Knowledge-Based Systems 24 (2011) 1048–1056

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

A case study on a hybrid wind speed forecasting method using BP neural network

Zhen-hai Guo [a], Jie Wu [b,*], Hai-yan Lu [c], Jian-zhou Wang [b]

[a] State Key Laboratory of Numerical Modeling for Atmospheric Sciences and Geophysical Fluid Dynamics, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China
[b] School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000, China
[c] Faculty of Engineering and Information Technology, University of Technology Sydney, Australia

ARTICLE INFO

Article history:
Received 19 May 2010
Received in revised form 28 April 2011
Accepted 28 April 2011
Available online 4 May 2011

Keywords:
Wind speed forecasting
Kolmogorov–Smirnov test
Year-ahead daily average wind speed forecasting
Seasonal exponential adjustment
Back-propagation neural network

ABSTRACT

Wind energy, which is intermittent by nature, can have a significant impact on power grid security, power system operation, and market economics, especially in areas with a high level of wind power penetration. Wind speed forecasting has been a vital part of wind farm planning and the operational planning of power grids with the aim of reducing greenhouse gas emissions. Improving the accuracy of wind speed forecasting algorithms has significant technological and economic impacts on these activities, and significant research efforts have addressed this aim recently. However, there is no single best forecasting algorithm that can be applied to any wind farm due to the fact that wind speed patterns can be very different between wind farms and are usually influenced by many factors that are location-specific and difficult to control. In this paper, we propose a new hybrid wind speed forecasting method based on a back-propagation (BP) neural network and the idea of eliminating seasonal effects from actual wind speed datasets using seasonal exponential adjustment. This method can forecast the daily average wind speed one year ahead with lower mean absolute errors compared to figures obtained without adjustment, as demonstrated by a case study conducted using a wind speed dataset collected from the Minqin area in China from 2001 to 2006.

1. Introduction

Wind energy is one of the most promising renewable energy sources. Forecasting wind speed is important for wind energy utilisation in general and for wind energy generation in particular. When connected with a power grid, wind energy can have a significant impact on power grid security, power system operation, and market economics due to its intermittent nature, especially in areas with high wind power penetration. As wind speed directly influences wind energy generation, the improvement of the accuracy of wind speed forecasting algorithms has significant technological and economic impacts on wind energy generation and, furthermore, on wind farm planning and the operational planning of power grids associated with the increase of renewable energies.

Significant research efforts have been devoted to developing efficient forecasting methods for the prediction of wind speed [1,4,9,10,24,36] or wind power generation [13,17,23]. The aims of these forecasting methods roughly fall into three main categories [16]: the first category of forecasting methods [6,26] targets wind speed, direction, or simulation and can be used to design wind generation control systems, wind generation protection systems, and mechanical parts of a wind turbine; the second category [2,11] targets the average wind speed and power generation in small time intervals (up to 10 min) to assist power dispatching; and the third category [3,18] targets the average wind speed of an hour on a typical day of a month and the average wind speed of a month to guide medium and/or long-term power generation planning and reserved generation capacity. This study falls into the third category of forecasting and presents a new hybrid forecasting method for predicting the daily average wind speed one year ahead.

Commonly used techniques for average wind speed forecasting include time series analysis or time series analysis combined with an artificial neural network [4,10,24,28,32]. Riahy and Abedi [21] presented a new linear prediction method for short-term wind speed forecasting. They utilised a linear prediction method in conjunction with filtering of the wind speed waveform to forecast wind speed based on the observation that filtering out less effective frequency components from a wind speed spectrum can increase the correlation between real and predicted winds. Sancho et al. [22] suggested a method of exploiting the diversity in input data using banks of artificial neural networks for accurate short-term wind speed prediction, which yields better results compared to those obtained from a system using single neural networks. Mohandes et al. [19] introduced a support vector machines (SVM) algorithm for wind speed prediction and compared its

* Corresponding author. Tel.: +86 931 8914050; fax: +86 931 8912481.
E-mail address: wuj19870903@126.com (J. Wu).

0950-7051/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.knosys.2011.04.019

performance with multilayer perceptron (MLP) neural networks. The results indicated that SVM compares favourably with the MLP model based on root mean square error testing between actual mean daily wind speed data from Madina city, Saudi Arabia and predicted data. Barbounis and Theocharis [3] proposed a locally feedback dynamic fuzzy neural network (LF-DFNN) model using spatial wind speed information from remote measurement stations at wind farms to estimate multi-step ahead wind speed from 15 min to 3 h ahead. Furthermore, they trained the LF-DFNN models using an optimal online learning scheme, the decoupled recursive prediction error algorithm (DRPE). It was shown that DRPE outperformed three gradient descent algorithms: the back-propagation through time, real-time recurrent learning, and recursive back-propagation algorithms, in training of recurrent LF-DFNN forecast models. Mohammad et al. [18] proposed a new strategy in wind speed prediction based on fuzzy logic and artificial neural networks. They trained their new strategy on real wind data measured in Rostamabad in northern Iran from 2002 to 2005. The experimental results demonstrated that the proposed method not only provided significantly less rule base but also increased the estimated wind speed accuracy when compared to traditional fuzzy and neural methods.

Moghram and Rahman [12] reviewed five forecasting methods: (1) Multiple linear regression; (2) time series; (3) general exponential smoothing; (4) state space and Kalman filtering; and (5) knowledge-based approaches. Each method was described briefly and implemented to predict the hourly load of a southeastern utility. Summer and winter loads were modelled separately, and the results were compared in terms of the per cent error. No method was determined to be superior. The transfer function approach was the best predictor over summer months, but it was the second worst predictor over winter months. Thus the authors concluded that because of its strong dependency on historical data, the transfer function approach did not respond as well to abrupt changes as did the knowledge-based approaches. The finding that there was no one best approach suggested that model performance under specific conditions should be analysed and understood, and incremental improvements should be made based on the knowledge gained.

Based on a literature review, we found that although wind speed prediction has been extensively studied, there is no single best forecasting algorithm that can guarantee good accuracy for all forecasting cases due to the fact that wind speed patterns can be very different between wind farms and that wind speed is usually influenced by many factors that are difficult to control and location specific. Hence, new efficient forecasting methods with improved accuracy under specific circumstance are still highly desirable.

