The 11th International Conference on
Computer Science & Education (ICCSE 2016)
August 23-25, 2016. Nagoya University, Japan
WdP4.1

Short-term Wind Speed Forecasting by Combination of Masking Signal-based Empirical Mode Decomposition and Extreme Learning Machine

Qiu Jihui, Shen Shaoping*, Xu Guangyu
Department of Automation,
School of Aeronautics and Astronautics ,
Xiamen University, Xiamen 361005, P. R. China
Email: xmqiujihui@163.com, shen_shaoping@163.com

***Abstract*—According to the requirement of accurate prediction of the short-term wind speed series, this paper proposes a new short-term combination prediction model of the wind speed series by means of the masking signal-based empirical mode decomposition (MS-EMD) and the extreme learning machine (ELM). Firstly, because of the non-stationary characteristics of the wind speed series, the wind speed series is decomposed into several components with different frequency bands by the MS-EMD to reduce the non-stationary characteristics. Secondly, in order to avoid the randomness of input dimensionality selection of the ELM, the phase space of each component is reconstructed. Thirdly, the ELM model of each component is established to predict the wind speed series. Finally, the predicted results of each component are superimposed to get the final result. The simulation result verifies that the proposed combination forecasting model is able to excavate the wind speed series features effectively and has relatively high prediction accuracy.**

***Index Terms*—Wind speed series, prediction, masking signal-based empirical mode decomposition, phase space reconstruction, extreme learning machine.**

I. Introduction

The stratosphere airship is an important floated platform which works in about 20 km altitude for a long time with a lot of payload. It can play a role similar to man-made satellite in the fields of communications relay, remote sensing to the ground, and air traffic control [1]-[2].

Due to the earth's rotation, there is a strong west wind in mid latitude stratosphere while the wind speed changes along with the height, the latitude and longitude, and the season [3]. In the station-keeping control for stratosphere airship, the wind is a disturbance factor. The propulsion control system of the stratosphere airship needs to adjust the speed of the motor to keep the position of the stratosphere airship in a predetermined region according to the wind speed change.

Therefore, wind speed forecasting is an important subject of the stratosphere airship, and it can make a great difference. According to the predicted wind speed at the next moment, the motor speed of the stratosphere airship is adjusted in advance, so that the stratosphere airship can overcome the influence of the wind speed change with smaller power, which can reduce the loss of energy.

In wind speed forecasting methods, a lot of work have been done at home and abroad. It mainly includes time series model [4], artificial neural network model [5], support vector machine model [6], etc. These methods not only have their characteristics, but have certain limitation. Time series method requires a lot of historical data. Artificial neural network is easy to fall into local minimum value, etc.

The extreme learning machine [7] is a new single hidden layer feedforward neural network, which is proposed by Huang Guangbin in 2006. It has greatly improved the learning speed and the generalization ability of the network. What's more, the extreme learning machine has a strong non-linear fitting ability, and has achieved good result in non-linear fitting prediction [8]. However, the extreme learning machine can only fit the non-linear part of the wind speed series. The non-stationary characteristics of the wind speed series will have a great influence on the predicted result. So, it is very important to reduce the non-stationary characteristics. The empirical mode decomposition (EMD) is very suitable for non-linear and non-stationary signal processing [9]. However, the mode-mixing phenomenon may exist in the process of decomposition and affect the accuracy of forecasting [10]. MS-EMD is an improvement on the empirical mode decomposition [11]. It can solve the mode mixing phenomenon of the EMD. Based on the above analysis, this paper proposes a new short-term combination prediction model of wind speed by means of masking signal-based empirical mode decomposition (MS-EMD) and extreme learning machine (ELM). The experimental results show that the method proposed in this paper has relatively high prediction accuracy.

II. Masking Signal-based Empirical Mode Decomposition

For processing the non-stationary characteristics of the wind speed series, the traditional empirical mode decomposition (EMD) method usually decomposes the wind speed series into several components with different frequencies. However, the mode-mixing phenomenon may exist in the decomposition process and affect the prediction accuracy. MS-EMD is an improvement on the empirical mode decomposition. It can

978-1-5090-2218-2/16/$31.00 ©2016 IEEE

solve the mode mixing phenomenon of the EMD. The main idea of MS-EMD is that the masking signal is inserted in the original wind speed signal, and then the EMD is used in the mixed-signal [11].

