Renewable Energy 78 (2015) 599–608

Contents lists available at ScienceDirect

Renewable Energy

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

Short-term wind speed forecasting based on spectral clustering and optimised echo state networks

Da Liu*, Jilong Wang, Hui Wang

School of Economics and Management, North China Electric Power University, Beijing 102206, China

ARTICLE INFO

Article history:
Received 16 August 2014
Accepted 10 January 2015
Available online 11 February 2015

Keywords:
Short-term wind speed forecasting
Spectral clustering
Echo state network
Wavelet transformation
Genetic algorithm

ABSTRACT

Predicting the wind speed at multiple time points over a time span between two and 4 h typically requires a multi-input/multi-output model. This study investigates a wind speed forecasting method based on spectral clustering (SC) and echo state networks (ESNs). A wavelet transformation was used to decompose the wind speed into multiple series to eliminate irregular fluctuation. The decomposed series were modelled separately. For every decomposed wind speed series, principal component analysis was used to reduce the number of variables and thus the redundant information among the input variables. SC was used to select similar samples from the historical data to form training and validation sets. An ESN was used to simultaneously predict multiple outputs, and a genetic algorithm was employed to optimise the ESN parameters and ensure the forecast accuracy and the generalisation of the model. The forecasts of the decomposed series were summed to get the wind speed. Tests based on actual data show that the proposed model can simultaneously forecast wind speeds at multiple time points with high efficiency, and the accuracy of the proposed model is significantly higher than that of the traditional models.

1. Introduction

Wind power is a clean, renewable energy source with great potential, and countries around the world have increasingly focused on wind power in recent years. Accurately forecasting the wind power or the wind speed of a wind farm can improve the integration of wind power into the grid and ensure the stability of the electric power system when wind power is included. Accurately forecasting the wind speed over periods of 2–4 h in advance is of great importance for integrating wind power into the grid and maintaining the stability of the power system. Countries such as China and India have proposed standards for short-term wind speed forecasting. For example, the National Energy Administration of China requires that wind farms report wind power forecasts up to 4 h in advance in 15-min intervals, and the forecast error cannot exceed 15%.

Wind speed is easily affected by weather and topographic factors, such as temperature, humidity, air pressure gradient, and evaporation, which means that wind speed can exhibit large variation. In addition, wind speed data contain random fluctuations due to measurement errors and random factors, making accurate wind speed forecasts difficult [1]. The wavelet transform (WT) is often used to reduce the noise contained in wind speed data before modelling is attempted, and this method produces fairly good results [2–4]. The wind speed has a relatively high degree of autocorrelation, and the majority of previous studies relied on historical wind speed data to build models. When using highly correlated historical wind speed data as the input to a model, there is a large amount of redundant information that can make modelling difficult. Many researchers have used dimension reduction techniques such as principal component analysis (PCA) to eliminate redundant information and improve the accuracy of the forecast [5,6].

The wind speed exhibits extremely complex variations, and it is therefore difficult to use a single statistical model such as classic time series analysis (e.g., auto-regressive and moving average model) and soft computing (e.g., artificial neural networks), to capture all wind speed variation patterns [1]. Traditional wind speed forecasting methods typically use all wind speed data as a single sample space for modelling and do not distinguish the data based on their characteristics, making the rationality of these methods questionable. If samples that are similar in terms of fluctuation patterns are chosen to build a model for wind speed forecasting, then the training requirements for the model can be greatly reduced and the forecasting accuracy of the model improved.

* Corresponding author.
E-mail addresses: liuda315@163.com (D. Liu), 65492201@qq.com (J. Wang), wanghuibd@163.com (H. Wang).

http://dx.doi.org/10.1016/j.renene.2015.01.022
0960-1481/

Several studies have proposed the use of the cluster analysis method, which is based on similarity theory, to forecast power system loads and the wind speed [7,8]. Traditional clustering methods such as the K-means method may converge to a local minimum, which affects their applicability to forecasting short-term wind speeds. Spectral clustering (SC) is a new type of clustering method based on spectrum division theory. Compared to traditional clustering algorithms such as K-means, SC has advantages such as low computational complexity, applicability to sample spaces of any shape, the ability to perform cluster analysis on a non-convex distribution, and convergence to a global optimal solution. SC is now widely used in the fields of pattern recognition and forecasting [9,10]. SC has been one of the most popular research topics in the field of machine learning. However, this method has not been used for sample categorisation in the field of short-term wind speed forecasting.

Methods such as the time series and intelligent forecasting have been widely used in developing wind speed forecast models [4,11,12,and13]. The time series forecasting method uses historical time series data to construct forecasting models. However, the existing low-order models are not highly accurate, and the high-order models have difficulty in estimating parameters and cannot maintain accuracy in multi-step forecasts. Because short-term wind speed forecasting requires simultaneous wind speed forecasts at multiple time points, a traditional time series method cannot be applied effectively. Artificial neural networks (ANNs) are widely used as an intelligent forecasting method. This method can produce nonlinear maps and perform parallel processing. Neural networks are therefore suitable for multi-input/multi-output learning and can simultaneously forecast wind speeds at multiple time points [1]. However, ANNs require a complex training process, they may converge to a local minimum, they have difficulty in determining the optimum network structure, and they experience fading memory (FM). Therefore, it is difficult to create a more accurate wind speed model using ANNs. Echo state networks (ESNs) are a novel type of ANN proposed by Jaeger in 2001 and are regarded as the closest representation of the learning process of the human brain [14]. ESNs can effectively solve all of the aforementioned problems with ANNs. The ESN model requires a simple training process and has short-term memory. Thus, this model has found widespread use and success in the field of time series forecasting for applications such as stock prices and power system loads [2,15–18]. An ESN can process multiple inputs and multiple outputs with such high efficiency that it is capable of meeting the requirements for real-time, multi-step, short-term forecasting of wind speeds. This study attempts to combine the WT and PCA for wind speed data processing and then use SC to select a proper sample set with which to construct an ESN model that can predict wind speeds over short periods. In addition, a genetic algorithm (GA) is used to optimise the ESN parameters.

The structure of this paper is as follows. Section 2 explains the basic modelling theory and introduces the principles and advantages of WT, PCA, GA, SC, and the ESN. Section 3 explains the development of the SC-ESN-GA model, and Section 4 describes the numerical examples used to verify the effectiveness of the methods proposed in this study. Section 5 discusses the results and presents conclusions.

2. Modelling theory

2.1. WT, PCA, and GA

The WT is introduced to eliminate noise within the wind speed data, which typically contain a high level of noise. Furthermore, because of the large number of inputs to the wind speed model, PCA is introduced to reduce the dimension of the input data. Through scaling and shifting, a WT performs multi-scale, multi-resolution analysis on signals. This method analyses a signal using a fixed-width time window, although the scale is adjustable. Therefore, a WT can effectively analyse transient and non-stationary signals. Selecting a wavelet function and determining the decomposition layers are vitally important when using a WT to process signals.

The basic concept behind PCA is to transform a given set of correlated variables into a set of uncorrelated variables through a linear transformation. The new variables are chosen based on their variances, which are arranged in descending order, to thus use the fewest number of variables possible to represent the information contained in the original variables.

