Renewable Energy 96 (2016) 727–737

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Short-term wind speed forecasting using empirical mode decomposition and feature selection

Chi Zhang [a], Haikun Wei [a,*], Junsheng Zhao [b], Tianhong Liu [a], Tingting Zhu [a], Kanjian Zhang [a]

[a] Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, PR China
[b] School of Mathematical Science, Liaocheng University, Liaocheng 252059, PR China

ARTICLE INFO

Article history:
Received 11 May 2015
Received in revised form
26 March 2016
Accepted 3 May 2016
Available online 18 May 2016

Keywords:
Wind speed prediction
Empirical mode decomposition
Feature selection
Hybrid model
Artificial neural networks
Support vector machines

ABSTRACT

Due to the non-linear and non-stationary characteristics of the wind speed time series, it is generally difficult to model and predict such series by single forecasting models. In this paper, two novel hybrid models, which combine empirical mode decomposition (EMD), feature selection with artificial neural network (ANN) and support vector machine (SVM), are proposed for short-term wind speed prediction. First, the original wind speed time series is decomposed into a set of sub-series by EMD. Second, the initial features (input variables) and targets are constructed from all the sub-series and the original series. Then, a feature selection process is introduced to constitute the relevant and informative features. Finally, a predictive model (ANN or SVM) is established using these selected features. The effectiveness of the proposed models has been assessed on the real datasets recorded from three wind farms in China. Compared with the single ANN, SVM, traditional EMD-based ANN, and traditional EMD-based SVM, the experimental results show that the proposed models have satisfactory performance, which are suitable for the wind speed prediction.

1. Introduction

As a clean and renewable energy resource, wind energy has been under large scale development throughout the world in the last decade [1]. Among the various renewable energy resources, wind energy showed the most rapid and consistent deployment of power generating capacities [2]. However, due to the intermittency and stochastic fluctuation of wind, the integration of wind power into power systems poses a number of challenges [3]. One of the possible solutions to the challenges is to improve the wind speed and power forecasting [4].

Many methods have been developed for the short-term prediction of wind speed and power generation in recent decades. Generally, these methods can be classified according to whether a numerical weather prediction (NWP) model is involved or not [5]. Two mainstream approaches exist: the physical and the statistical approach [4,5]. The former takes into account the physical description (such as local terrain) of the wind site and utilizes the output of NWP to predict the local wind speed. Contrary to the physical method, the statistical method usually constructs the statistical models based on a number of historical data to predict wind speed. Conventional statistical models, especially the autoregressive moving average (ARMA) models have been widely applied to the short-term wind speed forecasting [6,7]. These models can explicitly reveal the linear relationship in the time series, but the predictive results will be unsatisfactory if the non-linear characteristics of wind speed series are prominent. Apart from the traditional time series models, machine learning models, such as artificial neural networks (ANNs) [8,9] and support vector machines (SVMs) [10,11] have also been frequently adopted for wind speed forecasting. In general, machine learning models can capture non-linear relationships in wind speed series and thus offer better predictive performance.

In recent years, there has been a growing trend of combining signal decomposition algorithms and machine learning models, forming a hybrid model for wind speed prediction. The procedure of building these hybrid models mainly includes three steps: (1) the original wind speed time series is decomposed into several sub-series by some signal decomposition algorithm;

* Corresponding author.
E-mail address: hkwei@seu.edu.cn (H. Wei).

http://dx.doi.org/10.1016/j.renene.2016.05.023
0960-1481/

(2) the forecasting models are built for each sub-series, and the predictions are made in each sub-series; (3) the prediction results of each individual model are aggregated to generate the final forecast of the original time series. For example, Liu et al. [12] developed a hybrid model for wind speed prediction based on the EMD and the ANNs. In this method, the EMD was utilized to decompose the original time series into a set of different sub-series and for each sub-series, the predictive ANN model was built. The final forecast for the original time series was obtained by conducting an aggregate calculation for the predicting results in the sub-series. The results showed that the proposed method was robust in dealing with jumping samplings in non-stationary wind series. Hu et al. [13] used the ensemble empirical mode decomposition (EEMD) and the SVM to improve the quality of wind speed forecasting. The simulation results indicated that the proposed hybrid model was quite efficient to predict wind speed with the feature of high volatility and irregularity. Wang et al. [14] developed a hybrid model based on EMD and Elman neural network (ENN) [15] to forecast wind speed. The results showed that the hybrid EMD-ENN model had better forecasting performance than the persistence model, back-propagation neural network and single ENN. Liu et al. [16] used wavelet packet and wavelet algorithms to decompose the original wind speed data into several sub-series, and then both autoregressive integrated moving average (ARIMA) and ANN models were built for each sub-series. A more comprehensive study can be found in Ref. [17], in which four decomposing algorithms (wavelet, wavelet packet, empirical mode decomposition and fast ensemble empirical mode decomposition) and two representative neural networks (multilayer perceptron and adaptive neuro fuzzy inference system) were adopted to realize the wind speed predictions. Ren et al. [18] conducted a comparative study on EMD-based hybrid models for wind speed forecasting. In their study, the decomposition algorithms included ensemble EMD, complementary EEMD, and complete EEMD with adaptive noise; the predictive models included ANN and SVM.

From the studies mentioned above, it can be found that these hybrid models are different mainly in the decomposition algorithms or the predictive models, but they have nearly the same building process. These models are referred to as the decomposition forecasting aggregation (DFA) models in the sections below. In this paper, we will combine the signal decomposition algorithms and machine learning models in a different way. Here, after decomposing the original wind speed time series into a set of sub-series, the initial input-output pairs (i.e. training examples) are constructed from all the sub-series and the original series. Then a feature selection process is introduced to constitute the relevant and informative features. Finally, a predictive model is established using these selected features. This kind of hybrid model is referred to as the decomposition selection forecasting (DSF) model. Unlike the DFA, the proposed method does not build a forecasting model for each sub-series. In view of the wide use in wind speed prediction, the decomposition algorithm adopted for this study is the EMD and the forecasting model employed as the final predictor is the ANN or SVM.

The reminder of the paper is organized as follows. Section 2 introduces the EMD and the forecasting models used in the study. Section 3 presents the proposed hybrid models for wind speed prediction. The experimental procedure and results are presented and discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Empirical mode decomposition and forecasting models

2.1. EMD

EMD is a data-driven algorithm which can deal with non-linear and non-stationary signals. The main idea of EMD is to decompose a complicated signal into a finite and small number of oscillatory modes based on the local characteristic time scale by itself. Each oscillatory mode is expressed by an intrinsic mode function (IMF), which has to satisfy the following two conditions [19]: (1) In the whole dataset, the number of extrema and the number of zero-crossings must either be equal or differ at most by one; (2) At any point, the mean between the upper and lower envelopes, which are defined by the local maxima and minima, must be zero. Let $y(t)$ be a given original wind speed time series, the computational steps of EMD is described as follows [19]:

Step 1 Identify all the local extrema of $y(t)$ and then connect all the local maxima and local minima with an interpolation method (e.g. cubic spline) to generate an upper envelope $y_{\mathrm{up}}(t)$ and a lower envelope $y_{\mathrm{low}}(t)$, respectively.

