One Hour Ahead Prediction of Wind Speed Based on Data Mining

Liu Dejun
Faculty of Mechanical and Electronic Engineering
China University of Petroleum-Beijing,
102249 Beijing, China
liudj65@163.com

Li Hui, Ma Zhonghua
Faculty of Mechanical and Electronic Engineering
China University of Petroleum-Beijing,
102249 Beijing, China
toobage@163.com

Abstract — **Wind speed forecasting is very important to the utilization of wind energy in wind farm. In order to improve the forecast precision, a forecasting method based on empirical mode decomposition (EMD) and wavelet decomposition combine with least square support vector machine (LSSVM) is proposed in this paper. The wind speed time series was decomposed into several intrinsic mode functions (IMF) and the trend term. In order to reduce the nature of non-stationary, the high frequency band was decomposed and reconstructed by wavelet transform (WT). The different LSSVM models to forecast each IMF and trend term were built up. These forecasting results of each IMF and trend term were combined to obtain the final forecasting results. The simulation experiment shows the MAPE is 4.53% about wind speed forecasting and the prediction accuracy is improved considerably.**

Keywords-wind speed forecasting; empirical mode decomposition; wavelet transform; least square support vector machine

I. INTRODUCTION

Wind energy as a non-polluting, renewable energy, has been attached great attention to all over the world and become one of the rapid development of the new energy. It is expected that about 12% of the total world electricity demands to be supplied from wind energy resources by 2020[1]. Recently, the capacity of stand-alone generating units and the large-scale wind farm have been increased highly. As the development of technology, the cost of the new energy resources has already approached the conventional energy. The main method of exploitation and utilization the wind energy is the large-scale power generation connected with grid at present. Wind speed prediction become the essential issues for wind farms maintenance, optimal power flow between power network and wind farms, electricity marketing bidding, power system scheduling, and energy storages planning and scheduling.

Wind power as an intermittent energy has strong randomness and difficult to be controlled, usually, the fluctuation range of the wind speed is very large. Because of these reasons, it is difficult to control the voltage, reactive power and peak adjusting. At present, a number of different approaches have been applied to forecast wind speed and the power produced by wind farms. Such as: physical method [2], statistical method [3], spatial correlation model [4], kalman filter method, time series method [5], artificial neural network [6], fuzzy logic method [7], and so on. The forecast ability of wind speed above mentioned methods are not very excellent and the mean absolute percentage error (MAPE) usually located from 25% to 40%.

Wind speed can be considered as a non-stationary time series, it is difficult to construct the model for accurate forecast by traditional methods which based on the stationary signal. Considering the characteristics of wind speed, a forecasting method based on EMD, WT and LSSVM is proposed in this paper. Firstly, the wind speed was decomposed into several frequency bands. Secondly, the high frequency band was decomposed and reconstructed by WT. LSSVM is used to construct the high frequency and low frequency forecasting models. The last, forecasting results of the different frequency bands are combined to obtain the final forecasting results.

II. THEORY AND METHOD

A. Empirical Mode Decomposition

Empirical mode decomposition (EMD), proposed by Huang *et al*. [8], is a new method for decomposing multi-component signals. It has been found many immediate applications in a variety of problems covering geophysical and biomedical engineering. And, the method has been improved by Huang *et al* at 1999. EMD utilizes empirical knowledge of oscillations intrinsic to a time series in order to represent them as a superposition of components with well defined instantaneous freque- ncies. These components are called intrinsic mode functions (IMF). Huang *et al* have defined IMFs as a class of functions that satisfy two conditions: (1) In the whole data set, the number of extrema and the number of zero-crossings must be either equal or differ at most by one; (2) At any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The detail of the decomposition process is introduced by reference [8].

B. Wavelet Transform

Wavelet transform (WT) can be expressed as a basic wavelet or mother wavelet $\psi(t)$ through stretching factor a and translation factor b generating a function race $\{\psi_{a,b}(t)\}$.

$$\psi_{a,b}(t) = a^{-1/2}\psi\left(\frac{t-b}{a}\right), \qquad a>0, b\in R \tag{1}$$

The wavelet transform of the signal can be defined as:

978-1-4244-5848-6/10/$26.00 ©2010 IEEE

$$WT_x(a,b) = a^{-1/2}\int_{+\infty}^{-\infty} x(t)\psi^*\left(\frac{t-b}{a}\right)dt \qquad (2)$$

In Eq. (2), $\psi^*(t)$ is the complex conjugate function of the $\psi(t)$. Because of the time series is an ordered set of discrete data, so discrete wavelet transform (DWT) can be used to decomposition and reconstruction of the time series. In practice, Mallat algorithm [9] is the best method, it can be applied to discrete dyadic wavelet transform algorithm. The summary of the decomposition process is given by reference [9].

