September 2018



**Featured Application:** Wind speed forecasting.

**Abstract:** Wind speed forecasting plays a crucial role in improving the efficiency of wind farms, and increases the competitive advantage of wind power in the global electricity market. Many forecasting models have been proposed, aiming to enhance the forecast performance. However, some traditional models used in our experiment have the drawback of ignoring the importance of data preprocessing and the necessity of parameter optimization, which often results in poor forecasting performance. Therefore, in order to achieve a more satisfying performance in forecasting wind speed data, a new short-term wind speed forecasting method which consists of Ensemble Empirical Mode Decomposition (EEMD) for data preprocessing, and the Support Vector Machine (SVM)—whose key parameters are optimized by the Cuckoo Search Algorithm (CSO)—is developed in this paper. This method avoids the shortcomings of some traditional models and effectively enhances the forecasting ability. To test the prediction ability of the proposed model, 10 min wind speed data from wind farms in Shandong Province, China, are used for conducting experiments. The experimental results indicate that the proposed model cannot only improve the forecasting accuracy, but can also be an effective tool in assisting the management of wind power plants.

**Keywords:** cuckoo search algorithm; support vector machine; ensemble empirical mode decomposition; wind speed forecasting; forecasting validity

## 1. Introduction

### 1.1. Background and Motivation

In recent years, a demand for clean and renewable energy resources has increased significantly because of the air pollution caused by traditional fossil fuels. Wind power, which is one of the most promising renewable resources, proved to be an ideal alternative. Currently, wind energy has been successfully employed in many countries, representing approximately 10% of energy consumption in Europe and more than 15% in countries like America and Spain [1]. However, wind speed, which is one of the most essential factors in wind power generation, is difficult to forecast due to many natural factors such as pressure, temperature, and the rotation of the planet.

Therefore, in the development of wind energy, wind speed forecasting is rather important. The accuracy of the forecasting result can influence the wind rotating equipment, operation cost,

and the limitations of wind power penetration. With a precise prediction of wind speed, the dispatching department is able to make adjustments to the program efficiently, so that the influence of wind energy itself on the power station and the adverse effect of the wind farm on the power system can be minimized, making wind power more competitive in the global electricity market.

### 1.2. Existing Models

Currently, developing an accurate and effective wind speed forecasting model is a priority in many countries. These models can be divided into four categories based on the specific time horizon: Very short-term, short-term, medium-term, and long-term forecasting [2]. Very short-term and short-term forecasting can provide aids and references when making economic load dispatch plans, while medium and long-term forecasting are employed for the maintenance of the grid, wind farm planning, and providing information for site selection [3,4]. Using computational methods, these forecasting methods can be further categorized into four different types: Physical, statistical, intelligent, and hybrid methods. Physical methods are usually based on numerical weather prediction (NWP), which simulates the physical characteristics of the atmosphere through applying physical rules and geographical conditions, though there are still many difficulties in employing this model for wind speed forecasting directly—for example, the forecasting accuracy, resolutions of space and time, and the importance of the physical procedures. Statistical methods are used for wind speed forecasting through setting mathematical models, which are similar to the direct random time-series models [5]. As one of the most widely employed statistical methods, the Autoregressive Integrated Moving Average (ARIMA) model is adopted to predict wind speed. Moreover, an Autoregressive Fractional Integrated Moving Average (ARFIMA) model is used in wind speed prediction because of its superior ability to select valid information from the past time series more effectively, compared with the traditional ARIMA model. In recent years, intelligent approaches such as artificial neural networks (ANN) and back propagation neural networks (BP), with the aim of reducing errors by employing past time series, are being applied to forecast future wind speed because of the rising popularity of artificial intelligence [6–12].

Due to the complexity of the raw wind speed series, forecasting the traits in the time series accurately is very difficult to achieve with individual models. Therefore, hybrid models which adopt multiple approaches to increase the prediction ability of the raw time series, or which assemble multiple forecasting methods to obtain the traits of the raw data to forecast wind speed, have been proposed in many studies. Guo et al. combined the Seasonal Autoregressive Intergrated Moving Average model (SARIMA) and the Least Square Support Vector Machine model (LSSVM) to obtain a more accurate forecasting model [13]. This new approach can effectively catch the seasonal and nonlinear features in the input data. Similarly, Guo et al. used a back-propagation neural network, and relied on the theory of using a seasonal exponential adjustment to reduce seasonal influence from the raw data, to create a new hybrid approach [14]. Experimental results showed that this new model is able to effectively increase the forecasting precision, as compared to the results of an independent back propagation neural network (BPNN) with no adjustment of the seasonal exponent. From the studies of Pourmousavi et al., a new and very short-term wind speed hybrid model was proposed and proven to significantly increase the prediction range [15,16]. Liu et al. combined Empirical Model Decomposition to create a new hybrid Empirical Mode Decomposition-Artificial Neural Network (EMD-ANN) model. The proposed approach was effective in capturing jumping samples in raw data with high noise [17]. Mohammadi et al. proposed a Stackelberg game technique to improve the efficiency of electric grid [18]. Kianoosh et al. introduced a new multi-time-scale approach modeled by historical time-series for electric data forecasting [19]. Haque et al. put forward the idea of combining multiple soft computing models with a data preprocessing technique to perform short-term wind speed forecasting [20]. Li and Shi conducted an experiment to make a comprehensive comparison among three ANN using one h ahead of forecasting; namely, ADLINE (adaptive linear element), RBFNN (radial basis function neural network), and BPNN (back-propagation neural network) [21].

Moreover, to reduce the influence of data diversity, Ortiz-García et al. introduced several novel SVM structures, which achieved a better performance than a similar multilayer perception [22].

### 1.3. Introduction of the Proposed Model

However, these individual models mentioned above have certain drawbacks. The disadvantages are summarized as followed: (1) They usually require a huge amount of wind speed data in order to build a model for a precise prediction, because of the inner irregularity and instability within the raw data. If the raw data suddenly changes due to environmental factors, the error of the forecasting result can be relatively high [23]; (2) Some models only try to match the fitness of the model closer to the original data. As a result, when processing the wind speed data with high noise and irregularity, it is difficult to fit the individual model to it by using only traditional physical or statistical methods, causing poor forecasting accuracy and low efficiency; (3) Individual forecasting methods ignore the importance of data preprocessing and the necessity of model parameter optimization. Because of this, the accuracy of the result is not satisfying; (4) While some new models take advantage of artificial intelligence to enhance their forecasting ability, they still have the problem of over-fitting and a low convergence rate [24].

Therefore, in order to solve the problems mentioned above, a new hybrid model for more precise wind speed forecasting and better evaluation, achieved by adapting the Ensemble Empirical Mode Decomposition (EEMD), the Cuckoo Search Optimization (CSO) algorithm, and the Support Vector Machine (SVM) model, is introduced in this paper.

The main contributions and innovations of this paper compared to other studies in the field of wind speed forecasting are summed up as follows:

- The newly proposed approach in this paper takes advantage of the data preprocessing method and the algorithm of parameter optimization to enhance the performance of the SVM model. In this paper, the raw time series is first decomposed into several sub-signals, among which signals with high frequency ones are removed and the rest are restructured to obtain a stationary time series, with which the intrinsic characteristics of the wind speed data can be better captured and analyzed so that the forecasting accuracy can be greatly improved.
- This paper employs the Cuckoo Search algorithm to optimize the parameters of the SVM before training. The CS algorithm, which has the advantage of a powerful capability in terms of global optimization, requires few parameters and has strong multi-objective problem solving ability, and therefore can significantly improve the accuracy of the forecasting. The Support Vector Machine can overcome the difficulties of traditional models, such as the curse of dimensionality, falling into local optima easily, and over-learning.
- To verify the forecasting ability of the proposed approach, conventional models like BPNN, RBFNN, and ARIMA are used for comparison. A more comprehensive evaluation is conducted, including multi-step forecasting experiments and performance evaluation metrics such as six indexes and a DM(Diebold-Mariano) test, to assess and analyze the performance of the newly developed method.

The model used the raw data from wind turbines in a large wind farm. The result of this paper shows that the newly developed approach effectively improves the forecasting accuracy and can be successfully applied in wind grids to provide statistical support for making operation plans and managing the power station.

### 1.4. Structure of the Paper

This paper is organized as follows. Section 2 concisely introduces the required techniques. Forecasting performance evaluation criterion and numerical experimentation are introduced in Sections 3 and 4. In Sections 5 and 6, the results and the superiority of the proposed approach

are discussed by using comparisons with other conventional methods. Finally, Section 7 concludes this paper and introduces the direction for future studies.

## 2. Materials and Methods

In this section, main theories about the proposed model will be introduced first. Then, the proposed hybrid model is introduced. DM tests and forecasting effectiveness are also introduced.