Step 2 Calculate the mean envelop $m(t)$ from the upper and lower envelopes $m(t) = [y_{\mathrm{up}}(t) + y_{\mathrm{low}}(t)]/2$ and then subtract it from the original time series to obtain a detailed component $d(t) = y(t) - m(t)$

Step 3 Check whether $d(t)$ is an IMF. If $d(t)$ is an IMF then set $c(t) = d(t)$ and meantime replace $y(t)$ with the residual $r(t) = y(t) - c(t)$. Otherwise, replace $y(t)$ with $d(t)$ and repeat Steps 1–2 until the following termination criterion is satisfied:

$$\sum_{t=1}^{l} \frac{\left[\mathrm{d}_{j-1}(t) - \mathrm{d}_j(t)\right]^2}{\left[\mathrm{d}_{j-1}(t)\right]^2} \le \delta \; (j = 1, 2, \ldots; t = 1, 2, \ldots, l) \tag{1}$$

where l is the length of the signal and j denotes the number of iterative calculation. A typical value for δ is usually set between 0.2 and 0.3.

Step 4 Repeat Steps 1–3 until all the IMFs and the residual are obtained. Finally, the original time series $y(t)$ can be decomposed as follows:

$$y(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{2}$$

where $c_i(t)(i = 1,2,\ldots,n)$ represents different IMFs and $r_n(t)$ is the final residual.

2.2. ANN

Nowadays, ANNs have been widely used for time series modeling and forecasting [20,21]. In the present study, the ANN employed is a multi-input, single-output feedforward network, consisting of an input layer, a hidden layer, and an output layer. The inputs of ANN consist of previous observations and the output is the forecasted value. The relationship between the output (y_t) and the inputs ($y_{t-1}, y_{t-2}, \ldots, y_{t-p}$) has the following mathematical representation:

$$y_t = v_0 + \sum_{j=1}^{q} v_j g\left(w_{0,j} + \sum_{i=1}^{p} w_{i,j} y_{t-i}\right) \tag{3}$$

where g is the activation function, $w_{i,j}(i = 0,1,2,\ldots,p; j = 1,2,\ldots,q)$ is the connection weight from input node i to hidden node j, and $v_j (j = 0,1,2,\ldots,q)$ is the weight from hidden node j to the output node; p is the number of input nodes, and q is the number of hidden nodes. The activation function of the hidden node selected in this paper is a sigmoid function, that is,

$$g(x) = \frac{1}{1+e^{-x}} \tag{4}$$

The choice of q is data dependent and there is no theory that can be used to guide the selection of this parameter [20]. The usual method is to test numerous networks with varying numbers of hidden units and select the one with the lowest estimated generalization error [21,22]. Another important parameter is the number of lagged observations p. In this study, the partial autocorrelation function (PACF) [8,23] is utilized to find its best value. Once the structure of ANN (i.e. p,q) is specified, it is ready for training (i.e., finding the optimal values of the connection weights). In this study, the Levenberg-Marquardt backpropagation training algorithm [24] is adopted for this purpose.

2.3. SVM

SVM is another popular non-linear model developed by Vapnik and his co-workers based on statistical learning theory [25]. Like ANN, SVM has also been used extensively for time series forecasting [26,27]. The basic idea of SVM for regression is to map the input data into a high dimensional feature space via a non-linear mapping and to perform a linear regression in this feature space [26]. Specifically, consider a given training set $\{\mathbf{x}_i, y_i\}, i = 1,2,\ldots,N$, with input $\mathbf{x}_i \in \mathbb{R}^p$ and output $y_i \in \mathbb{R}$. The constructed regression model can be expressed by using nonlinear mapping function φ as

$$y = f(\mathbf{x}) = \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}) + b \tag{5}$$

where $\mathbf{w}$ is the weight vector and b is the bias term. The regression is calculated by minimizing a risk functional

$$R[f] = \frac{1}{2}||\mathbf{w}||^2 + \frac{1}{2}C\sum_{i=1}^{N} L(y_i, f(\mathbf{x}_i)) \tag{6}$$

where C is regularization constant and $L(y_i, f(\mathbf{x}_i))$ is the ε-insensitive loss function [28], which is defined as follows

$$L(y_i, f(\mathbf{x}_i)) = |y_i - f(\mathbf{x}_i)|_\varepsilon = \begin{cases} |y_i - f(\mathbf{x}_i)| - \varepsilon & \text{for } |y_i - f(\mathbf{x}_i)| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where ε controls the width of the error margin allowed. The regression problem can be solved by the following constrained optimization problem [29].

$$\begin{aligned} &\min \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \\ &\qquad \text{subject to} \\ &y_i - \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i) - b \leq \varepsilon + \xi_i \\ &-y_i + \mathbf{w}^{\mathrm{T}}\varphi(\mathbf{x}_i) + b \leq \varepsilon + \xi_i^* \\ &\xi_i, \xi_i^* \geq 0, \ \ i = 1,\ldots,N \end{aligned} \tag{8}$$

where ξ_i, ξ_i^* represent slack variables to make constraints feasible. By introducing the Lagrange multipliers α_i, α_i^*, the final form of function $f(\mathbf{x})$ can be obtained as follows

$$f(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}) + b \tag{9}$$

where k is the kernel function. In this study, the commonly used radial basis function (RBF) is used which is defined as

$$k(\mathbf{x}_i, \mathbf{x}_j) = e^{-\gamma||\mathbf{x}_i - \mathbf{x}_j||^2} \tag{10}$$

where $\|\cdot\|$ denotes the 2-norm and γ is an adjustable parameter to determine the RBF kernel width.

3. Proposed EMD-based DSF model

In this paper, an EMD-based decomposition selection forecasting (DSF) model is proposed for wind speed prediction. The purpose of feature selection is on one hand to improve the predictive accuracy, and on the other hand to reduce the computation time of model training (as there is only one forecasting model in this hybrid framework). The framework of the model is shown in Fig. 1 and the building procedure is described as follows:

Step 1 The original wind speed time series is decomposed by EMD into a number of IMFs and a residual.
Step 2 The original features are constructed from all the IMFs and the residual, and they constitute the potential input variables for the predictive models.
Step 3 The optimal feature subset is selected from various possible feature combinations.
Step 4 The forecasting model (ANN or SVM) is built using the selected feature subset to perform the final wind speed prediction.