This study aims to forecast the daily average wind speed over a long period of time, such as one year ahead. Existing methods for this purpose tend to yield results with poor accuracy because they cannot properly account for seasonal effects over the long term. To improve the accuracy of daily average wind speed forecasting, we present a novel hybrid forecasting method based on addressing original datasets to obtain superior forecasting performance. This method is different from those employed in previous related studies, which mainly focussed on the optimisation of parameters and further analysis of influential factors. It combines the seasonal exponential adjustment (SEA) with a back-propagation (BP) neural network to address seasonal effects over a long term. BP neural networks represent one of the most active methods in the realm of intelligence control, especially in uncovering nonlinearity between inputs and the outputs, even in the absence of sufficient information about the relationship between them [29]. BP networks are extensively employed in back analysis because of their simplicity and power to extract useful information from patterns. They allow specification of multiple input criteria and the generation of multiple output recommendations without prior assumptions regarding the form of functions related to input and output variables. The BP model eliminates the limitations of traditional regression methods and accurately establishes mapping between input and output variables. It can approximate an arbitrary nonlinear function with satisfactory precision [35]. It is simply a gradient descent method designed to minimise the total error or mean error of the output computed by the network [33]. Although there are many similar forecasting approaches in which neural networks have been built on deseasonalised time series data, they have mainly been used to predict aggregate retail sales [8], temperature [7], streamflow [27], total production value and soft drink [25] and total new privately owned housing units started [34], rather than wind speed.

As an applicative case of the proposed method, a daily average wind speed data series was collected from 1 January 2001 to 31 December 2006 from the Minqin region of China. The daily average wind speed in March, June, September and December 2006 are predicted using the daily average wind speed data series in the corresponding months from 2001 to 2005. Simulations show that the proposed method can significantly improve the accuracy of one-year-ahead daily average wind speed forecasting compared with a method that uses a BP network without seasonal adjustment.

The rest of this paper is organised as follows: Section 2 presents the applicative case in which the input dataset is described, and the forecasting problem is defined. Section 3 explains the proposed method in detail, including the theory of the K–S test, the design of the BP network and other related methods adopted in this paper. Simulation results are presented and discussed in Section 4. Finally, Section 5 highlights the findings and concludes the paper.

2. Applicative case study

The Minqin region, which is located in the Gansu Province in China, has abundant wind resources due to its geographical characteristics. It is situated in an area that measures 206 km from east to west and 156 km north to south, including a total area of 15,900 km^2. Its three main terrain types are desert, hills and plains. The average altitude is 1400 m, with the lowest and highest elevations being 1298 and 1936 m, respectively. This region has the potential to be a valuable wind farm site. To investigate actual wind power potential, it is highly desirable to forecast daily average wind speeds one year ahead of time for a period of time, such as a month.

With the aim of developing an effective forecasting method for daily average wind speed, we collected data on the daily average wind speed in this region from 1 January 2001 to 31 December 2006. Daily average wind speed curves for the Minqin area from 2001 to 2006 are plotted in Fig. 1.

The forecasting method under development will predict the daily average wind speed for a period of time, such as a month, one year ahead of time, by using the daily average wind speed values of the corresponding periods in the past 5 years. The algorithm should be able to predict the daily average wind speed for a particular month in 2006 by using the daily average wind speed in the corresponding months from 2001 to 2005 with a better accuracy compared to a single BP neural network. As examples, the daily average wind speeds for March, June, September and December 2006 will be predicted.

3. Hybrid forecasting method with a BP network and SEA

A flow chart of the hybrid forecasting method is shown in Fig. 2.

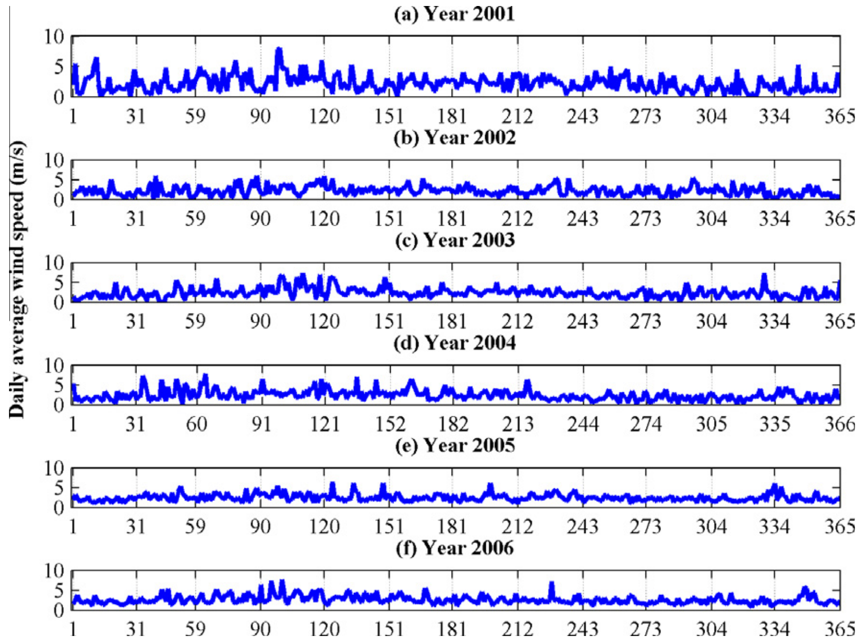

Fig. 1. Daily average wind speed of Minqin area from 1 January 2001 to 31 December 2006.