### 2.1. Empirical Mode Decomposition (EMD)

EMD, which has been proved to be an effective data preprocessing method, is usually applied to reduce the noise of a time series, which is non-stationary [25]. The main theory of EMD is that by local characteristic timescale filtering, the original series is decomposed into oscillatory modes (IMFs). The IMFs follow the rules as shown below: (1) The extremes and the zeros must have the same quantity or the difference of quantity must be within one in the whole time series; (2) The upper and lower bounds' mean must be equal to zero. Let the raw data be  $s(t)$  ( $t = 1, 2, \dots, l$ ). The main procedures of EMD are listed below:

- (a) Set all local maxima and minima of time series.
- (b) Connect all local extreme to produce the upper bound  $e_{up}(t)$  and the lower bound  $e_{low}(t)$  by applying a cubic spline.
- (c) Compute the mean value from the upper and lower bounds  $m(t) = [e_{up}(t) + e_{low}(t)]/2$ .
- (d) Compute the difference value between the raw data and the mean value  $h(t) = s(t) - m(t)$ .
- (e) Inspect if  $h(t)$  fits characteristics of IMF. If yes,  $h(t)$  is defined as the  $i$ th IMF and the residual  $r(t) = s(t) - h(t)$  will replace  $s(t)$ . If no,  $s(t)$  will be replaced by  $h(t)$ .
- (f) Repeat the above-mentioned procedures. Stop when the value of the two successive siftings' standard deviation is lower than the threshold set earlier.

By utilizing mentioned procedures, a group of IMFs is settled from the raw data, which is arranged based on the frequency from high to low.

**Definition 1.** The raw data, which is used for decomposition and consists of  $n$  IMFs and one residual is obtained as:

$$s(t) = \sum_{i=1}^n c_i(t) + r_n(t) \quad (1)$$

where  $n$  is the number of IMFs,  $r_n(t)$  is the residuals representing a trend in  $s(t)$  ( $t = 1, 2, \dots, l$ ), and  $c_i(t)$  ( $i = 1, 2, \dots, n$ ) represents the IMFs. When describing the local characters of a raw time series, each IMF is independent.

### 2.2. Ensemble Empirical Mode Decomposition (EEMD)

EEMD, which can successfully overcome the shortcomings of EMD, is first introduced by Wu and Huang [26]. The main theory of EEMD is that by using the features of the noise, the problem of mode mixing can be effectively solved. The original time series are combined with true signals and noise. Thus, to extract the true signals in the raw data, a white noise is added to the original data. Procedures of EEMD can be described as follows.

- Step (a). Add a white noise to the raw time series.
- Step (b). Based on the method of EMD, decompose the time series with the added white noise to  $n$ IMFs.
- Step (c). Repeat the mentioned two steps, but add the white noise at different scales each time.
- Step (d). Calculate the means of each IMF of decomposition to constitute the final IMFs.

Through the mentioned process, the whole series of white noise, which are added into the raw data, can provide a unified reference range to assist the process of noise reduction. As a result, the true IMFs can be obtained from the original data.

**Definition 2.** *The relationship among the ensemble number, the error tolerance, and the added noise level is described based on the research of Wu and Huang:*

$$N_\epsilon = \frac{\epsilon^2}{\epsilon_n^2} \tag{2}$$

where  $\epsilon$  represents the amplitude of the added noise,  $\epsilon_n$  is the error's standard deviation, and  $N_\epsilon$  is value of ensemble members. Generally, it is suggested that an amplitude fixed at 0.2 will result in an accurate result [26]. In this study, the value of ensemble members are set to 100 and the optimal standard deviation of white noise series are settled from 0.1 to 0.2.

### 2.3. Cuckoo Search Optimization (CSO) Algorithm

The CSO, which was recently introduced by Yang et al., is a meta-heuristic search algorithm [27]. The main principle of the algorithm is originated from the application of cuckoo birds' obligating brood parasitic behavior and the behavior of Lévy flight. Generally, by the switching/discovery probability ( $p_a$ ), Cuckoo Search algorithm is able to perform more powerful capability of optimization [28]. In order to successfully adopt the CS algorithm, three basic assumptions are illustrated as follows:

- The egg which is generated by each cuckoo bird represents a solution in a time period, and it is dumped randomly in the nest.
- The nests which contain better eggs (better solution) are described as the best nests and they will be passed to the next generation.
- The available host nests' number is restricted to  $n$ , and each host bird is able to recognize the cuckoo bird's egg with a probability  $P_a \in [0, 1]$ . As a result, the host bird has two possible choices, which are either throwing away the egg or giving up the whole nest and finding a new location to build a new nest.

**Definition 3.** *Based on the Lévy flight behavior of the cuckoo bird's nest-seeking nature, set the current solution as  $x_i^t$  for  $i$ th cuckoo, then new solution  $x_i^{t+1}$  is generated as follows:*

$$\begin{aligned} x_i^{t+1} &= x_i^t + \alpha \oplus L(\lambda), i = 1, 2, 3 \dots n \\ L(\lambda) &= t^{-\lambda}, (1 < \lambda \leq 3) \end{aligned} \tag{3}$$

where  $\alpha > 0$  is the size of each step which relates to the optimization.  $\oplus$  is the entry wise multiplications.  $L(\lambda)$  stands for a Lévy distribution which has an infinite mean and variance. Lévy flight essentially provides a random walk while the random step length is drawn from  $L(\lambda)$ , which can produce a random walk process. Around the best solution, the local search process can be faster because of the new solution generated by Lévy walk [29]. The flowchart of CSO is shown in Figure 1.

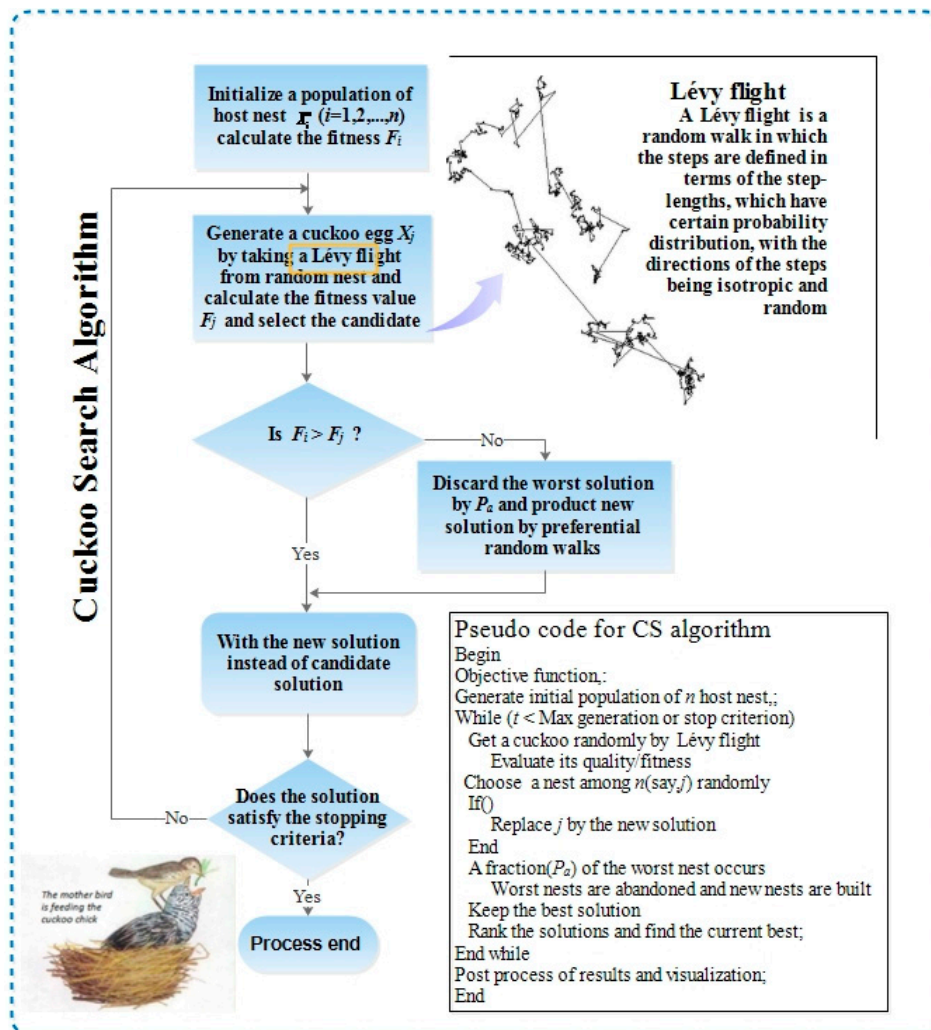


Figure 1. The flowchart of the Cuckoo Search Algorithm.

#### 2.4. Support Vector Machine (SVM)

The Support vector machine, which is introduced by Vapnik, is one of the newest algorithms [30]. Different from the conventional models, SVM follows the rule of statistical machine learning process and structural risk minimization to obtain the minimal upper bound generalization error, which is the biggest advantage of SVM compared with other models. Due to this, SVM has a great popularity and a wide use in the area of pattern recognition, classification, and regression analysis and forecasting [31,32].