The original features are constructed based on the potential relationship between the original time series and their lags. This relationship is determined according to the Box-Jenkins methodology [30]. Specifically, if the autoregressive order of a time series $y(t)$ is p, and $y(t)$ is decomposed into n IMFs and a residual (Eq. (2)), then the original feature vector at time period t is constructed as:

$$\begin{aligned}\mathbf{x} = [&c_1(t), c_2(t), \ldots, c_n(t), r_n(t), \ \ c_1(t-1), c_2(t-1), \ldots, c_n(t-1), \\ &r_n(t-1), \ldots, c_1(t-p+1), c_2(t-p+1), \ldots, c_n(t-p+1), \\ &r_n(t-p+1)]\end{aligned} \tag{11}$$

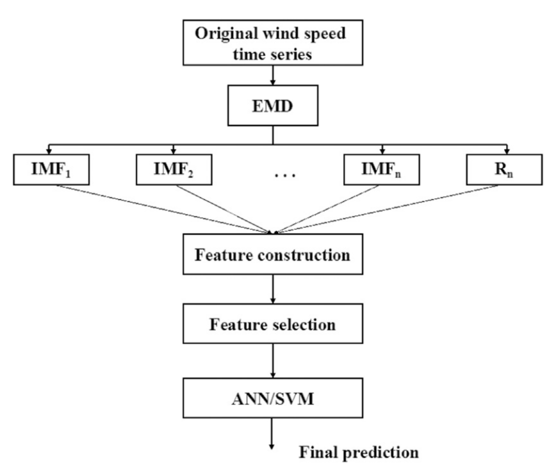

Fig. 1. Framework of the proposed EMD-based DSF model.

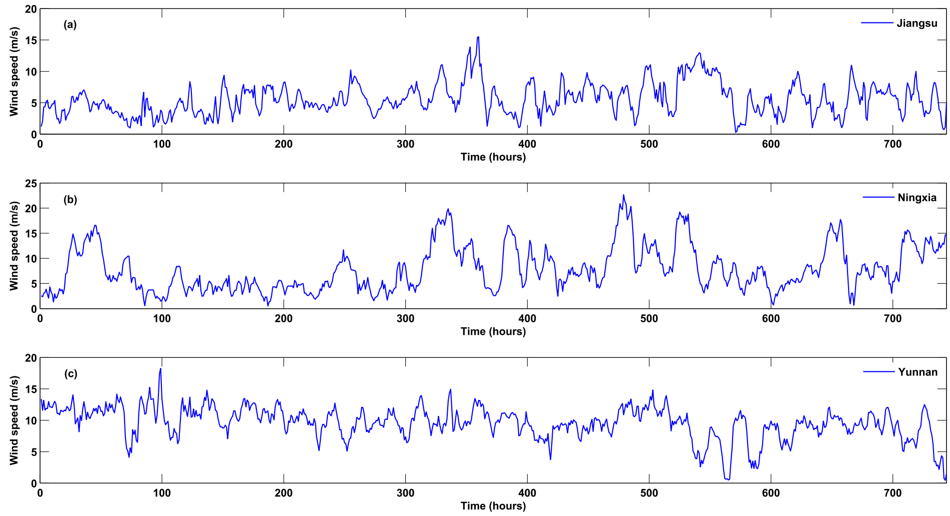

Fig. 2. The hourly mean wind speed data in: (a) Jiangsu, (b) Ningxia, and (c) Yunnan.

To select the optimal feature subset, the training data is splitted into two parts: learning set and validation set. Learning set is used to calculate the parameters of a predictive model, and validation set is used to estimate the predictive performance. The feature combination that generates the smallest validation error is selected as the best one. To avoid over-fitting, the predictive model used for feature selection is linear regression which has the form

$$y = w_0 + \sum_{i=1}^{m} w_i x_i = \mathbf{w}^{\mathrm{T}}\mathbf{x} \tag{12}$$

where w_i's are unknown coefficients. Typically, given a set of training examples $\{(\mathbf{x}_1,y_1)\ \ldots\ (\mathbf{x}_N,y_N)\}$, the least square estimation for $\mathbf{w}$ is

$$\mathbf{w} = \left(\mathrm{X}^{\mathrm{T}}\mathrm{X}\right)^{-1}\mathrm{X}^{\mathrm{T}}\mathbf{y} \tag{13}$$

where X is a $N \times (m+1)$ matrix with each row an input vector, and $\mathbf{y}$ is a $N \times 1$ vector of outputs in the training set.

4. Case studies

4.1. Wind speed datasets

The hourly mean wind speed data of three wind farms in China are employed to evaluate the proposed hybrid model. These data were collected from March 1, 2012 to March 31, 2012. The three wind farms are located in different provinces of China (Jiangsu, Ningxia and Yunnan). Fig. 2 shows the hourly wind speed time series corresponding to the three sites. From Fig. 2, it can be seen that the wind speed fluctuates severely and presents apparent non-stationarity. The descriptive statistics of the three wind speed datasets, including the mean, the standard deviation, the minimum and the maximum velocities, is shown in Table 1. In this table, it can be observed that the statistical measures are different for the three wind speed datasets. To evaluate the performance of the forecasting models, each dataset is partitioned into two parts: the first 624 samples for training and the remaining 120 samples for testing. The training set is used to establish the predictive models, and the testing set is employed to validate the effectiveness of the models.

It is important that the complexity of the forecasting models should correspond to the number of data. Usually, the more data we have, the more flexible models (e.g. ANNs with more hidden neurons) we can use to capture more information from data. In this study, we have already considered the complexity of the forecasting models to avoid underfitting and overfitting. Moreover, the number of data (1 month of hourly data) used in this study is also adopted in the similar work by other researchers [12,14,31]. However, it should be pointed out that so far, there is no unified standard or clear definition about the exact amount of data for the method validation. And other different sizes of data set (smaller or larger than 744) can also be found in the literature [13,32]. Note that no matter how many data are available (the more the better), the framework and the modeling process proposed in this paper remain unchanged.

4.2. Forecasting performance evaluation

In order to assess the performance of the predictive models quantitatively, three error measures: mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE), are employed. All of them are measures of the deviation between the actual and predicted values. In general, smaller values of these measures indicate a better forecasting performance. These measures are defined as follows:

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - \hat{y}_i\right| \tag{14}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}\left(y_i - \hat{y}_i\right)^2}{N}} \tag{15}$$

Table 1
Descriptive statistics of three wind speed datasets (m/s).

	Mean	St. dev.	Min.	Max.
Jiangsu	5.55	2.55	0.27	15.52
Ningxia	7.83	4.47	0.53	22.68
Yunnan	9.59	2.57	0.45	18.25

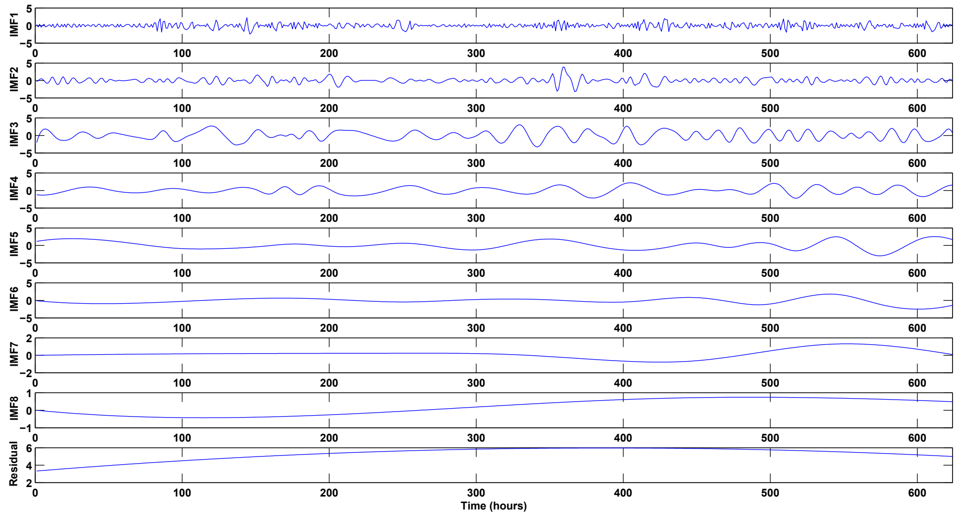

Fig. 3. EMD results in Jiangsu.