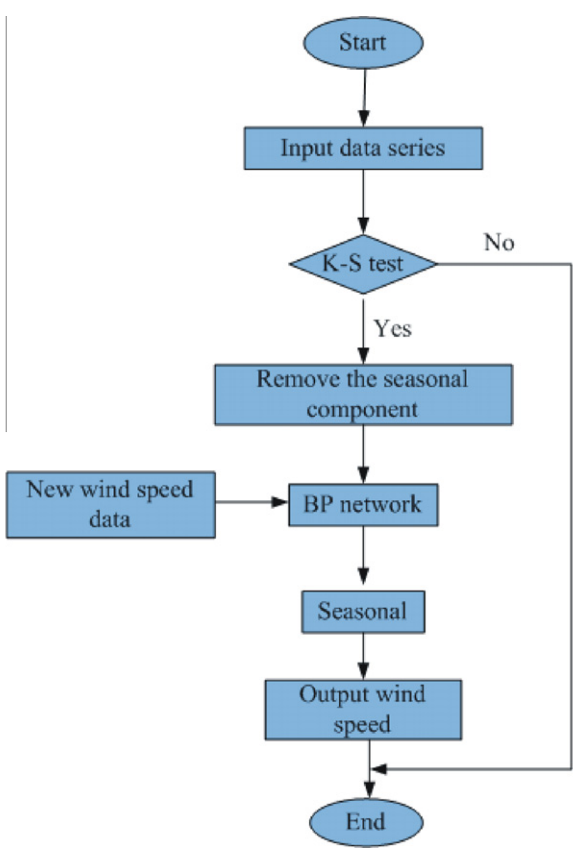

Fig. 2. Flow chart of the proposed method.

First, the Kolmogorov–Smirnov (K–S) test method is applied to test whether the input wind speed datasets exhibit significant differences in their distributions. If the distributions are significantly different, this method is not applicable. Second, for the samples that present very similar distributions, SEA is applied to eliminate the seasonal components from the actual wind speed datasets. Third, a BP neural network is constructed and trained using the training datasets and then applied to predict the trend component of the new wind speed series. Finally, the predicted trend component is adjusted by multiplying it by the seasonal index or by adding the seasonal index to the predicted trend component function.

The following sub-sections will describe the techniques used in the proposed method.

3.1. Kolmogorov–Smirnov (K–S) test method

The K–S test [30] is a nonparametric testing method named after the Russian mathematicians Keermoge and Smirnov. It is a goodness-of-fit test method and is suitable for exploring the distributions of continuous random variables.

The K–S test can test not only whether the probability distribution of a sample exhibits significant differences from a reference probability distribution (named one-sample K–S test), but also whether there is significant difference between the distributions of two samples (known as a two-sample K–S test). Compared with a one-sample K–S test, which uses actual variable values, the two-sample K–S test uses the rank of the variable values. The two-sample K–S test is considered to be one of the most useful nonparametric methods for comparing two samples because it is sensitive to differences in both the location and shape of the empirical cumulative distribution functions of the two samples. The null hypothesis of a two-sample K–S test can be used to verify whether the two samples are drawn from the same distribution; i.e., if the null hypothesis is validated, the two given samples are drawn from the same distribution, whereas if it is not validated, the two samples present significant differences in their distributions.

The two-sample K–S test method is briefly outlined below:

The K–S statistic $D_{n,n'}$ of two samples with n and n' observations, respectively, is defined as follows:

$$D_{n,n'} = \sup_x |F_{1,n}(x) - F_{2,n'}(x)| \quad (1)$$

where $F_{1,n}$ and $F_{2,n'}$ are the empirical distribution functions of the first and the second sample n and n' observations, respectively. The null hypothesis is rejected at level of significance α if

$\sqrt{\frac{nn'}{n+n'}}D_{n,n'} > K_\alpha$, where K_α is found from $\Pr(K \leqslant K_\alpha) = 1-\alpha$ and $K = \sup_{t\in[0,1]}|B(t)|$, and $B(t)$ is the Brownian bridge. A detailed introduction of this method is presented in [5].

Applying this test method to the wind speed datasets in our case study.

Suppose that we are to predict the daily average wind speed for March in 2006 in the Minqin region, with the input datasets being the measured daily average wind speed data during March in each year from 2001 to 2005. The number of data series is 5 (2001, 2002, 2003, 2004, and 2005), and the number of data items in each series is 31.

First, we performed a two sample K–S test against the five pairs of daily average wind speed samples to determine whether the distributions between the paired samples were significantly different (null hypothesis should be rejected). In the experiments, the two-sample K–S test method in the SPSS software tool [30] was used to calculate the differences between the distributions of the paired samples (years 2001 and 2002, years 2002 and 2003, years 2003 and 2004, and years 2004 and 2005) and their probabilities. The differences between the distributions of each of the paired samples and the probabilities are tabulated in Table 1.

It can be seen from Table 1 that all four probability values, 0.079, 0.253, 0.408 and 0.815, are larger than the predefined significance level of 0.05. Consequently, we can assume that the differences among the distributions of the daily average wind speed series in March from 2001 to 2005 are not significant. Hence, the strategy of using the wind speed of March from 2001 to 2005 to predict the wind speed in the corresponding month of 2006 is feasible.

3.2. Seasonal exponential adjustment (SEA)

The seasonal component and the trend component actually coexist in the daily average wind speed series. The models can be composed using the two components in the form of addition or multiplication. According to Zhang and Qi [34], the classical multiplicative decomposition model is appropriate for many time series with increasing seasonal variations. However, if the seasonal variation is relatively consistent with the trend, the additive decomposition model should be used. In our case, the seasonal variation in each wind speed series is not constant over the time; therefore, it is difficult to determine which model is more suitable. As a result, we adopted both additive and multiplicative decomposition models [20] in our case study to forecast wind speed.

Table 1
K–S test results for Minqin's wind speed in March from 2001 to 2005.

Years	Most extreme differences of wind speed (m/s)			Probability
	Absolute	Positive	Negative	
2001 and 2002	0.323	0.065	−0.323	0.079
2002 and 2003	0.258	0.258	−0.097	0.253
2003 and 2004	0.226	0.226	−0.032	0.408
2004 and 2005	0.161	0.097	−0.161	0.815

The multiplicative form of the wind speed at time t can be expressed as

$$x_t = \textit{Trend component} \times \textit{Seasonal component} \tag{2}$$

Then, the seasonal index can be obtained by

$$I_s = x_t/\textit{Trend component}. \tag{3}$$

Because the *Trend component* was unknown, we used the average of x_i in each cycle as its approximation.