**Definition 4.** Suppose a set of data  $\{x_i, d_i\}_i^n$ , the  $n$ -dimensional input vector is represented as  $x_i$ , and the output is expressed as  $d_i$ . The estimating function of SVM can be represented as:

$$f(x) = w \cdot \phi(x) + b \tag{4}$$

where  $\phi(x)$  is a nonlinear mapping while  $w, b$  is the weight vector and scalar, separately, and they are estimated through the equation:

$$R_{SVMs}(C) = \frac{1}{2} \|w\|^2 + C \frac{1}{n} \sum_{i=1}^n L(x_i, d_i) \tag{5}$$

where  $C$  represents the penalty factor of the error. The  $L(x_i, d_i)$  is the loss function  $C \frac{1}{2} \sum_{i=1}^n L(x_i, d_i)$  represents the empirical error.

**Definition 5.** Setting the upper and lower excess deviation  $\xi_i$  and  $\xi_i^*$  as the positive slack variables, the optimization problem is obtained as the follows:

$$\text{Minimize } R_{SVMs}(w, \xi^{(*)}) = \frac{1}{2} \|w^2\| + C \frac{1}{n} \sum_{i=1}^n (\xi_i + \xi_i^*) \tag{6}$$

$$\text{Subject to } \begin{cases} d_i - w\phi(x_i) - b_i \leq \varepsilon + \xi_i \\ w\phi(x_i) + b_i - d_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, \dots, l \end{cases} \tag{7}$$

where  $w^2/2$  represents the regularization term.  $\varepsilon$  stands for the loss factor, and its value is related to the approximate accuracy of the input sample.  $l$  represents the quantity of elements in the sample data series.

Equation (4) can be explained through the Lagrange function, which is described as follows:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) K(x, x_i) + b \tag{8}$$

where  $\alpha_i$  and  $\alpha_i^*$  are the multipliers of Lagrange function.  $K(x, x_i)$  represents the kernel function and  $K(x_i, x_j) = \phi(x_i)\phi(x_j)$ , which is the dot product of the two inner vectors  $x_i, x_j$  in  $\phi(x_i)$  and  $\phi(x_j)$ .

Gaussian function, which is known for its features of simplicity, efficiency, and reliable computing ability, has proved to be one of the most effective core functions [30]. SVM model, which employs it as the core function, can effectively obtain the complex features of nonlinearity of the original data and result sample by matching input data into a higher-dimensional feature space.

**Definition 6.** Gaussian function, which is employed as core function of SVM model, is described as follows:

$$K(x_i, x_j) = \exp\left(-\gamma \|x_i - x_j\|^2\right) \tag{9}$$

where the  $\gamma$  stands for the parameter of the kernel function.  $x_i, x_j$  are vector quantities in the input space.

In this paper, parameters ( $\gamma, C$ ), which are considered as the two most valuable figures that influence the performance of wind speed prediction. The CSO is used to optimize them.

### 2.5. Introduction of the EEMD-CSO-SVM Model

Based on the above-mentioned methods, the EEMD-CSO-SVM model is proposed for wind speed forecasting. Taking one of the datasets as an example, the main procedures of the proposed model is shown in Figure 2. Firstly, the EEMD noise reduction technique is applied to the raw wind speed data in order to obtain a stationary time series. In this paper, the raw data is decomposed into 10 IMFs, among which the first IMF is considered as noise and removed. The rest are recombined for the forecasting. Then, each set of data from the four wind turbines are used to perform three-step forecasting to testify the validity of the proposed model. The de-noised wind speed data is applied to train the SVM model whose key parameters ( $\gamma, C$ ) are optimized by the Cuckoo Search algorithm. The results of forecasting are recorded for further analysis.

### Procedures of the proposed hybrid model



Figure 2. The flowchart of the proposed forecasting method.

#### 2.6. DM Test and Forecasting Effectiveness

The DM test, which focuses on the difference in the forecasting precision between the proposed hybrid model and other traditional methods, is described as follows:

$$\begin{aligned}
 H_0 &: E(d_n) = 0, \forall n \\
 H_1 &: E(d_n) \neq 0, \exists n
 \end{aligned}
 \tag{10}$$

The value of the DM test is described as follows:

$$DM = \frac{\sum_{h=1}^k \left( L(\varepsilon_{t+h}^{(i)}) - L(\varepsilon_{t+h}^{(j)}) \right) / k}{\sqrt{S^2/k}}
 \tag{11}$$

where  $\varepsilon_{t+h}$  is the forecasting error.

$S^2$  represents the estimation value for the variance of  $d_h = L(\varepsilon_{t+h}^{(i)}) - L(\varepsilon_{t+h}^{(j)})$ .  $L$  is the loss function, which represents the accuracy of the forecasting.

Among the loss functions, absolute deviation error loss and square error loss are widely applied. They are described as follows:

Absolute deviation error loss:

$$L(\varepsilon_{t+h}^{(i)}) = \left| \varepsilon_{t+h}^{(i)} \right|
 \tag{12}$$

Square error loss:

$$L(\varepsilon_{t+h}^{(i)}) = \left( \varepsilon_{t+h}^{(i)} \right)^2
 \tag{13}$$

If no significant differences are found between the performances of the included methods, the null hypothesis will be rejected.

The null hypothesis is described as follows:

$$|DM| > z_{\alpha/2}
 \tag{14}$$

where  $z_{\alpha/2}$  represents the critical value of the standard value distribution when the value of significance is  $\alpha$ .



Forecasting effectiveness, which evaluates the performance of the proposed model using the sum of the squared errors and the mean squared deviation of the forecasting results is also applied. The forecasting effectiveness is described as follows [33].

The  $k$ th order forecasting effectiveness unit is described as:

$$m^k = \sum_{n=1}^N Q_n A_n^k \tag{15}$$

$$\sum_{n=1}^N Q_n = 1$$

where  $Q_n$  stands for the discrete probability distribution when the time is  $n$ .  $A_n$  represents the forecasting accuracy.

The  $k$ -order forecasting effectiveness is presented as:

$$H(m^1, m^2, \dots, m^k) \tag{16}$$

The first order forecasting effectiveness is defined as  $H(m^1) = m^1$  while the second-order forecasting effectiveness is the difference between the standard deviation and expectation, which is shown as:

$$H(m^1, m^2) = m^1 \left( 1 - \sqrt{m^2 - (m^1)^2} \right) \tag{17}$$

### 3. Performance Evaluation Criterion

In this study, to test the accuracy of the proposed approach, six evaluation indexes including MAE, MAPE, RMSE, WI,  $E_{NS}$ , and  $E_{LM}$  are applied for evaluating forecasting accuracy. These indexes are shown as follows in Table 1:

**Table 1.** The description of the error evaluation indexes.

Metric	Definition	Equation
MAE	The mean absolute error of $N$ forecasting results	$MAE = \frac{1}{N} \sum_{i=1}^N  yp_i - y_i $
MAPE	The average of $N$ absolute percentage error	$MAPE = \frac{1}{N} \sum_{i=1}^N \left  \frac{yp_i - y_i}{y_i} \right  \times 100\%$
RMSE	The square root of the average of the error square	$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (yp_i - y_i)^2}$
WI	Willmott's Index	$WI = 1 - \left[ \frac{\sum_{i=1}^N (yp_i - y_i)^2}{\sum_{i=1}^N ( yp_i - \bar{y}  +  y_i - \bar{y} )^2} \right], 0 \leq WI \leq 1$
$E_{NS}$	Nash-Sutcliffe coefficient	$E_{NS} = 1 - \left[ \frac{\sum_{i=1}^N (yp_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right], \infty \leq E_{NS} \leq 1$
$E_{LM}$	Legates and McCabe Index	$E_{LM} = 1 - \left[ \frac{\sum_{i=1}^N  y_i - yp_i }{\sum_{i=1}^N  y_i - \bar{y} } \right], \infty \leq E_{LM} \leq 1$

Respectively, where  $N$  is the total output samples;  $y_i$  represents the actual series; and  $yp_i$  stands for the prediction results. For WI,  $E_{NS}$ , and  $E_{LM}$ , when the values of them are closer to 1, the model achieves a higher performance.

