

# Short-term Wind Power Prediction Based on Wavelet Transform-Support Vector Machine and Statistic Characteristics Analysis

Jie Shi

North China Electric  
Power University  
2 Beinong Road,  
Changping  
Beijing, 102206, China  
shijie0921@gmail.com

Yongqian Liu

North China Electric  
Power University  
2 Beinong Road,  
Changping  
Beijing, 102206, China  
yongqianliu@gmail.com

Yongping Yang

North China Electric  
Power University  
2 Beinong Road,  
Changping  
Beijing, 102206, China  
yyp@ncepu.edu.cn

Wei-jen Lee

Fellow, IEEE  
University of Texas at  
Arlington  
701 S. Nedderman  
Arlington, TX 76019,  
U.S.A  
wlee@uta.edu

**Abstract** -- The prediction algorithm is an important key factor in wind power prediction. However, there are pros and cons on different forecasting algorithms. Based on the principles of wavelet transform (WT), support vector machine (SVM) as well as characteristics of wind turbine generation systems, two prediction methods are presented and compared in this paper. In method 1, the time series of model input are decomposed into different frequency composes and models are set up separately based on SVM. The results are combined together to obtain the final wind power output. In method 2, the wavelet kernel function is applied in place of RBF kernel function in SVM training. To supply more valuable suggestions, the means of evaluating prediction algorithm precision is proposed. The operation data from two wind farms both in North China and U.S.A are used to test the usability of the method. The mean relative error of WT-SVM model (method 1) is less than that of traditional SVM model.

**Index Terms**—Prediction methods, support vector machines, wind power generation, wavelet transforms, uncertainty.

## I. INTRODUCTION

The development of renewable energy, especially wind energy is among top national policies by countries all over the world. Due to the intermittence nature of the wind generation, the increasing wind power penetration level will affect the operation of the grid. A reliable and accurate wind power prediction is one of the most effective solutions to deal with this problem. In general, the wind capacity forecasting methods includes Multiple Linear Regression (MLR), Auto-Regressive and Moving Average (ARMA), Artificial Neural Network (ANN), Fuzzy Logic and so on [3][4][7][9][16]. Though researchers have devoted significant efforts in developing algorithms on wind capacity forecasting, there are rooms for improvement. Figure 1 shows the difference between forecasting and actual output of wind generation in May 16, 2009 in the Electric Reliability Council of Texas (ERCOT). This forecasting error was due to a cold front approached and passed through the region. The difference presents challenges to the system operators.

In [5], M. A. Mohandes utilized support vector machine to predict wind power and compared the results to the ANN methods. It is indicated that the errors of support vector machine are lower than ANN.

To optimize the prediction model and improve the

prediction precision and calculation speed, this paper compares two wavelet transform-support vector machine models for wind capacity forecasting. The effectiveness of the models are verified with the actual output of wind farms in North China and Texas. Study shows that the Mean Relative Error on both proposed models are better than the traditional approach. By making use of statistical characteristics of prediction errors, the prediction scope can be obtained with certain confidence degree to support the wind farm operation.

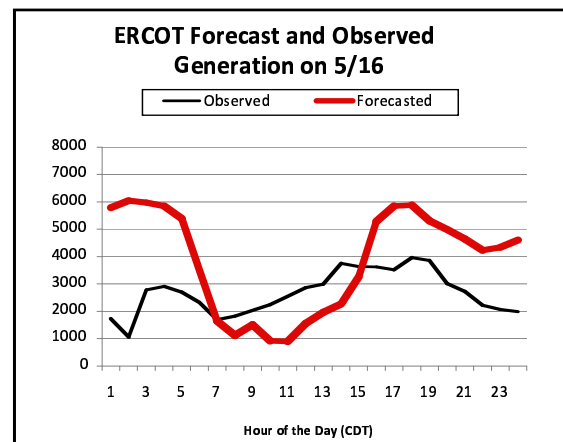


Fig. 1 ERCOT forecasted and observed wind generation in May 16, 2009 [1]

## II. SUPPORT VECTOR MACHINES (SVM)

Support Vector Machine (SVM) is a general learning method developed from Statistical Learning Theory with better performance than many other Routine methods. Statistical Learning Theory is based on a set of harder theory foundation, that provides a unite frame in order to solve the problem of limited sample learning. The basic idea of SVM applied to regression prediction is described as follows [12][15]:

Given the observations sample set  $P(x, y)$ ,  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n) \in R^n \times R$ , suppose the regression function is :

$$F = \{f \mid f(x) = \omega^T \cdot x + b, \omega \in R^n\} \quad (1)$$

Introduce the structure risk function:

$$R_{reg} = \frac{1}{2} \|\omega\|^2 + C \cdot R_{emp}[f] \quad (2)$$

Where,  $\|\omega\|^2$  is describing function;  $f(\cdot)$  is complexity term;  $C$  is constant which determines the trade-off between the empirical risk and the model complexity.