GAs are heuristic, parallel, global search methods based on the principles of natural selection and genetics. Typically, GAs are used to solve multi-parameter optimisation problems and consist of steps such as parameter initialisation, member selection, and chromosome crossover and mutation. It is critical that appropriate fitness functions be chosen when applying a GA to solve a parameter optimisation problem [12].

2.2. SC

SC transforms a clustering problem into an optimal spectral partitioning problem. Based on an eigenvector matrix, the SC method performs spectral partitioning according to the weights among the vertices. The basic concept behind SC is to cluster individual data points from data sets $\mathbf{D}=\{d_1,d_2,...d_n\}$ at the vertices of an undirected, weighted graph $G(\mathbf{D},\mathbf{W})$. The distances between data points d_i and d_j are then converted into a weighted graph edge w_{ij}. In this fashion, the data clustering problem is transformed into a spectral partitioning problem.

The partitioning strategy greatly affects the results given by SC clustering. Classic partitioning strategies include the normalised-cuts algorithm [19] and the Ng-Jordan-Weiss (NJW) algorithm [20]. Implementations of these algorithms vary slightly, although all are essentially continuous relaxations of the spectral partitioning problem. In short, SC can be summarised by the following three steps:

Step 1: Construct a similarity matrix **Z** that represents the sample data.
Step 2: Calculate the first k eigenvalues and eigenvectors and construct the eigenvector space.
Step 3: Cluster the eigenvectors within the eigenvector space by applying a clustering algorithm such as k-means.

2.3. ESNs

A standard ESN is composed of three layers: an input layer, a hidden layer, and an output layer. The network structure is shown in Fig. 1 [21]. The ESN forms an “input-state-output” system. As indicated in Fig. 1, the structure of an ESN is composed of a large number of randomly generated, sparsely connected neurons that form a large dynamic reservoir (DR). This DR is a core component of an ESN. The DR is excited by the input signal and generates a continuous state variable signal (i.e., a network “echo”). Through a linear combination of the echo signal and the target output signal, the ESN output weights can be determined to forecast a chaotic time series.

At time$(n+1)$, $\mathbf{U}(n+1)$ is the input vector, the hidden layer state is $\mathbf{X}(n+1)$, and the network output $\mathbf{Y}(n+1)$ can be obtained using Equations (1) and (2):

$$\mathbf{X}(n+1) = F(\mathbf{W}_{\text{in}}\mathbf{U}(n+1) + \mathbf{W}\mathbf{X}(n) + \mathbf{W}_{\text{back}}\mathbf{Y}(n)) \quad (1)$$

$$\mathbf{Y}(n+1) = G\left(\mathbf{W}_{\text{out}}\left[\mathbf{U}^{\mathbf{T}}(n+1), \mathbf{X}^{\mathbf{T}}(n+1)\right]^{\mathbf{T}}\right) \quad (2)$$

In Equations (1) and (2), $\mathbf{X}(n)$ and $\mathbf{Y}(n)$ are the state and output vectors, respectively, at time n, F and G are the neuron activation functions of the hidden layer and the output layer, respectively, $\mathbf{W}_{\text{in}}$ is the weighting matrix for the input layer connections, $\mathbf{W}$ is the weighting matrix for the connections within the DR, $\mathbf{W}_{\text{back}}$ is the feedback weighting matrix for the connections from the output layer to the hidden layer, and $\mathbf{W}_{\text{out}}$ is the weighting matrix for the output layer connections. The weighting matrices$\mathbf{W}_{\text{in}}$,$\mathbf{W}$ and $\mathbf{W}_{\text{back}}$ are randomly generated before the network is trained and do not change. The weighting matrix $\mathbf{W}$ ensures that the number of connections is sparse (typically 1%–5%), and the spectral radius is less than 1. The value of the DR spectral radius ensures asymptotic stability and transient dynamic memory. The ESN can be used to approximate nonlinear systems. Equations (1) and (2) show that the DR state at the current time is always correlated with the states at previous times, which reflects the unique characteristics of the ESN memory and also makes the ESN model very suitable for performing short-term wind speed forecasts.

The ESN is to be trained to determine the output weights $\mathbf{W}_{\text{out}}$. The output weights are the only parameters that must be determined from training and can be obtained from the solution of an optimisation problem. From Equation (2), an expression for $\mathbf{W}_{\text{out}}$ can be derived using the linear regression algorithm:

$$\mathbf{W}_{\text{out}} = \left(\mathbf{X}^{-1}\mathbf{Y}\right)^{\mathbf{T}} \quad (3)$$

The DR is randomly generated in the classic ESN. Generally, the size of the DR and the spectral radius are determined based on the specific problem. In this study, a GA is used to optimise the parameters.

3. SC-ESN-GA model

3.1. SC-ESN-GA modelling process

The short-term wind speed forecasting models incorporating SC and ESN (hereafter referred to as the SC-ESN-GA model) are constructed as follows. Prior to constructing the models, a WT is used to decompose the data to obtain an approximation series and several detail series for modelling respectively. The following steps are then performed to model these series individually. First, PCA is used to extract the principle components of the original input variables such as wind speed and temperature data. The SC method is then used to sort the data according to the degree of similarity among the samples. The samples are divided into sample families, where samples in each family had a high degree of similarity. An ESN is then formed for training on the sample family of the same type as those at the chosen forecast time points. A GA is used to optimise the ESN parameters. The optimised ESN is then used for forecasting. The final forecast is obtained by summing the forecasts for the various series obtained from the WT decomposition.

Fig. 2 shows the process for generating the SC-ESN-GA forecast model.

In Fig. 2, $\mathbf{x}$ and $\mathbf{T}$ represent the historical wind speeds and the temperatures, respectively, at various time points over the previous hours, $\mathbf{M}$ represents data such as the character matrix (maximum, minimum, and mean) of the previous wind speeds and the temperatures.

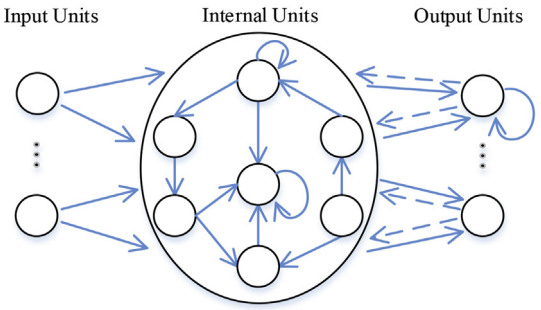

Fig. 1. ESN structure.

3.2. WT and data normalisation

The wind speed series used for forecasting are typically discrete data, so a discrete wavelet transform (DWT) is usually used. The Haar wavelet, the Daubechies wavelet, and the Symlets wavelet are the most commonly used wavelet functions. As a family of orthogonal wavelets, the Daubechies wavelets are widely used in solving a broad range of problems. Here, we process wind speed series with the Daubechies wavelet. The number of decomposition layers from the wavelet is normally determined by the characteristics of the wind speed series.

The sample data are normalised before the ESN is trained to ensure that the inputs to the model are not subject to differences in scale, which can adversely affect the training of the ESN. We map the inputs and the outputs, such as the temperature and the decomposed wind speed data, onto the interval [−1, 1] using Equation (4). After training, the inverse mapping is used to transform the output back to the original data range.

$$x_i^* = \frac{2 \times (x_i - x_{\min})}{x_{\max} - x_{\min}} - 1 \quad (4)$$

In Equation (4), x_i and x_i^* represent the values before and after data normalisation, respectively, and $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values in the sample data, respectively.