C. Least Squares Support Vector Machines

Support vector machines (SVM) for classification and nonlinear function estimation, introduced by Vapnik [10] and further investigated by many others, is an important methodology in the area of neural networks and nonlinear modeling. A modified version of SVM classifier, least squares SVM (LSSVM) classifier [11], was proposed by Suykens and Vandewalle (1999). A two-norm was taken with equality instead of inequality constraints so as to obtain a linear set of equations instead of a quadratic programming problem in the dual space. The formulation of LSSVM is introduced by reference [11].

III. MODELING PROCESS

Wind speed, as a kind of non-stationary time series, in order to improve the forecast precision, the forecast method based on EMD, WT and LSSVM is proposed in this paper. Firstly, the wind speed was decomposed into several IMF and trend term. Secondly, the high frequency band was decomposed and reconstructed by WT. The last, different LSSVM models to forecast each IMF and trend term are built up. These forecasting results of each IMF and trend term are combined to obtain the final results. Fig. 1 shows the detail process of the modeling.

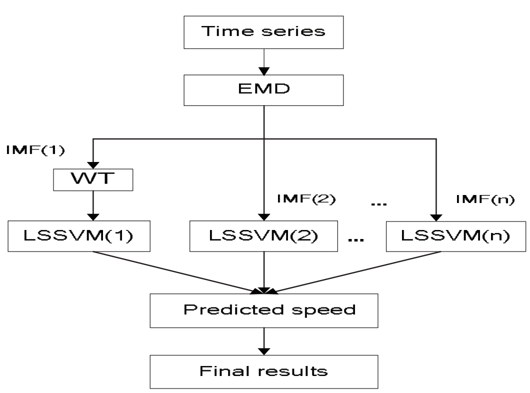


Figure 1. Prediction modeling process

IV. APPLICATION INSTANCE

A. Data

Hexi Corridor in northwest of China, stores a lot of wind energy. The Development and Reform Commission in China has been started to construct a typical base which would be produce 10 million kilowatts wind energy per year in Jiuquan. Until 2010, the installed capacity of the wind power base will be reach to 5 million kilowatts, and until 2015 the installed capacity will reach or beyond to 12 million kilowatts. So, it is very significant to forecast accurately the wind speed and power of Jiuquan wind farm. In short-term wind speed forecasting, the data used in the research were collected at a large wind farm in Jiuquan. The data contains totally 240 groups and each group stands for the minute average wind speed. The 180 groups are considered as the training samples, and the other 60 groups are considered as the verification samples.

B. Decomposed the Time Series Based on EMD

According to the algorithm of the EMD, MATLAB7.0 software is used as a simulate tool, so the time series can be decomposed into several IMFs and trend term R7. Fig. 2 shows the decomposed results.

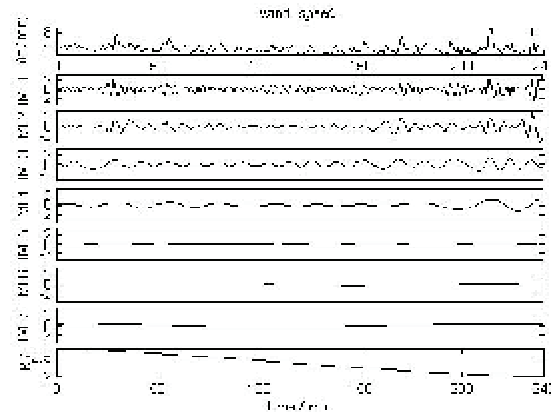


Figure 2. Wind speed and its EMD decomposition

C. LSSVM Model Prediction

According to the theory of LSSVM, model should be built under the training samples. Utilize the 180 groups minute average wind speed (12:00 pm to 15:00 pm, May 1st, 2009) as the training set. And, the other 60 groups minute average wind speed (15:00 pm to 16:00 pm, May 1st, 2009) as the test set. After grid search and cross validate, RBF kernel function was selected (*σ=0.039*, *γ=100*). Fig. 3 shows the final forecasting results.

