A robust probabilistic wind power forecasting method considering wind scenarios

***Jie Yan*[1,2], *Yongqian Liu*[1], *Shuang Han*[1], *Chenghong Gu*[2], *Furong Li*[2]**

[1]State Key Laboratory of Alternate Electrical Power System with Renewable Energy Sources, North China Electric Power University, Changping district, Beijing, 102206, China (yanjie_freda@163.com)
[2]Department of Electronic and Electrical Engineering, University of Bath, Bath BA2 7AY, U.K

Keywords: wind power, robust, probabilistic forecast, power curve, power generation process.

Abstract

Wind power forecasting is one of the cheapest and direct methods to alleviate negative impacts on power system reliability and stability from intermittent wind generation. Compared with deterministic forecasts, probabilistic forecasts can provide additional information concerning wind uncertainty for economic operation and efficient trading. However, it is far from ideal with respect to the accuracy, reliability and sharpness, since the wind shows strong variable property. In this paper, a robust probabilistic wind power forecasting method is proposed as (RPWPF) that can reflect the variability of wind generation under different wind conditions. The wind scenarios are identified concerning wind generation process and dominance of wind direction in a wind farm. And then, forecasting models for each scenario can be established and executed separately so that model parameters, such as kernel function and kernel width will be adjusted with the changing external wind conditions, wind speed and direction, in a real time operation. In this way, the forecasting model will provide more fined information on power outputs and their variabilities under different wind conditions. This proposed model is validated through comparison between the simulated power outputs and their variabilities under differing wind speeds and directions with the actual outputs on a practical 183 MW wind farm in northwest China. The results show that RPWPF achieves lower root mean square error comparing with artificial neural network model, while higher skill score for forecasting interval comparing with quantile regression. Finally, a sensitivity analysis is carried out to investigate the contribution of individual input (NWP variables) to help optimize the dimension of forecasting model.

1 Introduction

Wind power has gained significant attention in many countries worldwide in the context of a move towards cleaner and more sustainable energy. However, the intermittent nature of wind energy can introduce significant risks and thus additional costs in maintaining power system reliability, stability and power quality [1].

Wind power forecasting technology is one of the most effective ways to mitigate these negative impacts when the reliable prediction results are available. Many researchers have made significant efforts to develop and optimize wind power forecasting (WPF). As for short-term wind power deterministic forecasting, artificial neural network (ANN) method is the mainstream one which obtains competitive performance and various optimization scheme [2,3]. As for probabilistic wind power forecasting (PWPF), quantile regression [4,5] method is one of the most commonly used benchmark method. However, with these methods the characteristics of the power generation process are simulated by a fixed model, which makes the forecasting model less robust towards variable wind conditions. The problems are even more complex if the spatial and temporal complexity [6,7,8] is considered. So, it is significant to improve the robustness of WPF model and PWPF model.

To solve the above problem, a robust probabilistic forecasting method (RPWPF) has been proposed, where model parameters are adjusted according to changing external wind conditions in a real time. First, the wind scenarios are identified concerning wind generation process and dominance of wind direction in a wind farm. And then, it could accurately project wind generation with corresponding model for different situations. A case study for a wind farm in northwest China is presented where the performance of the proposed model is compared with deterministic predictions based on ANN. The reliability and skill score of the probabilistic forecasting are also assessed comparing with benchmark method - quantile regression (QR). Results show that the proposed probabilistic model has better prediction accuracy, better fits in instantaneous wind, and better robust ability. Moreover, the uncertainty estimates are provided in a more reliable way.

The organization is outlined as follows. In section 2, the mathematical theory of relevant vector machine forecasting model is introduced. The way to identify wind situation and model description is followed in section 3. The case study is in section 4. Finally, the conclusion is summarized in section 5.