If we rearrange the dataset $x_1, x_2, \ldots, x_T$ to be $x_{11}, x_{12}, \ldots, x_{1l}, \ldots, x_{k1}, x_{k2}, \ldots, x_{ks}, \ldots, x_{2l}, \ldots,$ and $x_{m1}, x_{m2}, \ldots, x_{ml}$ $(k = 1, 2, \ldots, m;\ s = 1, 2, \ldots, l)$, where m and l represent the number of cycles and the number of data items in each cycle, respectively, and $T = m \times l$;, then the average can be derived by

$$\bar{x}_k = (x_{k1} + x_{k2} + \cdots + x_{kl})/l \quad (k = 1, 2, \ldots, m) \tag{4}$$

Normalising the data items x_{ks}, we have

$$I_{ks} = \frac{x_{ks}}{\bar{x}_k} \quad (k = 1, 2, \ldots, m;\ s = 1, 2, \ldots, l) \tag{5}$$

I_j is defined as

$$I_j = \frac{I_{1j} + I_{2j} + \cdots + I_{mj}}{m} \quad (j = 1, 2, \ldots, l) \tag{6}$$

This definition of I_j conforms to the normalisation, as demonstrated by

$$\sum_{j=1}^{l} I_j = \frac{1}{m}\sum_{k=1}^{m}\sum_{s=1}^{l} I_{ks} = \frac{1}{m}\sum_{k=1}^{m}\left(\sum_{s=1}^{l} x_{ks}/\bar{x}_k\right) = \frac{1}{m}\sum_{k=1}^{m} l = l$$

Using the value of I_s, the series without the impact of the seasonal component can be obtained by

$$x'_{ks} = \frac{x_{ks}}{I_s} \quad (k = 1, 2, \ldots, m;\ s = 1, 2, \ldots, l) \tag{7}$$

If we re-arrange the data items $x'_{11}, x'_{12}, \ldots, x'_{1l}; \ldots; x'_{m1}, x'_{m2}, \ldots, x'_{ml}$ back to $x'_1, x'_2, \ldots, x'_T$, we then obtain the new data series without the seasonal component.

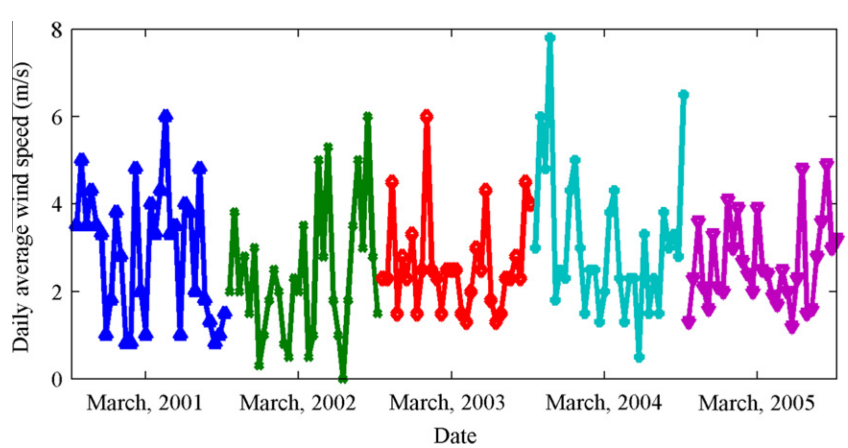

Fig. 3. Raw daily average wind speed of the Minqin area in March from 2001 to 2005.

Table 2
Daily average wind speed values in March from 2001 to 2005 after eliminating the seasonal component by the additive and multiplicative decomposition models.

Date	Daily average wind speed results after using the additive decomposition model (m/s)						Daily average wind speed results after using the multiplicative decomposition model (m/s)					
	Seasonal indices	2001	2002	2003	2004	2005	Seasonal indices	2001	2002	2003	2004	2005
3-1	−0.2806	3.7806	2.2806	2.5806	3.2806	1.5806	0.8875	3.9435	2.2535	2.5915	3.3802	1.4647
3-2	1.1794	3.8206	2.6206	1.1206	4.8206	1.1206	1.4172	3.5282	2.6814	1.623	4.2338	1.623
3-3	0.9794	2.5206	1.0206	3.5206	3.8206	2.6206	1.3484	2.5956	1.4832	3.3372	3.5597	2.6698
3-4	0.9994	3.3006	1.8006	0.5006	6.8006	1.1006	1.3233	3.2496	2.116	1.1336	5.8945	1.587
3-5	−0.4606	3.9606	1.9606	3.2606	2.2606	2.0606	0.828	4.2271	1.8116	3.3817	2.1739	1.9324
3-6	0.1794	3.1206	2.8206	2.1206	2.3206	3.1206	1.0769	3.0643	2.7858	2.1358	2.3215	3.0643
3-7	−0.9006	1.9006	1.2006	4.2006	3.2006	3.0006	0.6592	1.5169	0.4551	5.0059	3.489	3.1856
3-8	−0.5806	2.3806	1.5806	2.0806	4.8806	2.5806	0.7587	2.3725	1.318	1.9771	5.6676	2.6361
3-9	0.7394	3.0606	1.0606	1.7606	4.2606	3.3606	1.2486	3.0434	1.4416	2.0022	4.0045	3.2837
3-10	0.7594	2.0406	1.7406	5.2406	2.2406	2.2406	1.2952	2.1618	1.9302	4.6324	2.3162	2.3162
3-11	−0.5606	1.3606	2.5606	3.0606	2.0606	4.4606	0.8117	0.9856	2.4639	3.0799	1.8479	4.8046
3-12	−0.8806	1.6806	1.6806	3.1806	3.3806	3.5806	0.6687	1.1963	1.1963	3.4393	3.7383	4.0374
3-13	−0.3606	5.1606	0.8606	1.8606	2.8606	2.7606	0.8396	5.7171	0.5955	1.7866	2.9777	2.8586
3-14	−0.6806	2.6806	2.9806	3.1806	1.9806	2.6806	0.7652	2.6138	3.0058	3.2672	1.6989	2.6138
3-15	−0.4206	1.4206	2.4206	2.9206	2.4206	4.3206	0.8584	1.165	2.33	2.9125	2.33	4.5435
3-16	0.5594	3.4406	2.9406	1.9406	3.2406	1.9406	1.2085	3.3099	2.8961	2.0687	3.1444	2.0687
3-17	−0.3006	3.6006	0.8006	1.8006	4.6006	2.7006	0.8516	3.875	0.5871	1.7614	5.0493	2.8182
3-18	−0.5406	4.8406	1.5406	1.8406	2.8406	2.4406	0.7808	5.5069	1.2807	1.6649	2.9455	2.4333
3-19	0.4994	5.5006	4.5006	1.5006	0.8006	1.2006	1.2147	4.9395	4.1162	1.6465	1.0702	1.3995
3-20	0.0794	3.2206	2.7206	2.9206	2.2206	2.4206	1.0406	3.1712	2.6907	2.8829	2.2103	2.4025
3-21	0.4194	3.0806	4.8806	2.0806	1.8806	1.5806	1.1912	2.9382	4.4493	2.0987	1.9308	1.679
3-22	−0.9406	1.9406	2.7406	5.2406	1.4406	2.1406	0.679	1.4727	2.6508	6.3325	0.7363	1.7672
3-23	−0.2206	4.2206	1.2206	2.0206	3.5206	2.5206	0.8938	4.4753	1.1188	2.0139	3.6921	2.5733
3-24	−0.4206	4.2206	0.4206	1.7206	1.9206	5.2206	0.8272	4.5941	0	1.5717	1.8135	5.803
3-25	−0.8806	2.8806	2.6806	2.3806	3.1806	2.3806	0.6728	2.9728	2.6755	2.2296	3.4187	2.2296
3-26	0.0394	4.7606	3.4606	2.2606	1.4606	1.5606	1.0314	4.6539	3.3935	2.23	1.4544	1.5513
3-27	0.4394	1.3606	4.5606	1.8606	3.3606	2.3606	1.1889	1.5139	4.2054	1.9345	3.1961	2.355
3-28	0.0394	1.2606	2.9606	2.7606	2.9606	3.5606	1.0303	1.2617	2.9116	2.7175	2.9116	3.494
3-29	0.7594	0.0406	5.2406	1.5406	2.5406	4.1406	1.3297	0.6016	4.5123	1.7297	2.4818	3.6851
3-30	0.1194	0.8806	2.6806	4.3806	2.6806	2.8806	1.0654	0.9386	2.6282	4.2238	2.6282	2.8159
3-31	0.6394	0.8606	0.8606	3.3606	5.8606	2.5606	1.2074	1.2423	1.2423	3.3128	5.3834	2.6503

