

A hybrid system for short-term wind speed forecasting

Qingqing He^a, Jianzhou Wang^{a,*}, Haiyan Lu^b

^a School of Statistics, Dongbei University of Finance and Economics, Dalian, China

^b School of Software, Faculty of Engineering and Information Technology, University of Technology, Sydney, Australia



HIGHLIGHTS

- Develop a novel similarity-based short-term wind speed forecasting system.
- Propose a valid strategy of sample selection.
- Improve the similarity of building sample in the case of continuity.
- Discuss the relationship between the similarity with the modelling requirement.

ARTICLE INFO

Keywords:

Kernel-based fuzzy c-means clustering
Ensemble empirical mode decomposition
Wavelet neural networks
Short-term wind speed forecasting

ABSTRACT

Wind speed forecasting is important for high-efficiency utilization of wind energy. Correspondingly, numerous researchers have always focused on the development of reliable forecasting models of wind speed, which is often noisy, unstable and irregular. Current approaches could adapt to various wind speed data. However, many of these usually ignore the importance of the selection of the modeling sample, which often results in poor forecasting performance. In this study, a hybrid forecasting system is proposed that contains three modules: data preprocessing, data clustering, and forecasting modules. In this system, the decomposing technique is applied to reduce the influence of noise within the raw data series to obtain a more stable sequence that is conducive to extract traits from the original data. To extract the characteristic of similarity within wind speed data, a kernel-based fuzzy c-means clustering algorithm is used in data clustering module. In the forecasting module, a sample with a highly similar fluctuation pattern is selected as training dataset, and which could reduce the training requirement of model to improve the forecasting accuracy. The experimental results indicate that the developed system outperforms the discussed traditional forecasting models with respect to forecasting accuracy.

1. Introduction

With the depletion of conventional resources in recent years and with the lack of enforced uses of emission treatment methodologies, environmental pollution has become a global concern. Therefore, accelerating the exploration and exploitation of alternative renewable energy is a pivotal step in settling the pollution problem [1]. Correspondingly, it has acquired considerable importance for the sake of future sustainable energy development. As one of the most active renewable resources, wind power, owing to its abundance and cleanliness, has attracted widespread attention around the world during the last few decades. Although the widespread use of wind power elicits many benefits, nevertheless, there are also some challenges for wind farms. For example, the development of a scheme for the balanced supply of power based on the demand of electricity is a difficult problem [2]. Furthermore, wind speed plays an important role in wind-

power generation [3]. Therefore, efficient applications of wind energy based on accurate wind speed forecasting is a vital component of the operations of wind farms that concurrently provide the solution to the energy supply and demand balance problem [4]. However, wind speed is noisy, irregular, and easily affected by weather and geographical factors that drive the fluctuations of wind speed and increase the difficulties of precise forecasting [5]. Moreover, the accuracy of wind speed forecasting is not only determined by the forecasting cycle and the wind speed traits of the forecasted geographical locations, but is also dependent on the forecasting methods [4]. Many forecasting methods have been proposed in prior studies that have focused on the improvement of the methodological accuracy based on the different forecasting periods and diverse traits of forecasting sites. In general, the literature shows that the mainstream forecasting methods can be roughly divided into two categories, namely, physical models and statistical models [6].

* Corresponding author at: School of Statistics, Dongbei University of Finance and Economics of China, Dalian 116000, China.
E-mail address: wangjz@dufe.edu.cn (J. Wang).

The physical forecasting models use the detailed physical properties and historical wind speed data to construct forecasting models. This kind of methods assumes that the historical relationship between wind speed data and related physical information will remain valid in the future [7]. They mainly include mesoscale numerical models and computational fluid dynamic methods, and the input variables of these models are commonly physical or meteorology information. As such, they require considerable amounts of observed data at limited simulation scales, and consume excessive computing resources, which are expensive and difficult to obtain. Conversely, these methods are more suitable for long-term forecasting rather than short-term forecasting [8].

Statistical models have been extensively employed in recent studies for forecasting short-term wind speed. Methods, such as conventional statistical models, and statistical learning methods, have been extensively used in developing wind speed forecasts. Conventional statistical forecasting models, which only use prior wind speed data series, build statistical models based on existing statistical equations [9]. Statistical learning models usually build a high dimensional non-linear function to fit the wind speed, and the factors they influenced by minimizing the training error [10].

The most commonly used conventional statistical models include the autoregressive (AR), autoregressive moving average (ARMA), and the autoregressive integrated moving average (ARIMA) models. AR is a special form of ARMA, and the generalization of the ARMA model is considered to be the ARIMA for the sake of straightforward implementation of ARMA. Schlink and Tezlaff applied the AR for wind speed forecasting tasks at an airport [11,12], and the results showed that the width of intervals produced by AR were narrower than the intervals generated by the persistence model [12]. Lydia et al. forecasted wind speed based on linear and nonlinear ARMA models with and without external variables [13]. In [14], the fractional-ARIMA model was used to forecast wind speed for upcoming two-day periods with a higher accuracy compared to the persistence model. However, the limitations of these time series approaches concluded poor extrapolation effects and narrowed the forecasting scale.

With the rapid development of computer technology, the ability to perform complex calculation has significantly improved. A rapid increase of statistical learning models has been documented during the last few years, and these have led to the development of mature theoretical systems. The well-known ANN is one of extensively used statistical learning methods for wind speed forecasting, which can produce nonlinear maps and perform parallel processing. These approaches mainly include back propagation (BP) [15], radial basis function (RBF) [16], Elman neural network (ENN) [17] and wavelet neural network (WNN) [18], and others. They commonly consist of an input layer, one or more hidden layers, and an output layer. Each layer has some artificial neurons which are connected with the neurons of the previous layer with a connection function. The results show that the different structures of the network lead to different wind speed forecasting performances [19,20]. Thus, the instability and the dependence on data thus become an obvious defect for these models. Statistical machine learning methods emphasize the ability of generalization, the importance of regression perception machines, and the design the learning algorithms based on the generalization indices. They commonly contain support vector machines (SVM) [21], least-squares support vector machines (LSSVM) [22], Gaussian processes [23], and others. These simulation methods can be easily understood and combined with other approaches. In addition, the machine learning approaches elicit a more accurate performance when compared to ANNs, but they usually suffer the perplexity of over-fitting. The fuzzy logic is a research area based on the principles of approximate reasoning and computational intelligence [24]. Using fuzzy relations, the appropriate index is obtained for the assessment of system reliability [25]. The author in Ref. [26] used an adaptive Nero-fuzzy inference system (ANFIS) which was implemented as a hybrid fuzzy logic system with ANN. This system was also

employed for Sugeno fuzzy modeling, and had two inputs with four fuzzy rules [27]. Nevertheless, this type of method has a relatively low accuracy and lacks systematization.

However, owing to the inner irregularity, instability and noise of the raw wind speed series, the individual models listed above generally depend on a large amount of historical data to build models that would elicit precise forecasting results. As such, this large data volume could lead to a complex calculation dilemma of the model that causes poor forecasting performance. Conversely, they are often limited in achieving a better fitness to the original data by resorting to tedious and time-consuming trials. As a result, it is very difficult to fit a model to them based on the single use of the conventional physical or statistical methods, in the effort to forecast short-term wind speed from these noisy and irregular datasets.

Therefore, in this study, a powerful hybrid system is developed that comprises three modules and primarily adopts a sample with a highly similar fluctuation pattern to establish forecasting model. The ensemble empirical mode decomposition (EEMD) technique is used in data preprocessing module, and a kernel-based fuzzy c-means clustering (KFCM) algorithm is employed as a data clustering module. Selecting of a sample with distinctive traits of similarity to establish wavelet neural network (WNN) model to conduct the final forecasts defines the forecasting module. Specifically, EEMD technology is efficient to analyze nonlinear and non-stationary data. Furthermore, Shouxiang Wang et al. [28] showed that the hybrid method EEMD-GA-BP is more accurate than the traditional GA-BP model in wind speed forecasting. The KFCM algorithm, which is an extension of a conventional fuzzy c-means clustering method, mainly alters perfectly the input data into a higher dimensional feature space [29], and thus overcomes the drawback of imprecise clustering caused by lower dimensions.

Generally, the innovations of this study can be summarized as follows:

- (1) This study proposes a hybrid system that successfully takes advantages of data preprocessing technique and data clustering algorithm to improve the forecasting ability of WNN model on short-term wind speed. The novelty of the EEMD mainly relates to the inference of the decomposition of the wind speed series, and is followed by the removal of high frequency signals to obtain a smoother series. As such, it can be conducive to extract data features, thereby allowing the improvement of the forecasting accuracy of the models.
- (2) The common approach adopted is the direct use of the denoising results to train the forecasting models. However, there is an innovation in this proposed methodology that employs the KFCM algorithm to extract data characteristics with similarities before the training.
- (3) The data clustering module obtains samples with highly similar fluctuation patterns, which could reduce the training requirements for the forecasting model to improve the accuracy of short-term wind speed forecasting. In addition, this feature has been discussed in Section 5.2.
- (4) It is the first time where the relationship between the degree of similarity within the training sample and the forecasting performance of the neural network model is explored, as stated in Section 5.3.
- (5) Selecting data points with similar frequencies of fluctuation to produce the inputs of the model discontinuously makes the forecasting results unreasonable. This is caused by the intermittent within input vector. In this case, the settlement in this proposed hybrid system is the choice of the constructed input vectors as the inputs of the model. As a result, the behavior of the clustering approach considers both the similarity within the sample and its inner continuity.

This study is organized as follows. Section 2 outlines the principles

of methods related to this proposed hybrid system. The developed approach EKW is shown in Section 3. In addition, experiments and evaluations are presented in Section 4. Furthermore, in Section 5, the factors related to model forecasts are discussed, in addition to the influences of similarities on the training requirement of the WNN model, the relationship between the degree of similarity and the accuracy of model forecasting, and the practical significance of the proposed forecasting system. Finally, the study's conclusions and future work are presented in Section 6.

2. Methods

In this section, the related methods and theories i.e., EEMD, KFCM, and WNN model, utilized in this article are described in detail as follows.

2.1. Principle of ensemble empirical mode decomposition denoising technique

The empirical mode decomposition (EMD) was formulated based on the Hilbert–Huang transform, which was developed by Huang et al. in 1998 [30]. It incorporates a strong adaptive ability to analyze the nonlinear and nonstationary signal. Additionally, the primary innovation embodied in the EMD is the introduction of the intrinsic mode functions (IMFs), which are relatively stationary sub-sequences, and which can be easily modeled. However, owing to the primary drawback of mode mixing of EMD, a new noise-assisted signal analysis technique was proposed, known as EEMD. In general, EEMD benefits from white noise, and constitutes an effective self-adaptive dynamic filter bank [31,32]. The IMF components are defined as the mean of an ensemble of trails, and consist of white noise of finite amplitude plus the decomposition results of the signal [33].

Each IMF can be defined as a hidden oscillation mode that reflects the dynamic characteristics of the original signal. According to Ref. [30], the IMF is constraint by the following two conditions:

- (1) In an entire dataset, the number of extrema (the number of maxima and the number of minima), and the number of zero-crossings must be equal, or differ by one at the most.
- (2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero.

The IMF could be obtained via the following steps:

Step 1: For wind speed signal $x(t)$, identify all local extrema of $x(t)$.

Step 2: Connect all maxima using a cubic spline to produce the upper envelop $\{x_1(t)\}$. In the same way, produce the lower envelope $\{x_2(t)\}$ through the connection of all minima.

Step 3: Calculate the mean value of the upper and the lower envelopes $\{m\}$, and define the difference between $x(t)$ and m as h .

$$h = x(t) - m \quad (1)$$

$$m = \frac{x_1(t) + x_2(t)}{2} \quad (2)$$

Step 4: Repeat Step 3 i times until h is an IMF. The criterion of IMF is expressed by the following equation:

$$\sum_{i=1}^N \frac{[m_{i-1} - m_i]^2}{[m_{i-1}]^2} \leq \delta \quad (i = 1, 2, \dots; t = 1, 2, \dots, N) \quad (3)$$

where N is the length of the signal, i denotes the number of iterative calculations, and δ is a predetermined value.