To construct the optimal hyperplane in the case when the data are linearly nonseparable, The main idea of nonlinear support vector regression is to map the input vector  $x$  into high dimensional feature space by nonlinear mapping function  $\phi(x)$  and then to perform linear regression in the feature space. In this higher space, there is higher possibility that the data can be linearly separated. Then the problem can be described as:

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^l \xi_i \quad (3)$$

Subject to:

$$y_i(\omega \cdot \phi(x_i) + b) \geq 1 - \xi_i, \xi_i \geq 0, i = 1, \dots, l, C > 0.$$

The inner products  $\phi(x_i)$  in the high-dimensional space can be replaced by some special kernel functions  $K(x_i, x_j)$ , which can be calculated. All the necessary computations can be performed directly in input space by calculation kernels. The popular kernels are shown as follows:

Radial basis function (RBF) kernel:

$$K(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \quad (4)$$

Polynomial kernel:

$$K(x, x_i) = (1 + x \cdot x_i)^d \quad (5)$$

Where  $\gamma$  and  $d$  are parameters. Different learning machines with arbitrary types of decision surfaces can be constructed by using various kind of kernel functions  $K(x_i, x_j)$ .

In Actual application, kernel function has the influence on realized effect. It is important to select proper kernel function to optimize the kernel function solution. As is mentioned above, Polynomial kernel function, RBF (Radial Basis Function) kernel function, and Sigmoid function are the three routine methods for kernel function.[8][17].

### III. WAVELET TRANSFORM (WT)

#### A. The Definition of Wavelet Transform [11]

Assume that  $\psi(t)$  is a Square integral function,  $\psi(t) \in L^2(R)$ , and its Fourier transform  $\hat{\psi}(w)$  satisfies the condition:

$$\int_R \frac{|\hat{\psi}(w)|^2}{|w|} dw < +\infty \quad (6)$$

Take  $\psi(t)$  as a Wavelet base or Mother wavelet function. The mother wavelet function  $\psi(t)$  can be extended and

translated into:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right), a, b \in R; a \neq 0 \quad (7)$$

Just Consider as a Wavelet series, where  $a$  is scale factor,  $b$  is translation factor. Consider any function  $f(t) \in L^2(R)$ , the definition of Continuous wavelet transform can be expressed as:

$$W_f(a, b) = \langle f, \psi_{a,b} \rangle = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(t) \bar{\psi}\left(\frac{t-b}{a}\right) dt \quad (8)$$

Its Reconstruction formula is:

$$f(t) = C_{\psi}^{-1} \iint_{\infty} W_f(a, b) \psi\left(\frac{t-b}{a}\right) \frac{dadb}{a^2} \quad (9)$$

Because the wavelet  $\psi_{a,b}(t)$  is generated by mother wavelet function,  $\psi(t)$  plays an observation window role on the analyzed signals,  $\psi(t)$  must satisfy the Ordinary function Constraints condition, also:

$$\int_{-\infty}^{\infty} \psi(t) dt < \infty \quad (10)$$

Therefore,  $\hat{\psi}(w)$  is a continuous function which is shown as:

$$\hat{\psi}(0) \int_{-\infty}^{\infty} \psi(t) dt = 0 \quad (11)$$

In order to be numerically stable, except the complete reconstruction condition, assure  $\psi(t)$  Fourier transform to satisfy the following stability condition:

$$A \leq \sum_{-\infty}^{\infty} |\hat{\psi}(2^{-j}w)| \leq B \quad (12)$$

Where,  $0 \leq A \leq B < \infty$ .

#### B. Mother Wavelet Function

Replace mother wavelet function by certain forms, the Kernel function can be obtained. The space  $L^2(R)$  multi-resolution means that  $L^2(R)$  satisfy the monotonicity, flexibility, translation invariance, a series space  $\{\psi_j\}_{j \in \mathbb{Z}}$  that assure Riesz base to exist [6].