3.3. Input variable selection and dimensionality reduction

The selection of input variables should be based on an analysis of their mutually influence and their comprehensive influence on the output. The forecasted wind speeds will be highly correlated with historical wind speed and temperature data [12], and these data can therefore be used as input variables for the model. Short-term wind speed modelling requires simultaneous forecasts of the wind speed at multiple time points, so the historical wind speeds, the maximum, minimum, and average wind speeds, and the temperature from the same time period are selected as original input variables for the model and are used for clustering by SC.

The ESN is a multi-input/multi-output model. The original input to the model and the output function relationship are used to perform wind speed forecasting at various time points for the next n hours, as shown in Equation (5):

$$\mathbf{y} = f(\mathbf{x}, \mathbf{T}, \mathbf{M}) \quad (5)$$

In Equation (5), $\mathbf{x}$ and $\mathbf{T}$ represent the historical wind speeds and the temperatures, respectively, at various time points over the previous n hours, $\mathbf{M}$ represents data such as the maximum, minimum, and mean wind speeds and the temperatures at various time points over the previous n hours, and $\mathbf{y}$ represents the actual wind speeds at various time points for the next n hours.

After determining the input variables, the original input matrix can be generated using **x**,**T**, and **M**, and the original output matrix is composed of **y**.

The new input matrix can be obtained by applying PCA to reduce the dimensions of the original input variable matrix, eliminating redundant information.

3.4. Sample space clustering

The nature of an intelligent algorithm is that it allows the model to learn via a large amount of samples and to approximate the implicit relationship between the input and output variables based on the samples. The behaviour of the wind speed is complex, and different samples could be represented by different models. If only one model is constructed to fit all wind speed samples, then it is extremely difficult to establish a proper relationship between the input and output variables; therefore, the generalization of the model cannot be assured. Merging samples with a high degree of similarity will greatly reduce the difficulty in learning, which will improve the efficiency and the forecast accuracy of the model.

Fig. 3 shows the steps required to perform SC on input samples after dimension reduction has been performed by PCA on the various series obtained from the WT decomposition.

In Fig. 3, the check on the numbers of the sample space following clustering aims to ensure that the models on various series will all have adequate samples to fully capture the characteristics of the fluctuations. Following dimension reduction, the input matrix obtained from SC and the output are used to construct the sample space for ESN learning.

3.5. ESN modelling and GA optimisation

The series obtained from decomposition are divided into training and validation sets for ESN modelling. The ESN training and validation processes consist of the following four steps:

Step 1: Initialise the ESN by randomly generating the weighting matrices, $\mathbf{W}_{\text{in}}$, $\mathbf{W}$, $\mathbf{W}_{\text{back}}$, and initialise the hidden layer state $\mathbf{X}(n)$.

Step 2: Based on Equation (1), use the training data set to update the state of the DR, i.e., compute $\mathbf{X}(n+1)$.

Step 3: Based on Equation (3), calculate the output weighting matrix $\mathbf{W}_{\text{out}}$, examine the model on the validation set using Equation (2), and compute the normalised root mean square error (NRMSE, E_{NRMSE}) of the validation set, as given by Equation (6).

$$E_{\text{NRMSE}} = \sqrt{\frac{\frac{1}{k}\sum_{i=1}^{k}(\hat{y}(i)-y(i))^2}{\frac{1}{k}\sum_{i=1}^{k}(y(i)-\bar{y})^2}} \tag{6}$$

In Equation (6), k is the forecast data series length for each model, $y(i)$ is the original wind speed, $\hat{y}(i)$ is the forecast value of the wind speed, and $\bar{y}$ is the average value of the original wind speed over the forecast horizon.

Step 4: Optimise the ESN using GA with E_{NRMSE} of the validation set. Construct the optimum ESN forecast model. Apply Equation (2) to forecast the wind speeds for the next n hours at the chosen points.

In Step 4, a GA is used to optimise the parameters of the decomposition series such as the DR size and the spectral radius.

Fig. 4 shows the ESN model construction process.

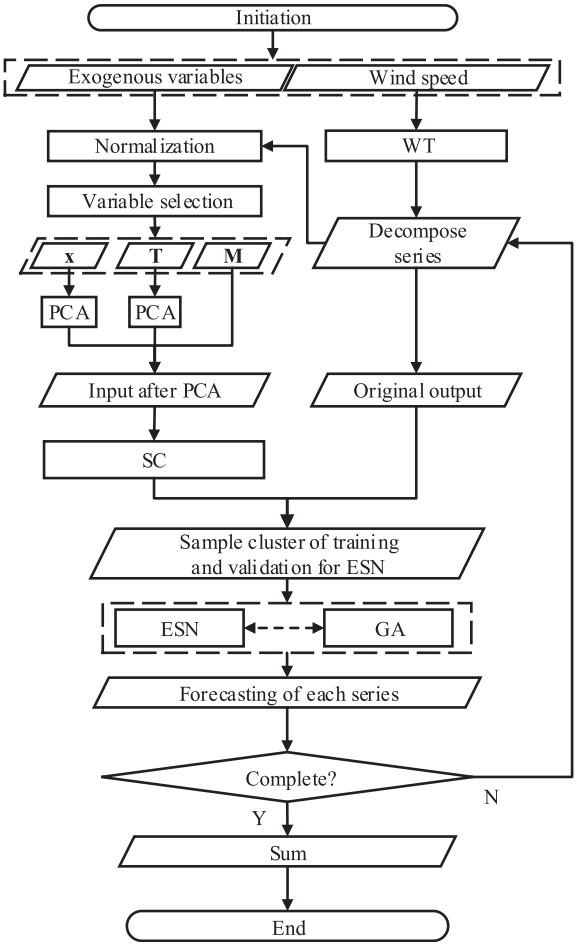

Fig. 2. Flowchart of the SC-ESN-GA model.

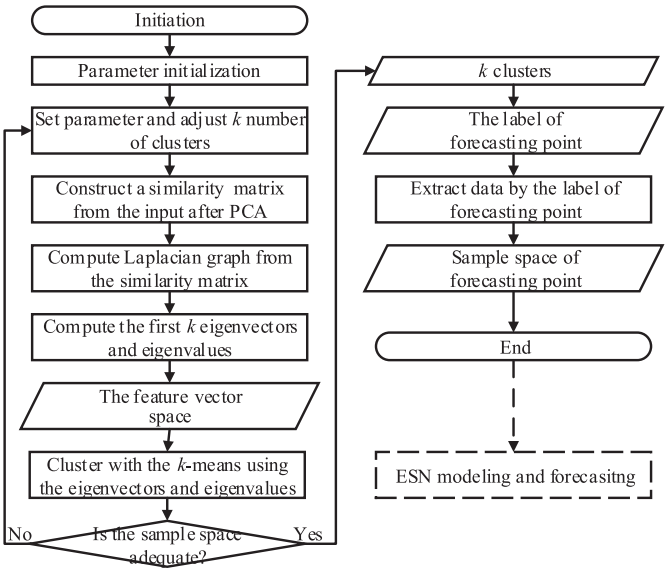

Fig. 3. Cluster the sample space with SC algorithm.