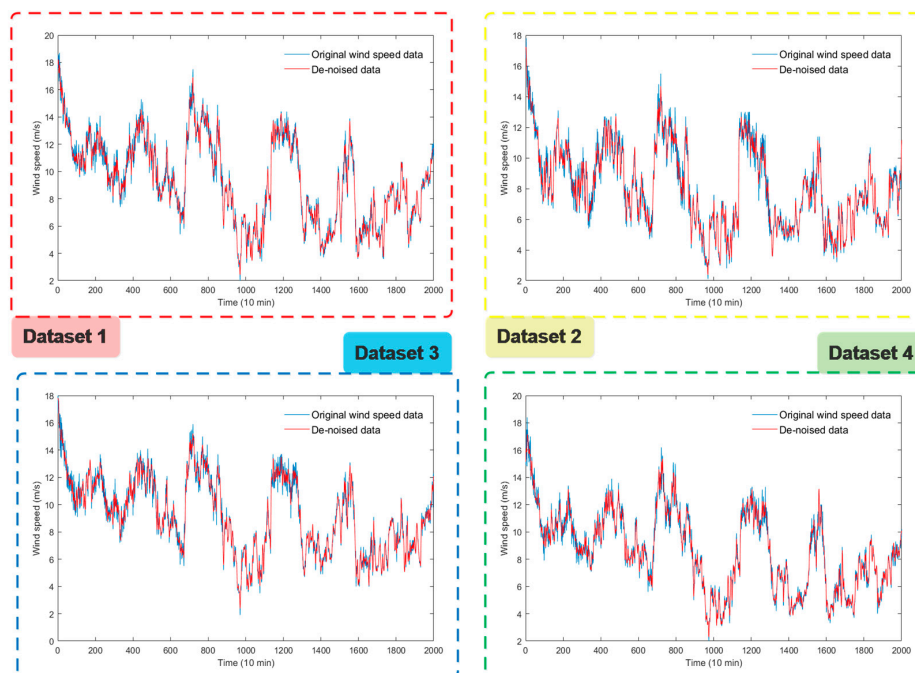
## 4. Numerical Experimentation

The Windows 7 Professional operating system was used to perform the experiments. The specific version of software, which was used to conduct the proposed model, is Matlab2016a. The details of the hardware are: Intel Core i5-3230M2.60 GHz CPU, and 4 GB RAM.

### 4.1. Introduction of Datasets

Penglai is located in the east coast of Shandong in China. It has a 3100 km coastline. Thus, although the whole region is not large, it is famous for its abundant wind power due to this unique geographical feature. The approximate wind power capacity of the region is 67 million KW. In this paper, datasets 1, 2, 3, and 4 were chosen from a wind farm in Penglai, with the latitude from  $120^{\circ}43'$  N to  $120^{\circ}47'$  N and longitude from  $37^{\circ}50'$  E to  $37^{\circ}37'$  E. These datasets are located in mountain and hilly areas whose altitude is from 100 m to 240 m. The features of the wind power generator are provided as follows: Rated power: 1500 KW. Height of measurement: 70 m. Sampling time period: 10 min. Scanning frequency: 144 times per day. All four datasets were included in the experiment, to help analyze the differences in the results. Multistep forecasting was also conducted in this paper.

Each group is divided into a training group and a testing group. The size ratio of the training and testing group is set to 9:1. The training sample included a total of 1350 10-min wind speed series, and the testing sample contains 150 wind speed data points. The ratio of input and output data is set to 4:1. Figure 3 shows the structure of each dataset.



**Figure 3.** The original wind speed data and de-noised data from four datasets.

### 4.2. Forecasting Model Parameter Setting

Wind speed series from four datasets are chosen to test the forecasting accuracy of the proposed hybrid model. The results of the proposed model are also used to compare with other conventional methods; namely, BP neural networks, the ARIMA model, and the RBF model. This paper followed the standard of energy industry NB/T31046-2013 and the rules for measuring wind sources, which were published and made by the National Energy Administration in 2013.

- (1) For BPNN, the newff function of the neural network toolbox is employed to build the network. The dimensions of the input, hidden, and output layers are 4, 5, and 1, respectively. The learning

rate is set to 0.1, the maximum number of iterations is set to 100, and the training precision is set to 0.00004.

- (2) For ARIMA, the forecasting results are influenced by the moving average and the order of auto-regressive. The observed value's fitting effect is measured by the AIC criterion, and the AIC also calculates the most suitable number for the parameters. When the AIC reaches the lowest value, the ARIMA method can achieve the best order.
- (3) For RBFNN, similar with BPNN, the newrb function of the neural network toolbox is employed to build the forecasting network. The same parameters as the BPNN are used in the RBFNN.

### 4.3. Experimental Results for Datasets

The original time series are first preprocessed through the EEMD. Figure 3 presents the preprocessed data obtained by the EEMD method for the four wind turbines. As indicated in Figure 3, for the #2, #3, and #4 wind turbines, 10 IMF sequences are obtained from the original training data of the time series. According to the principles of denoising, eliminating the high-frequency sequence from the IMF sequences can assist in obtaining a cleaner data sequence; that is, a data sequence with lower noise. For this paper, the first IMF sequence obtained by the EEMD method is eliminated from the original data sequence due to its high frequency, so that a stationary time series can be obtained to improve the accuracy of the prediction. Taking wind turbine #2's data as an example, the visualization of the de-noise preprocessing of the EEMD method is shown in Figure 4. The final result after the de-noise processing with the EEMD method is also presented in Figure 4. The preprocessing results of all four wind turbines are shown in Figure 3.

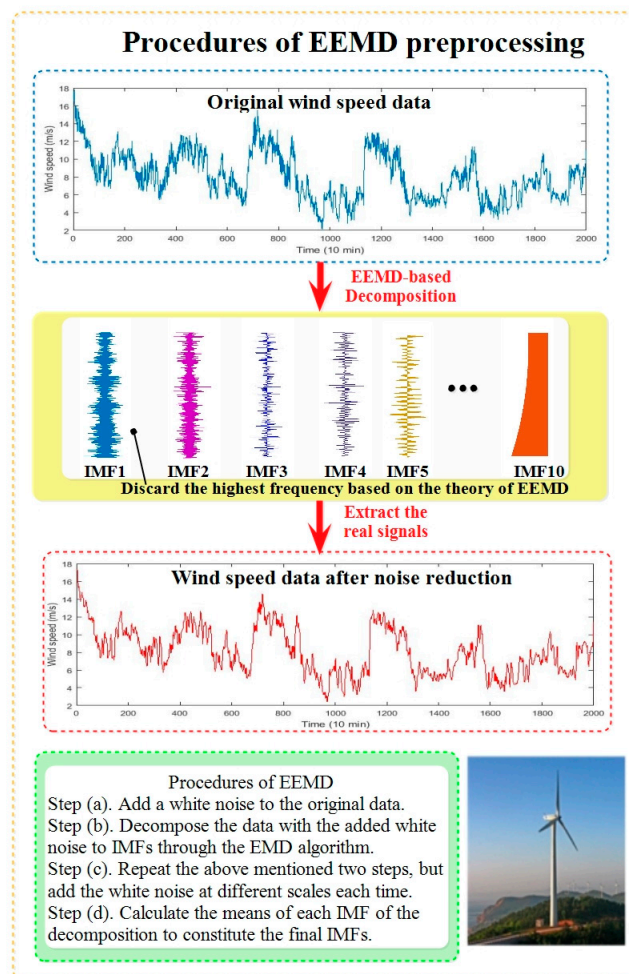


Figure 4. Wind speed data noise reduction using EEMD preprocessing process.

For the experiment, the proposed method is trained using the selected data from four datasets. Multistep forecasting is applied in the experiment, which performs the prediction through removing the old input data for each circulation. By applying the previous output values rather than the actual series, the multistep ahead method predicts the next wind speed value through this circulation [34]. The one-step forecasting result is calculated based on the observed values from  $x_1$  to  $x_n$ , where  $n$  is the number of output value. Two-step forecasting result is calculated based on the observed value from  $x_2$  to  $x_n$  and the one-step result. Additionally, three-step forecasting result is obtained based on the observed value from  $x_3$  to  $x_n$  and the previous two-steps results. The validity of the proposed hybrid model was analyzed based on the results of multistep ahead forecasting. Tables 2–5 show the forecasting results of different models in four wind turbines, respectively. The results of multistep ahead forecasting are shown Table 6.

The performance of hybrid and traditional methods in one-step forecasting are compared. To testify the effectiveness of the CSO, Particle swarm optimization (PSO) is used for comparison. PSO, which is similar with CSO, optimizes the parameters based on the velocity of particle and the method of position updating. The MAPE values of the proposed hybrid are 4.79%, 3.07%, 2.69%, and 3.91% in four wind turbines, respectively. The values of Willmott’s Index reach 0.9838, 0.9929, 0.9918, and 0.9839, respectively. The other values of forecasting error indexes also indicate that the hybrid approach performs better than traditional methods. The results are shown in Tables 2 and 3.

In the results of dataset 2, the MAPE values of the proposed model are 3.07%, 6.53%, and 10.70% in three-steps forecasting, separately, which shows that one step forecasting performs better. Multistep forecasting, which is based on the theory of removing old input data and adding new output data from the previous step, cannot achieve the same performance accuracy as single step forecasting.