According to the theory of Wavelet Transform and Support Vector Machine Kernel function constructed condition, the One-dimensional Wavelet Transform Kernel function can be constructed[1][13]:

$$K(s, t) = \sum_{j,k} \psi_{j,k}(s) \psi_{j,k}(t) \quad (13)$$

The multi-dimensional situation can be obtained according to One-dimensional Spreading for Tensor theory:

$$K^d(s, t) = \prod_{i=1}^d K(s_i, t_i) = \prod_{i=1}^d \sum_{j,k} \psi_{j,k}(s_i) \psi_{j,k}(t_i) \quad (14)$$

### IV. WAVELET TRANSFORM-SUPPORT VECTOR MACHINE MODEL FOR SHORT-TERM WIND POWER PREDICTION

#### A. Data

As mentioned earlier, SVM has advantages in predicting

certain samples. All the data in this paper come from two wind farms located in China (called WF1) and U.S.A (called WF2). The data from WF1 are the average of 10-min-data by hours, while the data from WF2 are the average of 1-min-data by hours. The influence of climate, temperature, and pressure are not considered in this paper, only the cosine and sine of wind speed, the hour average of wind power outputs are taken as input value. Though the kinds of variables are reduced and the error will be larger than before, the predicting errors are below 20% which is allowed. Besides, the scale of input sample is simpler and it is easier and faster to adjust parameter. MATLAB7.5 is used as the experiment platform in this paper [14]. A toolbox named SVM\_SteveGunn is used to perform training and testing the sample data after the decomposed process of wavelet transform.

### B. Wavelet Transform-Support Vector Machine (WT-SVM) Modeling Process

Two methods are introduced in this paper to establish WT-SVM model. They are illustrated as follows:

Method 1: The time series of wind speed and wind power outputs is decomposed into different frequency components according to wavelet transform. The different SVM models to predict the components from high frequency to low frequency are established. After model training and testing, these predicting results of the different frequency bands are combined to obtain the final results.

Method 2: The structure of wavelet transform-support vector machine is nearly the same as that of RBF support vector machine. We then replace the RBF kernel function with wavelet kernel one when mapping the input vector  $x$  into high dimensional feature space and perform linear regression in the feature space [2]. The regression function is shown as follows:

$$f(t) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(t_i, t) + b = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \sum_{j_{\min}}^{j_{\max}} K_j(t_i, t) + b \quad (15)$$

Where,  $K_j(t_i, t) = \sum_k \psi_{j,k}(s) \psi_{j,k}(t)$  is the multi-resolution wavelet kernel function.

## V. CASE STUDYEN

### A. Model Data Conformation

The wavelet transform-support vector machine model is established in use of the historical data from both wind farms. Take series of 1-hour-ahead historical data, 2-hour-ahead and 3-hour ahead ones before testing moment including wind power output, wind speed and cosine and sine of wind direction as model input. Take 1-hour-ahead wind power output as output.

### B. Modeling Process

#### (1). Characteristic Analysis on Wind Turbine Power Curve

The wind turbine output is defined as follows [14]:

$$P_s = \frac{1}{2} \rho v^3 f C_p \quad (16)$$

Where,  $P_s$  is the output of wind turbine, kW;  $\rho$  is the air density, kg/m<sup>3</sup>;  $v$  is wind speed, m/s;  $f$  is area, m<sup>2</sup>;  $C_p$  is utilization coefficient of wind energy, for level axis wind turbine, the maximum is 0.593.

Take the output data from No.10 wind turbine generator in WF1 for example; a relationship curve is shown in Figure. 2.

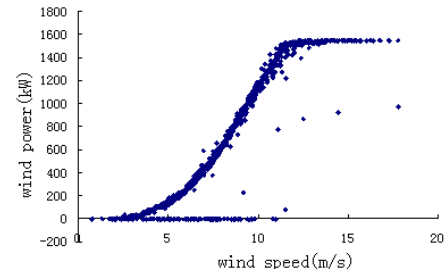


Fig. 2 The relationship between wind power and wind speed

From Figure 2, it can be seen that as the increase of wind speed the wind power gets larger and larger as a whole. There are two forms taking the point of 12m/s as the dividing point. That means there happens an inflexion at the point of 12m/s. When wind speed is below 12m/s, wind power increases like a curve. Whereas when wind speed is higher than 12m/s, the wind power stays steadily. As a result, the proposed model is divided into two parts for training. After judging whether the moment-ahead wind speed values are below or above 12m/s, the relating predicting model is chosen to obtain the more accurate output value. This piecewise support vector machine (PSVM) model [10] is utilized before both Method1 and Method 2 in this paper.