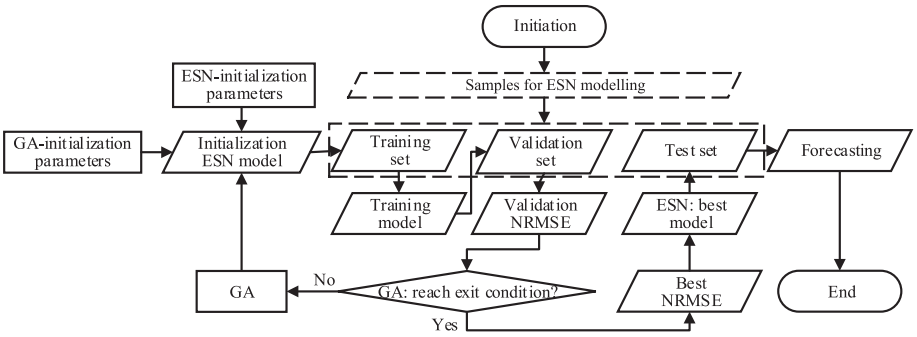

Fig. 4. Flowchart for GA optimisation of the ESN parameters.

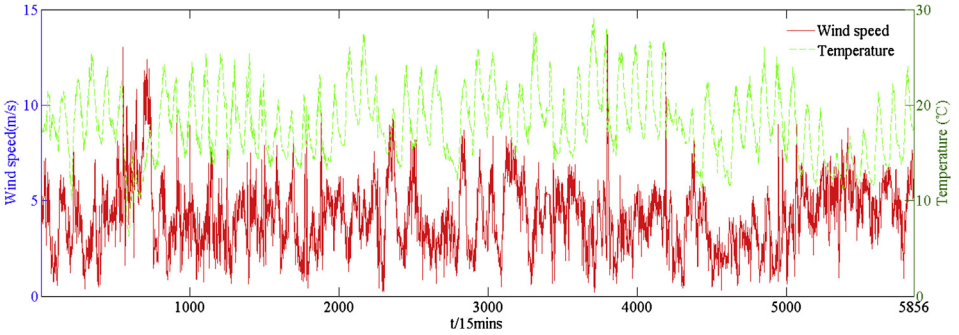

Fig. 5. Actual wind speed and temperature in July and August.

3.6. Forecasting model evaluation

The forecasts for the decomposition series are computed using the optimum ESN models for these series. The forecasts from the various decomposition series can be summed to reconstruct the final forecast value of the wind speed:

$$\hat{\mathbf{y}} = \boldsymbol{a}_j^* + \sum_{i=1}^{j} \boldsymbol{d}_i^* \tag{7}$$

In Equation (7), j is the number of decomposition layers, $\boldsymbol{a}_j^*$ are the forecast values for the approximation series, $\boldsymbol{d}_i^*$ are the forecast values for various detail series, and $\hat{\mathbf{y}}$ is the final forecast value.

The mean absolute percent error (E_{MAPE}), the mean absolute error (E_{MAE}), and E_{NRMSE} can be used to evaluate the accuracy of the forecasting model. The equations for computing E_{MAPE} and E_{MAE} are as follows:

$$E_{\mathrm{MAPE}} = \frac{1}{k}\sum_{i=1}^{k}\left(\frac{|\hat{y}(i) - y(i)|}{y(i)} \times 100\%\right) \tag{8}$$

$$E_{\mathrm{MAE}} = \frac{1}{k}\sum_{i=1}^{k}|\hat{y}(i) - y(i)| \tag{9}$$

4. Experiments study

4.1. Data and WT process

Data from a wind farm in northern China for the months of July and August 2011 are used as an example. The data consist of wind speeds and temperatures sampled at 15-min intervals. The model is constructed using data from July 11st to August 20th for both the

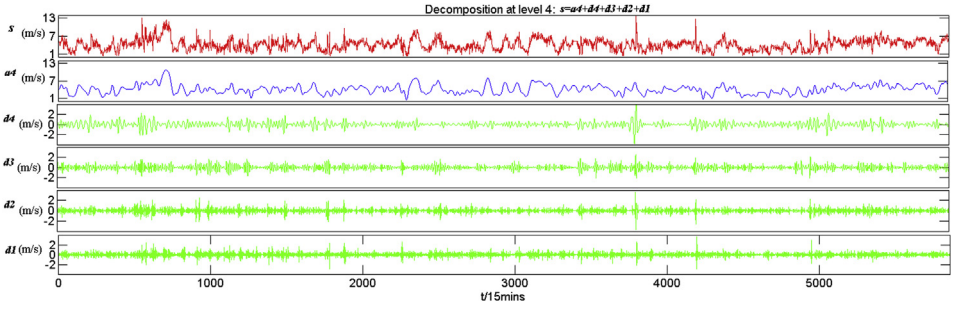

Fig. 6. Original speed and its decomposed series.

Table 1
Dimensions of the original input and output variables.

Input-output matrix	Notation	Dimension
x (historical wind speed)	$x_t,x_{t-1},\ldots,x_{t-15}$	16
T(historical temperature)	$T_t,T_{t-1},\ldots,T_{t-15}$	16
M (character of speed and temperature)	$\max(x_t,x_{t-1},\ldots,x_{t-15}),\min(x_t,x_{t-1},\ldots,x_{t-15})$ $\max(T_t,.T_{t-1},\ldots,T_{t-15}),\min(T_t,T_{t-1},\ldots,T_{t-15})$ $mean(x_t,x_{t-1},\ldots,x_{t-15}),mean(T_t,T_{t-1},\ldots,T_{t-15})$	6
y (future wind speed)	$x_{t+1},x_{t+2},\ldots,x_{t+16}$	16

Table 2
Component matrix.

Component	PC1	PC2	PC3	PC4	…
1	−0.7902	−0.3490	0.4119	…	
2	−0.8883	−0.1369	0.4158	…	
3	−0.9240	0.1011	0.3675	…	
…	…	…	…	…	
14	0.9242	0.0999	0.3674	…	
15	0.8882	−0.1379	0.4156	…	
16	0.7899	−0.3500	0.4118	…	

training and validation sets. Fig. 5 shows the original wind speed and temperature data. The maximum and minimum wind speeds are 13.68 m/s and 0.21 m/s, respectively, and the average wind speed is 4.04 m/s, clearly showing the extremely large variations in wind speed.

Generally, the peak load on the power system occurs every day between 8:00 and 14:00, and the wind speed peaks around 18:00. The load is at a minimum between 22:00 and 2:00. To validate the forecasting model, we choose the following five important periods for an electric power system that with wind power accommodating, i.e., peak load period (starting at 8:00 and 14:00), minimum load period (starting at 22:00 and 2:00) and peak wind speed period (starting at 18:00). For each forecasting period, the forecast time horizon is 4 h, and predictions are made in intervals of 15 min. In other words, 16 forthcoming wind speed values are forecasted for the 4-h horizon. For each of the five forecast periods, 16 wind speed values are simultaneously forecasted, a total of 80 forecasted values for five forecasting periods. In addition, to verify the generalization of the model, we also compute forecasts at 80 points per day (a total of 400 points) from August 22nd to August 26th, and take a further experiment study from UK.