Ref. [30] suggested that the top several IMF components usually mainly consist of noise, and can be filtered directly to achieve denoising. Therefore, in this study, the behavior of denoising is to remove

the top two IMF components, and then calculate the sum of the rest IMFs and residual to form a new series which is more stable than the original wind speed data series. In addition, according to [34], the standard deviation of added noise series equals 0.2, while the ensemble number is set to $M = 100$.

2.2. Principle of kernel-based fuzzy c-means clustering algorithm

Clustering techniques have been extensively used, among which the fuzzy clustering method has been gaining increasing attention. It is a multivariate statistical method which applies the fuzzy logic relationship. Moreover, fuzzy sets are viewed as powerful tools to handle uncertainty, and have already been applied to many fields, including wind speed forecasting [35]. Both fuzzy c-means and k-means clustering are applied to analyze data and treats observations of the data as objects based on the locations and distance between various input data points. The primary difference between these two clustering approaches is that the latter distance between the clustered object and the cluster center is not a spatial distance, while the former is [36]. However, owing to the drawback of the k-means, which depends on the initial clustering center, it lacks stability [37]. Additionally, the fuzzy c-means algorithm is more suitable for settling the issue of clustering, as shown in this study. Conversely, kernel-based fuzzy c-means clustering has emerged as an interesting visible alternative in fuzzy clustering, and constitutes an important tool for discovering the structure of the data [38]. This method maps the clustering problem into a higher dimension, and offers a clearer perspective for the difference among objects.

The kernel-based fuzzy c-means clustering method first uses a nonlinear function to map the data set into a potentially much higher dimensional feature space, and clustering is then performed. Herein, Gaussian function is chosen as the dimension conversion function. The use of Gaussian function in KFCM adopts a new kernel-induced distance in the data space to replace the original Euclidean distance [39]. As a result, the cluster prototypes still lie in the data space so that the clustering results can be reformulated and interpreted in the original space.

Given a dataset, $X = \{x_1, x_2, \dots, x_n\} \subset R^p$, the original FCM algorithm partitions X into c subsets by minimizing the following objective function.

$$\min J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|x_j - v_i\|^2 \quad (4)$$

where c is the number of clusters and given a specified value before processing, n represents the number of data points, u_{ij} is the membership of x_j in category i , satisfying $\sum_{i=1}^c u_{ij} = 1$, m is the quantity controlling clustering fuzziness, and V is the set of cluster centers or prototype ($v_i \in R^p$).

Now by considering the KFCM algorithm, define a nonlinear map as $\theta: x \rightarrow \theta(x) \in E$, where $x \in X$ denotes the data space, and E indicates the transformed feature space with a higher dimension. KFCM minimizes the following objective function:

$$\min J = \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m \|\theta(x_j) - \theta(v_i)\|^2 \quad (5)$$

$$\|\theta(x_j) - \theta(v_i)\|^2 = K(x_j, x_j) + K(v_i, v_i) - 2K(x_j, v_i) \quad (6)$$

where $K(x, y) = \theta(x)^T \theta(y)$ is an inner product kernel function, it does not have to be explicitly computed, and is specifically referred to as a kernel trick [40]. In this study, we adopt the Gaussian kernel function that is expressed by the following equation: $K(x, y) = \exp(-\|x-y\|^2/\sigma^2)$, then $K(x, y) = 1$. According to Eqs. (5) and (6), Eq. (5) can be written as:

$$\min J = 2 \sum_{i=1}^c \sum_{j=1}^n u_{ij}^m (1 - K(x_j, v_i)) \quad (7)$$

Minimizing Eq. (7) under the constraint of u_{ij} , and u_{ij} can be calculated as follow:

$$u_{ij} = \frac{(1/(1-K(x_j, v_i)))^{1/(m-1)}}{\sum_{k=1}^c (1/(1-K(x_j, v_k)))^{1/(m-1)}} \quad (8)$$

$$v_i = \frac{\sum_{j=1}^n u_{ij}^m K(x_j, v_i) x_j}{\sum_{j=1}^n u_{ij}^m K(x_j, v_i)} \quad (9)$$

The full description of KFCM algorithm is as follows:

- Step 1: Given c , $iter_{max}$, $m > 1$, and $\varepsilon > 0$ for some positive constant;
- Step 2: Initialize the memberships u_{ij}^0 ;
- Step 3: Update all prototypes v_i^{iter} with Eq. (9);
- Step 4: Update all memberships u_{ij}^{iter} with Eq. (8);
- Step 5: Compute $d^{iter} = \max_{ij} |u_{ij}^{iter} - u_{ij}^{iter-1}|$, if $d^{iter} \leq \varepsilon$.

2.3. Principle of wavelet neural network model

Wavelet theory is a mathematical theory that was developed by Grossman and Morlet in the 1980s. In terms of signal processing, the wavelet theory appears to be more effective than the Fourier transform that does not possess the ability to distinguish features in the time domain. Moreover, the continuous wavelet transforms decompose the non-stationary signals into a certain number of linear combinations of wavelets.

The continuous wavelet transform is defined as follows:

$$W_f(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} f(t) \psi\left(\frac{t-b}{a}\right) dt \quad (10)$$

where a is the scaling function used to stretch or compress mother wavelet $\psi(t)$ that is related to the frequency of the signal; b is the translation function used to shift mother wavelet $\psi(t)$ to a time domain of the signal; $f(t)$ is the input signal, the mother wavelet $\psi(t)$ is defined as follows.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (11)$$

In this study, the Morlet function is taken as the mother wavelet in the analysis process, which is shown as follows.

$$\psi(t) = Ce^{-t^2/2} \cos(1.75t) \quad (12)$$

To take advantages of the wavelet basis functions and neural network simultaneously, the wavelet neural network was proposed by Zhang in 1992 [41] and has been used with great success in a broad range of application ever since. WNN uses the wavelet function as the activation function instead of the sigmoid activation function. Furthermore, its topology is based on BP network and differs from the common network that the errors propagate backward, while the signals propagate forward. The simple architecture of WNN model is shown in Fig. 1 (e), x_1, x_2, \dots, x_k is the input vector, y_1, y_2, \dots, y_m is the relative output vector.

The wind speed data is a sequence of numeric observations from a particularly measured variable [42]. To represent the forecasting process, we consider herein a time series comprising N observations $X_N = (x_1, x_2, \dots, x_N)$. The r -step ahead forecast consists of forecasting $x_{k+1}, x_{k+2}, \dots, x_{k+r}$, where $k = 1, 2, \dots, N$, k and r are the indices corresponding to the number of points between the present and the future, usually referred to the prediction horizon [43]. The main objectives of a time series analysis are the construction of a model, and then use the model to conduct forecasts, whereby the future value is predicted only from the past. The data are described by a possible linear or nonlinear autoregressive process in the form as follow.

$$x_{k+r} = f(x_k, x_{k-1}, x_{k-2}, \dots, x_{k-n-1}) \quad (13)$$

where f is a function that represents the relationship between the past values of x and the present. In this study, $r = 1$ indicates that there is a one-step ahead prediction.

3. Proposed hybrid system EKW

Most previous forecasting studies mainly attempt to ensure that the fitting values were as close as possible to the observed values by using the experimental method, even though the experimental approach was time-consuming and repetitive. For example, owing to insurmountable disadvantage of each method, a combined wind-speed forecasting method was proposed that aimed to capitalize on the advantages and avoid the disadvantages of each forecasting model [44]. Even so, it is difficult to make up for the shortcomings associated with the training sample. Moreover, there is a major training requirement when the original wind speed data are used to build neural network models. As a result, it is very difficult to obtain the optimal performance of the model by employing the raw series and single model directly. This is mainly because of the chaotic nature and intrinsic complexity of the original wind speed data series. Obviously, the sample used for modeling would definitely influence the accuracy of the forecasting model. In this case, selecting appropriate samples or generating new sequences from the original data prior to the model becomes an effective method for improving the forecasting accuracy.

Therefore, in this study, a hybrid system is proposed for short-term wind speed forecasting, which commits to the selecting of an appropriate training sample to improve the precise of the prediction. In this system, a data preprocessing module primarily uses a signal decomposition technique to obtain a more stable sequence which contributes to the extraction of the real data traits from the original wind speed series. Furthermore, the data clustering module extracts data traits to enhance the inner similarity within samples using a clustering approach. Thus, modeling using these samples could reduce the training requirements of the models. Finally, the forecasting module mainly completes the selection of the training sample and the final forecasts.

3.1. Data preprocessing module

This module mainly uses a decomposition technique to eliminate noise from the raw wind speed data series. The common decomposition techniques mainly include the wavelet decomposition and EEMD. The wavelet decomposition is a traditional method and is able to obtain a relatively stable series. However, this method largely depends on the selection of wavelet basis function and the number of decomposition layers [45]. The EEMD technique was developed in recent years, and is more adaptive in analyzing nonlinear and non-stationary signals compared to the wavelet decomposition. As a result, EEMD is adopted in data preprocessing module to obtain a more stable and an easier modeling series, which is conducive in extracting data traits.

3.2. Data clustering module

The investigation of feature selection as applied in wind speed forecasting has received little attention [5]. Feature selection, which can remove certain irrelevant features and enhance the capability of the forecasting model in learning the nonlinear relationship of the time series data, is an effective technique for the selection of model inputs when performing forecasts [46]. Hence, according to the results of the data preprocessing module, this module employs the KFCM algorithm to cluster the constructed input vectors into a given number of categories. In terms of the behavior of the clustering algorithm in wind speed forecasting models, there is an extreme difference between the clustering of data points and datasets. Furthermore, data clustering was used to distinguish the fluctuation frequency of wind speed signal

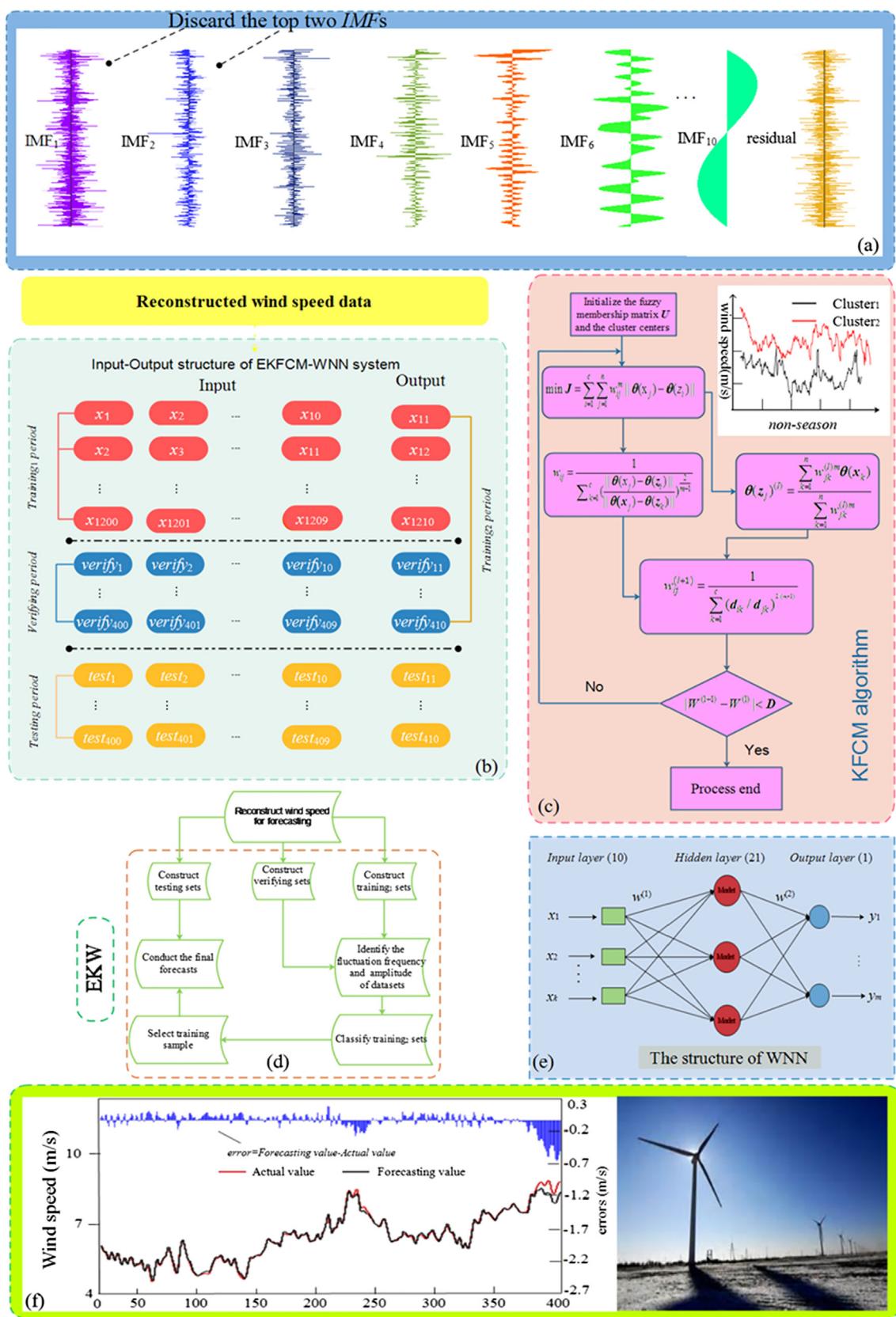


Fig. 1. The flowchart of EKW system across non-season datasets.