#### (2). Modeling Steps

The process of predicting wind power output:

Method1:

(a). The input data including wind speed, wind power output, cosine and sine of wind direction are taken as a non-stationary time series as well as training data and object, which are formed by wind power output at testing moment;

(b). The time series are decomposed into a stationary series in different frequency bands through wavelet decomposition;

(c). The decomposition results are dealt with using smoothing processing and normalization processing, that mean the historical data is changed between [-1,1];

(d). Different piecewise models are constructed using the PSVM model;

(e). The predicting results of different frequency bands models are combined to obtain the final predicting results;

(f). Anti-normalization processing is chosen after the predicting result is obtained, and then the mean relative error it is calculated.

Figure 3 displays the process of Method1 for predicting wind power output.

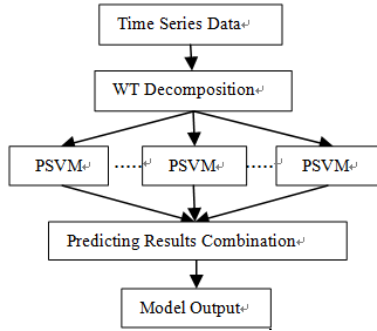


Fig. 3 The flow chart of predicting process using Method1

Method 2:

(a). According to the method mentioned in the last section, the training sample, training object, testing sample, and testing object are separately developed in MATLAB;

(b). Historical data are dealt with using smoothing processing and normalization processing, that mean the historical data is changed between  $[-1,1]$ ;

(c). Global variables are defined as  $PI$  and  $C$ , among which  $PI$  is the width of Kernel function while  $C$  is a coefficient. Both of them are given idiographic data according to different models;

(d). Kernel function wavelet and insensitive coefficient are chosen in this process;

(e). Toolbox in MATLAB7.5 is used to complete the training and testing;

(f). Anti-normalization processing is chosen after obtaining the predicting result. The mean relative error is calculated.

### C. Prediction Results and Uncertainty Analysis

#### (1). Prediction Results and Discussion

The historical data including wind power and wind speed as well as wind direction from both wind farms are utilized in this paper to predict 1 hour ahead wind power output. The prediction results are compared with true value, and the errors are illustrated as follows:

$$err = \frac{W_{pre} - W_{tr}}{W_{max}} \times 100\% \quad (17)$$

$$MRE = \frac{1}{N} \sum err \quad (18)$$

Where:  $W_{pre}$  is predicting value,  $W_{tr}$  is true value,  $W_{max}$  is the capability of wind farm which is the rated wind power summation of each wind turbine, 34500kW for WF1 and 120600kW for WF2,  $N$  is the sample scope.

Figure 4 and Figure 5 show the comparison between prediction value and true value using wavelet transform-support vector machine model with Method1 and Method2 model respectively.

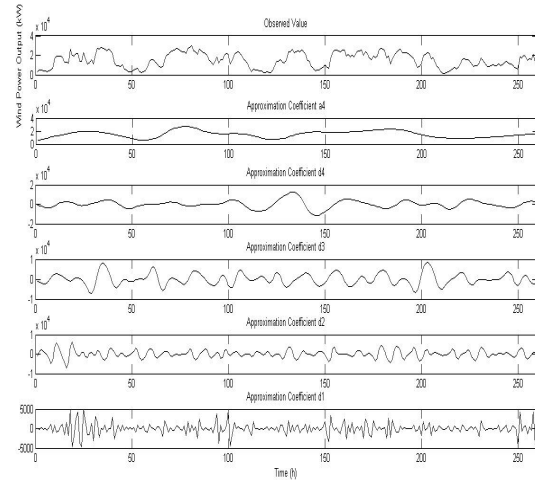


Fig. 4 Time series and wavelet decomposing (WF1)

Use the biorthogonal wavelet function to decompose the original wind power output series. The wind power output can be decomposed into four classes' time series. The low frequency approximate signal  $a4(k)$  (trend term) and each high frequency detail signal  $d_i(k)$  ( $i=1,2,\dots,4$ ) should be reconstructed to obtain the new  $a4(k)$  and  $d_i(k)$  ( $i=1,2,\dots,4$ ). Figure 3 shows the decomposition and reconstruction process by the wavelet. After PSVM training in every decomposition, the prediction results are combined together to obtain the final prediction results using method 1. The MSEs are 25.7% in WF1 and 27.1% in WF2 respectively.