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i)}{y_i}\right| \times 100\% \tag{16}$$

where N is the sample size, y_i and $\hat{y}_i$ are the observed and forecasted values at time period i, respectively.

4.3. EMD of wind speed time series

The original wind speed series is decomposed into several IMF components and one residual by using EMD, as shown from Figs. 3–5. All the IMF components are graphically illustrated in the order in which they are extracted, from the highest frequency to the lowest one. The first few IMFs represent the high time variant or noise in the original wind speed data, while the last few IMFs indicate the long-period characteristics. The last sub-series is the residual of sifting process, which generally represents the trend of the original time series. All the extracted sub-series are then used either to built their corresponding forecasting models or to select the appropriate inputs for the final forecasting model.

4.4. Determination of model order

Before training the forecasting models, training examples, which consist of the input-output pairs, have to be extracted from the time series data. In this study, the Box-Jenkins methodology is adopted to achieve this objective. The sample autocorrelation function (ACF) and PACF are used to make the first guess about the potential orders, and then the Bayesian information criterion (BIC) [33] is employed to select the best one. The ACF and PACF of the original wind speed time series in Jiangsu are provided in Fig. 6. From Fig. 6, it can be seen that the ACF graph shows an exponential decaying pattern, while the PACF graph shows a cutoff phenomenon after lag 2. The BIC values in Fig. 6 further verify that the suitable order for the time series is 2. Similarly, Figs. 7 and 8 display the results of order identification for the wind speed time series in Ningxia and Yunnan, respectively.

4.5. Results of feature selection

After determining the model order, the training examples, which consist of the original features and the target, are

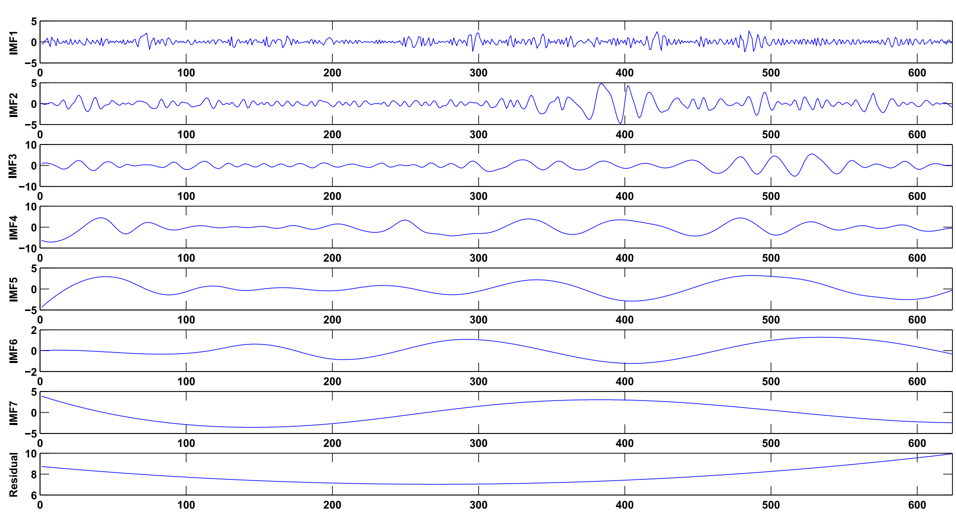

Fig. 4. Similar to Fig. 3 in Ningxia.

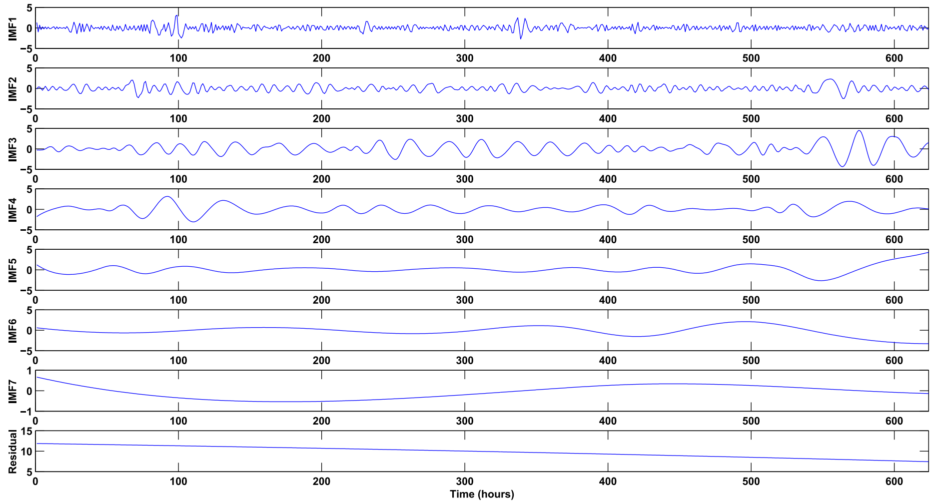

Fig. 5. Similar to Fig. 3 in Yunnan.

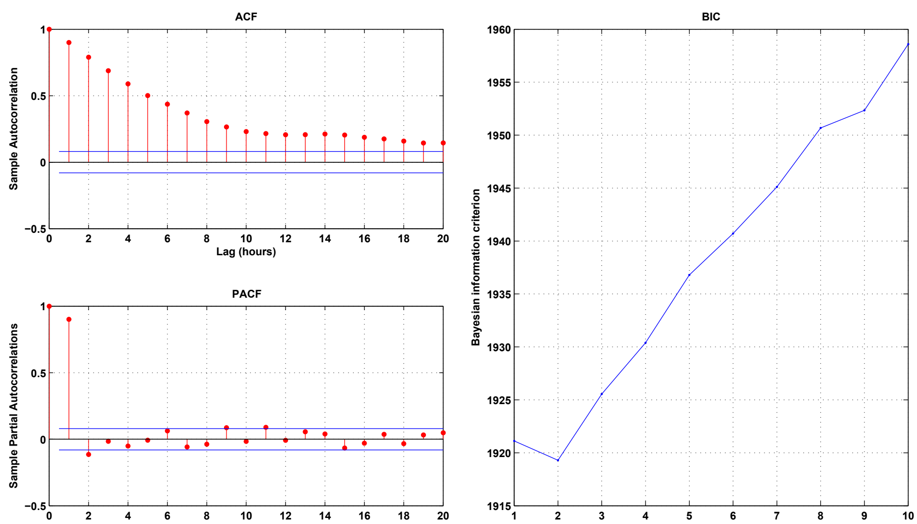

Fig. 6. Order identification for wind speed time series in Jiangsu.