Similar to the multiplicative decomposition model, for the additive decomposition model, we need to use the following equations to replace Eqs. (2), (3), (5), and (7), respectively:

$$x_t = f(t) + I_j \tag{2'}$$

$$I_j = x_t - f(t) \tag{3'}$$

$$I_{ks} = x_{ks} - \bar{x}_k \tag{5'}$$

and

$$x'_{ks} = x_{ks} - I_s \tag{7'}$$

Applying SEA to the input wind speed series in our case study, Fig. 3 shows the plot of the raw daily average wind speed curve in the Minqin region during March from 2001 to 2005. After eliminating the seasonal component using both the multiplicative and additive decomposition models, we obtained the two new wind speed series without the impact of the seasonal component, as shown in Table 2 (the values) and Fig. 4(a) and (b) (the curves).

3.3. BP neural network

We used a BP neural network to determine the trend in the daily wind speed series. The topology of the BP neural network is shown in Fig. 5.

This BP neural network consists of three layers: the input layer, the hidden layer and the output layer. The number of nodes in both the input and output layers is 31 because 31 data items, which are the known historical daily average wind speed values during March in the previous year, are used in each month as the input variables, and 31 predicted daily average wind speed values during March in the target year are used as the output [31]. In designing the hidden layer, we employed the Hecht-Nelson method [15] and chose the number of nodes in the hidden layer to be 63, which was determined by $2i + 1$, where i is the number of input nodes.

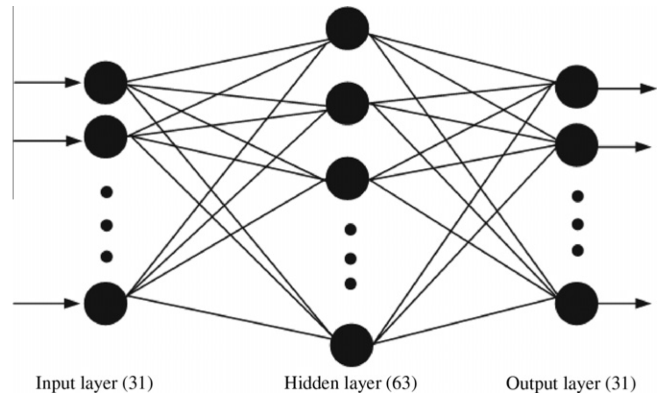

Fig. 5. Structure of the BP neural network.

To ensure that the input values were compatible despite significant differences in their magnitudes, we determined the input vector by normalising each input value using the following formula:

$$V = \{V_i\} = \frac{x_i - x_{i\min}}{x_{i\max} - x_{i\min}}, \quad i = 1, 2, \ldots, 31 \tag{8'}$$

where $x_{i\min}$ and $x_{i\max}$ are the minimal and maximal value of each input factor, respectively.

In training the BP neural network, the historical daily average wind speeds during March from 2001 to 2005 are used, and the transfer functions used in the hidden layer and the output layer are tan $sig(x)$ and log $sig(x)$, respectively.

The BP neural network is implemented using the NNET toolbox in Matlab [14].

