Short-term Forecasting for Wind Speed Based on Wavelet Decomposition and LMBP Neural Network

Wei WEI*, Guilian WU, Minghai YANG
Key Laboratory of Smart Grid of Ministry of Education
Tianjin University
Tianjin,China
weiw@tju.edu.cn

Yongwu ZHANG, Shengxiao QIU, Aiming SUN
Electric Power Dispatching Center
Weifang Electric Power Bureau
Weifang,China

Abstract—**In this paper, a wind speed forecasting method based on wavelet decomposition and BP neural network with Levenberg–Marquardt algorithm (LMBP) is proposed. Firstly, original wind speed seires is decomposed into one low-frequency component and several high-frequency components by wavelet decomposition method. Then different LMBP neural networks are built for the forecasting of every component respectively. Finally the predictions of components are reconstructed to obtain the prediction of original wind speed. As for the problem that the convergence rate is limited by the inversion calculation of large-scale matrix in the training process of LMBP network, super-memory gradient algorithm solving large linear equations is introduced to adjust weights and thresholds of the network. Meanwhile the structure of hidden layer neurons is optimized by least squares network pruning method. At the end of the paper, actual wind speed data from certain wind farm is used to verify the forecasting model and the results indicate that the model develops the precision of wind speed forecasting effectively.**

Keywords-wavelet decomposition; LMBP neural network; wind speed forecasting; super-memory gradient algorithm; least squares network pruning

I. INTRODUCTION

Wind power as a clean and new renewable energy, is paid more and more attention in recent years. Especially with the development of wind power generation technology, wind energy has been developed rapidly. However, because of wind power's intermittence and uncontrollable characteristics, large-scale wind power integrated in power system may cause several problems such as lowering power quality, making the operation of electric network instability and insecurity. Therefore it is necessary to forecast wind power ahead. If the power of a wind farm can be predicted reliably, the power dispatching plan would be made with more direction and also the disadvantages impact to the power system would be relived. Since the wind power depends mainly on the wind speed, it can be predicted through forecasting the wind speed.

Influenced by atmospheric conditions such as temperature, pressure, air density, wind speed has great fluctuations and wind speed series is a highly nonlinear series. Nowadays there are many approaches that used to forecast wind speed, like persistent model[1], time series model[2], Kalman Filters[3], Artificial Neural Networks[4], support vector machine[5] and so on. Each method has its advantages and disadvantages. Studies of wind speed behavior show that wind speed series is a Weibull distribution with two parameters[18]. It can be seen as the superposition of several different frequency components that are approximately periodic variation with similar frequency characteristic and consistent variation. In this paper, wavelet analysis theory is introduced to analyze wind speed in nature. Based on the multi-resolution analysis theory of wavelet analysis, the original wind speed can be decomposed into several components at different frequencies. The high-frequency components describe the detail information of wind speed and low-frequency components describe the approximate information. To forecast wind speed, Back Propagation neural network with Levenberg-Marquardt algorithm (LMBP) is used to forecast components. In order to avoid the inversion calculation of large scale matrix during the training process of LMBP network, super-memory gradient algorithm is introduced to solving large linear equations and computing time is shorten. To increase the network's generalization, least squares network pruning method is introduced. As the components decomposed by wavelet analysis are approximately periodic variation, they are easy to be forecasted with higher accuracy.

II. WAVELET ANALYSIS THEORY

The wavelet transform (WT) is an effective tool for time-frequency analysis of signals. By using WT, an original signal can be decomposed into several wavelet functions at different time and frequency levels, which are formed by scale expanding and translating a mother wavelet function. Compared with Fourier transform, WT has an advantage of time and frequency resolution can be adjusted automatically. WT provides better frequency resolution for the low-frequency components of signal while better time resolution for the high-frequency components. Because of the ability of time and frequency resolution adaptive adjusting, highly nonlinear wind speed can be analyzed by WT effectively.