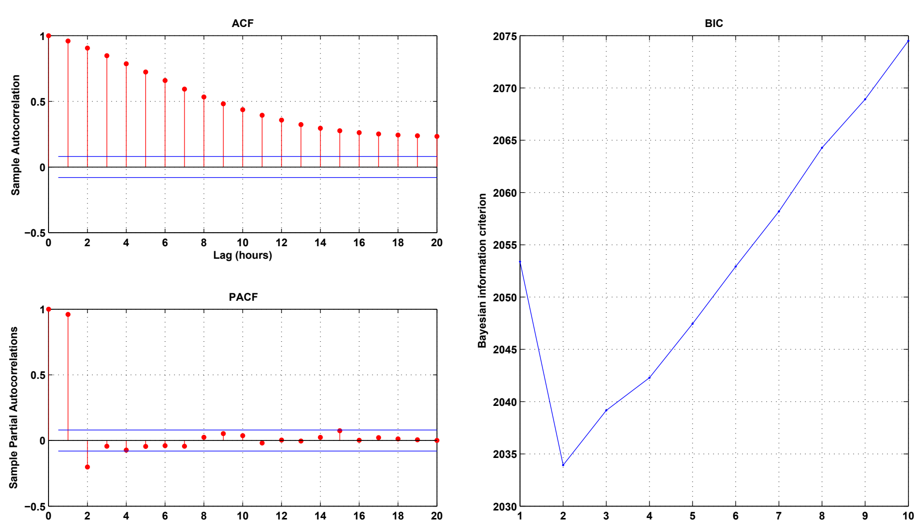

Fig. 7. Similar to Fig. 6 in Ningxia.

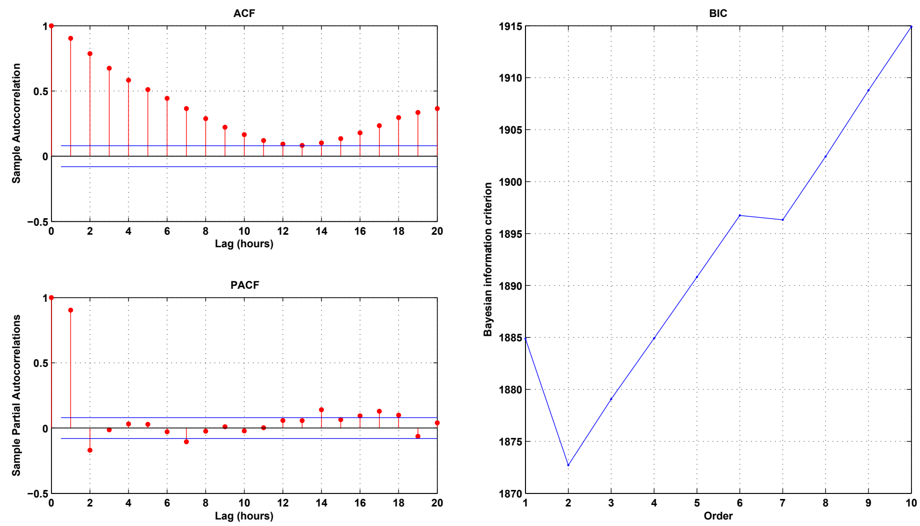

Fig. 8. Similar to Fig. 6 in Yunnan.

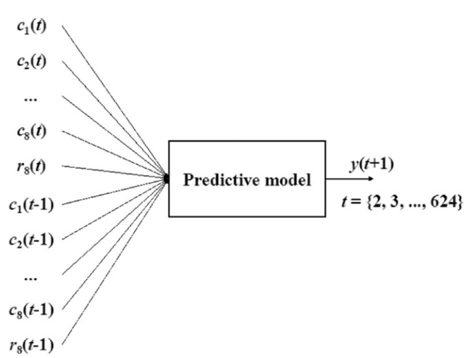

Fig. 9. Construction of training examples in Jiangsu.

constructed from all the sub-serie and the original series. For example, the wind speed time series in Jiangsu is decomposited into 8 IMFs and a residual, and the model order is identified as 2, then the training examples are extracted as Fig. 9:

In order to select the optimal feature subset, the training set is divided into two parts: the first 80% is used to estimate the parameters of the linear regression model (Eq. (13)), and the remaining 20% is employed to assess the performance of the feature subset. In view of the time relationship between the original features and the desired output, not all the combinations of the features are considered. In this study, a simple feature selection process is adopted. All the features are added to the current feature subset one by one, from the more recent ones to the less recent ones (e.g. from $c_1(t)$ to $r_8(t-1)$ in Fig. 9). The optimal feature subset is the one that generates the smallest error on the validation set. Fig. 10 displays the validation error against the feature number in Jiangsu. From Fig. 10, it can be seen that the best feature number is 14 in this case (from $c_1(t)$ to $c_5(t-1)$). Similarly, the same procedure

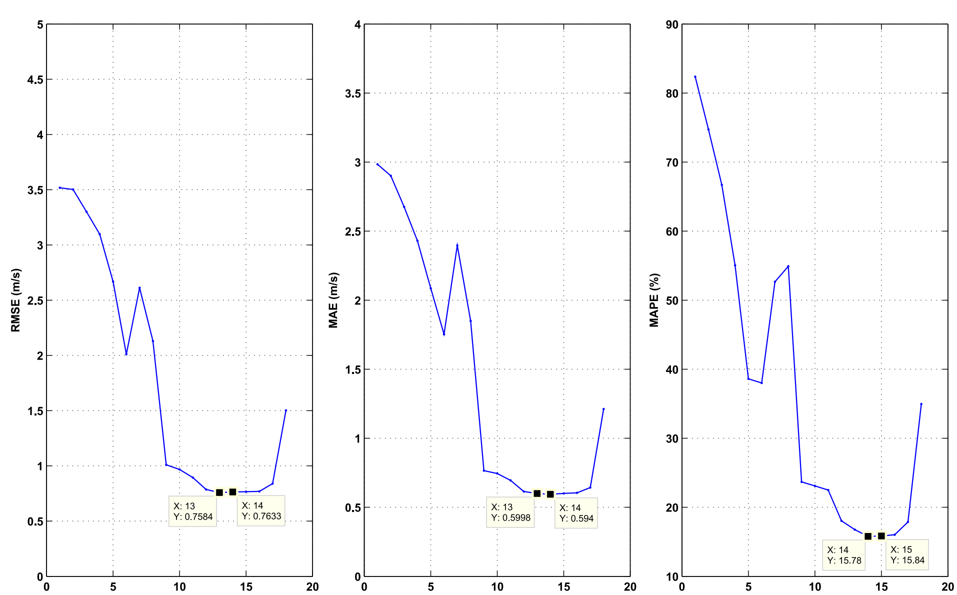

Fig. 10. Validation error of feature selection in Jiangsu.