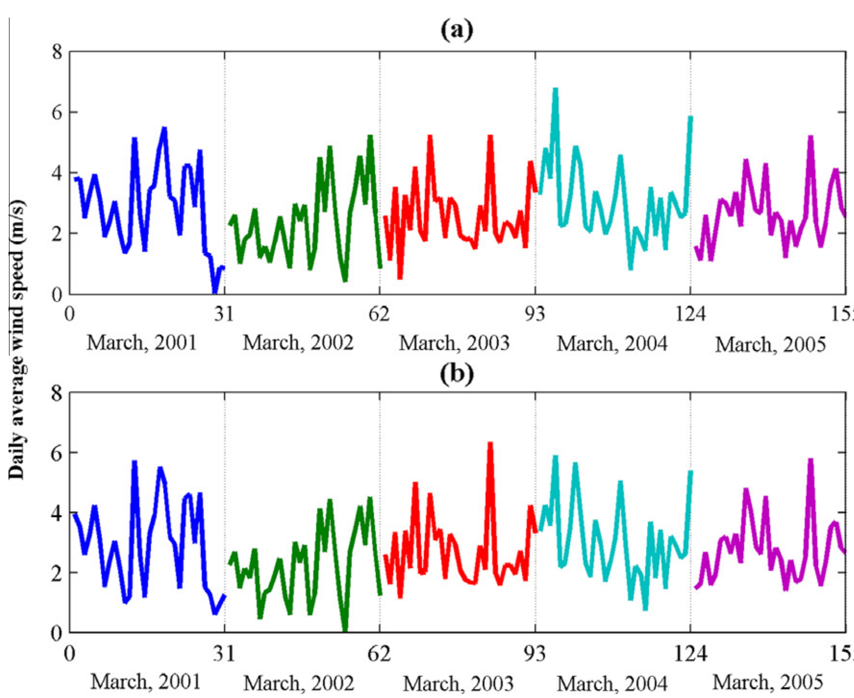

Fig. 4. Preprocessed daily average wind speed of the Minqin area in March from 2001 to 2005 by removing the seasonal component (a) using the additive decomposition model and (b) the multiplicative decomposition model.

Table 3
Forecasted daily average wind speed results using the BPPA and BPPM models.

Date	Forecasted wind speed by BPPA (m/s)	Forecasted wind speed by BPPM (m/s)	Date	Forecasted wind speed by BPPA (m/s)	Forecasted wind speed by BPPM (m/s)	Date	Forecasted wind speed by BPPA (m/s)	Forecasted wind speed by BPPM (m/s)
3-1	2.2384	2.3693	3-11	3.8164	3.8999	3-21	2.7787	3.0266
3-2	3.7971	3.7411	3-12	3.58	3.2128	3-22	0.6417	1.6688
3-3	3.5978	3.5999	3-13	2.3465	2.2863	3-23	4.9943	4.5887
3-4	2.1	1.9382	3-14	3.7257	3.675	3-24	2.4357	2.001
3-5	0.6735	2.2035	3-15	4.0158	4.1238	3-25	4.3348	1.1845
3-6	2.42	2.4943	3-16	3.26	3.3922	3-26	1.6949	3.8008
3-7	2.46	2.1646	3-17	2.3987	2.3868	3-27	4.5032	2.9828
3-8	1.6633	1.7573	3-18	2.0698	2.1884	3-28	1.6024	1.6164
3-9	2.9803	2.8932	3-19	1.7033	1.7138	3-29	3.8389	2.1476
3-10	3	2.9999	3-20	1.28	1.4624	3-30	1.6836	1.6528
						3-31	4.6493	1.873

Table 4
Forecasted daily average wind speed results using the BPR method.

Date	Forecasted wind speed by BPR (m/s)	Date	Forecasted wind speed by BPR (m/s)	Date	Forecasted wind speed by BPR (m/s)
3-1	2.4741	3-11	1.502	3-21	1.3887
3-2	1.2137	3-12	4.8884	3-22	4.8996
3-3	4.8301	3-13	4.8278	3-23	3.763
3-4	3.8763	3-14	3.9407	3-24	4.8959
3-5	1.201	3-15	3.0377	3-25	1.2001
3-6	3.293	3-16	1.2699	3-26	1.4345
3-7	2.7318	3-17	3.3332	3-27	1.942
3-8	3.1098	3-18	3.289	3-28	3.4787
3-9	4.1635	3-19	1.4372	3-29	1.5566
3-10	2.2055	3-20	4.4186	3-30	2.5459
				3-31	4.7312

4. Simulation results of the case study

To validate the proposed forecasting method, predicted the daily average wind speed values in four months (March, June, September and December, which are the representative months for each quarter of the year) in 2006 using the daily average wind speed during the corresponding months in the previous 5 years, i.e., 2001–2005.

In the following sub-sessions, the procedure of applying the proposed method to predict the daily wind speed in March 2006 is illustrated as an example, and the daily wind speeds in June, September, and December 2006 are then predicted using the same procedure.

4.1. Prediction of daily average wind speed in March 2006 using the proposed method

The procedure for applying the proposed method to predict the daily wind speed in March 2006 is described as follows:

Step 1. Conduct the two-sample K–S tests against the five input wind speed series.

The results of the tests are shown in Table 1. Because all of the probabilities in the test are greater than the significance level threshold of 0.05, the distributions of the input data series are not significantly different.

Step 2. Eliminate the seasonal component from the input wind speed data series.

The daily average wind speed values after eliminating the seasonal components using both the multiplicative and the additive decomposition models are shown in Table 2. The wind speed values after eliminating the seasonal component are referred to as the preprocessed daily average wind speed values hereafter.

Step 3. Train the BP neural network.

The training data are the preprocessed daily average wind speed values during March from 2001 to 2005. Because two models (additive and multiplicative) are used to eliminate the seasonal components from the input data series as preprocessing methods, for the ease of explanation, we refer to the BP networks trained by the preprocessed datasets by the additive and the multiplicative models as BPPA and BPPM, respectively.

The parameters of the BPPA and BPPM models are listed in Table 3.

Step 4. Predict the new wind speed values, and adjust the results by the seasonal index.

The daily average wind speed values for March 2006 are predicted using both the BPPA and BPPM models. The output values are adjusted by adding or multiplying the seasonal index, as shown in Table 2, and the final forecasted wind speed values of March 2006 are shown in Table 3.

4.2. Prediction of daily average wind speed values in March 2006 using the BP neural network trained by the raw input wind speed datasets

To demonstrate the effects of eliminating the seasonal components on the accuracy of prediction, we conducted an experiment to predict the daily average wind speed values of March 2006 using

Table 5
Comparison of the forecasting accuracies of the three models.