A. Empirical Mode Decomposition

EMD is an efficient method for signal decomposition. It has very good adaptability, which is very suitable for non-linear and non-stationary signal processing. It is based on the local characteristic scale of the signal, decomposing the waveform or trend of the different scales in any signal step by step, and generates a series of intrinsic mode functions (IMF) which are relatively stable and have different characteristic scales. There are 3 key steps to obtain the intrinsic mode functions (the original signal to be decomposed is $X(t)$):

Step 1: The author finds out the local maximum value of the original series. Here, in order to retain the characteristics of the original series, the local maximum value is defined as the value of a certain time in the series. The value of the previous time is not larger than it, and the value of the latter is also not bigger than it. Then the three order spline interpolation is used to obtain the value of the upper envelope series ($e_{up}(t)$) of the original series ($X(t)$). In the same way, we can get the lower envelope series ($e_{low}(t)$).

Step 2: It takes the average of $e_{up}(t)$ and $e_{low}(t)$ at each moment, so we can get their instantaneous average value $m_1(t)$.

$$m_1(t) = [e_{up}(t) + e_{low}(t)]/2 \tag{1}$$

Step 3: $X(t)$ minus $m_1(t)$ is $h_1(t)$. When $h_1(t)$=$X(t)$-$m_1(t)$ satisfies the IMF condition, $c_1(t)$=$h_1(t)$. Then the $c_1(t)$ is defined as IMF1 of the wind speed series, which contains the shortest periodic component of the original series.

The component $c_1(t)$ is separated from the original signal, and the residual component is obtained:

$$r_1(t) = X(t) - c_1(t) \tag{2}$$

The remaining component $r_1(t)$ is used as the new initial data, and we repeat the above steps to get the rest of the IMF and the last margin. The results are as follows:

$$\begin{cases} r_1(t) - c_2(t) = r_2(t) \\ r_2(t) - c_3(t) = r_3(t) \\ \qquad \vdots \\ r_{N-1}(t) - c_N(t) = r_N(t) \end{cases} \tag{3}$$

The original wind speed series $X(t)$ can be decomposed into:

$$X(t) = \sum_{i=1}^{N} c_i(t) + r_N(t) \tag{4}$$

In this paper, the termination condition of the EMD proposed by Rilling [12] is used, which is the improvement of the limited standard deviation criterion proposed by Huang.

If e_{max} and e_{min} are the upper and lower envelope respectively,

$$\delta(t) = \frac{|e_{max} + e_{min}|}{|e_{max} - e_{min}|} \tag{5}$$

We set the three threshold values θ_1, θ_2 and α, and there are two corresponding termination conditions:
condition 1):

$$\frac{s\{t \in D | \delta(t) < \theta_1\}}{s\{t \in D\}} \geq 1 - \alpha \tag{6}$$

where D is the duration range of the signal, $s(A)$ is the number of the elements in the set A, θ_1=0.05, α=0.05.
condition 2): for every moment there are

$$\delta(t) < \theta_2, \theta_2 = 10\theta_1 \tag{7}$$

B. Masking Signal-based Empirical Mode Decomposition

MS-EMD is proposed in the reference [11]. The main idea of this method is to insert a row of sine wave s(t) (masking signal) into the original signal to prevent the mode-mixing phenomenon which is caused by the low frequency components mixing into the IMF in the process of the EMD. The specific steps of the method are as follows:

Step 1: According to the frequency of the original wind speed series, the sine wave s(t) (masking signal) is constructed;

Step 2: We use the EMD to decompose $X_+(t) = X(t)+s(t)$ to get the IMF which are defined as $Z_+(t)$; The EMD is also used to decompose $X_-(t) = X(t) - s(t)$ to get $Z_-(t)$.

Step 3: The final IMF are defined as $Z(t) = (Z_+(t) + Z_-(t))/2$. $Z(t)$ are a series of intrinsic mode functions (IMF). They are relatively stable and have different characteristic scales, and are also the results of the MS-EMD.