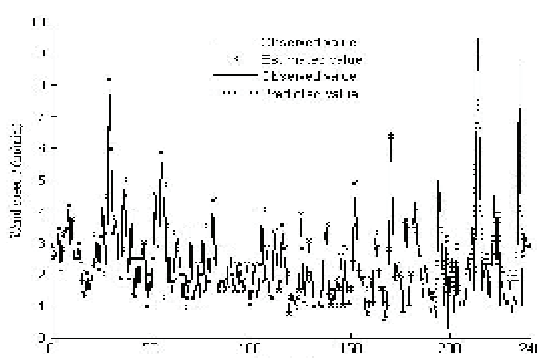


Figure 3. Forecasting results based on LSSVM

D. *EMD-WT-LSSVM Hybrid Model Prediction*

1) IMF1 and its wavelet transform

Use the *db2* wavelet function to decompose the IMF1, and IMF1 can be decomposed into three class time series. The low frequency approximate signal and high frequency detail signal should be reconstructed. So, the new low frequency approximate signal and high frequency detail signal can be obtained. Fig. 4 shows the decomposition and reconstruction process by WT.

2) Predicted results of the IMF1

After wavelet decomposing, different LSSVM models should be built to forecast each low frequency band and high frequency band. Forecasting results of the different frequency bands are combined to obtain the final forecasting results of IMF1. The RBF kernel function was selected to construct the LSSVM models. The forecasted results of IMF1 can be displayed in Fig. 5.

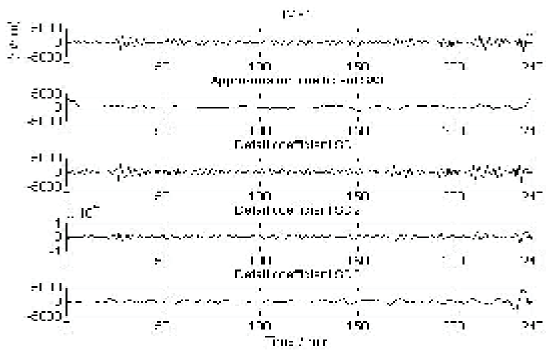
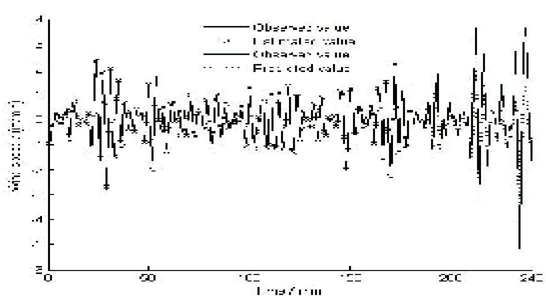


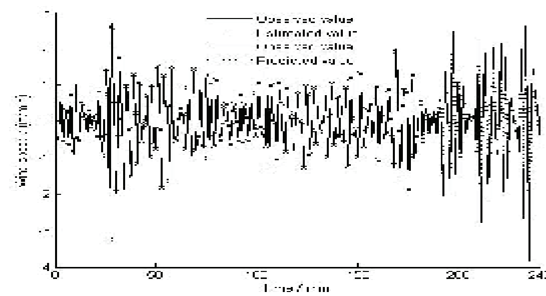
Figure 4. IMF1 and its wavelet decomposing



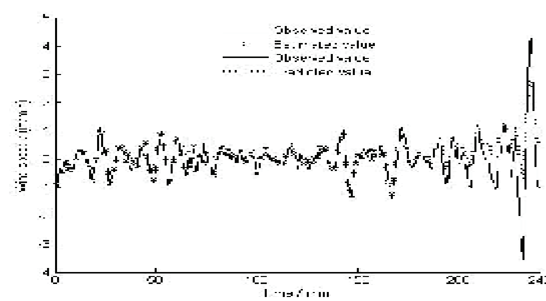
Detail coefficient SD2

3) EMD-LSSVM model Prediction

According to the different characteristics of IMFs and trend term, different LSSVM kernel functions should be selected. It is very important to choose the suitable kernel function because it can strongly influenced the performance of LSSVM. In order to make the computation process convenient, the same kernel parameters ($\sigma=0.039$, $\gamma=100$) would be selected. When forecast the trend term, polynomial kernel function would be selected and the parameter $d = 3$. Fig. 6 shows the forecasting results.