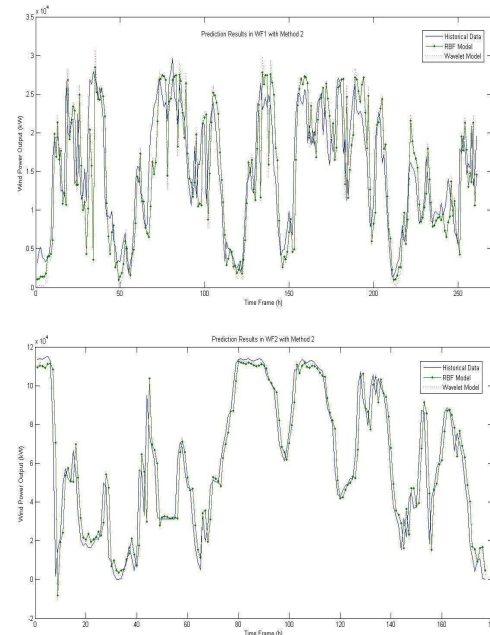


Fig. 5 The prediction results of method 1 in WF1 and WF2

It can be seen from Figure 5 that the prediction results in WF1 and WF2 using wavelet transform-support vector

machine model have better precision than that with RBF model. The MREs of prediction results vary from 10.72% (in WF1) and 7.15% (in WF2) to 10.23% (in WF1) to 7.05% (in WF1), which shows that wavelet kernel function performs better than RBF kernel function. The reason why model setting in this paper is better than the traditional one is that the wavelet has the characteristics like multi-resolution to observe data from large scope to small one. Therefore, it is useful and meaningful to make use of wavelet kernel for model setting.

Compared to two methods using above, one can see that method 2 is better than method 1 and RBF SVM model. Prediction results from method 1 are not practical, and the precision is worse than method 1 and RBF SVM model. That because the wind power output data, which can be seen as time series, has not so many similarities with biorthogonal characteristics. Method 2 offers better accuracy and faster calculating speed. Among wavelet transform support vector machine model with two methods and traditional RBF support vector machine model, the model in method 2 is the best option.

(2). *Uncertainty analysis*

The prediction results of wavelet transform-support vector machine model meet the precision requirements of grid and wind farm operation, but it is not enough for the need of analyzing market risk in wind power. The probability prediction based on statistical characteristics of prediction errors can make grid understand the uncertainties of wind power in the future. Therefore, it can help grid decision making. In addition, wind farm is able to hold priority in the hard competition according to the prediction errors. Therefore it is of great significance to study the distributing of prediction errors and perform probability prediction. The whole steps are described as follows:

(a). The relationship between wind speed and wind power is obtained based upon the historical data. In this paper, the historical data of No. 10 wind turbine from 2006-2-21 to 2006-2-28 are chosen to form the figure 5. The scope of wind power output is from 0 to 1600kW.

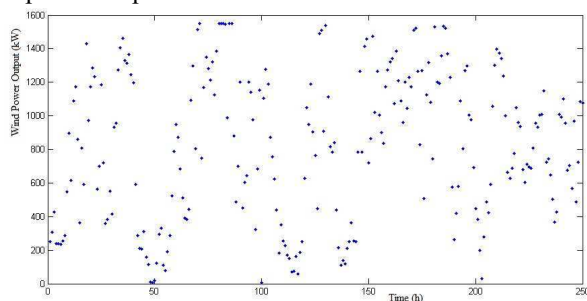


Fig. 6 The relationship between wind power and wind speed

(b). According to the wind power output, the data samples are divided into 4 groups between maximum value (1551kW) and minimum value (4kW) to set up probability density in each scope. It is shown in Figure 6 that the data distribute nearly equally in the four parts. That means that in each part there are plenty of points, which meet the truth that

the frequency histogram, which is nearly the same as the probability density of population distribution, can be utilized as probability density.