Basis of multi-resolution analysis theory, a fast discrete wavelet transform is developed known as Mallat algorithm[6]. Mallat algorithm decomposes signal into different resolution levels with the process of decomposition and reconstruction. In the decomposition process, original signal S is made to pass the high pass and low pass filters, then two coefficients $a1$ and

* Corresponding author: weiw@tju.edu.cn

$d1$ are obtained. Coefficient $a1$ got from low pass filter is a low-frequency component describing the approximate information of the signal. The other coefficient $d1$ got from high pass filter is a high-frequency component describing the detail information of the signal. Since the length of coefficients is halved compared with original signal, it's necessary to recover the length of coefficients by reconstruction. Send the coefficients $a1$ and $d1$ pass reconstruction filters, then a low-frequency series $A1$ and a high-frequency series $D1$ which have the same length with original signal are obtained.

If the decomposition level is more than one, then the decomposition process should be continued after the original signal is decomposed into two coefficients $a1$ and $d1$. In the following decomposition, only the approximate coefficient is needed to be decomposed. So the coefficient $a1$ is decomposed into another detail coefficient $d2$ and approximate coefficient $a2$ and $a2$ is continuing to be a third decomposed and so on. Suppose the decomposition level is J, then the original signal is decomposed of J times and high-frequency coefficients $d1, d2, ...dJ$ and low-frequency coefficient aJ are obtained. Made those different frequency coefficients pass reconstruction filters with different times and then decomposed series $D1, D2, ..., DJ$ and AJ are obtained. Series $D1, D2, ..., DJ$ are high frequency series and AJ is low frequency series. All of the series have the same length with original signal. The relationship between original signal and decomposed series meets as following:

$$S = D1 + D2 + ... + DJ + AJ\,. \tag{1}$$

Fig.1 illustrates the process of a signal been decomposed into three resolution levels. As shown in Fig.1, Mallat algorithm focus on a careful decomposition for low-frequency space of signal, making low-frequency series better resolution.

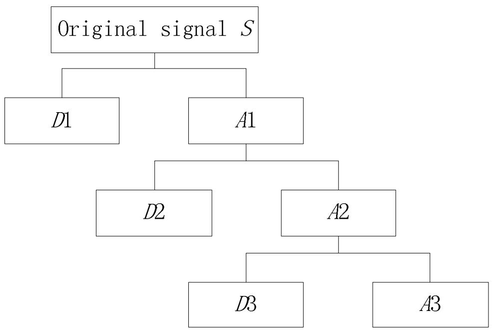

Figure 1. Process of a signal been decomposed into three resolution levels

Compared with original signal, series decomposed by Mallat algorithm reduce non-stationary greatly. As series decomposed by wavelet decomposition are approximate stationary series, wavelet decomposition is useful to improve forecasting accuracy.

III. LMBP Neural Network and Improvement

A. LMBP Neural Network

Back Propagation (BP) neural network is a kind of feed-forward neural network based on the error back-propagation learning algorithm. Through the learning process of network, complex relation between input and output variables is extracted and the generalization ability of the network is formed. Funahashi has theoretically proved that BP network with three layers and sigmoid function as transform function in hidden layer can approach any nonlinear function with any precision so long as it provides enough hidden layer neurons[17]. Due to the great generalization, BP network has been applied successfully in different problems such as power load forecasting, pattern recognition and so on.

BP network with three layers consist of an input layer, one hidden layer and an output layer. Each layer has several neurons. Neurons in one layer attach to the neighborhood layer through connection weights and thresholds. By connection weights and thresholds, information transmission is accomplished. Another important factor of building a neural network is transform function. In general, the sigmoid function is chosen for hidden layer and Purelin function for output layer for the purpose of forecasting. After the structure of network is established, it's time to train the network with samples. The training process includes information transmitted from input neurons to output layer by layer and error back transmitted from output neurons to input to adjust weights and thresholds. With the information transmitted forward and error transmitted back again and again, the error is reduced to an acceptable range and the training process is stopped. The connection weights and thresholds are well adjusted and the network generates strong generalization ability so that it can be used to prediction.