**Remark 1.** *The hybrid approach includes more parameters so that better performance can be achieved. As for the multistep forecasting, its process is complex and leads to high error because less historical data is employed and the forecasted result of each step will be included as input in the next circulation. Therefore, with more steps of forecasting being conducted, more errors will appear, leading to poor performance.*

**Table 2.** Comparison of the value of RMSE, MAE, and MAPE between the proposed model and some related models.

Model	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
Wind turbine 1			Wind turbine 3			
EEMDCSOSVM	0.3513	0.1815	4.79	0.2463	0.2120	2.69
EEMDPSOSVM	0.4652	0.2159	5.63	0.3027	0.2436	2.82
CSOSVM	0.8294	0.2975	10.67	0.5121	0.4251	5.08
PSOSVM	1.0135	0.7100	12.18	0.7928	0.6362	9.70
Wind turbine 2			Wind turbine 4			
EEMDCSOSVM	0.2342	0.1841	3.07	0.2679	0.3028	3.91
EEMDPSOSVM	0.2744	0.2191	3.24	0.3147	0.1869	4.44
CSOSVM	1.1381	0.1927	7.63	0.9204	0.6137	9.27
PSOSVM	0.6680	0.5165	8.58	1.1478	0.8190	10.15

**Table 3.** Comparison of the value of WI,  $E_{NS}$ , and  $E_{LM}$  between the proposed model and some related models.

Model	WI	$E_{NS}$	$E_{LM}$	WI	$E_{NS}$	$E_{LM}$
Wind turbine 1			Wind turbine 3			
EEMDCSOSVM	0.9838	0.9332	0.7600	0.9918	0.9670	0.8268
EEMDPSOSVM	0.9723	0.9251	0.7148	0.9744	0.9310	0.7201
CSOSVM	0.9152	0.6858	0.4460	0.9667	0.8720	0.6525
PSOSVM	0.8622	0.5367	0.3599	0.8239	0.4950	0.1197
Wind turbine 2			Wind turbine 4			
EEMDCSOSVM	0.9929	0.9713	0.8388	0.9839	0.9350	0.7565
EEMDPSOSVM	0.9876	0.9424	0.7950	0.9786	0.9234	0.7159
CSOSVM	0.9206	0.7011	0.5134	0.8274	0.4885	0.2514
PSOSVM	0.9174	0.6852	0.4812	0.7490	0.3355	0.1253

## 5. Analysis of the Forecasting Results

This section analyzes the experimental results of the hybrid approach, and the effectiveness of the proposed method is verified. First, based on the results of single step forecasting, the hybrid approach is compared with some existing conventional models. Then, the results of multi-step forecasting are used for further analysis.

### 5.1. Single-Step Forecasting

This is divided into two parts to verify the effectiveness of the hybrid approach. First, the hybrid method is tested by comparing it with conventional methods. Then, the results of four different datasets are analyzed.

#### 5.1.1. Analysis of the Proposed Method and Conventional Models

The forecasting results of the proposed model and conventional models are shown in Tables 4 and 5. The original time series and preprocessed data obtained by the EEMD technique is shown in Figure 3. By comparing the results illustrated above, the following conclusions are made:

The fluctuation and instability of the original time series are clearly shown in Figure 3. Additionally, in Figure 3, the reconstructed wind speed data in which the signals with high frequency have been removed obviously shows reduced noise of the raw data. By comparing the forecasting results of the new preprocessed data with the original data from Tables 2 and 3, the model achieves a high performance when using the de-noised wind speed data. The MAPE values of the CSOSVM model decrease by 5.88%, 5.56%, 2.39%, and 7.36%, respectively. Using the PSOSVM model for comparison, the MAPE values also decrease by 6.55%, 6.88%, 5.34%, and 5.71% in four datasets, respectively. Therefore, the EEMD technique has good validity.

The proposed hybrid approach achieves better performance than the conventional methods in wind speed prediction. From Table 4, the MAPE value of BPNN, RBFNN, ARIMA, and the proposed model in wind turbine 1 are 8.68%, 10.32%, 8.41%, and 4.79%, respectively. In the other three datasets the proposed method also achieves higher performance than other conventional methods. The values of the evaluation metrics shown in Tables 4 and 5 all suggest that the hybrid model achieves higher accuracy. The new hybrid approach can be regarded as more effective than traditional methods.

**Remark 2.** *The data preprocessing technique is a reliable method to reduce the fluctuation in the raw data, and the EEMD proves to be an effective method to achieve this purpose. Therefore, the proposed hybrid approach outperforms the conventional methods in forecasting.*

**Table 4.** Comparison of the value of RMSE, MAE, and MAPE between the proposed model and some traditional models (BP RBF ARIMA).

Model	RMSE	MAE	MAPE (%)	RMSE	MAE	MAPE (%)
Wind turbine 1			Wind turbine 3			
BPNN	0.5939	0.4984	8.68	0.4794	0.3286	5.28
RBFNN	0.8867	0.5924	10.32	0.6587	0.4468	6.52
ARIMA	0.5784	0.4672	8.41	0.4472	0.2907	5.13
EEMDCSOSVM	0.3513	0.1815	4.79	0.2463	0.2120	2.69
Wind turbine 2			Wind turbine 4			
BPNN	0.3998	0.3232	5.42	0.4585	0.3385	6.34
RBFNN	0.4847	0.3806	6.41	1.2374	0.5543	10.32
ARIMA	0.3627	0.2925	5.61	0.4122	0.3679	5.88
EEMDCSOSVM	0.2342	0.1841	3.07	0.2679	0.3028	3.91

**Table 5.** Comparison of the value of WI,  $E_{NS}$  and  $E_{LM}$  between the proposed model and some traditional models (BP RBF ARIMA).

Model	WI	$E_{NS}$	$E_{LM}$	WI	$E_{NS}$	$E_{LM}$
	Wind turbine 1			Wind turbine 3		
BPNN	0.9410	0.7822	0.5403	0.9708	0.8850	0.6806
RBFNN	0.8926	0.6017	0.4630	0.9429	0.7751	0.6128
ARIMA	0.9476	0.8137	0.6512	0.9755	0.9024	0.7148
EEMDCSOSVM	0.9838	0.9332	0.7600	0.9918	0.9670	0.8268
	Wind turbine 2			Wind turbine 4		
BPNN	0.9731	0.8889	0.6962	0.9412	0.7628	0.5877
RBFNN	0.9595	0.8320	0.6422	0.7035	0.3339	0.3248
ARIMA	0.9752	0.8933	0.7017	0.9503	0.7937	0.6128
EEMDCSOSVM	0.9929	0.9713	0.8388	0.9839	0.9350	0.7565

### 5.1.2. Analysis of the Four Different Wind Turbines

This paper employs four wind turbines to conduct the forecasting experiment of the wind speed. The original time series of the four datasets are presented in Figure 3. The results from the four different datasets using different approaches are shown in Tables 2–5.

From Figure 3, the general trend of the four time series is approximately the same, though there are still differences in some time points. The main reason for this difference is that in a general location, the wind speed is nearly the same. However, when narrowing down to a specific wind farm unit, the location and size of the wind farm varies, resulting in different wind speed data from one dataset to another.

When analyzing from the specific forecasting results, according to the data presented in Table 4 for instance, the MAPE values of the proposed model from the four wind turbines are 4.79%, 3.07%, 2.69%, and 3.91%. For the forecasting accuracy, dataset 3 achieves the highest performance.

**Remark 3.** *Because of the uncertainty of the magnitude and direction of the wind in different locations, the forecasting results of four different datasets vary. The wind speed also experiences enormous fluctuations in different time periods. As a result, the prediction precision also varies due to different locations and time periods.*

### 5.2. Multi-Step Forecasting

Multi-step forecasting is an effective way to test the accuracy of the forecasting method. Therefore, this paper employed multistep forecasting on the proposed hybrid model. The experimental results are used to verify the accuracy of the proposed approach. The results of the multi-step forecasting are presented in Table 6. Figures 5–7 show the output values of different methods using multi-step forecasting. In 1-step ahead forecasting, the MAPE values of the BP, ARIMA, RBF, and the proposed model are 5.46%, 5.22%, 4.83%, and 3.07%, respectively. In 2-step ahead forecasting, the MAPE values of these models are 6.70%, 6.42%, 5.97%, and 4.55%. Additionally, in 3-step ahead forecasting, the MAPE values of the four models are 7.83%, 7.15%, 6.92%, and 5.46%, which indicates that the proposed hybrid model achieves a better performance than other conventional models.

When comparing the results of three-step forecasting with each other, one-step forecasting clearly performs better than the others.

**Remark 4.** *These results can be summarized as follows: The proposed hybrid approach obtains higher accuracy compared with the traditional models used in our experiments in both single step and multi-step forecasting. Therefore, the proposed hybrid approach is valid.*

**Table 6.** Results of multi-step forecasting of the proposed model.

Dataset		MAE	MAPE (%)	RMSE	WI	E <sub>NS</sub>	E <sub>LM</sub>
Dataset 1	1-Step	0.1815	4.79	0.3513	0.9838	0.9332	0.7600
	2-Step	0.3622	6.19	0.4546	0.9720	0.8874	0.6699
	3-Step	0.5895	10.14	0.7354	0.9249	0.7115	0.4609
Dataset 2	1-Step	0.1841	3.07	0.2342	0.9929	0.9713	0.8388
	2-Step	0.3962	6.53	0.5034	0.9583	0.8355	0.6310
	3-Step	0.6579	10.70	0.8291	0.8887	0.6073	0.3925
Dataset 3	1-Step	0.2120	2.69	0.2463	0.9918	0.9670	0.8268
	2-Step	0.4814	5.97	0.5308	0.9529	0.8306	0.5820
	3-Step	0.8694	9.52	0.8329	0.8599	0.5705	0.3240
Dataset 4	1-Step	0.3028	3.91	0.2679	0.9839	0.9350	0.7565
	2-Step	0.4275	5.12	0.3652	0.9645	0.8604	0.6667
	3-Step	0.5310	8.03	0.5437	0.9189	0.6883	0.4797

## 6. Discussion

This section discusses the sample selection, experimental results, which consist of each element in the proposed hybrid approach, the error evaluation indexes, and the forecast performance based on the DM (Diebold-Mariano) test and the Forecasting Effectiveness test.