(c). The distribution of wind power is analyzed; the probability density curves of predicting wind power value in each wind power scope are established. The historical data from 2006-2-1 to 2006-2-20 are taken as training sample. After model training, the predicting values from 2006-2-21 to 2006-2-28 are obtained. Because there are some errors coming from the input value and NWP (Numeral Weather Prediction, which is also the input of model), the prediction value deviate the true value to an extent. Thus, in each wind power group, this deviation is taken as x-axis, while the frequency of it is taken as y-axis. The probability distribution of the deviation between prediction value and true value can be obtained. The relating probability density curves of prediction wind power are got as well as discretization, which are shown in Figure 7.

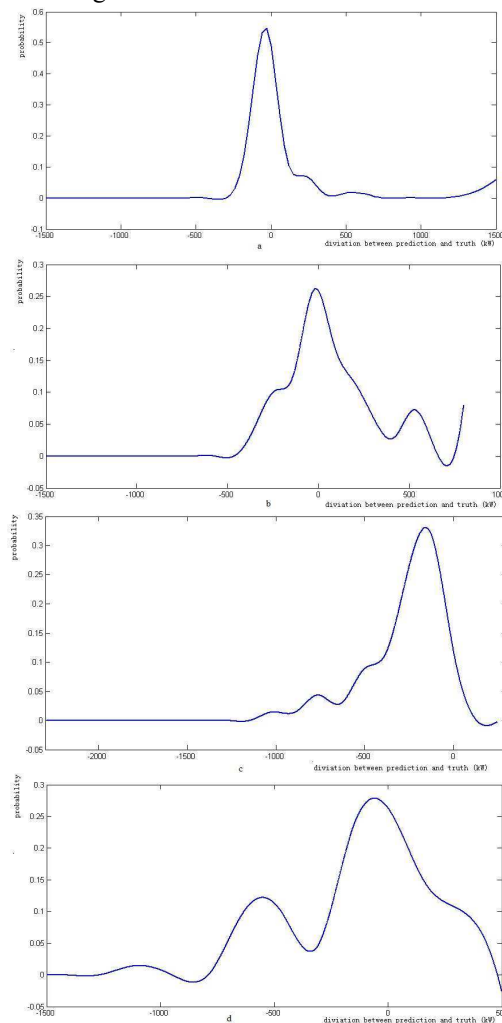


Fig. 7 Probability density curves in four wind power intervals

(d). The uncertainty analysis is proposed based on those curves. After judging which part the wind power belongs to, choose one curve for the analysis object; a positive number

'a' is generated via 0-1 random number generator, the corresponding x-axis stands for the deviation between prediction value and true value; with this deviations, the true value can be obtained. With enough sample space, the wind power true value should follow the probability density curve in Figure 7.

(e). Execute the step (d) for 1000 times to obtain 1000 true possible value.

(f). Sort the results of step (e) in descending order to get the scope of prediction value. When the confidence degree is 95%, the 950th prediction value is the upper limit, and the 50th one is the lower limit.

Statistical analysis is proposed to those prediction values to obtain the probability distribution of deviation between prediction value and true value in each wind power scope. The prediction value from 2006-3-20 to 2006-3-31 can be obtained through wavelet transform-support vector machine model. After uncertainty analysis, the confidence intervals of prediction wind power with 95% confidence degree are shown in figure 8.

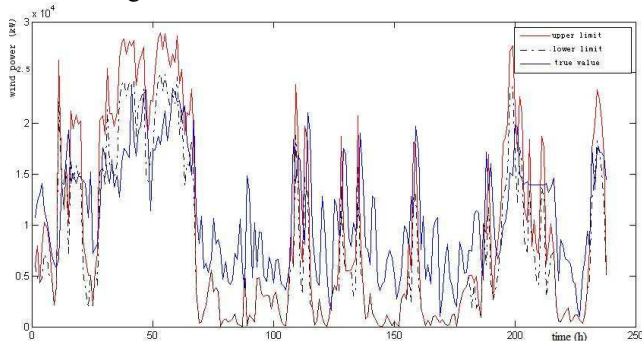


Fig. 8 Uncertainty analysis results (95% confidence degree)

Figure 8 shows that the exact prediction results are nearly in the middle of interval between upper limit and lower limit, and the wind power fluctuates seriously according to different wind speed and wind direction. Thus it is of great importance to use uncertainty analysis.