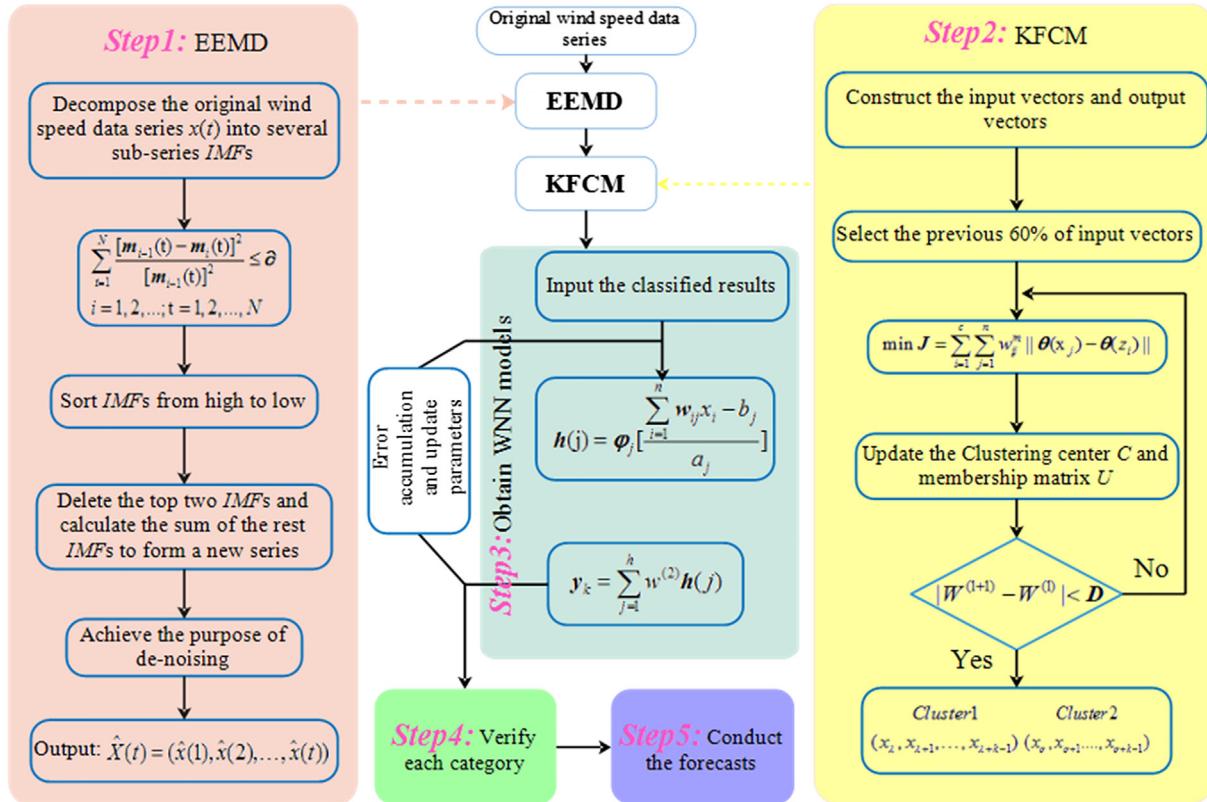


Fig. 2. The flowchart of EKW system.

points to form input samples for the prediction models. However, the main drawback of this practice is that the intermittent use of time points in input vectors, which contain multiple elements, will make the forecasting results unreasonable. To alleviate this dilemma of data discontinuity, we introduce an improvement that changes the objects of the clustering method from the data points to the constructed data vectors, whereby similar characteristics of fluctuation are considered simultaneously among the input vectors, and for the similarities among the input data points. As a result, this behavior does not maintain the continuity within each input vector, but also considers the similarity among training sets.

3.3. Forecasting module

Wenyu Zhang et al. [47] held that the input data selection has a great impact on the wind speed forecasting accuracy of models. In this case, in the forecasting module, a verifying dataset is used to distinguish which type of fluctuation patterns is suitable for modeling. Additionally, the test dataset is employed to conduct the final forecasts. Figs. 1 and 2 present the detailed flowchart of the proposed system.

This hybrid system conducts short-term wind speed forecasts as follows:

Step 1: Data segmentation

To evaluate the applicability, superiority and generality of the proposed novel hybrid model, wind speed data from Penglai, Shandong peninsula of China with a 10-min period are randomly assembled for 1-step ahead forecasting. 2000 data points are selected from each dataset and these observations are split into two subsets, which are the training set and the testing set. There are no explicit theories about how to select the number of training sample and testing sample currently. Too small sample will not yield well trained neural network, while too much sample will appear over-fitting. In practical terms, it is common to use

the remaining two-thirds for training and hold out one-third of the data for testing [48]. In this paper, the experience proportion between the training and testing sets is 4:1. That is, the initial 1600 data points act as the training sample and the remaining 400 data points act as the testing sample

Step 2: Data pre-processing

Employ the EEMD technology to remove the noise from the actual wind speed time series.

Step 3: Construct input-output structure

For WNN, the dimension of the input layer is 10, the dimension of the hidden layer is 21, and the dimension of the output layer is 1. The number of iterations is 100, the learning rate is 0.01, and the training precision is 0.00004. In addition, the input data is the denoising wind speed series, and the output data is also denoising wind speed time series.

Step 4: Feature extraction

In order to ensure that the model can maximally capture the fluctuations feature of wind speed data. KFCM is used to extract the features of the input data, and then adopted the extracted data as the training input of the forecasting model.

Step 5: Selection the optimal input set

Put the input data which processed by Step 4 into WNN model, and then compare the experimental results with the actual wind speed time series. Thus, the optimal training input set of WNN model is found by comparing MAPEs. At last, adopt the optimal input as the final training input to establish WNN model, and use the input set of Testing to

conduct the final forecasts.

Fig. 1(a) shows the decomposition results of the *non-season* short-term wind speed series. Correspondingly, **Fig. 1(b)** presents the input-output structure of the proposed hybrid system which contains four periods, namely, the training₁, verifying, testing, and training₂ periods. The detailed flowchart of the clustering approach of KFCM is shown in **Fig. 1(c)**, whereby the upper right picture shows the results of the corresponding output of the clustered input vectors across *non-season* featured datasets. **Fig. 1(d)** and (e) show the flowchart of this hybrid system of EKW and the WNN model, respectively. Finally, **Fig. 1(f)** presents the fitting results of the *non-season* featured wind speed with the EKW system.

4. Experiments and evaluations

The accurate prediction of wind speed is fairly important for the management of wind farms. Thus, in this study, we propose a hybrid system to improve the accuracy of the short-term wind speed forecasting, and the performance of the proposed hybrid system is evaluated in this section by applying the nonlinear dynamic system for forecasting short-term wind speed.

4.1. Datasets

The province of Shandong in China has abundant wind resources owing to its geographical characteristics. Because there is a tremendous wind power potential, it is very important to investigate its actual wind power based on the precise forecasting of wind speed. To evaluate the ability of this proposed hybrid system to handle short-term wind speed prediction, we collected two featured *non-season* and *season* wind speed datasets from Shandong province. The *season* datasets mainly contain four different features: *spring*, *summer*, *autumn*, and *winter*, and the principal description of these datasets are shown in **Fig. 3(c)**.

In effect, ten-minute wind speed data were collected from a wind farm in the province of Shandong, and we capture two types of featured datasets from the available data. The *non-season* data were collected as the mean values of the ten wind turbine generators (WTGs) from the early morning of 1th January to the early morning of 15th January. The *season* data were from four WTGs, and each entire year period was divided into four different zones before *season* wind speed data were

Table 1
Details of experimental data.

Features			
<i>Non-season</i>	Data Specific data	2/15–5/15 1/1(00:00)-1/15(00:00)	
<i>Season</i>	<i>Spring</i>	Data Specific data	1/1–3/31 1/1(00:00)-1/15(00:00)
	<i>Summer</i>	Data Specific data	4/1–6/30 4/1(00:00)-4/15(00:00)
	<i>Autumn</i>	Data Specific data	7/1–9/30 7/1(00:00)-7/15(00:00)
	<i>Winter</i>	Data Specific data	10/1–12/31 10/1(00:00)-10/15(00:00)

extracted. Specifically, *zone1*: from 1th January to 31th March; *zone2*: from 1th April to 30th June; *zone3*: from 1th July to 30th September and *zone4*: from 1th October to 31th December. The *spring* data are partly from *zone1*, *summer* data is derived from *zone2*, and the featured *autumn* and *winter* data are from *zone3* and *zone4*, respectively. In addition, the detailed extraction times of each featured data from every zone are shown in **Table 1**.

To further realize the inner characteristics within each featured wind speed dataset, we show the mean, standard deviation, maximum and the minimum velocities of the *season* data in WTG06 and *non-season* data in **Fig. 3(c)**. It is easy to observe that the fluctuation frequency of *summer* is the highest, while the lowest is the *autumn*, and that the *non-season* data exhibits a relatively smooth fluctuation frequency trend. Additionally, **Fig. 3(b)** shows the original *non-season*, *spring*, *summer*, *autumn* and *winter* data series, and reveals the fluctuation of the *non-season* wind speed data series which does not possess the characteristics of cycle fluctuations, like the *season* data.

4.2. Evaluation metrics

Prediction accuracy is an important criterion used to evaluate the ability of model forecasting. In order to identify the best methods, we choose three evaluation metrics to compare the proposed system to other discussed models in this study. These criteria mainly infer to the mean absolute percentage error (MAPE), mean square error (MSE), and mean absolute deviation (MAD). MAPE is a generally accepted metric

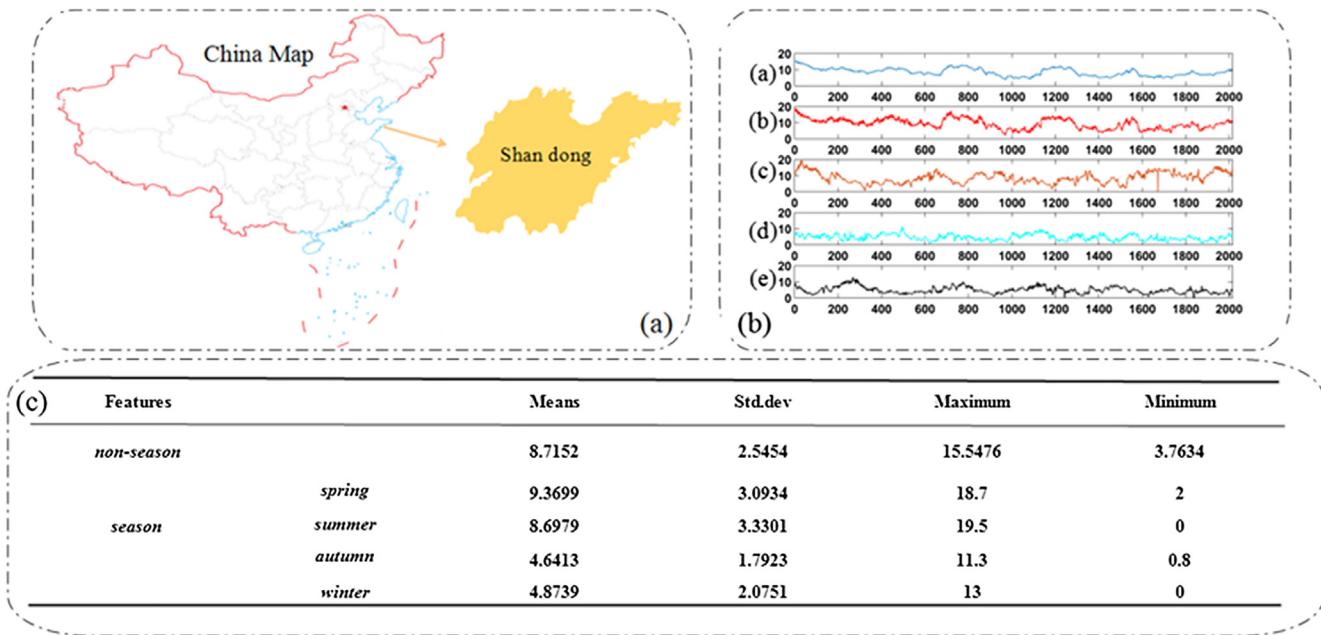


Fig. 3. Original data.