Typically, gradient descent algorithm is used to adjust weights and thresholds in the training process of BP neural network. In this paper, Levenberg-Marquardt (LM) algorithm is introduced to instead of gradient descent algorithm because the gradient descent algorithm learns slowly and may easily fall into local minimum leading to training failure. LM algorithm is a compromise algorithm of Gauss-Newton algorithm and gradient descent algorithm. It can solve convergence problem effectively. Equation (2) is the adjustment expression of weights with LM algorithm.

$$\Delta W = -(J^T J + \mu I)^{-1} J^T E \tag{2}$$

Where, ΔW is the amount of weight adjustment, J is the Jacobin matrix of error on the weight differential, I is the unit matrix, E is the error of network and μ is a scalar which can be adjusted automatically during the training process.

μ is an index of LM algorithm. When μ is large, LM algorithm is close to Gauss-Newton algorithm and μ is small closing to gradient descent algorithm. In the process of training, error can be reduced by increasing μ. So in order to reduce error and improve convergence rate, μ is adjusted as (3) during the training process.

$$\mu = \begin{cases} 10\mu & E(k+1) > E(k) \\ \mu/10 & E(k+1) <= E(k) \end{cases} \tag{3}$$

B. LMBP Neural Network's Improvement

According to (2), there is a calculation of large-scale inverse matrix. The convergence rate is limited because the calculation consumes a lot of time. To solve this problem, change the format of (2) to linear equations by moving the matrix $(J^T J + \mu I)$ to the left side of the equation. After that, the equations can be solved with numerical methods and the inverse calculation of $(J^T J + \mu I)$ is avoided. Compared with inverse matrix calculation, computing time is shortened effectively by using numerical methods.

There are some numerical methods used to solve linear equations, such as LU factorization algorithm, QR factorization algorithm, conjugate gradient algorithm and so on. As the forecasting for wind speed has lots of training samples and the structure of the network is complexity, there is a large of equations need to be solved, so the super-memory gradient algorithm[7] is introduced to solve the equations. Compared with other methods, super-memory gradient algorithm has faster speed to solve large linear equations and acquires well calculation results.

C. Least Squares Network Pruning

One of key factors affecting network performance is the neurons in hidden layer. If the number of neurons in hidden layer is small, it's difficult to train the network because the training process limited by initial conditions and parameters. If the number of neurons in hidden layer is large, although the training process is easier to accomplish, the generalization of network is reduced. It has proved that a network has achieved a given accuracy has a stronger generalization if the structure of network is less complication. So the generalization of the network can be increased by minimizing the number of neurons in hidden layer. Now yet how to define the number of neurons in hidden layer is still a big problem. Generally, it's determined according to designers' experience and lots of test experiments. In this paper, least square network pruning method is applied to define the number of neurons in hidden layer. The advantage of network pruning method is that it can minimize the number of neurons in hidden layer and maintain the performance of network well. Firstly, suppose eliminating one neuron in hidden layer and calculate the output error. If the error is minimum, the neuron should be eliminated, also with the weights connecting to this neuron. Secondly, adjust the weights and thresholds by using least squares algorithm so that the performance of the network can be maintained. Least squares network pruning method is a useful tool for optimizing the structure of network.

Before pruning the network, the structure of network should be ready by training process with training samples. It's necessary to initialize the number of neurons in hidden layer. According to Kolmogorov theorem, there is approximate relation between the number of neurons in input layer and hidden layer described as following:

$$m = 2n + 1 . \tag{4}$$

Where, m is the number of neurons in hidden layer and n is the number of neurons in input layer.

IV. Wind Speed Forecasting Model

A. Forecasting Model

In above sections, the theory of wavelet decomposition and LMBP neural network has been presented. Combining the advantages of wavelet decomposition and LMBP neural network, a forecasting model (WT-LMBP) for wind speed is proposed in this paper. Firstly, select a wavelet function and define the decomposition level of wind speed. Daubechies wavelet functions (db(N)) have been applied to deal with non-stability series widely since their characteristics of compactly supported orthogonal. This paper takes db(N) wavelet function to decompose original wind speed. After the original wind speed is decomposed into certain series at the given resolution level, different LMBP neural networks are established for the forecasting of every series respectively. Finally, the prediction results of each series are reconstructed to obtain the wind speed forecasting results.