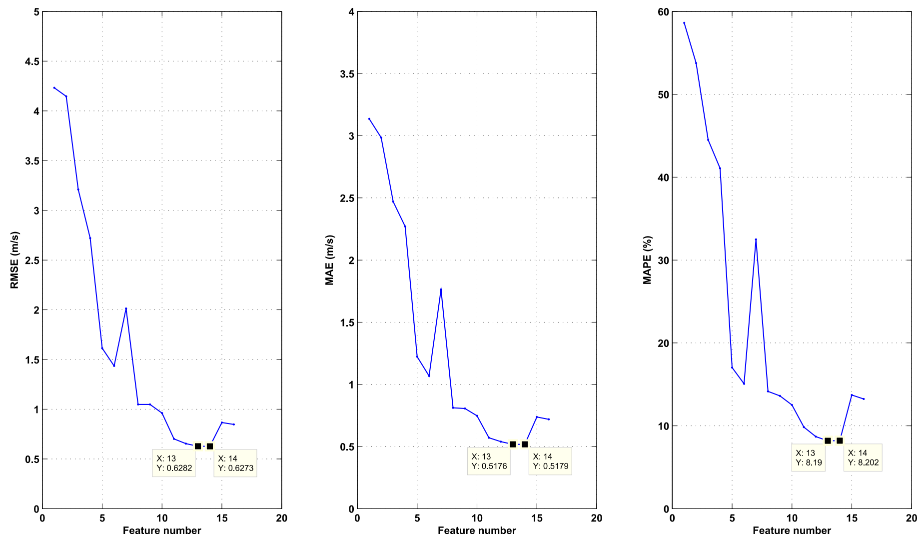

Fig. 11. Similar to Fig. 10 in Ningxia.

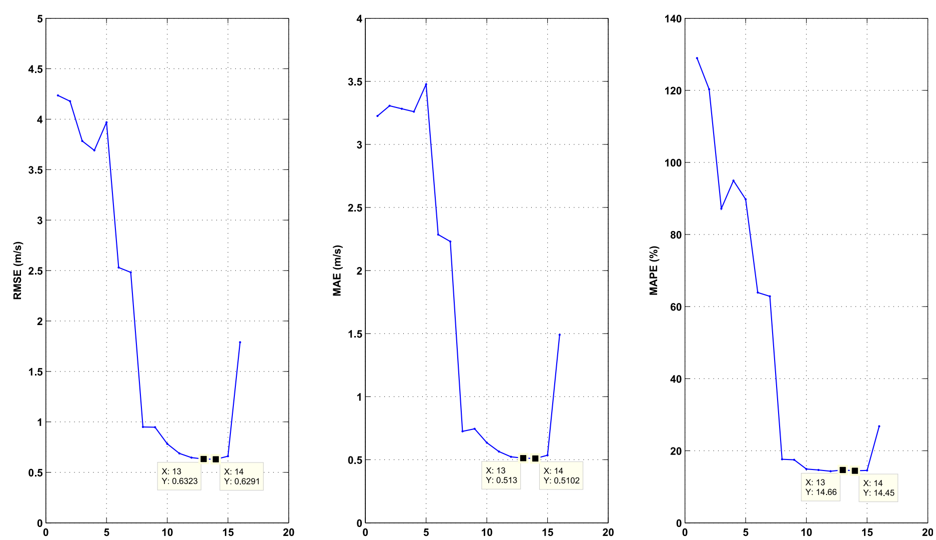

Fig. 12. Similar to Fig. 10 in Yunnan.

is conducted on the other two wind speed datasets, and the results of feature selection are shown in Figs. 11 and 12. The specific features selected for the three datasets are summarized in Table 2.

4.6. Forecast results and comparative analysis

For each wind speed dataset, six forecasting models are constructed: ANN, SVM, DFA-ANN, DFA-SVM, DSF-ANN, and DSF-SVM. The first two models are single forecasting models, which are built using the original wind speed time series. The last four models are EMD-based hybrid models, where DFA-ANN and DFA-SVM employ ANN and SVM for each decomposed sub-series; DSF-ANN and DSF-SVM use ANN and SVM as the final predictor, respectively.

In order to build the ANN model, it is necessary to know the following parameters: (1) the number of input, (2) the number of hidden layers, (3) the number of hidden neurons in each hidden layer. The number of input is determined by the result of model order identification. Note that, in the case of DFA-ANN, the procedure of model order identification is conducted for each decomposed sub-series. The number of hidden layers is 1 in this study, as it is accepted that a network with three layers connected towards ahead can approximate any continuous function in a reasonable way [34]. According to [35], the number of hidden neurons is

Table 2
Results of feature selection.

	Features
Jiangsu	$c_1(t)$,$c_2(t)$,$c_3(t)$,$c_4(t)$,$c_5(t)$,$c_6(t)$,$c_7(t)$,$c_8(t)$,$r_8(t)$, $c_1(t-1)$,$c_2(t-1)$,$c_3(t-1)$,$c_4(t-1)$,$c_5(t-1)$
Ningxia	$c_1(t)$,$c_2(t)$,$c_3(t)$,$c_4(t)$,$c_5(t)$,$c_6(t)$,$c_7(t)$,$r_7(t)$, $c_1(t-1)$,$c_2(t-1)$,$c_3(t-1)$,$c_4(t-1)$,$c_5(t-1)$,$c_6(t-1)$
Yunnan	$c_1(t)$,$c_2(t)$,$c_3(t)$,$c_4(t)$,$c_5(t)$,$c_6(t)$,$c_7(t)$,$r_7(t)$, $c_1(t-1)$,$c_2(t-1)$,$c_3(t-1)$,$c_4(t-1)$,$c_5(t-1)$,$c_6(t-1)$

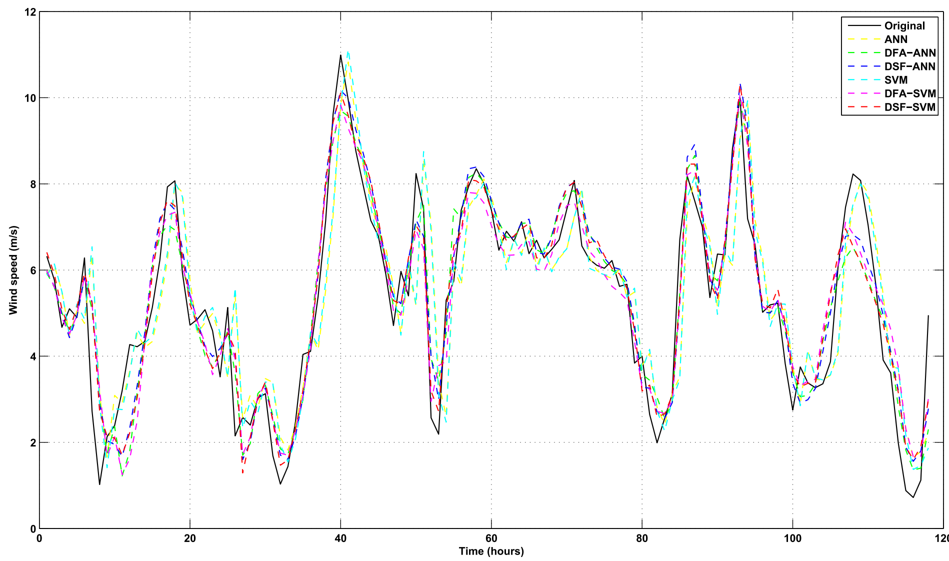

Fig. 13. The forecasting results of the wind speed series in Jiangsu.

determined as the integer number closest to log(T), where T is the number of training examples.