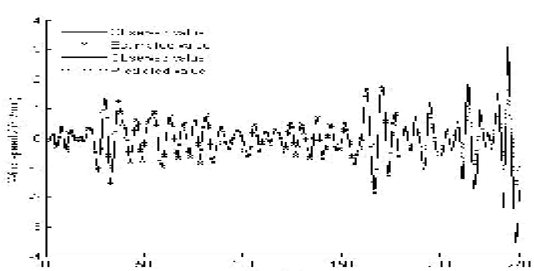


Detail coefficient SD1

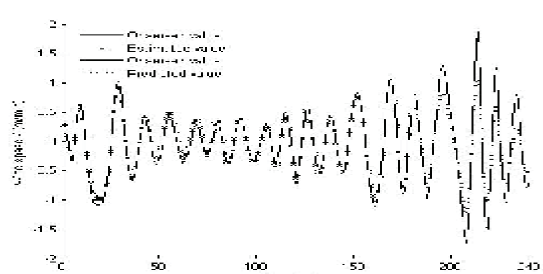


Detail coefficient SD3

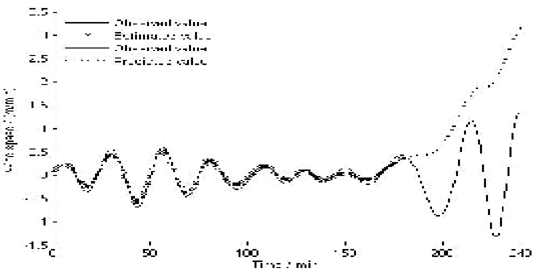
Figure 5. WT-LSSVM modeling and forecasting results



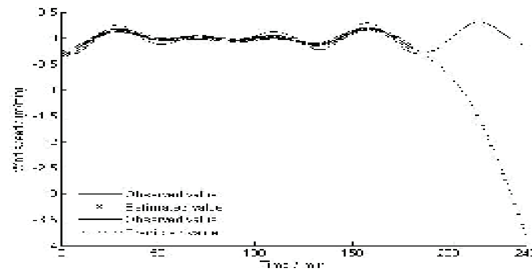
IMF 2



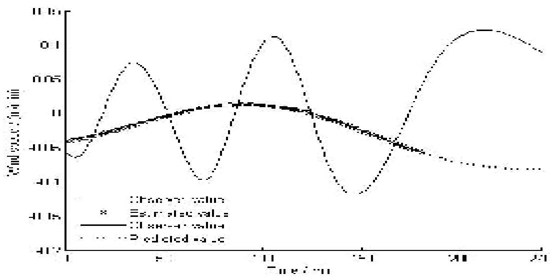
IMF 3



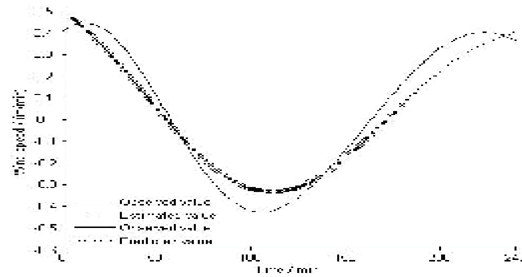
IMF 4



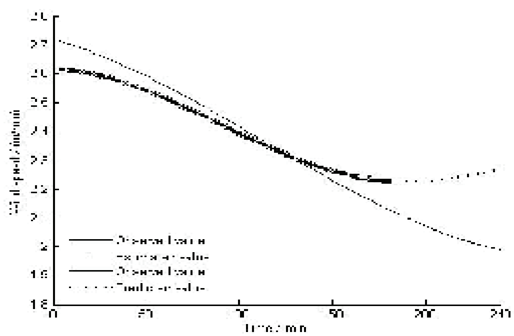
IMF 5



IMF 6



IMF 7



R 7

Figure 6. Forecasting results from IMF2 to R7

These forecasting results of each IMF and trend term are combined to obtain the final forecasting results. In order to demonstrate the effectiveness of the hybrid model, another forecasting model based on LSSVM has been built. Fig. 7 shows the final forecasting results of the LSSVM model and EMD-WT-LSSVM hybrid model.

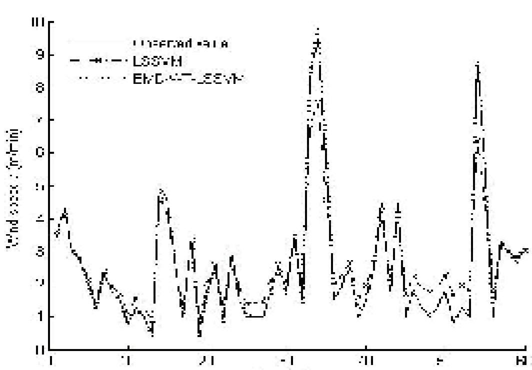


Figure 7. Final forecasting results of two models

E. Error Analysis

It is very significant to select the reasonable error analyze methods to judge the results. In this paper, the mean absolute percentage error (MAPE) and root-mean- squares of errors (RMSE) had been selected as the evaluation indicators. Fig. 8 shows the forecasting errors of the two models.