Models	MSE	MAPE (%)
BPPA	0.8766	23.03
BPPM	1.6541	21.13
BPR	3.4686	59.4

Table 6
Comparisons of the forecasting accuracies of the three models.

Month	MSE			MAPE (%)		
	BPR	BPPA	BPPM	BPR	BPPA	BPPM
June	1.8639	1.1078	1.0723	45.39	28.16	27.03
September	1.6483	0.4826	0.4762	56.53	26.03	24.60
December	3.1280	0.9783	0.6154	62.40	22.09	21.00

only the BP network and trained the BP network using the raw input wind speed series (without removal of the seasonal components). We refer to the BP network trained by the raw wind speed series as the BPR model. The predicted wind speed values for March 2006 obtained using BPR are listed in Table 4.

4.3. Comparisons of BPPA, BPPM and BPR

To validate the proposed method, especially the elimination of the seasonal components subject to passing the K–S test, we mainly compared the results of BPPA/BPPM and BPR. The mean square error (MSE) and mean absolute percentage error (MAPE) are used to measure the prediction accuracy of these three models. The MSE values can be calculated by

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(x_i(t) - \hat{x}_i(t))^2 \tag{9}$$

and the values of MAPE can be calculated by

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{x_i(t) - \hat{x}_i(t)}{x_i(t)}\right| \times 100\% \tag{10}$$

where $x_i(t)$ and $\hat{x}_i(t)$ represent the ith actual and forecasted values at time t, respectively.

Comparisons of the MSE and MAPE values for the BPPA, BPPM and BPR models are listed in Table 5.

It can be seen that the MSE values for the BPPA and BPPM models were 0.8766 and 1.6541, respectively, which are much smaller than that obtained in BPR, which was 3.4686. The MAPE values for BPPA and BPPM were 23.03% and 21.13%, respectively, which were also dramatically smaller than that obtained by BPR, which was 59.4%.

The comparison between BPPA and BPPM showed that the BPPA and BPPM models yielded similar results, and it was difficult to definitively conclude which model was better because the MSE value for BPPA was smaller (better) than that of BPPM, but the MAPE value for BPPA was larger (worse) than that of BPPM. This can be further demonstrated by the fact that the maximum relative error 69.39% obtained by BPPA and the minimum error 0.16% obtained by BPPM occurred on the same day: March 5.

4.4. Further simulations

The daily average wind speed values in the months of June, September and December 2006 were predicted using the three models, BPPA, BPPM and BPR. The MSE and MAPE values for the three models in forecasting the daily average wind speed values in these three months are listed in Table 6. It is worth noting that the numbers of nodes in the input, hidden and output layers are 30, 61 and 30, respectively, for both June and September because these are 30-day months.

The comparisons shown in Table 6 again reveal that the BPPA and BPPM models present much smaller MSE and MAPE values than BPR, indicating that the BPPA and BPPM models can provide better forecasting accuracy. However, the comparisons also show that all of the MSE and MAPE values obtained from BPPM in June, September and December are smaller than those obtained from the BPPA model. This finding does not agree with the one presented in Section 3.3. Therefore, it appears that BPPM performs better than BPPA, except for in March. In summary, it is clear that both the BPPA and BPPM models perform better than the BPR model, but it is still unclear when BPPM performs better than BPPA.

5. Conclusions

Numerous research efforts have been devoted to improving the accuracy of wind speed forecasting through the optimisation of parameters and further analysis of factors that have a significant impact on the final output in these models. However, fewer studies have attempted to improve forecasting performance by addressing original datasets. In this paper, a new hybrid method based on seasonal exponential adjustment and a BP neural network was proposed to forecast the daily average wind speed values for a period of time (e.g., a month) based on historical daily average wind speed values during the corresponding time period over a number of cycles (e.g., a year). This method first tests its applicability to the input datasets using the two-sample K–S test method. If the applicability is confirmed, the SEA will be applied to eliminate seasonal effects from the input datasets, and the resultant series will be the trend component. A BP neural network is used to forecast the trend component for the wind speed. The trend component of the historical daily average wind speed series is used to train the network. The predicted trend component of the new wind speed series will be combined with the seasonal exponential adjustment method to forecast the daily average wind speed. Additive and multiplicative decomposition models are considered in the seasonal exponential adjustment.

Wind speed forecasting remains a very challenging problem, and the mean absolute percentage errors associated with such forecasting usually range from 25% to 40% [32]. In general, the shorter the forecasting horizon or the smaller the variation in wind speed, the smaller the forecasting error. Simulation of daily average wind speed values for March, June, September and December of 2006 in the Minqin region in China showed that the proposed method is valid and can achieve a better forecasting accuracy compared to a forecasting method without seasonal exponential adjustment. Comparisons among the measured wind speed values, the simulated results using the proposed method and the BP neural network showed that the proposed method performed better than the single BP neural network. The maximum mean absolute percentage error of wind speed forecasting is 28.16%. Given that the forecasting horizon is one year ahead, the forecasting accuracy is substantially high. However, this method is only applicable to forecasting tasks with multiple datasets that share almost the same distribution.

Acknowledgements

This research was supported by the National Fundamental Research Program of China (Grant No. 2009CB421402).