### 6.1. Sample Selection

Currently, there is no explicit theory about how to select the number of training and testing sample. Too small and the input sample cannot train the model well, while too large and the sample will cause over-fitting. In terms of wind speed forecasting, how to select the number of input set is still a difficult and challenging issue [35,36]. It can only get the optimal input set by experiments. Besides, we organized and list the simulation results based on different ratio between training and testing samples in Table 7. Taking the one-step forecasting result of dataset one as an example, the MAPE values of the proposed model are 5.40%, 4.92%, 4.79%, 4.95%, and 5.11% when the ratio of training and testing sample is set to 7:3, 8:2, 9:1, 14:1, and 29:1, respectively. The forecasting accuracy of other ratios are all lower than the ratio 9:1's accuracy. Table 1 below shows that the experiment we took before choosing a suitable ratio of training sample and testing sample. It can be indicated that when the ratio of training and testing sample is set to 9:1, the performance of wind speed forecasting outperforms the other ratios used in the experiment. Thus, based on the experiments and experience sample data are selected.

**Table 7.** MAPE (%) values of different ratios between training and testing sample of the proposed model.

Datase	Forecasting Steps	Raito Between the Training Sample and Testing Sample (MAPE (%))				
		7:3	8:2	9:1	14:1	29:1
Dataset 1	1-step	5.40	4.92	<b>4.79</b>	4.95	5.11
	2-step	9.57	8.15	<b>6.19</b>	8.82	8.97
	3-step	13.94	13.46	<b>10.14</b>	14.84	15.73
Dataset 2	1-step	4.16	3.91	<b>3.07</b>	3.54	3.76
	2-step	9.32	7.52	<b>6.53</b>	7.32	8.44
	3-step	12.69	11.24	<b>10.70</b>	11.15	11.80
Dataset 3	1-step	3.48	3.28	<b>2.69</b>	3.41	3.97
	2-step	6.62	6.12	<b>5.97</b>	6.15	6.72
	3-step	10.73	9.82	<b>9.52</b>	9.84	10.15
Dataset 4	1-step	4.88	4.09	<b>3.91</b>	4.16	4.51
	2-step	7.26	6.65	<b>5.12</b>	5.52	5.84
	3-step	11.01	10.39	<b>8.03</b>	8.85	8.99

6.2. Forecasting Error Analysis

The proposed hybrid approach outperforms the conventional methods. The forecasting results and precision of the proposed model in four different wind turbines in multi-step forecasting are shown in Table 6. Then, Figures 5–7 show the forecasting results of multi-step forecasting from the proposed approach and different traditional methods. The following conclusions are obtained from these results: (1) From the results of three-step forecasting, we can see that the hybrid method is more accurate than other conventional models because the error is the lowest many times. (2) The degree of fitting between the output series and the actual data from different models is shown in Figures 5–7. The proposed approach is superior to the corresponding traditional methods, with a higher precision.

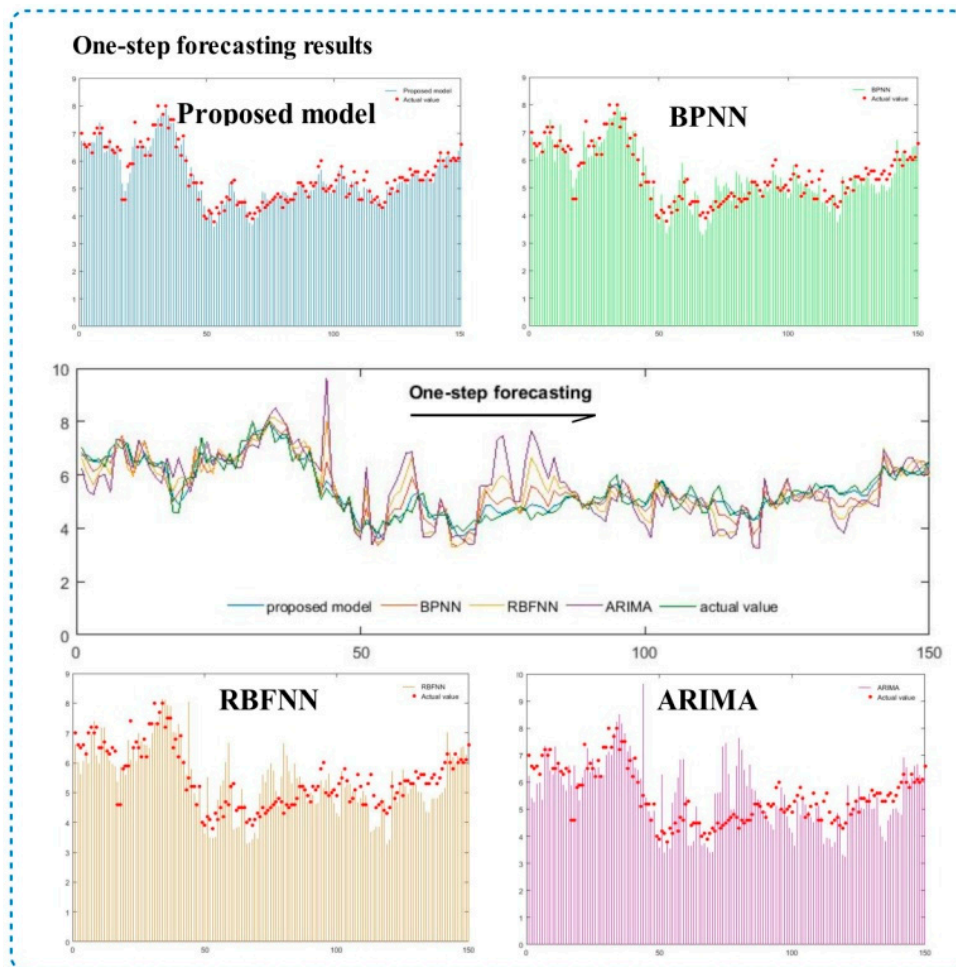
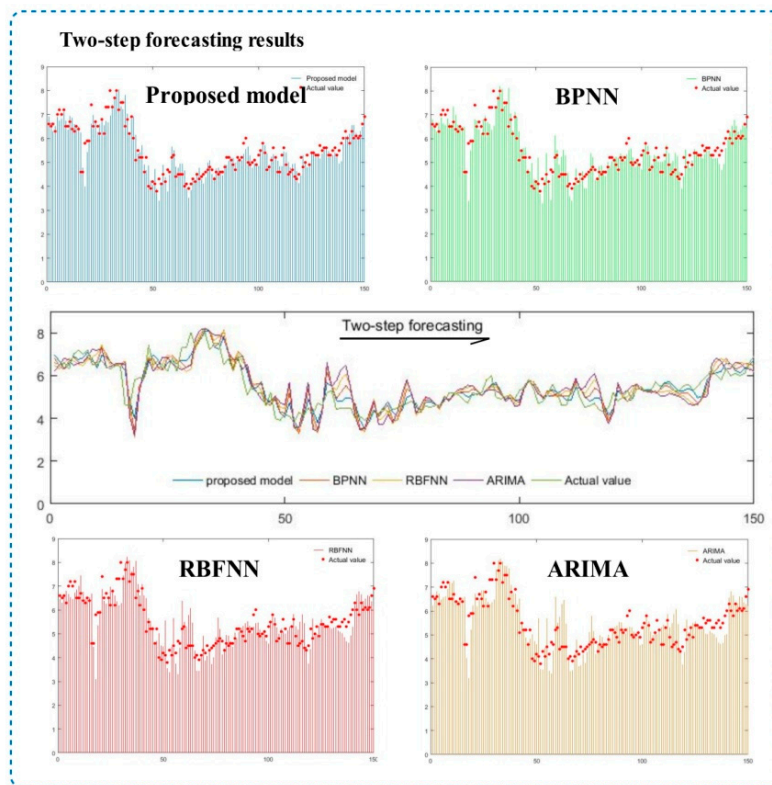
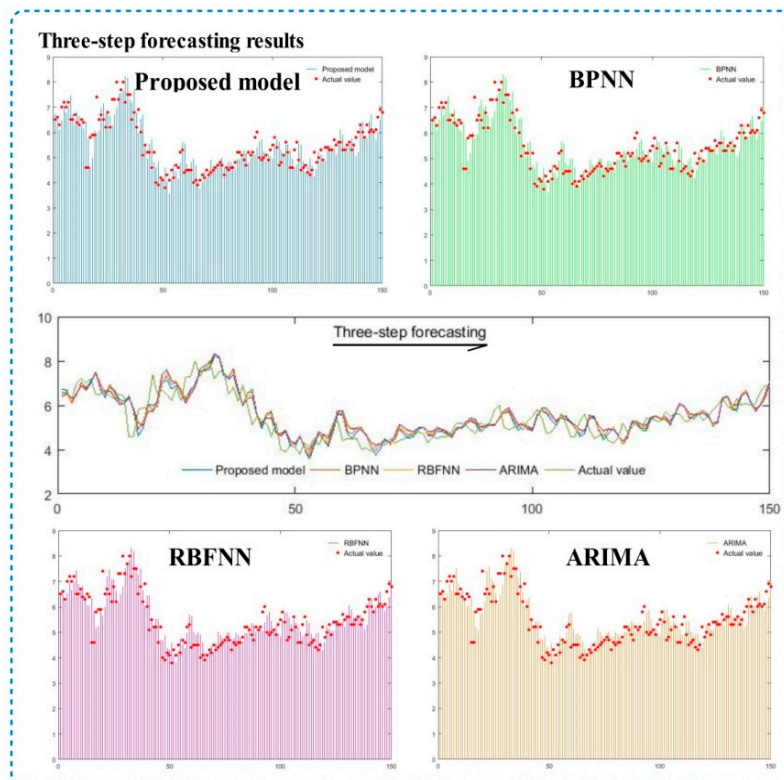