The 6th-order Daubechies wavelet with 4 decomposition layers is used for the wavelet transform. The original wind speed series data are decomposed into one approximation series (***a*4**) and four detail series(***d*1, *d*2, *d*3, *d*4**). The original speed in July and August and its decomposed series are demonstrated in Fig. 6.

It can be seen from Fig. 6 that the approximation series ***a*4** is a low-frequency signal, and its variations coincided with the original wind speed series. Detail series (***d*1, *d*2, *d*3**) are all high-frequency signals with relatively small amplitudes. These series primarily reflected the irregular fluctuations in the wind speed (i.e., noise) and are eliminated from the model construction process. Forecasts are computed using only ***a*4** and ***d*4**. The forecasts given by ***a*4** and ***d*4** are summed to form the final forecast.

4.2. Parameter initialisation

The threshold on the variance for the accumulated contributions from the PCA components is 0.99.

The number of nearest neighbours in the SC is set at 50, the block size for partitioning the data matrix is set at 10, the sigma value used in computing similarity is set at 20, the initial number of clusters is set at 10, the normalised-cuts algorithm is applied, and the minimum sample space is set at 200.

The connection sparsity weight for the ESN is set at 0.05. The hyperbolic tangent function is selected for the neuron activation function in the hidden layer, and the linear activation function (identity function) is used for the output layer neurons. The DR size and the spectral radius are determined using the GA.

The population size for the GA is 20, the generation termination is 100, the crossover rate is 0.8, the mutation rate is 0.05, and the fitness function is E_{NRMSE}, which is computed from the ESN validation sample set.

4.3. Selection of initial input variables

Wind speeds are forecasted for 16 points in a 4-h time span, i.e., at 15-min intervals, using the short-term wind speed model. The dimensions of the original input and output variables in Equation (5) are shown in Table 1.

In Table 1, the total dimensionality of the original input variable matrix is 38 and that of the original output variable matrix is 16. A long training time and poor modelling performance would be expected if we use the original input variable matrix directly because of its high dimensionality.

4.4. PCA, SC processing and GA optimization

As discussed in Section 4.3, the input matrix with 16 historical wind speeds and 16 historical temperatures, contain too much redundant information for modelling. We mine the major information containing in the speeds and temperatures using PCA respectively, as we discussed previously with Fig. 2.

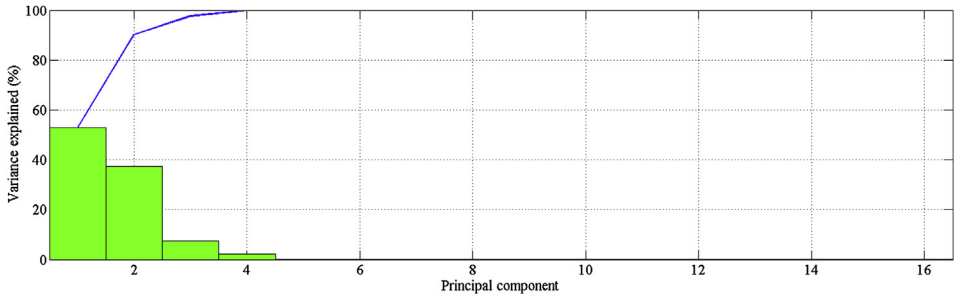

Fig. 7. Scree plot of ***d*4** of wind speed.

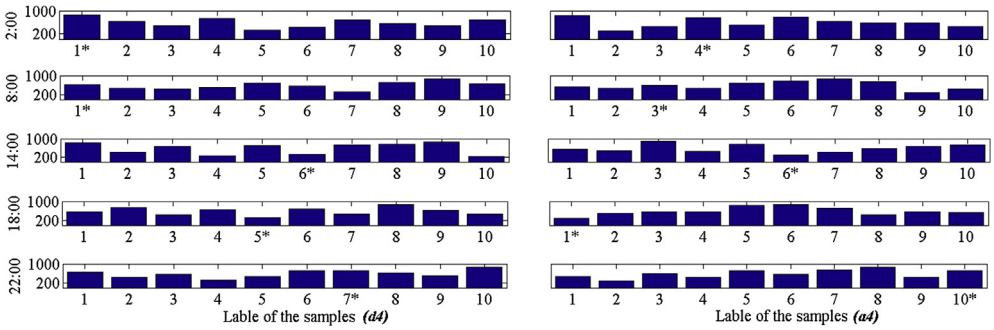

Fig. 8. Number of samples for the decomposed series after SC for each forecast period. Note : "*" represents the cluster to which the forecast points belong.

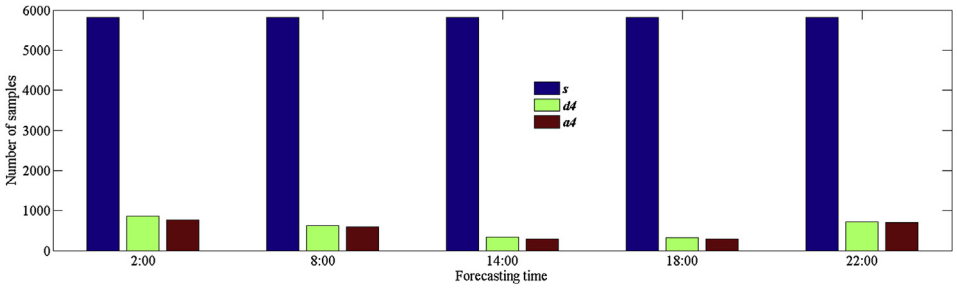

Fig. 9. Number of the samples reducing by SC (August 21st, 2011).

We take the PCA process result of wind speed for ***d***4 modelling as an example. The PCA dramatically reduces the number of the input, as Fig. 7 and Table 2 shows. It can be seen from Fig. 7 and Table 2 that the first three principal components explain nearly 100% of the original 16 wind speeds, so we can use these three principal components to replace the 16 historical speeds as a part of the input.

After reducing the number of input variables using PCA, the dimensions of the samples of all the five forecasting periods are significantly decreased. As for input of ***d***4 modelling, the number of wind speed data and temperature data are both reduced from 16 to 3 , so the total dimensions of ***d***4 are reduced from the 38 to 12. As for ***a4***, the number of wind speed and temperature in the model input matrix are both reduced to 4, so the total dimensions of are reduced to 14.

Next, SC is applied to group the samples. Samples from the historical sample space with similar characteristics at the various forecast times are selected for model training and validation, which greatly reduced the number of samples in the training and validation sets.

Fig. 8 shows the number of samples of each cluster for the five forecast periods for ***a4*** and ***d***4 after SC processing. It can be observed that SC greatly reduced the sample space of the model, and the number of ESN samples in the decomposition series decreases from 5840 (16 samples are deleted from the 5856 samples for modelling lagging period) to 400–1000 after SC. Therefore, the goals of simplifying the sample space and increasing the effectiveness of the training are achieved.

The efficiency of clustering similar samples by SC is presented by the case of August 21, as Fig. 9 shows. The original data for model is 5840 (the left bar), and the number of ***d4*** (the middle bar), ***a4*** (the right bar) for modelling after SC are less than 1000. Those samples with less similarity to the target sample are deleted from the sample set by SC, which greatly contributes to the efficiency and effectiveness of the modelling.