The key point of the MS-EMD is to find a suitable masking signal s(t). According to the energy mean method, the masking signal s(t) is determined in the literature [13]. The amplitude and the frequency of the s(t) is determined by the Hilbert envelope amplitude and instantaneous frequency of the IMF1 which is the highest frequency component of the original wind speed series (X(t)) in the EMD. The masking signal is $s(n) = a_0 sin(2\pi \bar{f} n/f_s)$, where f_s is the sampling frequency of the signal, and the a_0 is 1.6 times of the average amplitude of the signal component.

$$\bar{f} = \frac{\sum\limits_{i=1}^{K} a_1(i) f_1^2(i)}{\sum\limits_{i=1}^{K} a_1(i) f_1(i)} \tag{8}$$

Where $a_1(i)$ is the Hilbert envelope amplitude of the IMF1, $f_1(i)$ is the instantaneous frequency of IMF1, $\bar{f}$ is the average instantaneous frequency of the IMF1 at k sampling points.

III. Extreme Learning Machine

A. Extreme Learning Machine

Extreme learning machine is a new type of the single hidden layer feedforward neural network [7]. For N different samples (X_i, Y_i) , where $X_i = [x_{i1}, x_{i2}, \cdots, x_{in}]^T \in R^n$, $Y_i = [y_{i1}, y_{i2}, \cdots, y_{im}]^T \in R^m$, the mathematical model of the extreme learning machine which contains L hidden nodes and the activation function $f(x)$, can be expressed as

$$\sum_{i=1}^{L} \beta_i f(W_i \cdot X_j + b_i) = o_j \qquad j = 1, \cdots, N \tag{9}$$

where β_i is the weight vector connecting the ith hidden node and the output nodes, $W_i = [w_{i1}, w_{i2}, \cdots, w_{in}]^T$ is the weight vector connecting the ith hidden node and the input nodes,and b_i is the threshold of the ith hidden node. o_j is the output value of the jth node, and $W_i \cdot X_j$ denotes the inner product of W_i and X_j. The training of the network can approximate these N samples with zero error, where $\sum_{j=1}^{L} \|o_j - Y_j\| = 0$, and there exist $\widehat{W_i}$, $\widehat{b_i}$, $\widehat{\beta_i}$

$$\sum_{i=1}^{L} \widehat{\beta_i} f(\widehat{W_i} \cdot X_j + \widehat{b_i}) = Y_j \qquad j = 1, \cdots, N \tag{10}$$

The above equation can be written compactly as

$$H\beta = Y \tag{11}$$

where

$$H = \begin{bmatrix} f(W_1X_1 + b_1) & \cdots & f(W_LX_1 + b_L) \\ \vdots & & \vdots \\ f(W_1X_N + b_1) & \cdots & f(W_LX_N + b_L) \end{bmatrix}_{N\times L} \tag{12}$$

$$\beta = [\beta_1^T, \cdots, \beta_L^T]_{L\times m}^T \tag{13}$$

$$Y = [y_1^T, \cdots, y_N^T]_{N\times m}^T \tag{14}$$

H is called the hidden layer output matrix of the neural network, and the ith column of the H is the ith hidden node output with respect to inputs $x_1, x_2, \cdots, x_N$.

The reference [7] has proven: When the activation function is infinitely differentiable, the network parameters do not need to be fully adjusted. The input connection weights (W) and the bias of the node in hidden layer (b) can be randomly selected at the beginning of training, and the output connection weights can be obtained by solving the Least Square Solution of the linear equations (15).

$$min_\beta \|H\beta - Y\| \tag{15}$$

Its solution is

$$\widehat{\beta} = H^+Y \tag{16}$$

where H^+ is the Moore-Penrose generalized matrix inverse of the hidden layer output matrix H.

The learning algorithm of the extreme learning machine can be divided into 3 steps:

Step 1: Randomly assign the input weight W_i and the bias b_i, where $i = 1, \cdots, L$.

Step 2: Calculate the hidden layer output matrix H.

Step 3: Calculate the output weight $\widehat{\beta}$.

Compared with the traditional neural network, the extreme learning machine does not need to adjust the value of W_i and b_i in the training process. It can obtain a global optimal solution only by fitting the β, and the training speed can be improved significantly. What's more, it will not fall into local optimum.

B. Network Structure of The Extreme Learning Machine

The prediction performance of extreme learning machine is affected by the network structure [14]. For the extreme learning machine in this paper, the number of neurons in the input layer are the embedding dimension (m) in phase-space reconstruction, and the output layer has one neuron. However, the activation function and the number of neurons in the hidden layer can not be determined.