B. Normalization of Samples

One of sigmoid transform function's characteristics is that if the input data is far from the null range, then the output data will be close to saturation making little adjustment to weights and thresholds. In order to improve the convergence rate, the input data of network should be normalized to close to null range before being used to train the network.

Equation (5) and (6) show how to normalize the data. After the normalization, the data is mapped to the range from -1 to 1.

$$x_{mid} = (x_{\max} + x_{\min}) / 2 \tag{5}$$

$$x_i^{'} = 2(x_i - x_{mid})/(x_{\max} - x_{\min}) \tag{6}$$

Where, x_i is original data, $x_i^{'}$ is normalized data, $x_{\max}$ and $x_{\min}$ are the maximum and minimum of original data.

After the prediction calculation, the results should be anti-normalized to original data range according to (7).

$$x_i = \frac{1}{2} x_i^{'} (x_{\max} - x_{\min}) + x_{mid} \tag{7}$$

V. CASE ANALYSIS

This paper takes real wind speed data from certain wind farm as an example to analyze the proposed forecasting method at matlab7.0. The daily average wind speed of the farm is shown in Fig.2. Those data values collected from February 27, 2010 to November 5, 2010 without May 17 because the value in that day is bad. There are finally 261 data values. For testing the performance of the method, the first 251 data values are used to form training samples training the network and the last 10 data values are left for testing. Here one day ahead forecasting for wind speed is accomplished.

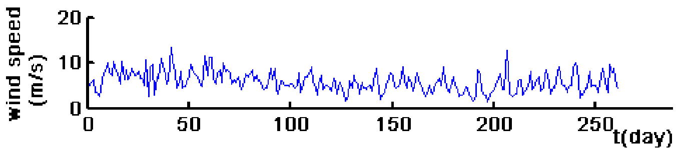

Figure 2. Orignal wind speed series

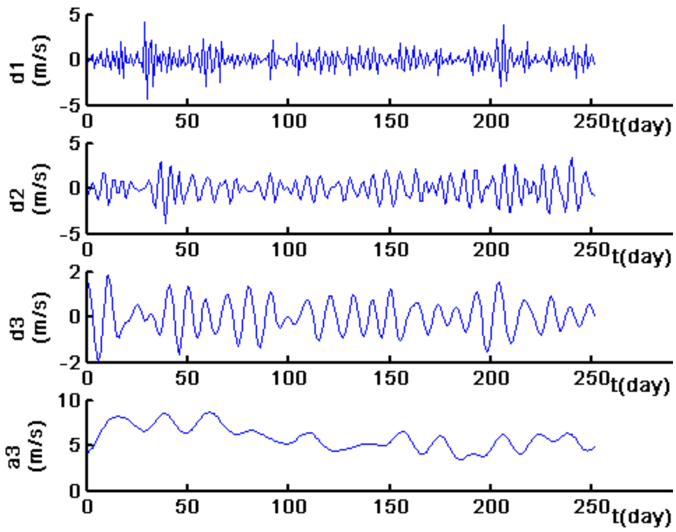

Figure 3. Decomposed series

A. Wavelet Decomposition

Firstly, select db30 as the wavelet function and define the decomposition level is three. Base on the theory of multi-resolution analysis, those wind speed values used to form training samples are decomposed into three high-frequency series d_1, d_2, d_3 and one low-frequency series a_3. Fig.3 shows the curve of those decomposed series.

To analyze the series' characteristics, five indicators including mean (*M*), variance (*VAR*), standard deviation (*STD*), skewness (*S*) and kurtosis (*K*) are introduced to illustrate the stationary of series. Here *M* measures the trends of data center, *VAR* measures the degree of deviation from the mean, *STD* measures the discrete degree of series, *S* measures the degree of asymmetry around the mean and *K* measures the degree of precipitous or flat at the top of frequency distribution curve. Those indicators are represented as following:

$$M = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{8}$$

$$VAR = \frac{1}{n}\sum_{i=1}^{n}(x_i - M)^2 \tag{9}$$

$$STD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - M)^2} \tag{10}$$

$$S = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - M}{STD}\right)^3 \tag{11}$$

$$K = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_i - M}{STD}\right)^4 . \tag{12}$$

x_i is the time series value and n is the length of time series.