References

[1] M.C. Alexiadis, P.S. Dokopoulos, H.S. Sahsamanoglou, Wind speed and power forecasting based on spatial correlation models, IEEE Transactions on Energy Conversion 14 (3) (1999) 836–842.
[2] N.N. Barberis, O. Holmstrom, B.J. Birgitte, Model of a synthetic wind speed time series generator, Wind Energy 11 (2) (2008) 193–209.
[3] T.G. Barbounis, J.B. Theocharis, A locally recurrent fuzzy neural network with application to the wind speed prediction using spatial correlation, Neurocomputing 70 (2007) 1525–1542.
[4] K. Cai, L.N. Tan, C.L. Li, X.F. Tao, Short-term wind speed forecasting combing time series and neural network method, Power System Technology 32 (8) (2008) 82–85.
[5] Chakravarti, Laha, Roy, Handbook of Methods of Applied Statistics, vol. I, John Wiley and Sons, 1967. pp. 392–394.
[6] H. Chen, J. Chen, X. Duan, Fuzzy modeling and optimization algorithm on dynamic economic dispatch in wind power integrated system, Dianli Xitong Zidonghua/Automation of Electric Power Systems 30 (2006) 22–26.
[7] O.H. Choon, J.L.T. Chuin, A comparison of neural network methods and box-Jenkins model in time series analysis, in: Proceedings of the 4th IASTED International Conference on Advances in Computer Science and Technology, 2008, pp. 344–350.
[8] C.W. Chu, G.P. Zhang, A comparative study of linear and nonlinear models for aggregate retail sales forecasting, International Journal of Production Economics 86 (2003) 217–231.
[9] I.G. Damousis, M.C. Alexiadis, J.B. Theocharis, A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation, IEEE Transactions on Energy Conversion 19 (2) (2004) 352–361.
[10] Y. Du, J.P. Lu, Q. Li, Y.L. Deng, Short-term wind speed forecasting of wind farm based on least square-support vector machine, Power System Technology 32 (15) (2008) 62–66.
[11] T.H.M. El-Fouly, E.F. El-Saadany, M.M.A. Salama, Grey predictor for wind energy conversion systems output power prediction, IEEE Transactions on Power Systems 21 (3) (2006) 1450–1452.
[12] L.N. Elvira, Annual electrical peak load forecasting methods with measures of prediction error, Diss. Abstr. Int. 62 Section: B, 4719, 2002.
[13] G. Fan, W. Wang, C. Liu, G. Fan, Wind power prediction based on artificial neural network, Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering 28 (2008) 118–123.
[14] Fei Si Technology Product Research Center, Neural Network Theory and its implement by MATLAB 7, Publishing House of Electronics Industry, 2005.
[15] R. Hecht-Nelson, Kolmogorov's mapping neural network existence theorem, in: Proceedings of 1st IEEE Annual International Conference on Neural Networks, San Diego, June 21–24, IEEE Press, Piscataway, NJ, 1987, pp. III.11–111.14.
[16] T.J. Jing, M.H. Yang, Equivalent wind speed model of wind generation, in: IEEE Transmission and Distribution Conference and Exposition, 2010, pp. 1–5.
[17] G.N. Kariniotakis, G.S. Stavrakakis, E.F. Nogaret, Wind power forecasting using advanced neural networks models', IEEE Transactions on Energy Conversion 11 (4) (1996) 762–767.
[18] M. Mohammad, R. Hasan, M.K. Hossein, A new strategy for wind speed forecasting using artificial intelligent methods, Renewable Energy 34 (2009) 845–848.
[19] M.A. Mohandes, T.O. Halawani, S. Rehman, A.A. Hussain, Support vector machines for wind speed prediction, Renewable Energy 29 (2004) 939–947.
[20] D.X. Niu, S.H. Cao, J.C. Lu, Load forecasting technology and its applications, second ed., Electric Power Press of China, 2009.
[21] G.H. Riahy, M. Abedi, Short term wind speed forecasting for wind turbine applications using linear prediction method, Renewable Energy 33 (2008) 35–41, doi:10.1016/j.renene.2007.01.014.
[22] S.S. Sancho, M.P.B. Ángel, G.O.G. Emilio, P.F. Antonio, P. Luis, C. Francisco, Accurate short-term wind speed prediction by exploiting diversity in input data using banks of artificial neural networks, Neurocomputing 72 (2009) 1336–1341, doi:10.1016/j.neucom.2008.09.010.
[23] G. Sideratos, N.D. Hatziargyriou, An advanced statistical method for wind power forecasting, IEEE Transactions on Power Systems 22 (1) (2007) 258–265.
[24] C. Sun, Y. Wang, X. Li, Vector autoregression model of hourly wind speed and its applications in hourly wind speed forecasting, Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering 28 (2008) 112–117.
[25] F.-M. Tseng, H.-C. Yu, G.-H. Tzeng, Combining neural network model with seasonal time series ARIMA model, Technological Forecasting and Social Change 69 (2002) 71–87.
[26] P. Wang, N. Tai, B. Wang, H. Zhai, J. Ye, L. Li, J. Zhu, L. Qi, Short-term load forecasting correction method considering climatic factors, Dianli Xitong Zidonghua/Automation of Electric Power Systems 32 (2008) 92–96.
[27] W. Wang, P.H.A.J.M.V. Gelder, J.K. Vrijling, J. Ma, Forecasting daily streamflow using hybrid ANN models, Journal of Hydrology 324 (2006) 383–399.
[28] X.H. Wu, H. Zhou, M. Huang, Wind speed and generated power forecasting based on pattern recognition in wind farm, Relay 36 (1) (2008) 27–32.
[29] Z. Xu, X.L. Song, X.M. Zhang, J. Bai, A Monte-Carlo-based network method for source positioning in bioluminescence tomography, International Journal of Biomedical Imaging (2007) 2007, doi:10.1155/2007/48989.
[30] W. Xue, Statistical Analysis Methods and Applications of SPSS, second ed., Publishing House of Electronics Industry, Beijing, 2009.
[31] B.A. Yang, L.X. Li, H. Ji, J. Xu, An early warning system for loan risk assessment using artificial neural networks, Knowledge-Based Systems 14 (2001) 303–306, doi:10.1016/S0950-705100110.
[32] X. Yang, Y. Xiao, S. Chen, Wind speed and generated power forecasting in wind farm, Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering 25 (2005) 1–5.
[33] B. Yu, D.H. Zhu, Combining neural networks and semantic feature space for email classification, Knowledge-Based Systems 22 (2009) 376–381.
[34] G.P. Zhang, M. Qi, Neural network forecasting for seasonal and trend time series, European Journal of Operational Research 160 (2005) 501–514.
[35] L.P. Zhang, K. Wu, Y.F. Zhong, P.X. Li, A new sub-pixel mapping algorithm based on a BP neural network with an observation model, Neurocomputing 71 (2008) 2046–2054, doi:10.1016/j.neucom.2007.08.033.
[36] Y. Zhang, L. Kang, S. Zhou, B. Cao, Wavelet analysis applied to wind speed prediction in predicate control system of wind turbine, Taiyangneng Xuebao/Acta Energiae Solaris Sinica 29 (2008) 520–524.