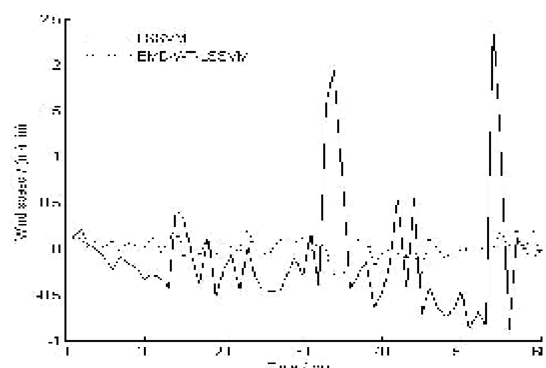


Figure 8. Forecasting errors of two models

The comparisons of the MAPE and RMSE for the proposed two models (EMD-WT-LSSVM model and LSSVM model), are showed in Tab. I.

TABLE I. COMPARISON OF FORECASTING ERRORS

Model	MAPE (%)	RMSE
LSSVM	16.38	0.626
EMD-WT-LSSVM	4.53	0.117

In Tab. I, the scope of the MAPE for two models changes from 4.53% to 16.38% and the RMSE changes from 0.117 to 0.626. According to the forecasting errors, it is clearly see that EMD-WT-LSSVM hybrid model is the optimal model because of the lowest errors. The simulation experiment demonstrates the hybrid model has the ability to improve the predict precision towards the non-stationary time series.

V. CONCLUSIONS

In this paper, the LSSVM model, EMD-WT-LSSVM model for forecasting wind speed were built up. The one hour ahead of wind speed in the wind farm has been predicted by each of the models. The error analysis shows that EMD-WT-LSSVM hybrid model is the optimal model for wind speed forecasting, because of the lowest MAPE and the RMSE. The simulation experiment shows that the hybrid model is an effective method to short term wind speed forecasting.

REFERENCES

[1] Shu Fan, James R. Liao, Ryuichi Yokoyama, et a1. "Forecasting the wind generation using a two-stage network based on meteorological information", IEEE Trans. on Energy Conversion, 2009, pp. 474-482.

[2] Landberg L. "Short-term prediction of the power production from wind farms", Journal of Wind Engineering and Industrial Aerodynamics 1999, pp. 207-220.

[3] Notis C, Trettel D, Aquino J, Piazza T, Taylor L, Trask D. "Learning to forecast wind at remote sites for wind energy applications", PNL-4318, Pacific Northwest Laboratory, 1983.

[4] Alexiadis M C,Dokopoulos P S,Sahsamanoglou H S. "Wind speed and power forecasting based on spatial correlation models", IEEE Trans. on Energy Conversion,1999, pp. 836-842.

[5] Guoyang W, Yang X, Shasha W. "Discussion about short-term forecast of wind speed on wind farm", Jilin Electric Power 2005, pp. 21-24(in Chinese).

[6] Wang X, Sideratos G, Hatziargyriou N, Tsoukalas LH. "Wind speed forecasting for power system operational planning", In: Proceedings of the 8th International Conference on Probablitistic Methods Applied to Power System, 2004.

[7] Damousis IG, Alexiadis MC, Theocharis JB, Dokopoulos PS. "A fuzzy model for wind speed prediction and power generation in wind parks using spatial correlation", Energy Conversion 2004, pp. 352-361.

[8] Huang N E, Shen Z, Long S R. "The empirical mode decompo- sition and the hilbert spectrum for nonlinear and non-stationary time series analysis", Proc R Soc London A, 1998, pp. 903-99.

[9] Stephane G. Mallat. "A theory for multiresolution signal decomposition: The wavelet representation", IEEE Trans. on Pattern Analysis and Machine Intelligence, 1989, pp. 674-693.

[10] Vapnik V N. "The Nature of Statiscal Learning Theory [M]", New York: Springer Verlag, l999.

[11] Suykens J A K, Vandewalle J. "Least squares support vector machine classifiers", Neural Processing Letters, 1999, pp. 293- 300.