For the construction of the SVM model, it is necessary to know the following parameters: (1) the number of input, (2) parameter C in Eq. 6, (3) parameter ε in Eq. 7, (4) parameter γ in Eq. (10). Similarly, the number of input is determined by the result of model order identification. C is fixed as $y_{\max} - y_{\min}$, according to [36], where $y_{\max}$ and $y_{\min}$ are the maximum and minimum values of the output of the training examples. According to [37], the remaining two parameters are determined by try-and-error experiments, which is similar to the process of above mentioned feature selection.

Figs. 13–15 show the one-step ahead forecasting results of the testing set for the three wind speed datasets. The corresponding error measures of the six models are tabulated in Tables 3–5, respectively. Figs. 13–15 and Tables 3–5 indicate that: (a) the EMD-based hybrid forecasting models outperform the single ones significantly in terms of RMSE, MAE and MAPE. This reveals the effectiveness of the EMD; (b) when comparing the performance of the proposed DSF models with that of the DFA models, the former models have higher accuracy than the latter ones except for the case in Yunnan (Table 5). For the dataset in Yunnan, these two kinds of EMD-based hybrid models have comparable performance; (c) the performance difference between the DSF-ANN and the DSF-SVM is not significant. Likewise, the performance difference between the DFA-ANN and the DFA-SVM, as well as the difference between the single ANN and the SVM is not significant too; (d) the EMD-based hybrid models (DFA and DSF) can generates satisfactory forecasts most of the time, but they sometimes underestimate the wind speed for those extrema.

Note that the proposed forecasting models are only applicable to the prediction for one month and cannot be generalized for a yearly period. In order to predict wind speed for other period of the year, more data are needed to build a specific model for each month [8]

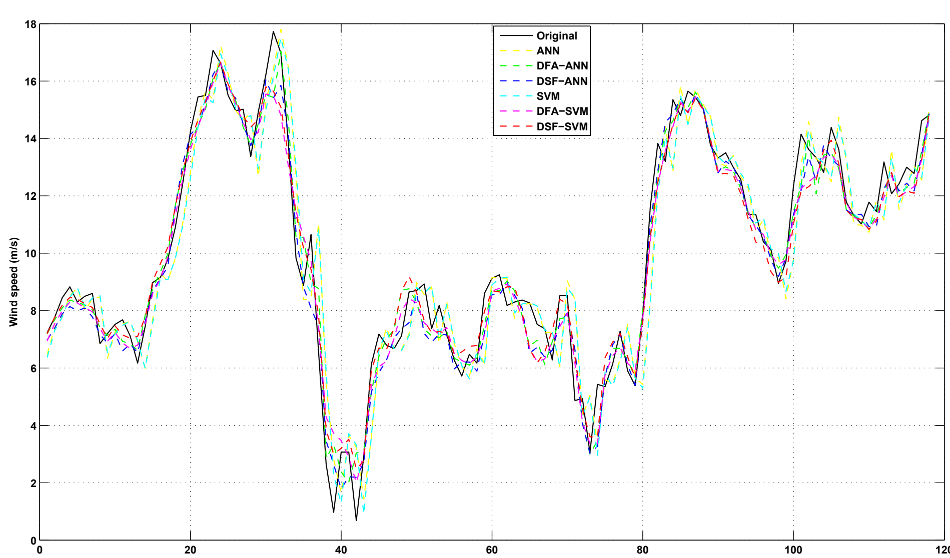

Fig. 14. Similar to Fig. 13 in Ningxia.

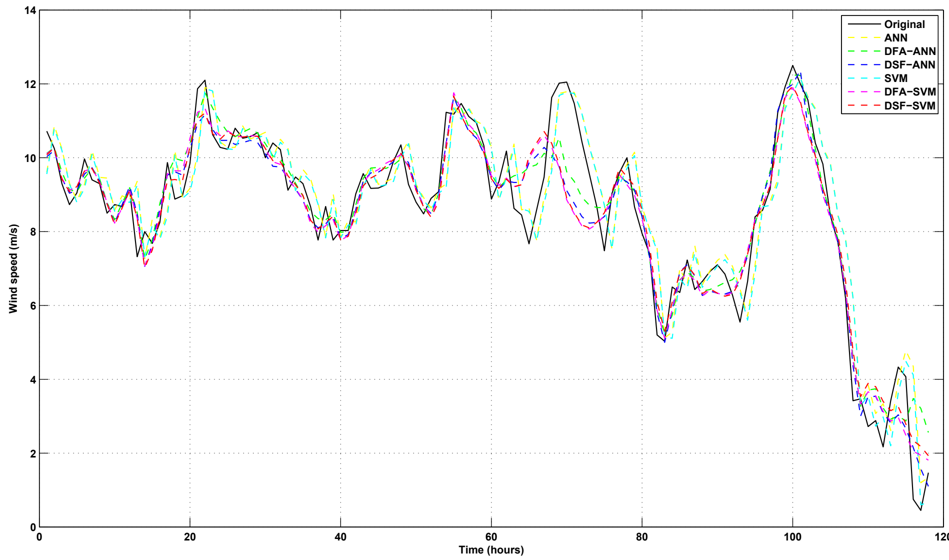

Fig. 15. Similar to Fig. 13 in Yunnan.

or each season [11]. However, no matter which method is used (building a model for each month or each season), the modeling strategy and process proposed in this study are the same. Only the parameters of the final models are different.

Table 3
Testing set results in Jiangsu.

Model	RMSE (m/s)	MAE (m/s)	MAPE (%)
ANN	1.25	0.92	25.24
DFA-ANN	0.85	0.64	18.10
DSF-ANN	0.80	0.62	17.98
SVM	1.24	0.91	24.20
DFA-SVM	0.86	0.65	19.40
DSF-SVM	0.79	0.60	17.30

Table 4
Similar to Table 3 in Ningxia.

Model	RMSE (m/s)	MAE (m/s)	MAPE (%)
ANN	1.35	1.03	17.26
DFA-ANN	0.82	0.63	11.82
DSF-ANN	0.80	0.61	10.29
SVM	1.36	1.02	17.13
DFA-SVM	0.84	0.65	11.11
DSF-SVM	0.81	0.63	10.91

Table 5
Similar to Table 3 in Yunnan.

Model	RMSE (m/s)	MAE (m/s)	MAPE (%)
ANN	0.93	0.71	14.67
DFA-ANN	0.80	0.58	12.90
DSF-ANN	0.82	0.59	11.25
SVM	0.92	0.70	13.46
DFA-SVM	0.81	0.57	11.69
DSF-SVM	0.81	0.58	12.72

5. Conclusions

This paper proposes to combine EMD with feature selection to construct novel hybrid models for wind speed prediction. Specifically, two hybrid models, DSF-ANN and DSF-SVM are developed, which employ ANN and SVM as the final predictor, respectively. The proposed models have been evaluated with the single models (ANN and SVM) and the traditional EMD-based models (DFA-ANN and DFA-SVM). The results show that: (1) the combination of EMD with ANN or SVM can improve the performance significantly; (2) both DSF-ANN and DSF-SVM have satisfactory forecasting results compared with the traditional DFA-ANN and DFA-SVM.