Figure 5. One-step forecasting results of the proposed model and other traditional models (BP RBF ARIMA).





**Figure 6.** Two-step forecasting results of the proposed model and other traditional models (BP RBF ARIMA).



**Figure 7.** Three-step forecasting results of the proposed model and other traditional models (BP RBF ARIMA).

Multi-step forecasting is employed to further test the validity of the new hybrid approach. From Table 6, the results of multi-step forecasting of four wind turbines are shown. For wind turbine 1, the MAPE values of multi-step forecasting are 4.79%, 6.19%, and 10.14%, respectively. The other three wind turbines obtain similar results. As the results show, the new hybrid model is valid in multi-step forecasting.

**Remark 5.** *From the analysis, the performance of the hybrid method is better than those of the traditional models. The proposed model is able to adapt to the fluctuation of the input data so that it can obtain a more accurate result. Therefore, the new hybrid approach is more effective in wind speed prediction.*

### 6.3. Validity of the Data Preprocessing Technique

The irregularity of the raw wind speed series often makes the input data contain high noise and fluctuation. Therefore, removing the noise from the original data is important to obtain a better forecasting accuracy. To test the validity of this data preprocessing technique, the results of the proposed model for four datasets using original and preprocessed data are listed together for comparison. From Tables 2 and 3, the MAPE values of the proposed model decrease by 5.88%, 5.56%, 2.39%, and 7.36% in four wind turbines compared with the CSO-SVM model without the EEMD process. The MAE values also decrease by 0.1160, 0.0086, 0.2031, and 0.3109 in four wind turbines, respectively. The WI value increases by 0.0686, 0.0723, 0.0251, and 0.1565, respectively. The results indicate that the EEMD effectively improves the accuracy of the forecasting. As Tables 2 and 3 show, all six metrics changed positively, and the results demonstrate that after reducing the noise of the raw data, the performance of the hybrid model improved significantly.

### 6.4. Significance of the Error Evaluation Indexes

Based on the results of six error evaluation indexes, namely MAPE, MAE, RMSE, WI,  $E_{NS}$ , and  $E_{LM}$ , the significance of the proposed hybrid model is estimated. Taking the forecasting results of dataset 2 as an example, the MAPE values of the proposed model, BPNN, RBFNN, and ARIMA are 3.07%, 5.42%, 6.41%, and 5.61%, respectively. The WI values are 0.9929, 0.9731, 0.9595, and 0.9752, respectively. Combining the results of these together, the hybrid model is proven to be superior to the traditional methods. In conclusion, the forecasting ability of the proposed model can be better analyzed with the help of these evaluation indexes, and the comparison between the proposed model and the other forecasting methods is more comprehensive.

### 6.5. Results of DM Test and Forecasting Effectiveness

Besides the error evaluation indexes, which have been applied to evaluate the forecasting ability of the proposed hybrid approach, this paper also used the DM (Diebold-Mariano) test and Forecasting Effectiveness to further study the forecasting accuracy.

In this paper, the DM test is used to study the significant differences in forecasting ability between the new hybrid model and some conventional methods. The results of the DM test are shown in Table 8. The DM statistical values far exceed the critical value at the 1% significance level, at which the value of  $|DM|$  is 2.1017. Thus, the proposed approach performs significantly different with the conventional methods at the 1% significance level. In conclusion, the proposed approach significantly outperforms the conventional methods.

**Table 8.** Results of DM test.

Datasets		BPNN	RBFNN	ARIMA
Dataset 1	1-step	7.2859	7.4392	8.0205
	2-step	6.8073	7.1682	7.5217
	3-step	6.5270	6.9463	7.2258
Dataset 2	1-step	6.9469	7.4245	6.2829
	2-step	6.7716	7.2839	6.5691
	3-step	6.8358	7.1475	6.2039
Dataset 3	1-step	7.1325	6.2161	6.2934
	2-step	7.0413	6.4755	6.7390
	3-step	6.9427	6.3586	6.4803
Dataset 4	1-step	6.8941	7.2790	7.1783
	2-step	7.2564	7.2387	7.0492
	3-step	6.4396	6.9257	6.5864

The results of forecasting effectiveness is shown in Table 9. The first-order forecasting effectiveness is based on the expected value of the forecasting accuracy sequence, while the second-order forecasting effectiveness is related to the difference between the standard deviation and expectation of the forecasting accuracy sequence. In Table 8, the forecasting effectiveness of the proposed model outperforms all the other models in both first order and second order.

**Table 9.** Forecasting Effectiveness of different models for four datasets.

Dataset		BPNN	RBFNN	ARIMA	EEMDCSOSVM
Dataset 1	First order	0.8951	0.8889	0.8874	<b>0.9146</b>
	Second order	0.8322	0.8257	0.8283	<b>0.8638</b>
Dataset 2	First order	0.9130	0.9029	0.9072	<b>0.9367</b>
	Second order	0.8384	0.8362	0.8436	<b>0.8842</b>
Dataset 3	First order	0.9283	0.9177	0.9220	<b>0.9409</b>
	Second order	0.8517	0.8582	0.8625	<b>0.8975</b>
Dataset 4	First order	0.8956	0.8934	0.9031	<b>0.9269</b>
	Second order	0.8344	0.8360	0.8492	<b>0.8816</b>

### 7. Conclusions

With a rapidly rising demand for clean energy obtained from renewable power resources, more attention and resources have been focused on the development of effectively utilizing these energy resources. Among those resources, wind power has the most promising future. In the area of the prediction, not only the accuracy, but also the stability, should be seen as key factors of the forecasting model. However, due to the uncertainty and intermittence of the raw wind speed data, the forecasting results obtained by conventional models cannot meet this goal. Additionally, with the poor results of forecasting, the power grids cannot make adjustments to the wind plan timely, causing low productivity and increasing the operation cost of the wind farms. Therefore, it is necessary to develop a short-term forecasting method that can achieve satisfactory accuracy and stability at the same time [37]. This paper suggests a new hybrid model that combines a parameter optimization algorithm (CSO) and a forecasting module (SVM). The key weight coefficients of the forecasting module are optimized by the proposed algorithm. To reduce the noise in the raw data, a data preprocessing technique (EEMD) is used to obtain a stationary time series. To verify the validity of the proposed model, multi-step ahead forecasting and several performance evaluation metrics are applied in this paper. Traditional forecasting methods are used for comparison. In one-step ahead forecasting, the MAPE values of the BP, ARIMA, RBF, and the proposed model are 5.46%, 5.22%, 4.83%, and 3.07%, respectively. In two-step ahead forecasting, the MAPE values of these models are 6.70%, 6.42%, 5.97%,

and 4.55%, while in three-step ahead forecasting, the MAPE values of the four models are 7.83%, 7.15%, 6.92%, and 5.46%. Not only the figures of MAPE, but also other five variables and the two tests, show the excellent performance of the proposed approach. Based on the experimental results, the newly proposed approach achieved the highest performance in wind speed prediction in both one-step forecasting and multi-step forecasting, in comparison with the other three methods used in the experiment. In conclusion, based on the experimental results, the proposed approach achieved significant improvement in both forecasting accuracy and stability. The proposed a hybrid model, which effectively improved the accuracy and the stability of wind speed forecasting, can be a great tool in managing wind farms. Since this paper used the wind speed data from the power grids of China, the proposed model can adapt to the system smoothly, which can reduce the costs and risks of the power station. When employed in the dispatch of the power system, the proposed hybrid model can generate benefits economically. For instance, the wind station is able to make timely adjustments of the operating plan, and the accurate forecasting results can reserve the capacity of the wind farm, saving unnecessary costs, and avoiding generating waste. As for the focus of future studies, this proposed hybrid approach could be employed in other areas relating to trend forecasting, namely extreme weather prediction, profit forecasting, and traffic condition forecasting.