Table 2
Forecasting performance indices and their definition.

Metrics	Definition	Calculation
MAPE	Mean absolute percentage error	$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{ e_i }{y_i} \times 100\%$
MSE	Mean square error	$MSE = \frac{1}{n} \sum_{i=1}^n e_i^2$
MAD	Mean absolute deviation	$MAD = \frac{1}{n} \sum_{i=1}^n e_i $

Table 3
MAPE criteria.

MAPE (%)	Forecasting power
< 10	Excellent
10–20	Good
20–50	Reasonable
> 50	Incorrect

for prediction accuracy, while MAD and MSE both measure the average magnitude of the forecasting errors. But the latter imposes a greater penalty on a large error than several small errors. The smaller the values of these criteria are, the closer the predicted values to the actual values are. These three criteria are defined in the following table as follows (see Table 2).

Where $e_i = \hat{y}_i - y_i$, and n corresponds to the size of tested sample. In addition, \hat{y}_i and y_i represent the expected and predicted values at time i respectively, and the criterion of MAPE correspond to the forecasting power, as shown in Table 3 [43].

4.3. Experimental setup

In order to evaluate the ability of the EKW hybrid system on short-term wind speed forecasting, and explore the effect of similarity on short-term wind speed forecasts, we divided the experiments into three parts, namely, *Experiment I*, *Experiment II*, and *Experiment III*. *Experiment I* compares the performance of EKW system with that of WNN and the EEMD-WNN models based on the data in WTG06. *Experiment II* compares the performance of the EKW system with those of well-known wind speed forecasting methods across different featured data. *Experiment III* compares the performance of EKW systems trained by the two categories with different characteristic of fluctuation based on the data in WTG06.

The *non-season* and *season* featured datasets, which have been described in detail in Section 4.1, were both used in our experiments. The purpose of *Experiment I* was to compare the EKW system with the other two basic WNN models. In this proposed system, we subdivided the available sample into four parts before establishing the wavelet neural network model. *Part I* included the first 60% of the data of the available sample, and *part II* contained the succedent 20% data. *Part III* was formed by combining *parts I* and *II*, and contained the first 80% data of the available data. The remaining 20% of the data of the available sample constituted *part IV*. *Parts I, II, III, and IV*, are known as the

training₁, *verifying*, *training₂*, and *testing* periods, respectively in this study. *Parts I* and *II* were used to identify the type of category that can be used to train the wavelet neural network to obtain accurate prediction results. *Part III*, namely *training₂* period, was used to select the training sample for this proposed system. Finally, *part IV* was adopted to conduct the final forecasts of the short-term wind speed.

Experiment II was designed to compare the EKW with other five common wind speed forecasting methods. For full consideration of the comparison of model prediction ability, we chose commonly used models from time series models, neural network models, meta-heuristically optimized neural networks, and machine learning models. Respectively, ARIMA, Elman, Elman optimized based on particle swarm optimization algorithm (PSO-Elman), RBF, BP, and SVM models were used. The orders of p and q in the ARIMA model were determined based on AIC comparisons. The lowest AIC values corresponding to p and q were set as the parameters of ARIMA model.

Experiment III aimed to compare the performances of wavelet neural network models obtained based on the use of various categories with different fluctuation characteristics, and thus explored the role of their similarities on enhancing the ability of WNN on short-term wind-speed forecasting. Shown on the top right corner of Fig. 1(c), is the clustered result of *non-season* datasets. In other words, the purpose of *experiment III* was to explore the validity differences of these types of fluctuation frequencies on training. The number of clusters in this study was assigned to two, and the EKW system conducted the final forecasts by choosing a constructed model which was trained by the selected data from the *training₂* period. In Section 5.3, we will discuss the relationship between the forecasting accuracy and the degree of similarity within a training sample.

4.4. Experiment I

Table 4 and Fig. 4 show the comparison results of the evaluated metrics on the WNN, EEMD-WNN, and the EKW hybrid system based on the *season* data in WTG06 and *non-season* data. It is obvious that the proposed EKW hybrid system obtains the best overall prediction accuracy across most of the featured datasets. Regarding the comparison between the original WNN and the EEMD-WNN model, the performance of the EEMD-WNN model is better than that of the WNN model in terms of MAPE in all cases. This means that data preprocessing is necessary for short-term wind speed forecasting. In this case, we will discuss the result of the comparison of the EEMD-WNN model and the proposed hybrid system EKW. For the *non-season* datasets, the EKW system yielded the best forecasting accuracy with MAPE at 1.24%. Similarly, the MSE and the MAD were also the lowest with the respective values of 0.022 and 0.09.

For the *season* datasets, the forecasting performance of EKW system was the best, and three of the four zones elicited the best performance in the comparison of the evaluation metrics. For *summer* dataset, the EEMD-WNN model yielded the best MAPE with 2.12%, while the performances on the other two metrics were also greater than the EKW system. However, in consideration of the *spring*, *autumn* and *winter* featured datasets, the proposed hybrid system obtained the best

Table 4
Comparison of the evaluation metrics of the WNN, EEMD-WNN, and EKW model based on the data in WTG06.

Data	WNN				EEMD-WNN				EKW			
	MAPE (%)	MSE	MAD	Time (s)	MAPE (%)	MSE	MAD	Time (s)	MAPE (%)	MSE	MAD	Time (s)
<i>Non-season</i>	3.96	0.099	0.022	58.7	2.46	0.04	0.157	58.1	1.24	0.022	0.09	28.1
<i>Spring</i>	9.69	0.773	0.73	58.3	3.72	0.086	0.254	57.4	2.98	0.174	0.234	27.2
<i>Summer</i>	> 50	1.613	0.782	58.2	2.12	0.123	0.237	58.1	3.36	0.357	0.362	26.8
<i>Autumn</i>	14.62	0.4	0.488	59	7.5	0.087	0.259	58.4	4.07	0.024	0.131	29.1
<i>Winter</i>	> 50	0.476	0.517	58.4	14.83	0.285	0.473	59	5.44	0.33	0.134	32.8

Note: The bold in the table refer to the best forecasting performance.

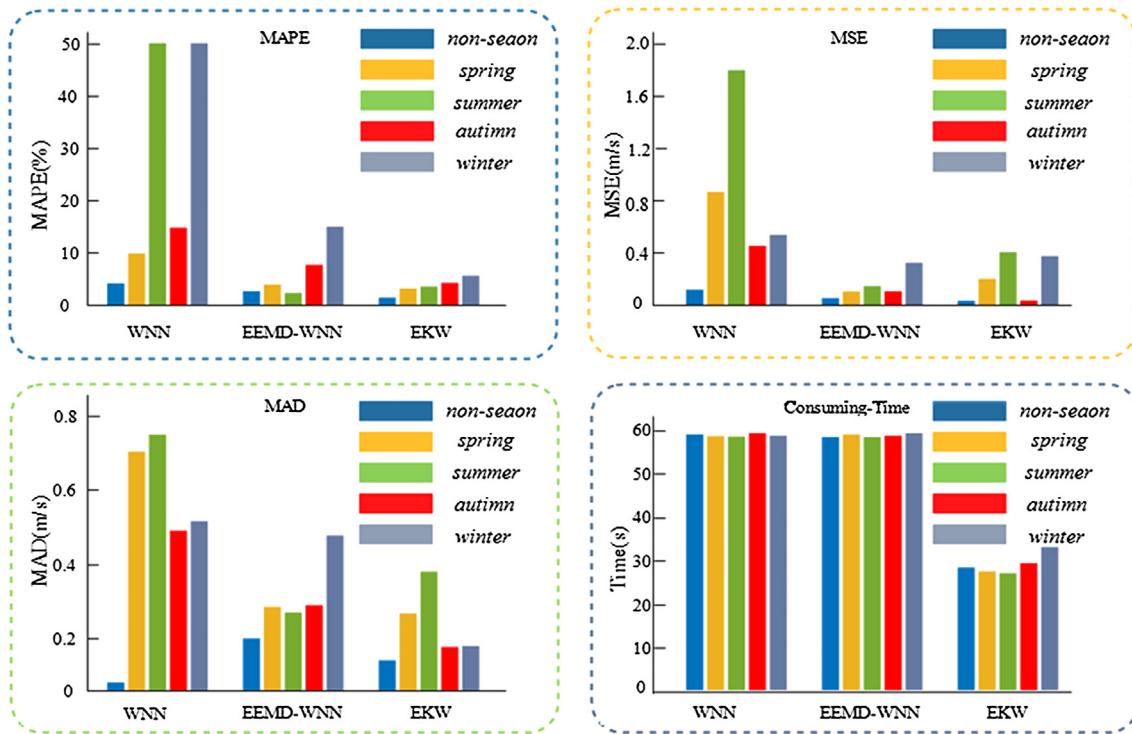


Fig. 4. Forecasting results of the WNN, EEMD-WNN models, and the EKW system.

forecasting performance with MAPEs 2.98%, 4.07%, and 5.44%, respectively. Additionally, the values of MSE and MAD were lower than the other two models. In terms of the consumption time, the hybrid system proposed in this study were executed in almost half the time compared to the other two models in the case of the featured datasets. This advantage was mainly attributed to the behavior of sample selections before the modeling was conducted.

4.5. Experiment II

In this part, the *season* data was collected from four WTGs, WTG06, WTG03, WTG08, and WTG09, respectively. Table 5 shows the comparison of the results of the different evaluation metrics in accordance to commonly used approaches based on the *season* in WTG06 and *non-season* data. Additionally, the comparison results of prediction performance between the common methods and basic WNN models with the proposed hybrid system based on the data in WTG03, WTG08, and WTG09, are shown in Table 6, Table 7, and Table 8, respectively. Moreover, Fig. 5 presents the forecasting performance of five compared methods with the proposed hybrid system using *non-season* featured datasets. In general, the proposed EKW system in this study outperformed the compared forecasting approaches of short-term wind

speed. In addition, the forecasting performances of the EKW system proposed in this study for *non-season* and *season* datasets in WTG06 are shown in Fig. 6.

For the *non-season* dataset, the PSO-Elman, BP, ARIMA, and the EKW system, all elicited excellent precision results on all evaluation metrics. Additionally, the RBF neural network model also achieved good MAPE outcomes that were less than 20%, but the other two models performed poorly. The ARIMA model is always viewed as an effective model for forecasting nonlinear time series, and it is logical that its MAPE is lower than those of other commonly used models discussed herein. However, its performance was worse than our proposed system with a MAPE value of 1.24%.

For the *season* data, in general, the proposed system elicited the best performance. In regards of *spring* feature data, EKW system obtained the lowest value of MAPE in three WTGs, respectively, 2.98%, 2.48%, and 4.89%. As for compared methods, ARIMA, BP, WNN, and EEMD-WNN models all performed well relatively, among which ARIMA still did not perform as good as the EKW model in dealing with *spring* feature data while it is a well-known wind speed forecasting model. Because of the amount of training sample, the EKW yielded little higher value of MAPE in WTG09 than the EEMD-WNN model.

Regarding of *summer* featured data, the EKW obtained the best

Table 5

Comparison of the evaluation metrics of the EKW model with the other common wind speed forecasting methods based on the data in WTG06.