## VI. CONCLUSIONS

Study and analysis on short term wind power prediction based on historical data are processed in this paper, and the main content and conclusions are shown as follows:

(1). The algorithm of wavelet transform-support vector machine is set up with the utilization of wavelet transform and support vector machine principles;

(2). The method 1 of wavelet transform-support vector machine model is set up according to the characteristics of wind turbine and the principles of wavelet transform, The method 2 model is built up based on the characteristics of wind turbine and the replacement of RBF kernel function;

(3). The case studies in WF1 and WF2 show that the wavelet transform-support vector machine models (method 2) outperform the RBF support vector machine for 1 hour ahead prediction.

(4). The exact prediction results are changed into intervals, and the proper evaluation is proposed with certain confidence interval. The method proposed in this paper reflects potential risk factor in prediction, which supplies precondition for reliability study.

## VII. ACKNOWLEDGMENT

The authors gratefully acknowledge the support from National High Technology Research and Development Program 863: wind power prediction method study and system development, whose serial number is 2007AA05Z428 and National Science and Technology Support Program grant 2008BAA14B03 to Yongqian Liu and Jie Shi. The work described in this paper was supported by a grant from National Natural Science Foundation of China (Project no. 50976034) and the Fundamental Research Funds for the Central Universities (Project no. 09MG17). The authors are also grateful to the valuable advice of the reviewers.

## REFERENCES

### Technical Reports:

- [1] John W. Zack, "Overview of the Current Status and Future Prospects of Wind Power Production Forecasting for the ERCOT System," the Electric Reliability Council of Texas (ERCOT), Austin, TX, Jun. 2009.

### Periodicals:

- [2] A.R., "Non-parametric regression with wavelet kernels." *John Wiley and Sons Ltd*, vol. 21(2): pp. 153-163, 2005.
- [3] A., S., "A comparison of various forecasting techniques applied to mean hourly wind speed time series." *Renewable Energy*, vol. 21: pp. 23-35, 2000.
- [4] Liu Y.Q., H. S., Yang Y.P., "Study on Combined Prediction of Three Hours in Advance for Wind Power Generation." *ACTA ENERGIAE SOLARIS SINICA*, vol. 28(8): pp. 839-843, 2007.
- [5] M.A., M., "Support vector machines for wind speed prediction." *Renewable Energy*, vol.29: pp. 939-947, 2004.
- [6] Rakotomamonhy A., C. S. A. F., "Reproducing kernel, regularization and learning." *Journal of Machine Learning Research*, vol. 6: pp. 1485-1515, 2005.
- [7] V.N., V., "The nature of statistical learning theory." *John Wiley and Sons Inc*: 156-160, 1999.
- [8] X.G., Z., "Statistical Learning theory and support vector machine." *ACTA AUTOMATICA SINICA*, vol. 26 (1): pp. 32-42.
- [9] Z.S., H. Z. a. C., "Use of time-series analysis to model and forecast wind speed." *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 56: 311-322, 1995.
- [10] Yongqian Liu, Jie Shi, Yongping Yang, Shuang Han, "Piecewise Support Vector Machine Model for Short Term Wind Power Prediction." *International Journal of Green Energy*, vol. 6 (5): pp. 479-489, 2009.

### Books:

- [11] D.F., Z., *MATLAB wavelet analysis and industrial application*, Beijing: National Defense Industry Press, 2007, p. 81.
- [12] L., D. H. B. C. J. C. A. K., "Support vector regression machines," *Advances in Neural Information Processing Systems*, 1997, p.24.
- [13] Liu G.Z., D. S. L., "wavelet analysis and application," Xi'An, XI DIAN University Press, 1992, p. 56.
- [14] W. G., L., "MATLAB programming course," Beijing: China Water Power Press.2005, p. 81.
- [15] Y.J., D. N. Y. A. T., "New data mining method-Support vector machine," Beijing: science press. 2004, p. 75.

*Papers from Conference Proceedings (Published):*

- [16] Y.P., H. S. L. Y. Q. A. Y., "Wind speed prediction model of neural network based on tabu search algorithm." in *Proc. 2007 2nd IEEE Conference on Industrial Electronics and Applications*, pp. 23-25.
- [17] Theodore B., T. A. H. I., "Support vector machine for regression and application to financial forecasting." In *Proc. 2000 IEEE-INNS-ENNS international joint conference on neural networks*, vol. 6: pp. 348-353,