**Author Contributions:** Conceptualization, T.L.; Methodology, T.L.; Software, Y.G.; Validation, T.L. and J.H.; Formal Analysis, S.L.; Investigation, T.L.; Resources, S.L.; Data Curation, Y.G.; Writing-Original Draft Preparation, T.L.; Writing-Review & Editing, T.L.; Visualization, J.H.; Supervision, S.L.; Project Administration, S.L.

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

## Abbreviation

Variables	Meaning
$s(t)$	Raw data
$c_i(t)$	Residuals in raw data
$r_n(t)$	IMFs of the raw data
$\varepsilon$	Amplitude of the added noise
$\varepsilon_n$	Standard deviation of the error
$N_\varepsilon$	Value of ensemble member
$x_i^{t+1}$	New solution of cuckoo search
$x_i^t$	Current solution of cuckoo search
$\alpha$	Size of each step in cuckoo search
$\oplus$	Entry wise multiplications
$L(\lambda)$	Lévy distribution
$\phi(x)$	Nonlinear mapping
$w$	Weight vector of SVM
$b$	Scalar of SVM
$C$	Penalty factor of the error
$L(x_i, d_i)$	Loss function
$C \frac{n}{2} \sum_{i=1}^n L(x_i, d_i)$	Empirical error
$w^2/2$	Regularization term
$l$	Quantity of elements in the sample data series
$\alpha_i, \alpha_i^*$	Multipliers of Lagrange function
$K(x, x_i)$	Kernel function of SVM
$\gamma$	Parameter of the kernel function
$N$	Total output samples
$y_i$	Actual series
$yp_i$	Predicting results

## References

- Salcedo-Sanz, S.; Pérez-Bellido, Á.M.; Ortiz-García, E.G.; Portilla-Figueraa, A.; LuisPrieto, L.; Correosoc, F. Accurate short-term wind speed forecasting by exploiting diversity in input data using banks of artificial neural networks. *Neurocomputing* **2009**, *72*, 1336–1341. [[CrossRef](#)]
- Soman, S.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 2010, Arlington, TX, USA, 26–28 September 2010.
- 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](#)]
- Mohandes, M.A.; Halawani, T.O.; Rehman, S.; Hussaina, A.A. Support vector machines for wind speed prediction. *Renew. Energy* **2004**, *29*, 939–947. [[CrossRef](#)]
- Liu, H.; Tian, H.-Q.; Chen, C.; Li, Y. A hybrid statistical method to predict wind speed and wind power. *Renew. Energy* **2010**, *35*, 1857–1861. [[CrossRef](#)]
- Liu, H.; Tian, H.Q.; Pan, D.F.; Li, Y.F. Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. *Appl. Energy* **2013**, *107*, 191–208. [[CrossRef](#)]
- Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320. [[CrossRef](#)]
- Barbounis, T.G.; Theocharis, J.B.; Alexiadis, M.C.; Dokopoulos, P.S. Long-term wind speed and power forecasting using local recurrent neural network models. *IEEE Trans. Energy Convers.* **2006**, *21*, 273–284. [[CrossRef](#)]
- Cheng, C.H.; Wei, L.Y. A novel time-series model based on empirical mode decomposition for forecasting TAIEX. *Econ. Model.* **2014**, *36*, 136–141. [[CrossRef](#)]
- Zhou, J.Y.; Shi, J.; Li, G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers. Manag.* **2011**, *52*, 1990–1998. [[CrossRef](#)]
- Liu, D.; Niu, D.X.; 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](#)]
- Hu, Q.H.; Zhang, S.G.; Xie, Z.X.; Mi, J.S.; Wan, J. Noise model based v-support vector regression with its application to short-term wind speed forecasting. *Neural Netw.* **2014**, *57*, 1–11. [[CrossRef](#)] [[PubMed](#)]
- 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](#)]
- Guo, Z.H.; Zhao, J.; Zhang, W.Y.; Wang, J.Z. A corrected hybrid approach for wind speed prediction in Hexi Corridor of China. *Energy* **2011**, *36*, 1668–1679. [[CrossRef](#)]
- Pourmousavi Kani, S.A.; Aredhali, M.M. Very short-term wind speed prediction: A new artificial neural network-Markov chain model. *Energy Convers. Manag.* **2011**, *52*, 738–745. [[CrossRef](#)]
- Pourmousavi Kani, S.A.; Riahy, G.D.; Mazhari, D. An innovative hybrid algorithm for very short-term wind speed prediction using linear prediction and markov chain approach. *Int. J. Green Energy* **2011**, *8*, 147–162. [[CrossRef](#)]
- Liu, H.; Chen, C.; Tian, H.Q.; Li, Y.F. A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks. *Renew. Energy* **2012**, *48*, 545–556. [[CrossRef](#)]
- Mohammadi, A.; Dehghani, M.J.; Ghazizadeh, E. Game Theoretic Spectrum Allocation in Femtocell Networks for Smart Electric Distribution Grids. *Energies* **2018**, *11*, 1635. [[CrossRef](#)]
- Kianoosh, G.; Boroojeni, M.; Hadi, A.; Bahrami, S.S.; Iyengar, A.I.; Sarwat, O.K. A novel multi-time-scale modeling for electric power demand forecasting: From short-term to medium-term horizon. *Electr. Power Syst. Res.* **2017**, *142*, 58–73.
- Haque, A.U.; Mandal, P.; Kaye, M.E.; Meng, J.L.; Chang, L.C.; Senjyu, T. A new strategy for predicting short-term wind speed using soft computing models. *Renew. Sustain. Energy Rev.* **2012**, *16*, 4563–4573. [[CrossRef](#)]
- Li, G.; Shi, J. On comparing three artificial neural networks for wind speed forecasting. *Appl. Energy* **2010**, *87*, 2313–2320.
- Ortiz-García, E.G.; Salcedo-Sanz, S.; Perez-Bellido, A.M.; Gascon-Moreno, J.; Portilla-Figuera, J.A.; Prieto, L. Short-term wind speed prediction in wind farms based on banks of support vector machines. *Wind Energy* **2011**, *14*, 193–207.

23. 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](#)]
24. Vapnik, V. *Statistical Learning Theory*, 2nd ed.; Wiley: New York, NY, USA, 1998.
25. Vapnik, V. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1999.
26. Wu, Z.; Huang, N.E. Ensemble empirical mode decomposition: A noise-assisted data analysis method. *Adv. Adapt. Data Anal.* **2009**, *1*, 1–41. [[CrossRef](#)]
27. Yang, X.S.; Deb, S. Engineering optimization by cuckoo search. *Int. J. Math. Model. Numer. Optim.* **2010**, *1*, 330–343.
28. Yang, W.; Wang, J.; Wang, R. Research and application of a novel hybrid model based on data selection and artificial intelligence algorithm for short term load forecasting. *Entropy* **2017**, *19*, 52. [[CrossRef](#)]
29. Yang, X.S.; Deb, S. Cuckoo search via Lévy flights. Presented at 2009 World Congress on Nature & Biologically Inspired Computing (NaBIC), Coimbatore, India, 9–11 December 2009.
30. Pai, P.; Lin, C. A hybrid ARIMA and support vector machines model in stock price forecast. *Omega* **2005**, *33*, 497–505. [[CrossRef](#)]
31. Huang, C.; Davis, L.; Townshend, J. An assessment of support vector machines for land cover classification. *Int. J. Remote Sens.* **2002**, *23*, 725–749. [[CrossRef](#)]
32. Sung, A.H.; Mukkamala, S. Identifying important features for intrusion detection using support vector machines and neural networks. In Proceedings of the 2003 Symposium on Applications and the Internet, Orlando, FL, USA, 27–31 January 2003.
33. Xiao, L.; Shao, W.; Wang, C.; Zhang, K.; Lu, H. Research and application of a hybrid model based on multi-objective optimization for electrical load forecasting. *Appl. Energy* **2016**, *180*, 213–233. [[CrossRef](#)]
34. 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](#)]
35. Chao, R.; Ning, A.; Wang, J.Z.; 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.
36. Wang, J.; Zhu, S.; Zhao, W.; Zhu, W. Optimal parameters estimation and input subset for grey model based on chaotic particle swarm optimization algorithm. *Expert Syst. Appl.* **2011**, *38*, 8151–8158. [[CrossRef](#)]
37. Zhou, J.Z.; Sun, N.; Jia, B.J.; Peng, T. A Novel Decomposition-Optimization Model for Short-Term Wind Speed Forecasting. *Energies* **2018**, *11*, 1752. [[CrossRef](#)]



© 2018 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/>).