Model	Non-season			Spring			Summer			Autumn			Winter		
	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD
Elman	22.24	0.001	0.028	39.8	4.036	1.628	> 50	83.608	8.794	> 50	6.705	2.302	> 50	6.492	2.077
PSO-Elman	6.07	0.286	0.408	21.27	2.492	2.984	29.65	16.662	3.577	> 50	2.711	1.399	> 50	3.777	1.644
ARIMA	2.23	0.037	0.146	7.66	0.519	0.526	2.37	0.165	0.243	4.56	0.036	0.144	5.63	0.032	0.129
RBF	12.42	2.549	0.828	> 50	32.539	4.374	> 50	75.891	7.702	28.04	1.763	0.985	> 50	452.075	13.85
SVM	33.33	5.942	2.088	37.06	7.807	2.311	22.79	10.332	2.743	> 50	4.658	1.705	> 50	6.405	2.037
EKW	1.24	0.022	0.09	2.98	0.174	0.234	3.36	0.357	0.362	4.07	0.024	0.131	5.44	0.033	0.134
BP	2.05	0.033	0.137	5.79	0.296	0.404	> 50	1.608	0.694	13.98	0.348	0.455	> 50	0.313	0.392

Note: The bold in the table refer to the best forecasting performance.

Table 6

Comparison of the evaluation metrics of the EKW model with the other common wind speed forecasting methods based on the season data of WTG03.

Model	Spring			Summer			Autumn			Winter		
	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD
Elman	57.11	17.552	3.406	> 50	21.116	4.011	> 50	3.805	1.53	44.59	4.57	1.457
PSO-Elman	15.14	1.327	0.915	21.59	4.014	1.451	35.5	1.81	1.011	26.73	1.346	0.931
ARIMA	8.56	0.376	0.498	9.27	0.704	0.633	14.59	0.285	0.423	13.27	0.283	0.432
RBF	33.35	13.737	2.013	> 50	71.34	2.567	—	—	—	—	—	—
SVM	> 50	20.035	4.243	> 50	68.614	7.714	> 50	10.45	2.945	> 50	9.766	2.799
EKW	2.48	0.038	0.153	5.62	0.259	0.366	5.79	0.088	0.178	5.43	0.082	0.206
BP	6.58	0.282	0.407	8.34	0.67	0.597	13.4	0.343	0.447	—	—	—
WNN	7.27	0.381	0.468	10.55	1.455	0.882	15.81	0.399	0.498	> 50	0.684	0.626
EEMD-WNN	2.85	0.045	0.177	5.83	0.629	0.538	7.22	0.066	0.215	8.47	0.136	0.321

Note: The bold in the table refer to the best forecasting performance.

forecasting performance in three WTGs, respectively WTG03, WTG08, and WTG09, when compared with other compared methods. The corresponding MAPEs are 5.62%, 3.63%, and 5.49%, respectively. As for *autumn* featured data, WTG06, WTG03, and WTG08 yielded the best accuracy with the MAPEs value of 4.07%, 5.79%, and 4.04%, respectively. However, with respect to the *winter* featured data, EKW system could behavior the best forecasting performance overall four WTGs, and which corresponding MAPEs are 5.44%, 5.43%, 3%, and 3.82%, respectively. In general, the proposed approach in this study performed little worse than some of compared methods in some scenarios. However, it is still the most favorable approach to predict short-term wind speed in consideration of the synthesized performance overall four WTGs.

4.6. Experiment III

Table 9 shows the comparison of the evaluation metrics during the verification period, where the season data come from WTG06. Fig. 7 presents the fitting performance of each category for different featured datasets during the testing period, where (a) shows the fitted results of the EEMD-WNN models trained by different categories, and (b) presents the prediction errors. In general, there was a major gap between the two types of datasets in terms of their different fluctuation characteristics. For the *non-season*, the EEMD-WNN model yielded excellent performance when the forecasts were conducted based on the second category. However, the prediction produced by the first category was particularly poor. In regard to the *spring*, *summer*, and *winter* datasets, the elicited differences of the respective comparisons of the results between the first category with the second were similar to the *non-season* data. In regard to the *autumn* data, the discrepancy between these two categories was different from the former four featured datasets. Conversely, the forecasting performance of the EEMD-WNN model established by the first category was better than the second.

5. Discussion

In this section, the experimental *season* data were collected from WTG06, and we will discuss the factors related to statistical models that would influence the forecasting performance. In addition, we will also investigate the influence of similarity within samples on the training requirement of the WNN model, and explore the relationship between the degree of similarity within training samples, and the forecasting performance of WNN models. Finally, the practical significance of the proposed forecasting system in this study is stated in Section 5.4.

5.1. Related factors of statistical models

To our knowledge, there are numerous applications of statistical models to the forecasting of time series data that are unstable, noisy, and irregular. According to the result of *Experiment II*, ARIMA outperformed other discussed models. The configurations of the arguments of p , d , and q , influenced the forecasting performance of the ARIMA model considerably. For our experiments, the arguments p and q were chosen based on the autocorrelation functions plots. Specifically, in our experiments, conducted for the *non-season* data, including the data of *spring*, *summer*, *autumn* and, *winter*, the optimal configurations of the arguments, and the optimal differences of each featured dataset were elicited by ARIMA (2,1,7), ARIMA (1,1,1), ARIMA (3,1,5), ARIMA (2,1,4) and ARIMA (8,1,9), respectively.

Neural network models have the ability of self-learning and are self-adaptive, and have gained tremendous popularity. They have always played an important role in nonlinear problem forecasting. However, there are a lot of parameters to configure, and the types of neural network models used also determine to a great extent their abilities on short-term wind speed prediction. There were no established rules for choosing the appropriate parameters [43] and the fitting models on short-term wind speed forecasting. As a result, the usual solution involves repetitive trials to obtain the appropriate parameters, and the

Table 7

Comparison of the evaluation metrics of the EKW model with the other common wind speed forecasting methods based on the season data of WTG08.

Model	Spring			Summer			Autumn			Winter		
	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD
Elman	42.33	9.429	2.754	24.81	8.073	1.197	27.71	2.184	0.882	45.4	7.157	2.05
PSO-Elman	14.69	1.415	0.951	19.99	3.338	1.338	35.04	1.864	1.03	25.8	1.603	0.974
ARIMA	9.09	0.513	0.572	7.79	0.513	0.533	18	0.386	0.497	12.79	0.339	0.463
RBF	31.26	12.179	1.924	24.02	14.956	1.166	12.85	1.3	0.396	—	—	—
SVM	> 50	20.063	4.218	> 50	54.197	6.966	> 50	11.33	3.083	> 50	11.604	3.148
EKW	4.89	0.131	0.311	3.63	0.113	0.247	4.04	0.03	0.123	3	0.033	0.118
BP	6.9	0.367	0.459	8.06	0.634	0.567	14.99	0.37	0.471	> 50	0.56	0.465
WNN	7.13	0.408	0.482	12.67	1.599	1.024	16.59	0.455	0.535	> 50	0.823	0.629
EEMD-WNN	6.67	0.102	0.237	6.04	0.535	0.548	8.57	0.087	0.253	11.47	0.295	0.483

Note: The bold in the table refer to the best forecasting performance.

Table 8

Comparison of the evaluation metrics of the EKW model with the other common wind speed forecasting methods based on the season data of WTG09.

Model	Spring			Summer			Autumn			Winter		
	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD	MAPE (%)	MSE	MAD
Elman	17.1	2.225	1.166	24.47	6.602	1.464	> 50	6.382	1.6	43.5	2.972	1.439
PSO-Elman	14.93	1.725	1.017	21.07	4.199	1.505	34.3	1.963	1.077	27.18	1.196	0.857
ARIMA	8.66	0.508	0.571	10.65	1.138	0.837	17.9	0.458	0.52	15.54	0.309	0.458
RBF	37.73	28.88	2.623	> 50	107.044	3.079	19.88	3.845	0.579	—	—	—
SVM	> 50	26.42	4.895	> 50	56.049	7.058	> 50	11.877	3.126	> 50	10.042	2.945
EKW	4.02	0.15	0.296	5.49	0.277	0.418	7.37	0.078	0.206	3.82	0.058	0.146
BP	6.03	0.33	0.435	8.96	0.846	0.684	15.6	0.468	0.505	> 50	0.388	0.406
WNN	8.21	0.56	0.592	10.86	1.376	0.908	17.08	0.528	0.565	> 50	0.667	0.61
EEMD-WNN	3.19	0.236	0.425	5.65	0.668	0.504	6.29	0.057	0.202	10.99	0.184	0.386

* The symbols of ‘—’ in these above tables imply that the prediction of methods is invalid, and when the value of MAPE yield exceeds 50% means the prediction is incorrect.

Note: The bold in the table refer to the best forecasting performance.

choice of neural network models that lead to the best forecasting performance.

In this study, four NN models were used to compare the proposed system, namely, the Elman, WNN, PSO-Elman, and the RBF models. We focused on the first three models (Elman, WNN, PSO-Elman) and examined the performance with different configurations in the case of the *non-season* featured dataset, with the consideration of three key parameters, namely, the train-to-verify ratio (3:1, 4:1 and 5:1), input layers (5–15), and hidden layers (10–25). However, we did not identify any rule to determine the optimal configuration of parameters which would elicit the best performance on short-term wind speed forecasting.

Additionally, during our experiment on NN models, we trained the three selected models 20 times with the same configuration of the selected parameters.

Fig. 8 shows the difference of forecasting performances among these three studied models. It is shown that there are considerable differences in the forecasting values among the NNs trained with the same configuration. The volatility of Elman is the largest, and its maximum MAPE was 39.37%, while the minimum error was 5.23%. This indicated that the Elman model was unable to handle short-term wind speed data compared to the WNN or the PSO-Elman models. In contrast, adopting particle swarm optimization algorithms to seek the

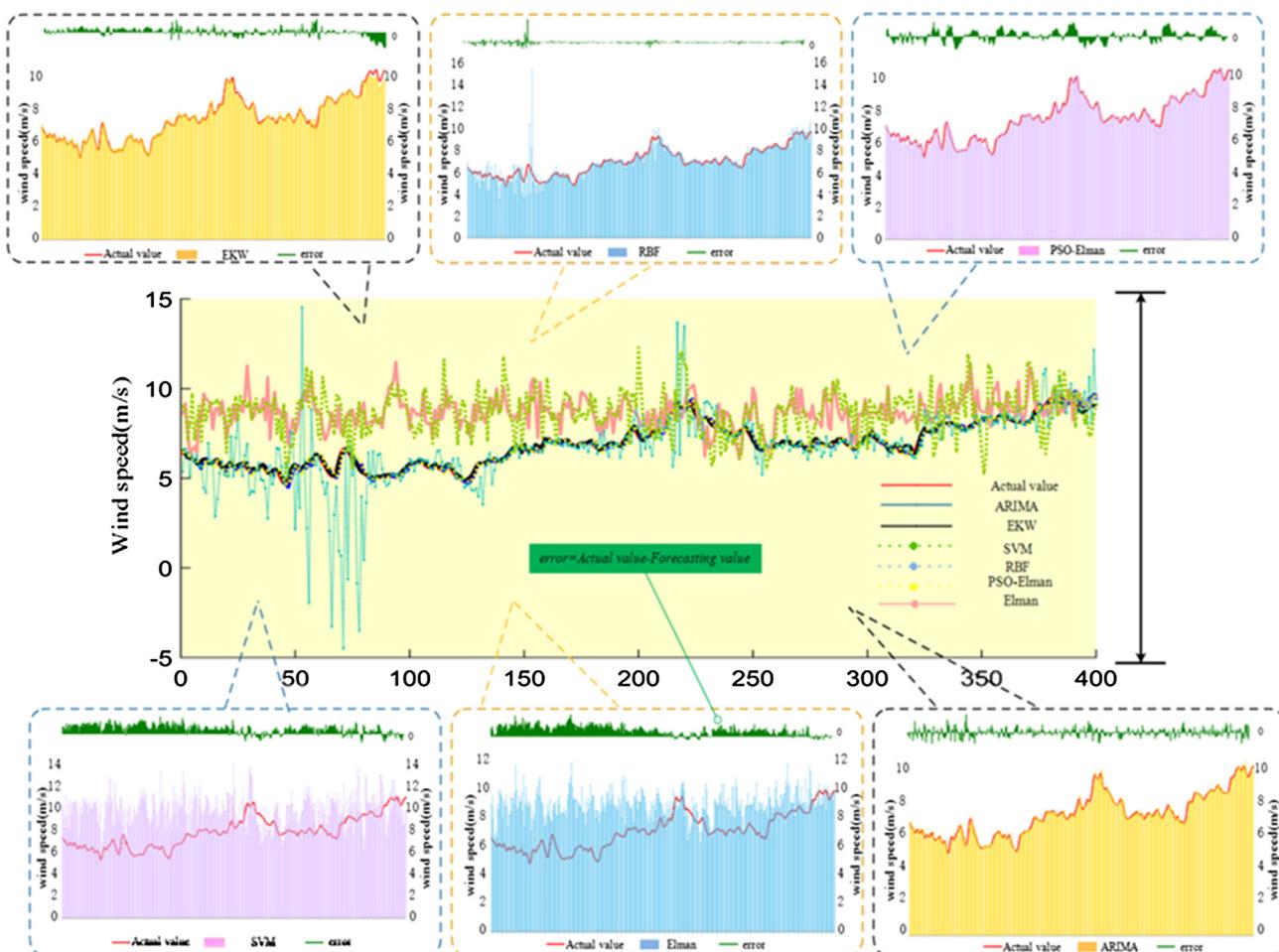


Fig. 5. Forecasting results of the six studied methods. *If the values of MAPE exceed 50%, they were set to 50% in this study.