In order to select the activation function and the number of neurons in the hidden layer, the author selects 4 kinds of activation functions (Hardlim, Sigmoidal, Sine [15] and radbas). The initial number of neurons in the hidden layer is 10, and the number is increased by 2 as a recycle, and the maximum number is set to 120. The root mean square error (e_{rms}) of the test sample is selected as the performance index. Finally, the different activation functions and different number of neurons in hidden layer are selected to analysis of the influence on the e_{rms}. When the e_{rms} takes the minimum, the optimal activation function and the number of neurons in the hidden layer are determined.

IV. Phase-space Reconstruction

Takens theorem has proved that the appropriate selection of the delay time and the embedding dimension will make the reconstructed phase space reflect the rule of the system state with the time evolution [16]-[18]. For the wind speed series $x(i)$, where $i = 1, \cdots, N$, the delay time τ and the embedding dimension m are the key of the phase-space reconstruction. There are many mature methods in phase-space reconstruction. In this paper, the mutual information method [19] is used to get the delay time (τ), and the embedding dimension (m) is obtained by the method of False Nearest Neighbours, FNN [20].

V. Case Analysis

A. Evaluation Index

In this paper, the mean absolute error (e_{mae}), the root mean square error (e_{rms}) and the prediction time (t) are selected as the performance indexes in the test sample. e_{mae} and e_{rms} are shown in (17) and (18).

$$e_{mae} = \frac{1}{N}\sum_{i=1}^{N} |y'(i) - y(i)| \tag{17}$$

$$e_{rms} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (y'(i) - y(i))^2} \tag{18}$$

where $y'(i)$ is the predictive value, $y(i)$ is the real value.

B. Sample Design

1) Wind Speed Series: The author can not obtain the wind speed series in the stratosphere, and the wind field in the troposphere is more complex than the stratospheric. So, the author chooses the 10 consecutive days of the horizontal wind speed series which is 155 meters above the wind farm in the Yancheng city of the Jiangsu province from November 30, 2015 to December 9, 2015. The horizontal wind speed series is measured by the wind measurement lidar, and the sampling period of the system is 10min. The collection of the wind speed series (1440 data points) are presented in Fig.1.

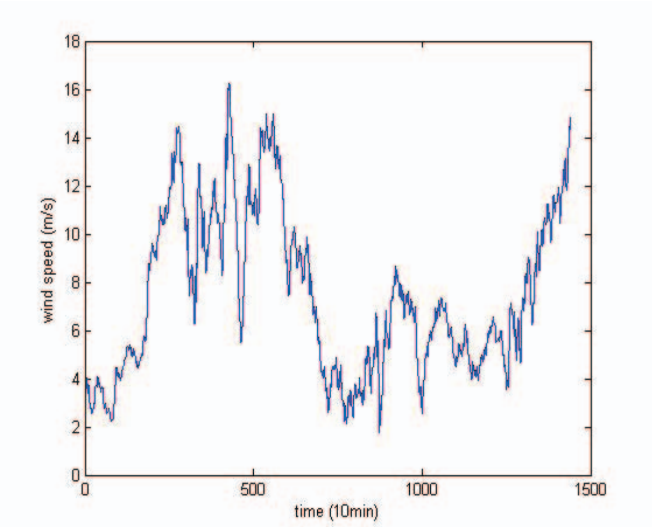

Fig. 1. Wind speed series

2) The EMD and MS-EMD of The Wind Speed Series: Firstly, the original wind speed series is decomposed by EMD, and the intrinsic mode is obtained. Secondly, the masking signal in the MS-EMD is $s(n) = 0.1455sin(0.438\pi n)$, and then the MS-EMD is used. Finally, the decomposition results of the EMD and the MS-EMD are obtained as shown in Fig.2.

3) Parameters of The Phase-space Reconstruction: The author use the mutual information method to obtain the delay time (τ) of each component, and the False Nearest Neighbours, FNN for the embedding dimension (m). The parameters are shown in Table I.