2 Theory of forecasting model

The forecasting function is commonly defined by a kernel function, given a pairs of input-target sample $\{x_n, y_n\}$.

$$y(x;w)=\sum_{i=1}^{M} w_i K(x,x_i)+w_0 \tag{1}$$

where, w_i is weights vector; $K(x,x_i)$ is the kernel function; M is the sample size.

The probability of the samples is defined as

$$p(y|w,\sigma^2)=(2\tau\sigma^2)^{\frac{M}{2}} e^{\frac{1}{2\sigma^2}\|y \;\; \Omega w\|^2} \tag{2}$$

where, $\Omega(x)=K(x,x_i)$ is the vector of kernel function.

Then, a 'prior' probability distribution is imposed.

$$p(w|\alpha)=\prod_{i=0}^{M} N(w_i|0,\alpha_i^{\;1}) \tag{3}$$

where, α is 'hyperparameters'; N represents a Gaussian distribution with variance σ^2.

The posterior probabilities over unknown samples are defined as.

$$p(w,\alpha,\sigma^2|y)=\frac{p(y|w,\alpha,\sigma^2)\times p(w,\alpha,\sigma^2)}{p(y)} \tag{4}$$

The aim is to search for the optimal α and σ^2 by setting derivatives of the following equation.

$$p(y|\alpha,\sigma^2)=(2\pi)^{\frac{M}{2}}|\sigma^2 I+\Omega A^{\;1}\Omega^{T}|^{\frac{1}{2}} exp\;\; \frac{1}{2}y^{T}(\sigma^2 I+\Omega A^{\;1}\Omega^{T})^{\;1}y \tag{5}$$

Assuming that new test target is y_*, new test input x_*. Then, the distribution of new predictions can be written as:

$$p(y_*|y)=\int p(y_*|w,\alpha,\sigma^2)p(w,\alpha,\sigma^2|y)\mathrm{d}w\mathrm{d}\alpha\mathrm{d}\sigma^2 \tag{6}$$

With maximization of $\alpha_{MP}, \sigma_{MP}^2$, the new prediction is calculated as follows [9]:

$$p(y_*|y,\alpha_{MP},\sigma_{MP}^2)=\int p(y_*|w,\sigma_{MP}^2)p(w|y,\alpha_{MP},\sigma_{MP}^2)\mathrm{d}w \tag{7}$$

3 Model description

In this section, different wind scenarios are identified according to characteristic of power curve of wind turbines and complexity of wind climate; and then the training samples are established for each scenario; when the prediction models are built, the input variables (NWP data in this case) would be consecutively judged to decide which prediction model it belongs to at each prediction timeslot.

3.1 Robust forecasting model

Numerical Weather Prediction (NWP) results are commonly used as model inputs [10] due to the fact that there is a clear and strong bond between wind speed and power generation via power curve (PC). It is clearly shown in Table 1 that this mapping bond is different throughout power curve, especially. Apart from the well-known non-linearity on theoretical PC, the things are even complex on empirical one, because the existence of topographic effects, wake effects, hysteresis, and curtailments would broaden the deterministic curve in the real operation at a wind farm. Specifically, when wind is below cut-in wind speed, the empirical PC is a little thicker than theoretical one; while between cut-in and rated it is much more scattered; and it narrows down on a small scale between rated on cut-off. All these above implicate that the relationship between wind speed and wind power generation is distinguishing under different wind speed segments on PC. Therefore, it is easy to conclude the necessity of building separate prediction model for different wind speed segments.