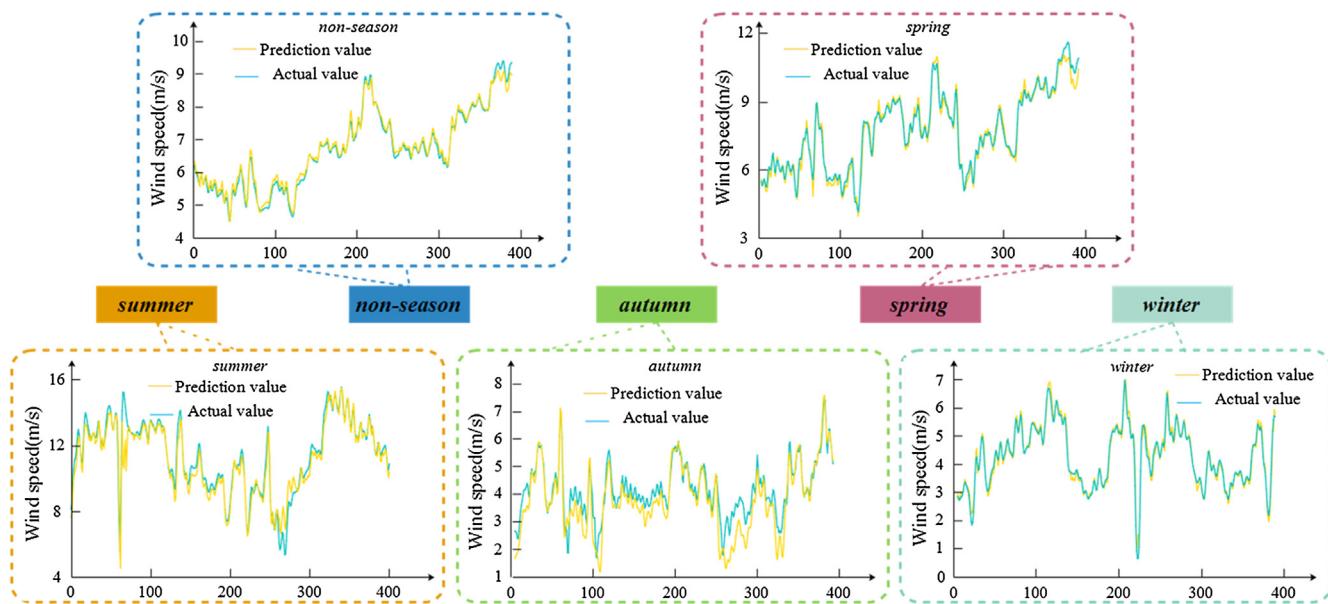


Fig. 6. Simulation results of the proposed hybrid system.

Table 9
Comparison of Evaluation Metrics with different categories.

	Non-season	Spring	Summer	Autumn	Winter
MAPE (%)	Cluster ₁	22.1	6.07	28.39	5.04
	Cluster ₂	5.1	3.06	11.52	22.79
MSE	Cluster ₁	2.741	0.495	3.842	0.135
	Cluster ₂	1.122	0.073	2.470	0.722
MAD	Cluster ₁	1.329	0.513	1.308	0.248
	Cluster ₂	0.523	0.209	1.025	0.603

Note: The bold in the table refer to the best forecasting performance.

optimal parameters of the Elman model improved significantly the forecasting performance of stability and precision. In addition, among these NNs, the WNN model performed best in terms of the prediction precision and stability. The selection of the most appropriate configuration of the parameters has been studied before, but it is beyond the scope of this study.

Additionally, the gains of excellent forecasting performance of these models largely depended on the amount of the original data. However, in this study, our available data spanned ten-minute wind speed data acquisitions, and the data were extremely unstable and noisy. Therefore, the MAPEs of these discussed models generally were much higher than the proposed system. The noise, instability and irregularity, were associated with the raw data series that ultimately defined the poor quality of the input data that are in turn used to establish the specific model. As a result, even though these statistical models are superior in theory and adopt a large amount of historical data to search for the appropriate parameters, they are still unable to handle these imperfect data. The common experience often encountered was the enlargement of the scale of the training samples to obtain relative good forecasting performance. However, in this study, there was no attempt to meet grand training requirements. Instead, imposed requirements were down-emphasized by enhancing the inner similarity. We investigate the influence of similarity on the training requirements of the WNN model in the next subsection.

5.2. Influence of similarity within sample on the training requirement of WNN model

It is well-known that the performance of neural network models is

determined by the amount of training sets, and that their dependence on a large amount of historical data always challenges the accuracy of short-term wind speed forecasts. In contrast, the instability and irregularity within the data which are selected as the training samples always trouble the development of the forecasting accuracy on the short-term wind speed. As a result, it is important to select a superior sample for modeling. As such, a powerful hybrid system was proposed in this study, which basically selected an appropriate sample for modeling in the effort to obtain the optimal forecasting performance. In addition, this system embodies the superiority of the sample by enhancing the inner similarity within it. The outcomes of *Experiment II* justified the experiment's superiority in forecasting accuracy, while the outcomes from *Experiment I* presented the increased efficiency of the experiment on short-term wind speed forecasting. In other words, the experiment was able to obtain equal or higher forecasting accuracies when a sample with higher similarity was used to establish the WNN model.

To obtain accurate results, the common practice is to expand the scale of the sample to fulfill the training requirement of the model when the available data is obtained from the original sample. The larger the training sample size is, the higher the prediction accuracy. However, in terms of the behavior of KFCM, the sample used for modeling was halved in size to enhance the inner similarity. Based on previous experiences, the prediction accuracy of this system became worse when the number of training sets was reduced compared to that obtained by a training sample when more samples were used. In other words, the EKW system should perform worse than the conventional EEMD-WNN model. In fact, the EKW system elicited better or similar forecasting performances for different featured datasets except for the summer dataset in *Experiment I*.

The experience tells us that the larger the scale of training sample is, the greater the probability is for obtaining the most appropriate parameters. However, the problem is that larger size of data could result in extremely complex computations in the effort to complete the training of the model, and could lead to poor forecasting performances. In this case, the sample was divided into two categories based on one type of data characteristics. One category was then adopted which possessed this characteristic to train the neural network model. In this way, the training was simple and directional, in other words, the extraction of data traits could use smaller scales of samples to obtain similar or better prediction performance than the larger size before clustering.

To further explore the influence of similarity within sample on the

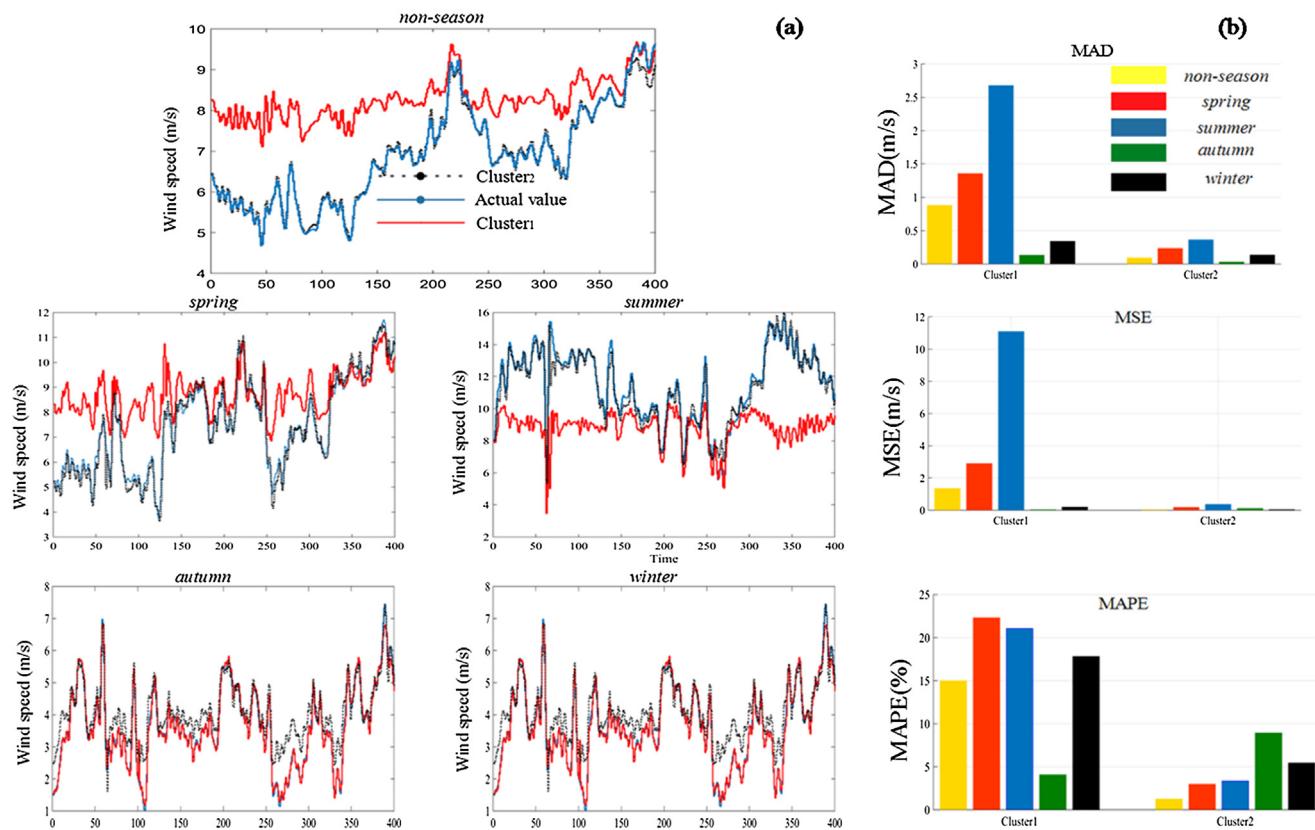


Fig. 7. The comparison of EKW model built by different categories across non-season data set.

training requirement of the WNN model, the KFCM-WNN model and WNN model are compared in this section. Both adopt the original data without processing the denoising procedure to conduct the modeling. The former model selected a part of the first 80% data as the training sample based on the results of clustering, and the latter used all 80% of data as its training sets. Fig. 9 shows that the KFCM-WNN model yielded better or similar forecasting performance within a shorter execution time. In other words, it is very wise to enhance the inner

similarity within the training sample at the expense of scale. Furthermore, in Fig. 10, the results of MAPE and MAD of KFCM-WNN and WNN models indicated that the samples with high similarity for modeling could reduce the training requirements of the WNN model considerably, and the forecasting accuracy was thus improved.