TABLE I
PARAMETERS OF THE PHASE-SPACE RECONSTRUCTION

Component	*Delay Time*	*Embedding Dimension*
$IMF1$	1	8
$IMF2$	2	8
$IMF3$	6	6
$IMF4$	12	3
$IMF5$	17	3
$IMF6$	11	5
$IMF7$	16	2
RES	6	2

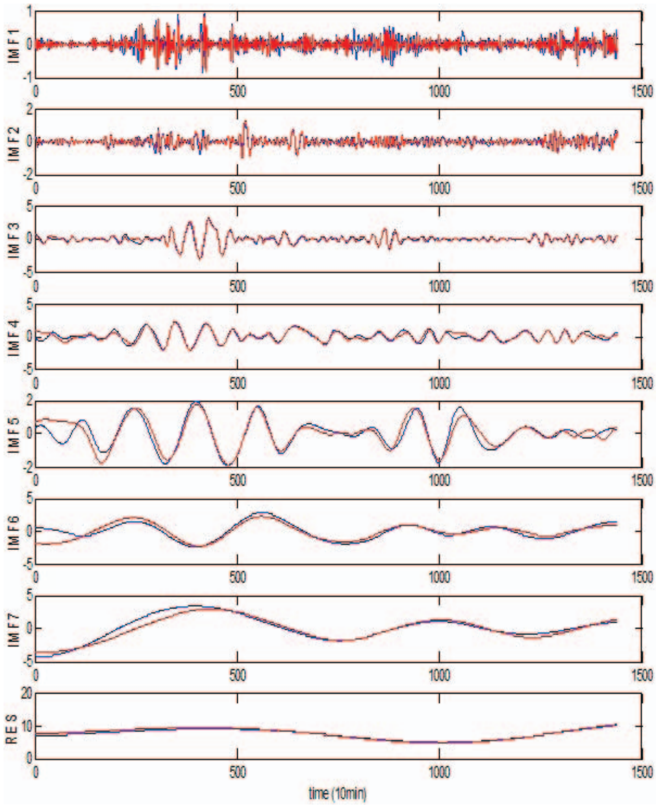

Fig. 2. Results of the MS-EMD (red line: MS-EMD; blue line: EMD)

4) Network Structure of The Extreme Learning Machine: After the phase-space reconstruction, each component can form T phase points, where $T = N - (m-1)\tau$ and N is the number of the original data samples. The number of the vectors are T in the phase-space reconstruction, and the original data samples were divided into T groups. In this paper, the author selects the first (T-144) groups of the data as the training samples, and the later 144 groups of the data are used as the testing samples. The wind speed prediction of 10min in advance is discussed in this paper. The activation function and the number of neurons in the hidden layer of the extreme learning machine are determined in each component, as shown in Table II.

TABLE II
THE ACTIVATION FUNCTION AND THE NUMBER OF NEURONS IN THE HIDDEN LAYER

$Component$	$ActivationFunction$	$NumberofTheNeurons$
$IMF1$	$Sigmoidal$	89
$IMF2$	$Sigmoidal$	66
$IMF3$	$Sigmoidal$	51
$IMF4$	$sine$	22
$IMF5$	$Sigmoidal$	52
$IMF6$	$Sigmoidal$	40
$IMF7$	$Sigmoidal$	12
RES	$Sigmoidal$	17

5) Forecast result: The predicted results of each component of the MS-EMD are shown in Fig.3. The predicted results of each component are superimposed to get the final result. Then the predicted result of the wind speed series is shown in Fig.4.

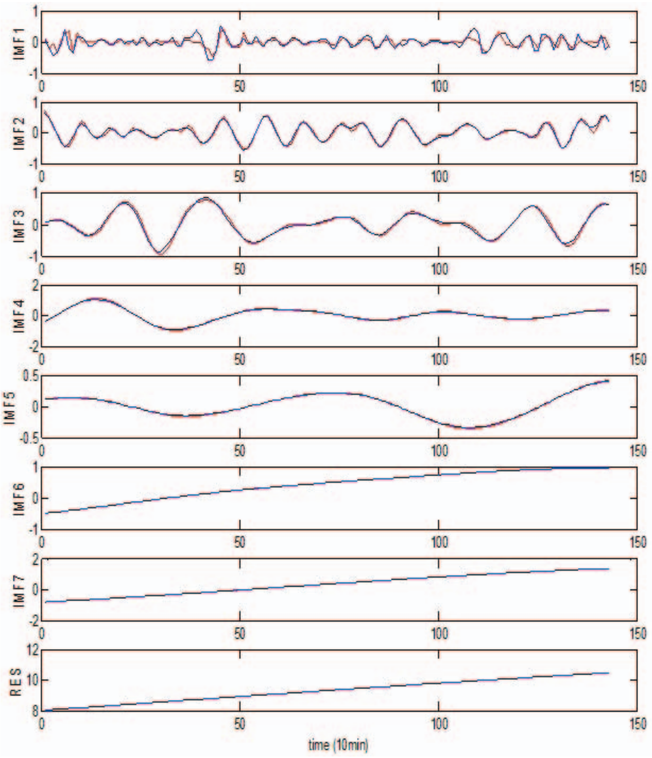

Fig. 3. Predicted results of each component (red line:predictive value; blue line:real value)