In this study, the five indicators values of original wind speed series and decomposed series are shown in Tab. I respectively.

TABLE I. INDICATORS OF ORIGINAL WIND SPEED SERIES AND DECOMPOSED SERIES

Series	**Indicators**				
	M	***VAR***	***STD***	***S***	***K***
wind speed	5.8833	4.8092	2.1972	0.5492	3.0717
d1	-0.0003	1.1640	1.0811	-0.0072	4.6732
d2	-0.0017	1.5117	1.2320	0.0299	2.9206
d3	0.0034	0.5159	0.7197	-0.0073	2.5883
a3	5.8500	1.6148	1.2733	0.3530	2.4188

Compared with original wind speed, the *VAR* and *STD* values of decomposed series are smaller, which means that decomposed series values are more close to the data center. Also the smaller values of *S* and *K* indicate that decomposed series have more symmetry and less volatility. From those indicators values, we found that decomposed series have less fluctuation and more stationary than original wind speed. Especially the lower the frequency is, the smoother the series curve is. Decomposed series d_1, d_2, d_3 and a_3 are approximate stationary series.

B. Establish LMBP Neural Network

As there are four series d_1, d_2, d_3 and a_3 after original wind speed series been decomposed, it need to establish four BP neural networks for the four series. For each series, the modeling process is the same. In this study, the prediction values are obtained in terms of the following steps:

Step 1: Normalize series d_1, d_2, d_3 and a_3 and map them to the range from -1 to 1 according to (5) and (6) respectively.

Step 2: For each series, arrange the network with 16 input neurons and one output neuron, then set up training samples with those series respectively. Since each series has 251 data values, there are 235 samples for each network.

Step 3: For the four networks, select the Tansig function as transform function for hidden layer and Purelin function for output layer. Then train the networks with improved LM algorithm introduced in above sections respectively. Also optimize the structure of network with least squares network pruning method.

In order to ensure the reliability of the network, divide the 235 samples into two parts: the training samples and the verification samples. Use the first 205 samples to train the network, after that use the other 30 samples to verify the prediction performance of the network. The training process is stopped until the error of verification samples is within the allowable range and minimize.

Step 4: With the network models which are built in Step 3 to calculate the next moment values of series d_1, d_2, d_3 and a_3 respectively. Anti-normalize the predictions to original data range according to (7). In this case, there are 10 values been calculated for each series. The predictions values of d_1, d_2, d_3 and a_3 are $d_1^{'}, d_2^{'}, d_3^{'}$ and $a_3^{'}$.

Step 5: Reconstruct the predictions of each series to obtain the prediction of original wind speed: $S^{'} = d_1^{'} + d_2^{'} + d_3^{'} + a_3^{'}$.

The predictions values of series $d_1^{'}, d_2^{'}, d_3^{'}$ and $a_3^{'}$ are shown in Fig.4 and the prediction values of original wind speed are shown in Fig.5.