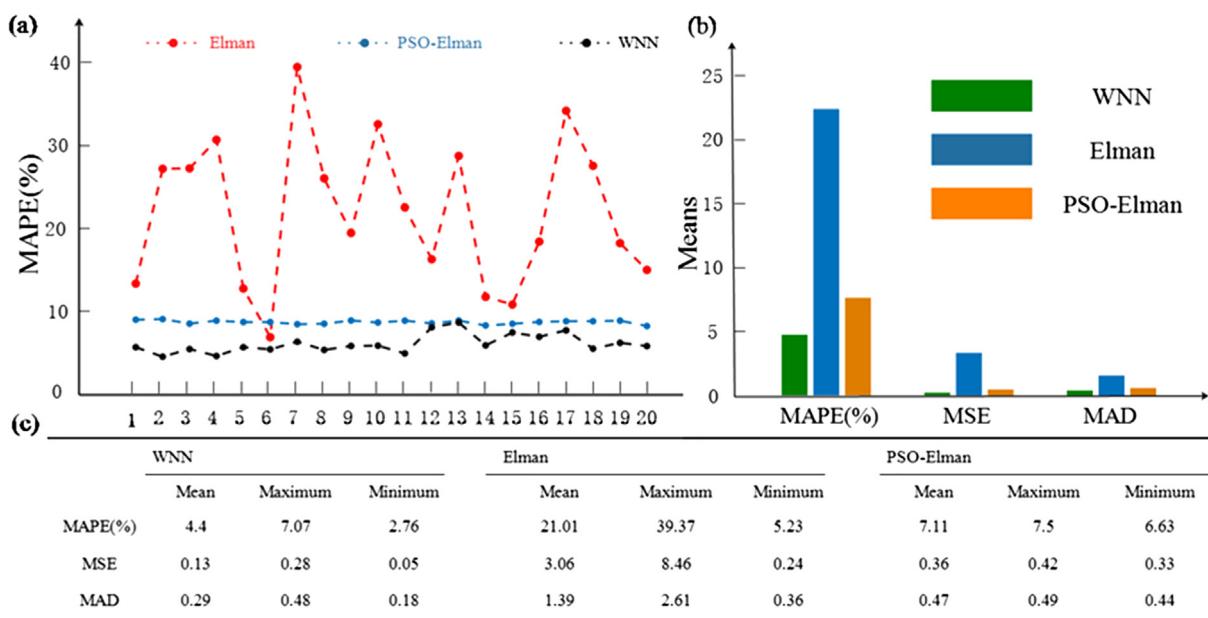


Fig. 8. The forecasting performance of different neural network models. *Conduct the networks for 20 times with the non-season datasets.

Featured datasets	Models	MAPE (%)	MSE	MAD	Consuming-time (s)	Scales
<i>non-season</i>	KFCM-WNN	2.74	0.06	0.19	28.1	722
	WNN	5.76	0.2	0.39	58.7	1600
<i>spring</i>	KFCM-WNN	7.48	0.54	0.54	27.2	750
	WNN	8.83	0.62	0.66	58.3	1600
<i>summer</i>	KFCM-WNN	>50	2.13	0.94	26.8	735
	WNN	>50	1.75	0.81	58.2	1600
<i>autumn</i>	KFCM-WNN	15.81	0.43	0.51	29.1	802
	WNN	14.86	0.44	0.51	59	1600
<i>winter</i>	KFCM-WNN	>50	0.5	0.5	32.8	886
	WNN	>50	0.51	0.54	58.4	1600

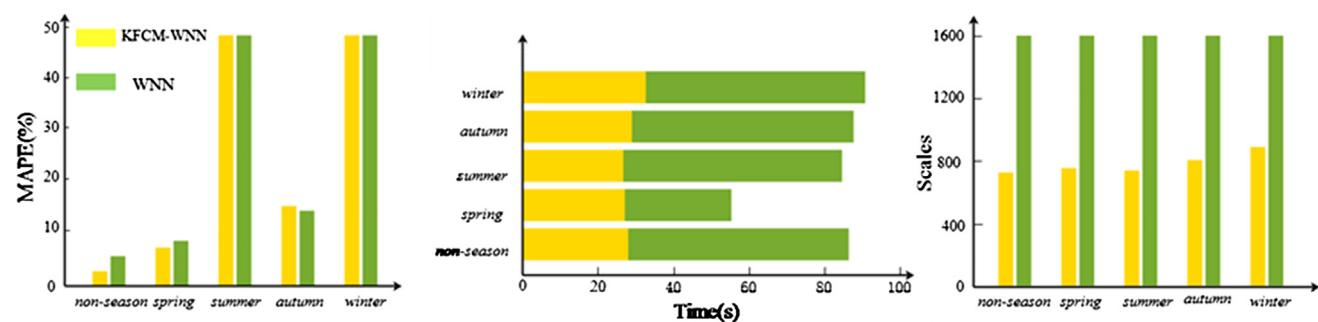


Fig. 9. Comparison of the KFCM-WNN and WNN models.

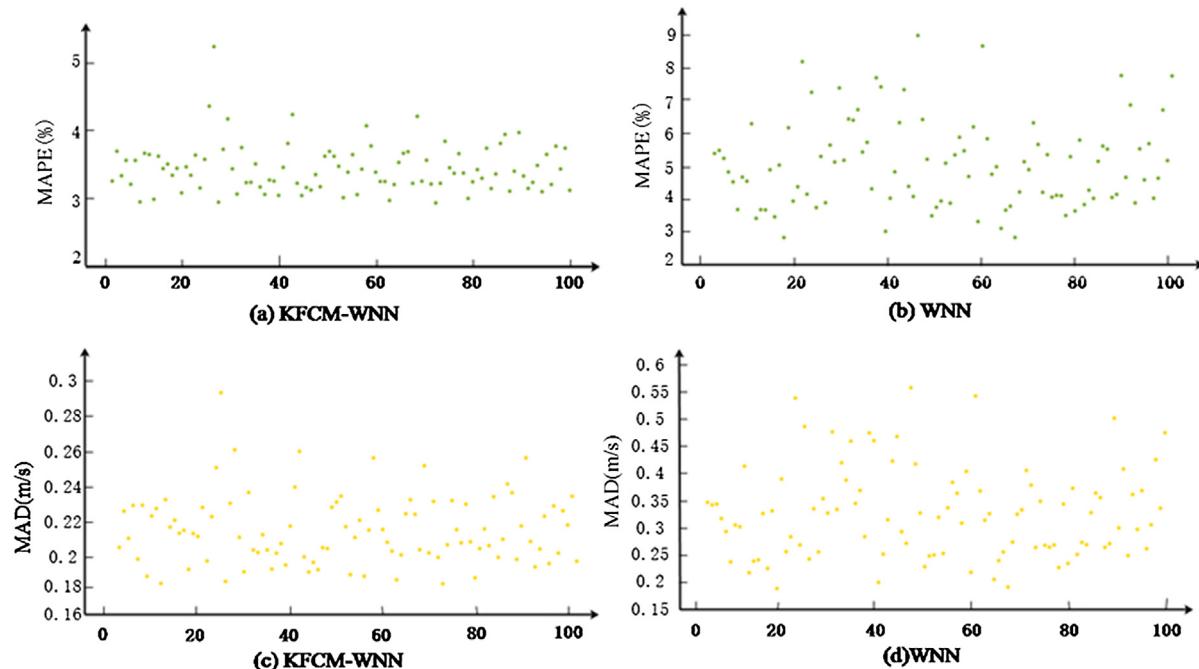


Fig. 10. An illustration showing the high degree of similarities on forecasting performances for the dataset of non-season when the neural network was used to construct the networks based on an experimental repetition of 100 times under the same configuration using 10 input layers and 21 hidden layers.

5.3. Relationship between the similarity within training sets and the performance of the WNN model

The similarity within a training sample in this study refers to similar frequency and amplitude fluctuations. However, its degree is always a vague concept and it is difficult to measure. The results of *Experiment III*

indicated that there was a considerable difference between the forecasting performances of the EEMD-WNN models built by Cluster₁ and Cluster₂. In regard to the similarity within the time series, Nirab Chandra Adhikary et al. [49] proposed that the measure of the self-similarity of electrostatic fluctuations could be obtained by calculating the Hurst exponent. The Hurst exponent mainly searches for the self-

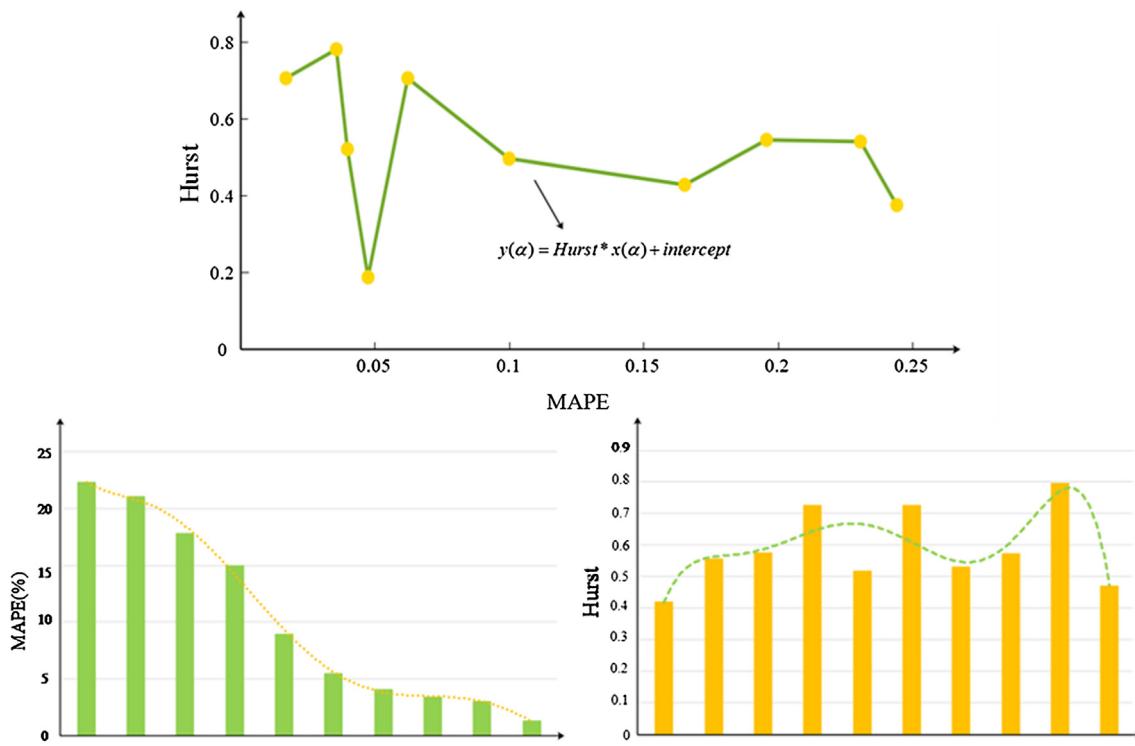


Fig. 11. The relationship between Hurst exponent and MAPE.

similarity within the time series. Because we setup the number of input neurons to ten, and the output neurons to one, herein, we adopted the traits of the corresponding output data to express the traits of the clustered input vectors. In other words, we will calculate the Hurst exponent of the corresponding output data series to represent the degree of similarity of different categories in this part. The corresponding output data series is still a time series, and as such, employing the Hurst exponent to measure the similarity within clustered input vectors is reasonable. The relationship between the Hurst exponents and MAPEs is shown in Fig. 11.

Fig. 11 shows the approximate relationship between the forecasting accuracy of the WNN model and the inner similarity within the training sample. We calculated the Hurst exponents of different datasets, which were used to train the WNN model, and then adopted the constructed models to conduct predictions. This trend indicated that as the inner similarity within the training sample enhanced gradually, and the prediction accuracy of the WNN model constructed using this sample was higher. In comparison, this relationship also indicated that the higher degree of similarity does not necessarily lead to a higher prediction accuracy, which is determined by so many other data traits. Therefore, there are tremendous accuracy differences among different models trained by different categories in *Experiment III*. However, the study of the exact quantification of the improvement of the prediction accuracy by the training sample similarity is beyond the scope of our study, and we will not elaborate in detail herein. The research conducted in this study on the measurement of similarity primarily aimed to provide a theoretical basis for our hybrid system EKW.