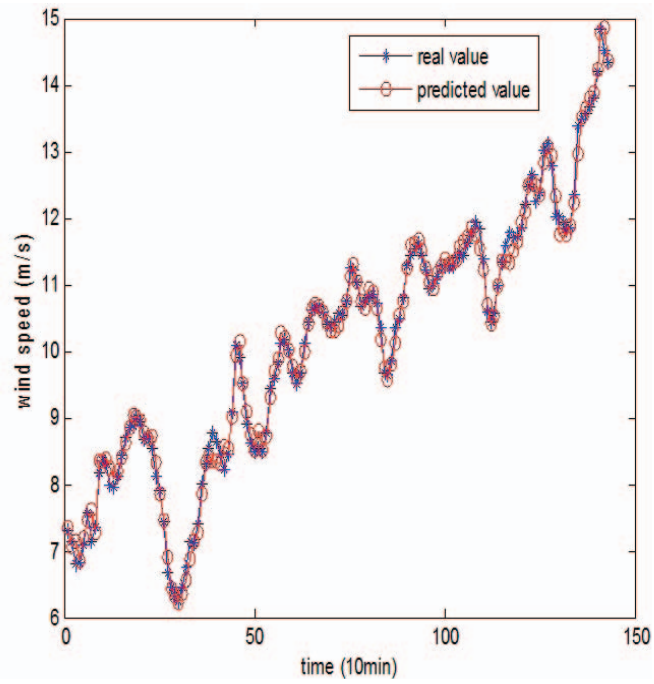

Fig. 4. Result of the wind speed forecasting

6) Comparison of the different combination model in the wind speed forecasting: Firstly, the MS-EMD and the EMD are compared in this paper to reduce the non-stationary of the wind speed series. Secondly, in order to explain the extreme learning machine being a relatively good method, the ELM neural network, the BP neural network and the Elman neural network [21] are compared in this paper. The comparison results are shown in Table III.

TABLE III
MODEL PERFORMANCE

$Decomposition$	$network$	e_{mae}	e_{rms}	$t(s)$
EMD	ELM	0.1188	0.1514	0.4099
EMD	BP	0.1257	0.1621	17.8404
EMD	$Elman$	0.1328	0.1712	34.1101
$MS-EMD$	ELM	0.1099	0.1471	0.3402
$MS-EMD$	BP	0.1112	0.1491	17.7556
$MS-EMD$	$Elman$	0.1184	0.1536	28.8373

As can be seen from the Fig.4, the model of the combination of MS-EMD and ELM in this paper can be used to track the wind speed series, and get a good predictive effect. As can be seen from the Table III, through the comparison of the mean absolute error (e_{mae}), the root mean square error (e_{rms}) and the prediction time (t) of the 6 kinds of wind speed forecasting models, the e_{mae} of the combination model of the MS-EMD and the ELM is minimum and is 0.1099 m/s, and the e_{rms} of the combination model of the MS-EMD and the ELM is minimum and is 0.1471 m/s, and the t of the combination model of the MS-EMD and the ELM is minimum and is 0.3402 s. Therefore, the combined model proposed in this paper has a good advantage in wind speed forecasting.

VI. CONCLUSION

This paper proposes a new short-term combination prediction model of the wind speed series by means of masking signal-based empirical mode decomposition(MS-EMD) and extreme learning machine(ELM). The following conclusions are drawn from the analysis of the predicted results. First of all, in the respect of reducing the non-stationary characteristics of the wind speed series, the MS-EMD is better than EMD. Secondly, the extreme learning machine has strong nonlinear learning ability, and has achieved good prediction effect on the wind speed forecasting. Finally, through the comparison of the forecasting results of the different models, it can be seen that the proposed model in this paper is effective and advanced.