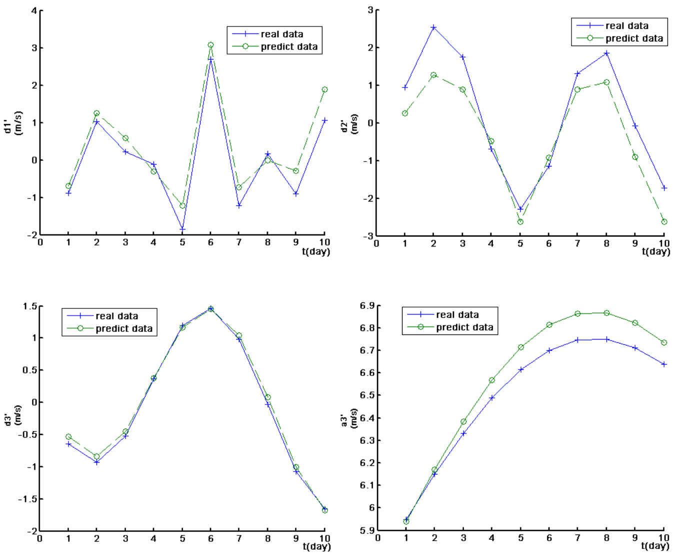

Figure 4. Predictions of decomposed series

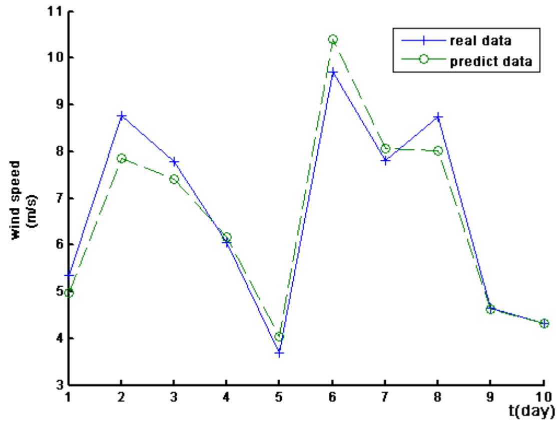

Figure 5. Prediction of original wind speed by proposed method in this paper

C. *Test the Performance of Proposed Method*

In this paper, BP neural network is used to instruct the performance of proposed method. Fig.6 shows the predictions of original wind speed by BP neural network and proposed method in this paper. It can be seen that the prediction curve of BP network model has large error in great fluctuation points. The proposed model can solve this problem effectively since the decomposed series have less fluctuation and more stationary.

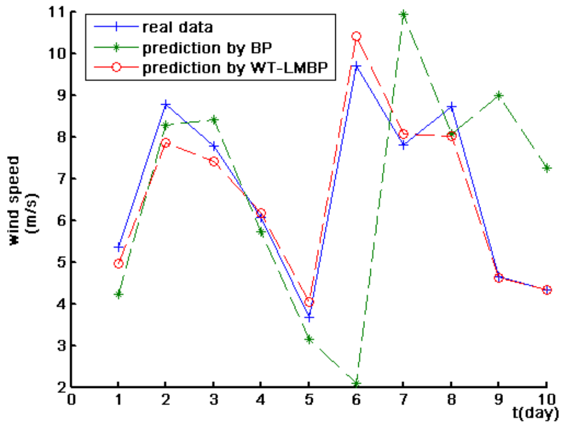

Figure 6. Predictions of two methods

The mean absolute percentage error (*MAPE*) and root mean square error (*RMSE*) are selected as the evaluation indicators.

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|x_i^{'} - x_i|}{x_i} \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i^{'} - x_i)^2} \tag{14}$$

Where x_i is the real value and $x_i^{'}$ is the predict value, n is the length of data series.

The prediction errors of BP model and WT-LMBP model proposed in this paper are shown in Tab.Ⅱ. Compared with BP model, the errors of WT-LMBP model are much smaller which illustrates that WT-LMBP model has improved the prediction precision effectively.

TABLE II. PREDICTION ERRORS OF BP AND WT-LMBP MODEL

Predict method	***MAPE(%)***	***RMSE(m/s)***
BP	0.032	3.128
WT-LMBP	0.0062	0.484

In the aspect of time consuming, the training process of each series can quickly convergence since the series are approximate stationary. Meanwhile by using super-memory gradient method to adjust weights and thresholds, computing time is shortened effectively. Compared with BP neural network, there is not increase of computing time significantly.

VI. CONCLUSION

In this paper, a forecasting model for wind speed is proposed based on wavelet decomposition and LMBP neural network. By wavelet decomposition, original wind speed is decomposed into certain series at different frequencies. Since the decomposed series are approximate stationary series, they are easy to be predicted with higher accuracy. In the process of network training, super-memory gradient algorithm and least squares network pruning are introduced to improve the convergence rate and optimize the structure of network. A case study of certain wind farm shows that the proposed model develops the precision of wind speed forecasting effectively.