5.4. Practical significance of the proposed forecasting system

Given the extensive use of wind energy, related research studies, have been receiving increasing attention, including, wind speed multi-step forecasting [50] and point forecasting. However, wind speed forecasting is always regarded as a challenging task owing to the uncertainty and randomness that ultimately results in negative influences on the robust scheduling and management of wind power systems in a

wind farm. In addition, the generation of wind power is determined by wind speed directly, and according to [51], the formula for converting wind energy into wind power is:

$$P_a = \left\{ \frac{\exp[-(v_c/c)^k] - \exp[-(v_f/c)^k]}{(v_r/c)^k - (v_c/c)^k} - \exp[-(v_f/c)^k] \right\} \times P_r$$

where v_c , v_r , v_f are the cut-in, cut-off, and nominal wind speed values (m/s) respectively. Additionally, c is the Weibull scale parameter (m/s), while P_a and P_r are the average power outputs of the wind turbine (kW), and the rated electrical power of the wind turbine (kW) respectively. It is obvious that improving the accuracy of wind speed forecasting plays a vital role in the wind turbine power generation. The specific contributions of the powerful wind speed forecasting system to the research and application of the power system are described as follows:

- It is mentioned in the first section that the problem pertaining to the supply electricity based on the demand is highly complex. The balance dilemma has been challenging the efficient management of wind power systems. Specially, the power overload will not only affect the quality of power supply, but will even damage the security and stability of the power system, and increase the cost of wind farms owing to the difficulties in electricity storage [52]. Therefore, a powerful wind speed prediction system is needed, which could contribute in the provision of precise forecasting results of electricity generation, and help the decision-makers reach appropriate decisions to avoid economic losses.
- In order to guarantee maximum use of wind energy, adjusting the amount and capacity of wind turbines based on the volume of wind speed is vital for the management of a wind farm. Furthermore, accurate wind speed forecasting models could provide very reliable references for managers to schedule the configuration of wind turbines to avoid damages and reduce operation cost.

6. Conclusions and future work

Short-term wind speed forecasting plays a vital role in the effective operation and optimal risk management of wind farms. Furthermore, the 10-min wind speed data generated by wind turbines are easily affected by random fluctuations. Thus, the accurate short-term wind speed forecasting appears to be particularly important. Given this challenge, many methods have been proposed. However, most of them seldom considered the importance of the selection of the training sample of the model. In consideration of this aspect, and motivated by recent progress in similarity-theory-based forecasting models, a hybrid system EKW was proposed in this study. The reasonable and effective forecasting performance of this system was owing to the data preprocessing and data clustering modules. Moreover, it is clear from *Experiment I* and Discussion 5.2 that data preprocessing and subsample selection before modeling could improve the forecasting accuracy largely. In addition, we evaluated and compared the EKW system with other commonly used, conventional forecasting methods. *Season* and *non-season* datasets were adopted to evaluate the effectiveness and the availability of this proposed system. The results of *Experiments I* and *II* showed that the EKW system generally outperformed other methods in terms of the forecasting accuracy. Furthermore, in the discussion about the certainty of similarities, the quantification of the role of similarity on the forecasting accuracy of the WNN model was emphasized. Therefore, our future work will focus on researching the quantification problem and the application of this promotion trend to other statistical models.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant Nos. 71671029 and 41475013).

References

- [1] Ma Xuejiao, Jin Yu, Dong Qingli. A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting. *Appl Soft Comput* 2017;54:296–312.
- [2] Sun Shaolong, Qiao Han, Wei Yunjie, Wang Shouyang. A new dynamic integrated approach for wind speed forecasting. *Appl Energy* 2017;197:151–62.
- [3] Santamaría-Bonfil G, Reyes-Ballesteros A, Gershenson C. Wind speed forecasting for wind farms: a method based on support vector regression. *Renew Energy* 2016;85:790–809.
- [4] Jiang Ping, Wang Yun, Wang Jianzhou. Short-term wind speed forecasting using a hybrid model. *Energy* 2017;119:561–77.
- [5] Liu Da, Wang Jilong, Wang Hui. Short-term wind speed prediction based on spectral clustering and optimized echo state networks. *Renew Energy* 2015;78:599–608.
- [6] Jung JS, Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renew Sustain Energy Rev* 2014;31:762–77.
- [7] Jung JS, Broadwater RP. Current status and future advances for wind speed and power forecasting. *Renew Sustain Energy Rev* 2014;31:762–77.
- [8] Zhao Jing, Guo Yanling, Xiao Xia, Wang Jianzhou, Chi Dezhong, Guo Zhenhai. Multi-step wind speed and power forecasts based on a WRF simulation and an optimized association method. *Appl Energy* 2017;197:183–202.
- [9] Xiao LY, Qian F, Shao W. Multi-step wind speed forecasting based on a hybrid forecasting architecture and an improved bat algorithm. *Energy Convers Manage* 2017;143:410–30.
- [10] Liu Da, Niu Dongxiao, Wang Hui, Fan Leilei. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. *Renew Energy* 2014;62:592–7.
- [11] Schlink U, Tetzlaff G. Wind speed forecasting from 1 to 30 minutes. *Theor Appl Ckinatol* 1998;60:191–8.
- [12] Gneiting T, Larson K, Westrick K, Genton MG, Aldrich E. Calibrated probabilistic forecasting at the Stateline wind energy center: the regime-switching space-time method. *J Am Stat Assoc* 2006;101:968–79.
- [13] Lydia M, Kumar SS, Selvakumar AI, Kumar GEP. Linear and nonlinear auto-regressive models for short-term wind speed forecasting. *Energy Convers Manage* 2016;112:115–24.
- [14] Kavasseri Rajesh G, Seetharaman Krithika. Day-ahead wind speed forecasting using f-ARIMA models. *Renew Energy* 2009;34:1388–93.
- [15] Guo ZH, Jie WuJ, Lu HY, Wang JZ. A case study on a hybrid wind speed forecasting method using BP neural network. *Knowl-Based Syst* 2011;24:1048–56.
- [16] Li G, Shi J. On comparing three artificial neural networks for wind speed forecasting. *Appl Energy* 2010;87(7):2313–20.
- [17] Jiang Ping, Liu Feng, Song Yiliao. A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. *Energy* 2017;119:694–709.
- [18] Xiao Liye, Shao Wei, Mengxia Yu, Ma Jing, Jin Congjun. Research and application of a hybrid wavelet neural network model with the improved cuckoo search algorithm for electrical power system forecasting. *Appl Energy* 2017;198:203–22.
- [19] Cadenas E, Rivera W. Short term wind speed forecasting in La Venta, Oaxaca, Mexico, using artificial neural networks. *Renew Energy* 2009;34:274–8.
- [20] Ramasamy P, Chandel SS, Yadav AK. Wind speed prediction in the mountainous region of India using an artificial neural network model. *Renew Energy* 2015;80:338–47.
- [21] Mohandes MA, Halawani TO, Rehman S, Hussain AA. Support vector machines for wind speed prediction. *Renew Energy* 2004;29(6):939–47.
- [22] Zhou J, Shi J, Li G. Fine tuning support vector machines for short-term wind speed forecasting. *Energy Convers Manage* 2011;52:1990–8.
- [23] Jianming Hu, Wang Jianzhou, Xiao Liqun. A hybrid approach based on the Gaussian process with t-observation model for short-term wind speed forecasts. *Renew Energy* 2017;114:670–85.
- [24] Maran Poongavanam Sardar, Ramalingam Ponnusamy. A meteorological tower based wind speed prediction model using fuzzy logic. *Am J Environ Sci* 2013;9:226–30.
- [25] Javad Shahram, Hojjatinia Zeinab. Wind speed modelling and prediction in wind farms using fuzzy logic. *Appl Math Electr Comput Eng* 2004:98–103.
- [26] Mohands M, Rehman S, Rahman SM. Estimation of wind speed profile using adaptive neuro-fuzzy inference system (ANFIS). *Appl Energy* 2011;88:4024–32.
- [27] Chang Fi-John, Chang Ya-Ting. Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Adv Water Resour* 2006;29:1–10.
- [28] Wang Shouxiang, Zhang Na, Lei Wu, Wang Yamin. Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method. *Renew Energy* 2016;94:629–36.
- [29] Vanisri D. A novel kernel based fuzzy c means clustering with cluster validity measures. *Int J Eng Trends Appl* 2014;3:1–8.
- [30] Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng QN, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc A* 1998;454:903–95.
- [31] Li G, Shi J. Applications of Bayesian methods in wind energy conversion systems. *Renew Energy* 2012;43:1–8.
- [32] Liu H, Tian HQ, Li YF. Four wind speed multi-step forecasting models using extreme learning machines and signal decomposing algorithms. *Energy Convers Manage* 2015;100:16–22.
- [33] Wu ZH, Huang NE. Ensemble empirical mode decomposition: a noise assisted data analysis method. *Adv Adaptive Data Anal* 2009;1:1–41.
- [34] Gaci Said. A new ensemble empirical mode decomposition (EEMD) denoising method for seismic signals. *ScienceDirect* 2016;97:84–91.
- [35] Li Tianhao, Zhang Liyong, Wei Lu, Hou Hui, Liu Xiaodong, Pedrycz Witold, et al. Interval kernel fuzzy c-means clustering of incomplete data. *Neurocomputing* 2017;237:316–31.
- [36] Ghosh Soumi, Dubey Sanjay Kumar. Comparative analysis of k-means and fuzzy c-means algorithm. *International Journal of Advanced Computer Science and Applications*. 2013;4:35–9.
- [37] Jipkate BR, Gohakar VV. A comparative analysis of c-means clustering and k-means clustering algorithm. *Int J Comput Eng Res* 2012. ISSN: 2250-3005.
- [38] Graves Daniel, Pedrycz Witold. Kernel-based fuzzy clustering and fuzzy clustering: a comparative experimental study. *Fuzzy Sets Syst* 2010;4:522–43.
- [39] Ayvaz M Tamer, Karahan Halil, Aral Mustafa M. Aquifer parameter and zone structure estimation using kernel-based fuzzy c-means clustering and genetic algorithm. *ScienceDirect* 2007;343:240–53.
- [40] Graves D, Pedrycz W. Kernel-based fuzzy clustering and fuzzy clustering: a comparative experimental study. *Fuzzy Sets Syst* 2010;161(4):522–43.
- [41] Zhang Q, Benveniste A. Wavelet networks. *IEEE Trans Neural Networks* 1992;3(6):889–98.
- [42] Doucoure Boubacar, Agbossou Kodjo, Cardenas Alben. Time series prediction using artificial wavelet neural network and multi-resolution analysis: Application to wind speed data. *Renew Energy* 2016;92:202–11.
- [43] Liu Li, Wang Qianru, Wang Jianzhou, Liu Ming. A rolling grey optimized by particle swarm optimization economic prediction. *Comput Intell* 2014.
- [44] Wang Jianzhou, Heng Jiani, Xiao Liye, Wang Chen. Research and application of a combined model based on multi-objective optimization for multi-step ahead wind speed forecasting. *Energy* 2017;125:591–613.
- [45] Salis Christos I, Malissovas Anastasios E, Bizoopoulos Paschalos A, Tzallas Alexandros T, Angelidis PA, Tsalikakis Dimitrios G. Denoising simulated EEG signals: a comparative study of EMD, wavelet transform and Kalman filter. *IEEE*; 2013.
- [46] Wang Jianzhou, Niu Tong, Haiyan Lu, Guo Zhenhai, Yang Wendong, Pei Du. An analysis-forecast system for uncertainty modeling of wind speed: a case study of large-scale wind farms. *Appl Energy* 2018;211:492–512.
- [47] Zhang Wenyu, Wu Jie, Wang Jianzhou, Zhao Weigang, Shen Lin. Performance analysis of four modified approaches for wind speed forecasting. *Appl Energy* 2012;99:324–33.
- [48] Witten Ian H, Eibe Frank. Data mining practical machine learning tools and techniques. 2nd ed. Morgan Kaufmann Press; 2011.
- [49] Adhikary Nirab Chandra, Pal Arup Ratan, Bailung Heremba, Chutia Joyanti. Self-similarity of electrostatic fluctuations in a linear magnetised plasma system. *Phys Lett A* 2006;350:380–5.
- [50] Wang Jianzhou, Song Yiliao, Liu Feng, Hou Ru. Analysis and application of forecasting models in wind power integration: a review of multi-step-ahead wind speed forecasting models. *Renew Sustain Energy Rev* 2016;60:960–81.
- [51] Wang Jianzhou, Du Pei, Niu Tong, Yang Wendong. A novel hybrid system based on a new proposed algorithm-Multi-Objective Whale Optimization Algorithm for wind speed forecasting. *Appl Energy* 2017.
- [52] Hamzacebi C, Es HA. Forecasting the annual electricity consumption of Turkey using an optimized grey model. *Energy* 2014;70:165–71.