The optimization of DR size and the spectral radius implemented by GA is shown in Fig. 10 (with the forecast case at 22:00, August 21.)

It can be seen from Fig. 10 that the mean fitness of GA, E_{NRMSE}, drops dramatically within 10 generations, after that it leaps around

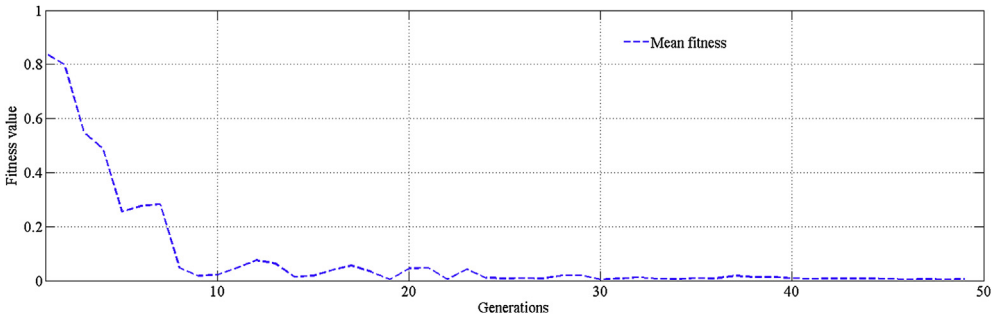

Fig. 10. The iterations of GA to find best parameter.

Table 3
Forecasting results from various models.

Forecast starting time	ARIMA			GRNN			ESN			SC-ESN-GA		
	MAPE (%)	MAE (m/s)	NRMSE	MAPE (%)	MAE (m/s)	NRMSE	MAPE (%)	MAE (m/s)	NRMSE	MAPE (%)	MAE (m/s)	NRMSE
2:00	14.28	0.45	1.35	13.24	0.50	5.71	16.12	0.47	1.07	13.68	0.42	0.89
8:00	24.24	0.67	2.32	12.11	0.40	1.79	13.78	0.51	8.34	13.52	0.37	2.78
14:00	36.58	1.84	2.55	17.73	0.63	1.45	29.81	1.52	1.65	12.56	0.36	0.46
18:00	16.18	0.74	2.11	26.35	1.01	3.31	5.64	0.26	0.53	5.71	0.26	0.93
22:00	13.84	0.65	2.50	20.86	0.81	6.04	33.01	1.04	2.99	13.34	0.59	1.85
Mean	21.02	0.87	2.17	18.06	0.67	3.66	19.67	0.76	2.92	11.76	0.40	1.38

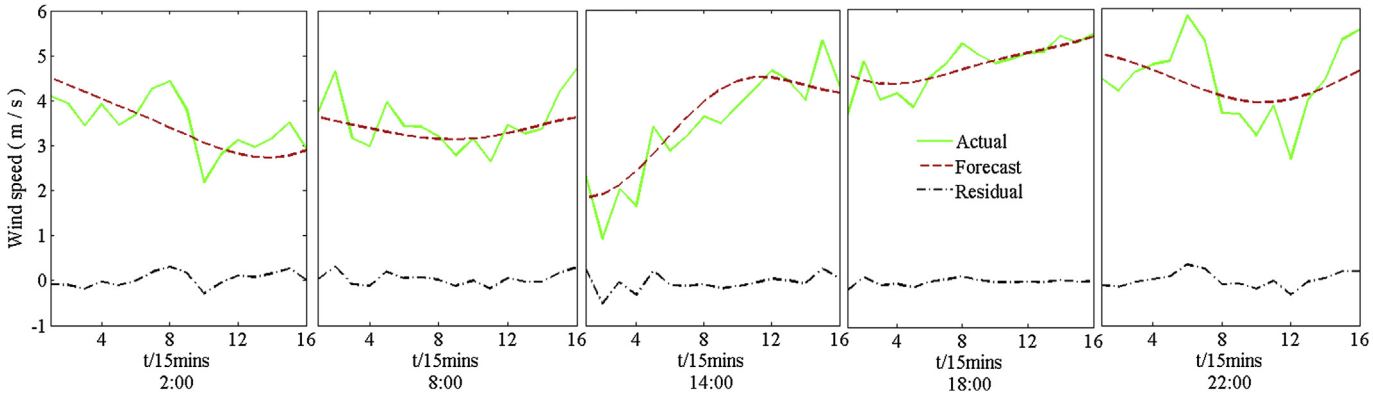

Fig. 11. Wind speed forecast from the SC-ESN-GA model.

Table 4
Forecast error for the SC-ESN-GA model for August 22nd – August 26th.

Forecast date	E_{MAPE} (%)	E_{MAE} (m/s)	E_{NRMSE}
Aug. 22nd	13.23	0.54	2.92
Aug. 23rd	13.79	0.61	1.59
Aug. 24th	23.93	0.51	1.88
Aug. 25th	13.96	0.49	1.14
Aug. 26th	12.27	0.42	1.43

0.07; it approaches to zero after 25 generations, which means that the optimum DR size and the spectral radius are found within 25 generations in GA.

4.5. Forecast evaluation and model comparison

It takes less than 5 min to forecast wind speed of the 16 points in the 4 coming hours with SC-ESN-GA for a laptop (Intel Core i5/4 GB DDR3). In real-time implementation, workstation with parallel computing would be applied and the modelling would be much faster.

To verify the performance of the SC-ESN-GA model, the results are compared with those of an autoregressive integrated moving average model (ARIMA) and a generalised regression neural network (GRNN) model. ARIMA uses the Akaike information criterion (AIC) to determine the model order. The input variables selected for the GRNN are the same as those for the SC-ESN-GA model. The spread in the GRNN is selected through experiment. For convenience in analysing the effect of sample grouping with SC, an ESN model without clustering is included in the comparison. The data for all of the models are processed using the same wavelet decomposition that used for the SC-ESN-GA model.

A comparison of the results generated by the ARIMA, GRNN, ESN, and SC-ESN-GA models is given in Table 3.

Overall, it can be observed from Table 3 that when used for forecasting wind speeds over a 4-h period at 15-min intervals, the accuracy of the SC-ESN-GA model is significantly higher than that of the ARIMA and GRNN models. The SC-ESN-GA

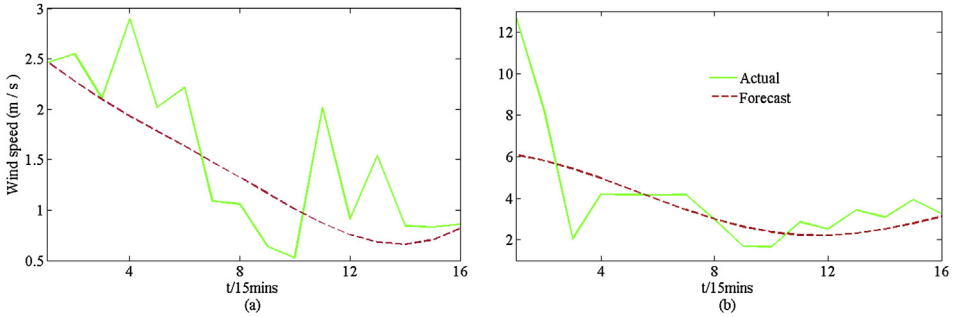

Fig. 12. Wind speed forecasts starting at 22:00 on August 23rd and 24th.