ACKNOWLEDGMENT

The work reported in this paper is partly supported by National Natural Science Foundation of China (50707019), Special Fund of the National Basic Research Program of China (2009CB219701, 2010CB234608), Tianjin Municipal Science and Technology Development Program (09JCZDJC25000), Specialized Research Fund for Doctor Discipline (20090032110064).

REFERENCES

[1] Wu Guoshang, Xiao yang, Weng Shasha, “Discussion about short-term forecast of wind speed on wind farm,” Jilin Electric Power, No.6, pp. 21-24, 2005.

[2] L. Kamal, YZ. Jafri, “Time series models to simulate and forecast hourly averaged wind speed in Quetta, Pakistan”, Solar Energy, vol. 61, pp. 23-32, 1997.

[3] Bossanyi E A, ”Short-term wind prediction using Kalman filters,” Wind Engineering, vol. 9, pp. 1-8, 1985.

[4] T. G. Barbounis, J. B. Theocharis, M. C. Alexiadis, and P. S. Dokopoulos, “Long-term wind speed and power forecasting using local recurrent neural network models,” IEEE Trans. Energy Conversion, to be published.

[5] Liu Bin, Su Hongye, Chu Jian, “Predictive control algorithm based on least squares support vector machines,” Control and Decision, vol. 19, pp. 1399-1402, 2004.

[6] Mallat S, “A theory for multiresolution signal decomposition:the wavelet representation,” IEEE Trans Pattern Anal And Mathchine Intell, vol. 11, pp. 674-693, 1989.

[7] Shi Zhenjun, “A supermemory gradient method for unconstrained optimization problem,” Chinese Journal of Engineering Mathematics, vol. 17, pp. 99-104, 2000.

[8] Bashir Z, El-Hawary M E, “Short term load forecasting by using wavelet neural networks,” Proc of the IEEE Conference onElectrical and Computer Engineering, Canada, 2007.

[9] Xiao Yongshan, Wang Weiqing, Huo Xiaoping, “Study on the time-series wind speed forecasting of the wind farm based on neural networks,” Energy Conservation Technology, vol. 25, No.2, pp. 106-108, 2007.

[10] Gu Xingkai, Fan Gaofeng, Dai Huizhu, “Summarization of wind power prediction technology,” Power System Technology, vol. 31, pp. 335-338, 2007.

[11] Fan Gaofeng, Wang Weisheng, Liu Chun, “Wind power prediction based on artificial neural network,” Proceedings of the CSEE, vol. 28, No. 34, pp. 118-123, 2008.

[12] Yang Jianguo, Wavelet Analysis and Its Engineering Applications, Beijing: China machine press, 2005.

[13] Fredric M. Ham, Ivica Kostanic, Principles of Neurocomputing for Science and Engineering, Beijing: China machine press, 2007.

[14] Zhang Ni, Tian Xueming, “Research on improved algorithm of LMBP neural network,” Computers and Applied Chemistry, vol. 27, No. 8, pp. 1111-1114, 2010.

[15] Geng Xiaoqing, He Jinsheng, Yu Baoqi, “Research on several improved BP algorithms and comparison in application ,” Computer Engineering and Applications, vol. 43, No.33, pp. 243-246, 2007.

[16] Yang Ying, Tan Ping, Wang Yuechao, “Neural network structure optimization based on improved LMBP algorithm,” Computer Engineering, vol. 34, No.1, pp. 215-217, 2008.

[17] Liu Kai, “Ji yu gai jin BP shen jing wang luo de duan qi fu he yu ce yan jiu,” Hohai University, Nanjing, 2005.

[18] Han Shuan, “feng dian chang gong lv duan qi yu ce fang fa yan jiu, ” North China Electric Power University, Beijing, 2008.

[19] Liu Yibing, Li Hu, Ma Zhiyong, Xin Weidong, “The nonlinear characteristics analysis of wind speed time series,” Journal of North China Electric Power University, vol. 35, No. 6, pp.99-102, 2008.