Acknowledgement

This research is supported by the National Natural Science Foundation of China (Grant No. 61374006), and by the Major Program of National Natural Science Foundation of China (Grant No. 11190015).

References

[1] X. Zhu, M.G. Genton, Short-term wind speed forecasting for power system operations, Int. Stat. Rev. 80 (1) (2012) 2–23.
[2] P. Pinson, et al., Wind energy: forecasting challenges for its operational management, Stat. Sci. 28 (4) (2013) 564–585.
[3] P.S. Georgilakis, Technical challenges associated with the integration of wind power into power systems, Renew. Sustain. Energy Rev. 12 (3) (2008) 852–863.
[4] J. Jung, R.P. Broadwater, Current status and future advances for wind speed and power forecasting, Renew. Sustain. Energy Rev. 31 (2014) 762–777.
[5] G. Giebel, R. Brownsword, G. Kariniotakis, M. Denhard, C. Draxl, The State-of-the-Art in Short-term Prediction of Wind Power: a Literature Overview, Tech. rep., ANEMOS. Plus, 2011.
[6] J.L. Torres, A. Garcia, M. De Blas, A. De Francisco, Forecast of hourly average wind speed with arma models in navarre (Spain), Sol. Energy 79 (1) (2005) 65–77.
[7] E. Erdem, J. Shi, Arma based approaches for forecasting the tuple of wind speed and direction, Appl. Energy 88 (4) (2011) 1405–1414.
[8] E. Cadenas, W. Rivera, Short term wind speed forecasting in la venta, oaxaca, méxico, using artificial neural networks, Renew. Energy 34 (1) (2009) 274–278.
[9] G. Li, J. Shi, On comparing three artificial neural networks for wind speed forecasting, Appl. Energy 87 (7) (2010) 2313–2320.
[10] M. Mohandes, T. Halawani, S. Rehman, A.A. Hussain, Support vector machines for wind speed prediction, Renew. Energy 29 (6) (2004) 939–947.
[11] J. Zhou, J. Shi, G. Li, Fine tuning support vector machines for short-term wind

speed forecasting, Energy Convers. Manag. 52 (4) (2011) 1990–1998.
[12] H. Liu, C. Chen, H.-q. Tian, Y.-f. Li, A hybrid model for wind speed prediction using empirical mode decomposition and artificial neural networks, Renew. Energy 48 (2012) 545–556.
[13] J. Hu, J. Wang, G. Zeng, A hybrid forecasting approach applied to wind speed time series, Renew. Energy 60 (2013) 185–194.
[14] J. Wang, W. Zhang, Y. Li, J. Wang, Z. Dang, Forecasting wind speed using empirical mode decomposition and elman neural network, Appl. Soft Comput. 23 (2014) 452–459.
[15] J.L. Elman, Finding structure in time, Cogn. Sci. 14 (2) (1990) 179–211.
[16] H. Liu, H.-q. Tian, D.-f. Pan, Y.-f. Li, Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks, Appl. Energy 107 (2013) 191–208.
[17] H. Liu, H.-q. Tian, Y.-f. Li, Comparison of new hybrid feemd-mlp, feemd-anfis, wavelet packet-mlp and wavelet packet-anfis for wind speed predictions, Energy Convers. Manag. 89 (2015) 1–11.
[18] Y. Ren, P. Suganthan, N. Srikanth, A comparative study of empirical mode decomposition-based short-term wind speed forecasting methods, IEEE Trans. Sustain. Energy 6 (1) (2015) 236–244.
[19] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.-C. Yen, C.C. Tung, H.H. Liu, The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis, in: Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, vol. 454, The Royal Society, 1998, pp. 903–995.
[20] G.P. Zhang, Time series forecasting using a hybrid arima and neural network model, Neurocomputing 50 (2003) 159–175.
[21] M. Khashei, M. Bijari, An artificial neural network (p, d,q) model for timeseries forecasting, Expert Syst. Appl. 37 (1) (2010) 479–489.
[22] G.P. Zhang, B.E. Patuwo, M.Y. Hu, A simulation study of artificial neural networks for nonlinear time-series forecasting, Comput. Oper. Res. 28 (4) (2001) 381–396.
[23] Z. Guo, W. Zhao, H. Lu, J. Wang, Multi-step forecasting for wind speed using a modified emd-based artificial neural network model, Renew. Energy 37 (1) (2012) 241–249.
[24] M.T. Hagan, M.B. Menhaj, Training feedforward networks with the marquardt algorithm, IEEE Trans. Neural Netw. 5 (6) (1994) 989–993.
[25] V. Vapnik, The Nature of Statistical Learning Theory, Springer Science & Business Media, 2000.
[26] K.-R. Müller, A.J. Smola, G. Rätsch, B. Schölkopf, J. Kohlmorgen, V. Vapnik, Predicting time series with support vector machines, in: Artificial Neural NetworksICANN'97, Springer, 1997, pp. 999–1004.
[27] N.I. Sapankevych, R. Sankar, Time series prediction using support vector machines: a survey, Comput. Intell. Mag. IEEE 4 (2) (2009) 24–38.
[28] A. Smola, N. Murata, B. Schölkopf, K.-R. Müller, Asymptotically optimal choice of ε-loss for support vector machines, in: ICANN 98, Springer, 1998, pp. 105–110.
[29] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Stat. Comput. 14 (3) (2004) 199–222.
[30] H. Liu, H.-q. Tian, Y.-f. Li, Comparison of two new arima-ann and arima-kalman hybrid methods for wind speed prediction, Appl. Energy 98 (2012) 415–424.
[31] W. Zhang, J. Wang, J. Wang, Z. Zhao, M. Tian, Short-term wind speed forecasting based on a hybrid model, Appl. Soft Comput. 13 (7) (2013) 3225–3233.
[32] D. Liu, D. Niu, H. Wang, L. Fan, Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm, Renew. Energy 62 (2014) 592–597.
[33] G. Schwarz, et al., Estimating the dimension of a model, Ann. Stat. 6 (2) (1978) 461–464.
[34] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. 2 (5) (1989) 359–366.
[35] N. Wanas, G. Auda, M.S. Kamel, F. Karray, On the optimal number of hidden nodes in a neural network, in: Electrical and Computer Engineering, 1998, IEEE Canadian Conference on, vol. 2, IEEE, 1998, pp. 918–921.
[36] V. Cherkassky, Y. Ma, Practical selection of svm parameters and noise estimation for svm regression, Neural Netw. 17 (1) (2004) 113–126.
[37] J. Shi, J. Guo, S. Zheng, Evaluation of hybrid forecasting approaches for wind speed and power generation time series, Renew. Sustain. Energy Rev. 16 (5) (2012) 3471–3480.