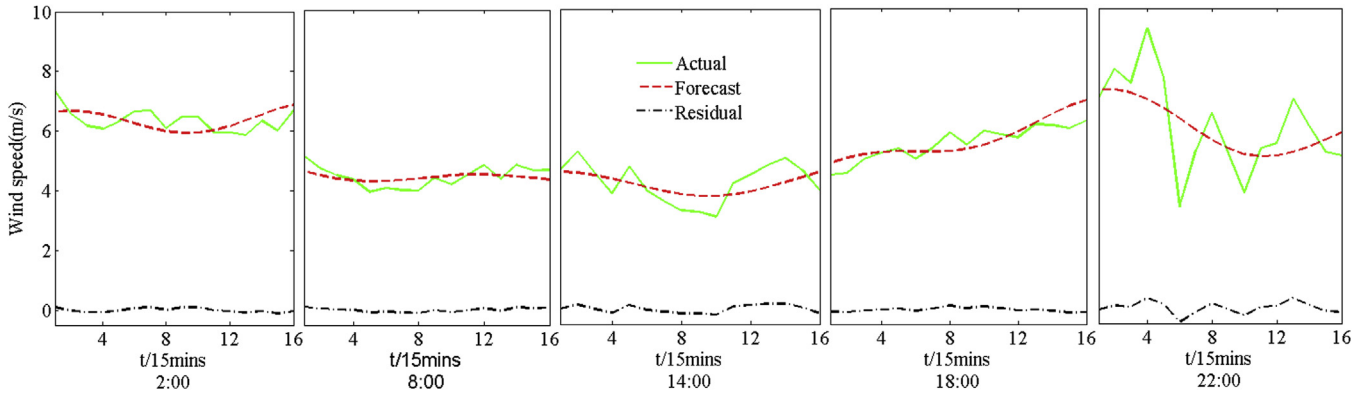

Fig. 13. Wind speed forecast from the SC-ESN-GA mode for CEDA, Nov. 21st, 2010.

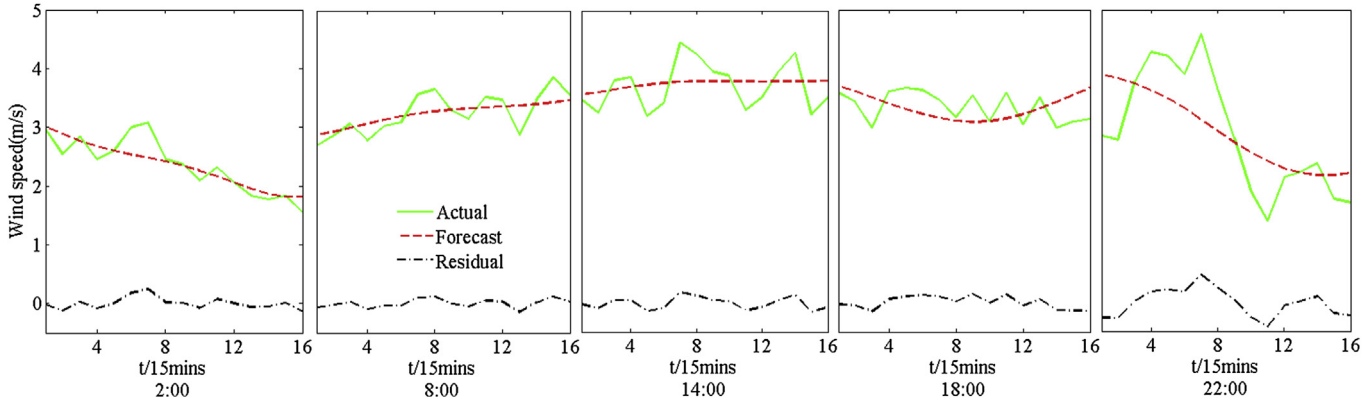

Fig. 14. Wind speed forecast from the SC-ESN-GA mode for CEDA, Feb. 21st 2011.

model is also stable. Because SC is used to form sample families with similar wind speed variations from the historical data, many samples with outlying behaviour are eliminated prior to training. Consequently, the performance and the efficiency of the learning model are ensured, and the forecasts from the SC-ESN-GA model are more accurate than those from the ESN model without clustering. Fig. 11 shows the forecast from the SC-ESN-GA model.

Fig. 11 shows that the forecasted wind speed from the SC-ESN-GA model is consistent with the actual wind speed. The forecast is most accurate when the wind speed variations are more uniform during the forecast period, as is the case for the period starting at 18:00. The performance is slightly worse when the variations in the wind speed are less uniform and large during the forecast period, as is the case for the period starting at 22:00.

To further examine the forecasting capability of the SC-ESN-CA model, we compute forecasts for the same periods using a different set of data, that from August 22nd to August 26th. The values of the forecast error, the E_{MAPE}, the E_{MAE}, and the E_{NRMSE}, are shown in Table 4.

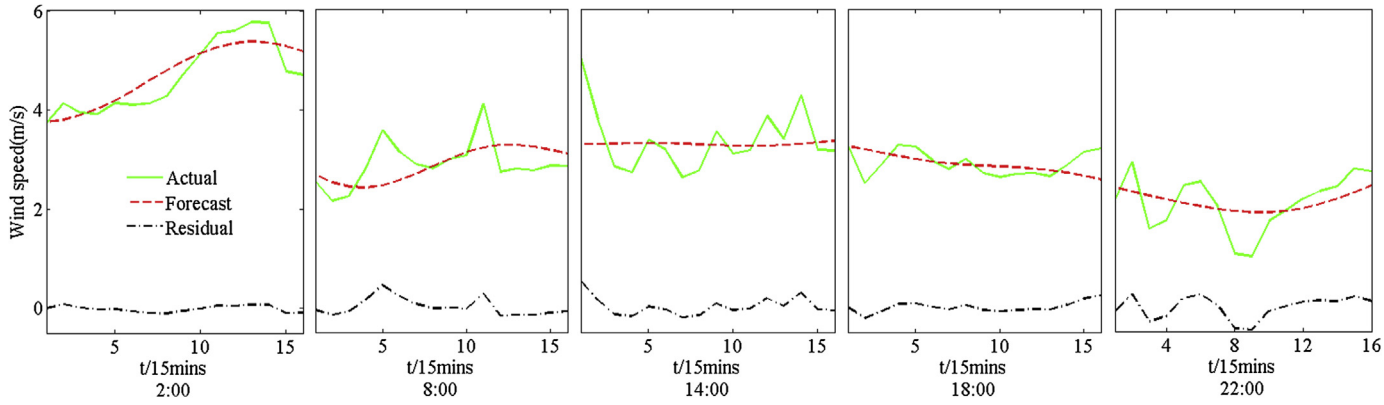

Fig. 15. Wind speed forecast from the SC-ESN-GA mode for CEDA, May 20th, 2011.

Table 5
Forecast error by the SC-ESN-GA model for CEDA.