ACKNOWLEDGMENT

The author would like to thank the teacher Jin Huiyu for his help and support in making this work possible. This work is supported by National Natural Science Foundation (NNSF) of China under Grant 61273199.

REFERENCES

[1] LI Zhibin, WU Lei, ZHANG Jingrui, et al. Review of dynamics and control of stratospheric airships, Advances in Mechanics, 42(4): 482-493, 2012.

[2] Fan Changxin. A kind of new developing mobile communication method–research of stratosphere communication, Modern Electronic Technique, 210(19): 1-3, 2005.

[3] JM Zhang, LU Yuping, J Liu. Research on trajectory control systems for stratospheric long duration balloon, Aircraft Design. 27(2): 21-24, 2007.

[4] YS Bao, WU Zhen-Sheng. Short-term wind forecasting based on SVM time-series method [J]. Electric Power, 2011, 44(9): 61-64.

[5] Gong L, Shi J, Zhou J. Bayesian adaptive combination of short-term wind speed forecasts from neural network models [J]. Renewable Energy, 2011, 36(7): 352-359.

[6] Yang Xiyun, Sun Baojun, Zhang Xinfang, et al. Short-term wind speed forecasting based on support vector machine with similar data [J]. Proceedings of the CSEE, 2012, 32(4): 35-41.

[7] Huang G B, Zhu Q Y, Siew C K. Extreme learning machine: theory and Applications [J]. Neuro Computing, 2006, 70(1-3): 489-501.

[8] GAO Guangyong, JIANG Guoping. Prediction of multivariable chaotic time series using optimized extreme learning machine [J]. Acta Physica Sinica, 2012, 61(4): 1-9.

[9] HAN Zhonghe, ZHU Xiaoxun, LI Wenhua, et al. Research of suppressing cross term of Wigner-ville distribution based on EMD [J]. Steam Turbine Technology, 2010(003): 211-214.

[10] Hu Aijun, Sun Jingjing, Xiang Ling. Mode Mixing in Empirical Mode Decomposition [J]. Vibration test and diagnosis, 2011, 31(4): 429-434.

[11] PRINCE A, SENROY N, BALASUBRAMANIAN R. Targeted approach to apply masking signal-based empirical mode decomposition for mode identification from dynamic power system wide area measurement signal data [J]. IET Generation, Transmission & Distribution, 2011, 5(10): 1025-103.

[12] Yang Weiqin, Gu lan. Time series analysis and dynamical modeling [M]. Beijing: Beijing University of Technology Press, 1988: 26-33.

[13] DEERING R, KAISER J F. The use of a masking signal to improve empirical mode decomposition [C]. IEEE International Conference on Acoustics, Speech, and Signal Processing. Philadelphia, Pennsylvania, USA, 2005: 485-488.

[14] FENG Guorui, HUANG Guangbin, LIN Qing-ping, et al. Error minimized extreme learning machine with growth of hidden nodes and incremental learning[J]. IEEE Trans on Neural Networks, 2009, 20(8): 1352-1357.

[15] LI Bin, LI Yibin. Chaotic time series prediction based on ELM learning algorithm [J]. Journal of Tianjin University, 2011, 44(8): 701-704.

[16] Sun Bin, Yao Haitao, Liu Ting. Short-term wind speed forecasting based on Gaussian process regression model [J]. Proceedings of the CSEE, 2012, 32(29): 104-109.

[17] Kim H S, Eykholt R, Salas J D. Nonlinear dynamics delay times and embedding windows [J]. Physica D, 1999, 127(1): 48-60.

[18] Luo Xin, Zhou Yuhui, Guo Hongyu. Short-term price forecasting based on chaotic property and phase space recostructed neural networks [J]. Relay, 2008, 36(1): 48-52.

[19] L Tao, TANG Wei, SUO Li. Prediction of short-term wind speed in wind farm based on chaotic phase space reconstruction theory [J]. Power System Protection and Control, 2010, 38(21): 113-117.

[20] LIAO Zhiqiang, LI Taifu, YU Dejun, et al. Short-term wind speed forecasting of neural network based on phase space reconstruction [J]. Journal of Jiangnan University: Natural Science Edition, 2012, 11(1): 15-18.

[21] CA Yao, JI Shao-Long, YU Yong-Chang. Short-term combination forecasting of wind speed based on wavelet transform and Elman neural network [J]. Renewable Energy Resources, 2012.