Forecast date		Forecast time	E_{MAPE} (%)	E_{MAE} (m/s)	E_{NRMSE}
2010	Nov. 21st	2:00	5.78	0.37	1.37
		8:00	5.64	0.25	3.01
		14:00	11.18	0.46	1.73
		18:00	5.48	0.31	0.60
		22:00	15.61	0.94	1.45
		Mean	8.74	0.47	1.63
2011	Feb. 21st	2:00	6.97	0.17	0.62
		8:00	5.95	0.19	1.32
		14:00	8.83	0.33	4.99
		18:00	9.01	0.30	1.63
		22:00	19.59	0.58	1.09
		Mean	10.07	0.32	1.93
	May. 20th	2:00	5.66	0.27	0.53
		8:00	13.30	0.38	1.43
		14:00	13.79	0.45	23.58
		18:00	7.98	0.23	1.59
		22:00	18.55	0.39	2.52
		Mean	11.86	0.34	5.94

It can be observed from Table 4 that the performance of the SC-ESN-GA model is satisfactory. The E_{MAPE} is less than 15%, except on August 24th. The performance of the SC-ESN-GA model is poor when the average wind speed is less than 2 m/s during the forecast period. For example, starting at 22:00 on August 24th, the average wind speed during the forecast period is 1.54 m/s and the E_{MAPE} is 30%; the actual data and the forecast are shown in Fig. 12(a). This poor forecast may have been due to insufficient low-speed wind data (data with wind speeds less than 2 m/s account for only approximately 11% of the total sample data). In addition, the performance is relatively poor when the wind speed exhibited momentary fluctuations, as is the case for the period starting at 22:00 on the 23rd, shown in Fig. 12(b). A possible explanation for this exception is that other factors affecting the wind speed are not captured by the SC-ESN-GA model.

4.6. Further experiment study

Another experiment is carried out with the data from the Centre for Environmental Data Archival (CEDA), UK [22]. We forecast the wind speed at the same starting point (2:00, 8:00, 14:00, 18:00 and 22:00) on November 21st in 2010, February 21st, May 20th in 2011, the data of previous 50 days are employed as training and validation set. The forecasting performance is shown in Figs. 13–15, and the forecasting errors are listed in Table 5.

It can be seen that the proposed method have a desirable capability for most of the forecasting. However it performs worse at 22:00. Another obviously undesirable foresting occurred at 14:00, May 20th, 2011. If we look into the actual speed and the forecasting speed, we can find that the actual speed at the following 4 h after 14:00 are around 3 m/s with little variations, showing extreme difference to the normal variation. So the variation of the actual speed approaches to zero, leading tremendous E_{NRMSE}.

5. Conclusions

This study developed an SC-ESN-GA model for short-term wind speed forecasting. The data were processed using WT to extract the relevant information, PCA to effectively select input variables and to reduce the number of input variables, and SC to group similar samples. An ESN model was constructed to approximate the input–output relationships, and a GA was used to optimise the ESN parameters. Finally, the wind speed was forecasted using the optimised ESN model.

The results of this study led to the following conclusions: 1) PCA reduced the correlation among the input variables, effectively reducing the number of inputs in the model and significantly reducing the complexity of model training; 2) by performing a clustering analysis on the original data via SC and using similar samples to train the model, the learning efficiency and ability of the model to generalise are improved; 3) the flexibility of the multi-input/multi-output ESN and the optimisation capability of the GA for determining the model parameters gave the SC-ESN-GA model exceptional performance for short-term wind speed forecasting; and 4) the performance of the model at low wind speeds is relatively poor due to far fewer samples with low wind speeds being available in the training set, leading to insufficient model learning.

Acknowledgements

The authors are very appreciated for the valuable suggestion and comments from the reviewers. This work was supported by the National Natural Science Foundation of China (70901025), the Social Science Funding Project of Beijing (13JDJGC055) and the Fundamental Research Funds for the Central Universities.

References

[1] Foley AM, Leahy PG, Marvuglia A, McKeogh EJ. Current methods and advances in forecasting of wind power generation. Renew Energy 2012;37(1):1–8.
[2] Niu D, Ji L, Wang YL, Liu D. Echo state network with wavelet in load forecasting. Kybernetes 2012;41(10):1557–70.
[3] Cao JC, Cao SH. Study of forecasting solar irradiance using neural networks with preprocessing sample data by wavelet analysis. Energy 2006;31(15):3435–45.
[4] Catal O JPS, Pousinho HMI, Mendes VMF. Short-term wind power forecasting in Portugal by neural networks and wavelet transform. Renew Energy 2011;36(4):1245–51.
[5] Taylor JW, de Menezes LM, McSharry PE. A comparison of univariate methods for forecasting electricity demand up to a day ahead. Int J Forecast 2006;22(1):1–16.
[6] Skittides C, Früh W. Wind forecasting using principal component analysis. Renew Energy 2014;69(0):365–74.
[7] Niu D, Liu D, Wu DD. A soft computing system for day-ahead electricity price forecasting. Appl Soft Comput 2010;10(3):868–75.
[8] Yadav V, Srinivasan D. A SOM-based hybrid linear-neural model for short-term load forecasting. Neurocomputing 2011;74(17):2874–85.
[9] Huang S. Integrating spectral clustering with wavelet based kernel partial least square regressions for financial modeling and forecasting. Appl Math Comput 2011;217(15):6755–64.
[10] Alzate C, Suykens JAK. Sparse kernel spectral clustering models for large-scale data analysis. Neurocomputing 2011;74(9):1382–90.
[11] Li G, Shi J, Zhou J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models. Renew Energy 2011;36(1):352–9.
[12] Liu D, Niu D, Wang H, Fan L. Short-term wind speed forecasting using wavelet transform and support vector machines optimized by genetic algorithm. Renew Energy 2014;62(0):592–7.
[13] Kavasseri RG, Seetharaman K. Day-ahead wind speed forecasting using f-ARIMA models. Renew Energy 2009;34(5):1388–93.
[14] Jaeger H, Haas H. Harnessing nonlinearity: predicting chaotic systems and saving energy in wireless communication. Science 2004;304(5667):78–80.
[15] Deihimi A, Orang O, Showkati H. Short-term electric load and temperature forecasting using wavelet echo state networks with neural reconstruction. Energy 2013;57(0):382–401.
[16] Güntürkün U. Sequential reconstruction of driving-forces from nonlinear nonstationary dynamics. Phys D Nonlinear Phenom 2010;239(13):1095–107.
[17] Lin X, Yang Z, Song Y. Short-term stock price prediction based on echo state networks. Expert Syst Appl 2009;36(3, Part 2):7313–7.
[18] Ongenae F, Van Looy S, Verstraeten D, Verplancke T, Benoit D, De Turck F, et al. Time series classification for the prediction of dialysis in critically ill patients using echo statenetworks. Eng Appl Artif Intell 2013;26(3):984–96.
[19] Shi J, Malik J. Normalized cuts and image segmentation. Pattern Analysis Mach Intell IEEE Trans 2000;22(8):888–905.
[20] Ng AY, Jordan MI, Weiss Y. On spectral clustering: analysis and an algorithm. Adv Neural Information Process Syst 2002;2:849–56.
[21] Jaeger H. The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. Bonn, Germany: German National Research Center for Information Technology; 2001. p. 34. GMD technical report 148.
[22] [EB/OL] http://www.ceda.ac.uk/browse/badc/ukmo-metdb; [accessed 11.11.14].