Time	*Wind speed (m/s)*	*Wind power (kW)*
0:00	5.96	221
1:00	4.31	94
2:00	5.12	89
3:00	5.97	89
4:00	4.93	87
5:00	5.1	59
6:00	2.58	0
7:00	2.58	0
8:00	1.73	0
9:00	3.98	87
10:00	4.21	87
11:00	7.10	476
12:00	5.62	328
13:00	5.62	132
14:00	6.64	109
15:00	6.64	25
16:00	3.67	12
17:00	4.05	0
18:00	0.94	0
19:00	1.54	0
20:00	1.64	0
21:00	2.28	0
22:00	4.61	18
23:00	4.12	33

Table 1: Operational data of wind turbine installed in wind farm in question on 6th May

In Table 1, wind power almost linearly changes throughout wind speed ranges from 3 m/s to 12 m/s. There is an inflexion at the data point of around 12 m/s, which is the rated wind speed. After the rated point, that is, when wind speed is above 12 m/s and below 25 m/s, wind power stays almost steady at the rated generation power. Considering this nonlinear feature, the fixed model parameters may not be able to capture different generation characteristics along the power curve so that leaving the precision and robustness a large space to improve with moving model parameters. Therefore, by training model and forecasting separately according to the characteristic of power curve, the performance is expected to be improved than the previous works.

Besides, wind direction also has a significant importance in wind turbine operation and generation, for instance the yawing, as well as other atmospheric factors. Although minor impacts on power output compared with wind speed, it is still observed different characteristic of power generation towards different wind direction. The direction dependent PC models can lead to better accuracy when power prediction is transferred from predicted wind speed.

Fig.1 draws the wind direction rose map which clearly shows the prevailing wind climate pattern at wind farm mentioned in section 2. Direction sectors of N, NW, W, SW are prevailing sectors while the rest of sectors are unprevailing. Note that the direction sector could be divided in smaller slots with the condition that the data for each defined sector is rich enough to train prediction model or the wind climate pattern is complex. In this case, only two direction sectors are defined due to a relative clear dominance of wind direction and limited data amount.

According to the features of empirical power curve and dominance of wind direction, six models are separately built up for forecasting within different ranges of wind speed and direction. Given a pair of historical training samples, the forecasting model is trained and the resulting model parameters are stored to integrate into one whole RPWPF model. The new forecasting would be processed after judging which group the external wind condition belongs to. By implementing forecasting model with specific wind conditions, the proposed model is expected to be much more accurate and satisfies the characteristics of the power curves of wind turbine.

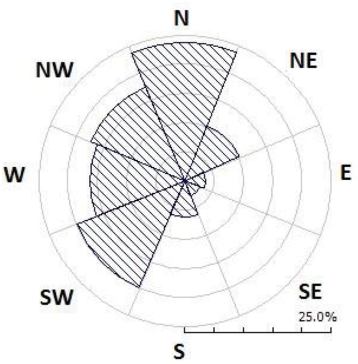

Fig.1: Wind direction rose map in 2010 year

Therefore, it is easy to conclude the necessity of building separate prediction model for different wind speed segments and direction sectors. Here, these speed segments along the power curve and direction sectors are referred as ‘wind scenarios’. As discussed above, wind speed is first to divide into three segments as “before cut-in”, “cut-in to rated”, “rated to cut-off”. Furthermore, considering the effects of wind direction variation, especially in wind farms with complex weather pattern, the samples are then divided by two wind direction sectors. The definition of different wind scenarios is shown in Table 2.

Direction sectors	Wind speed segments		
	Before cut-in	Cut-in to Rated	Rated to Cut-off
Prevailing	Wind scenario 1	Wind scenario 3	Wind scenario 5
Unprevailing	Wind scenario 2	Wind scenario 4	Wind scenario 6

Table 2: Definition of different wind scenarios

3.2 Implement steps

The process of building model is shown as follows:

1. To classify the training samples into groups according to the characteristic of power curve.
2. To establish the sample pairs for each classified group, including the training inputs, training targets; testing inputs, and testing objects.
3. Normalization processing: all the inputs data are integrated within the scope of [-1, 1] [11].
4. To calculate the posterior distribution over weights.
5 To update the mean and variance of posterior respectively.
6. To iterate until the convergence occurs.
7. To store the optimal model parameters for each group.
8. Anti-normalization processing.
9. To judge the forecasting group by the input wind speed.
10. To pick the corresponding model parameters and to make a new forecasting.

4 Case study

4.1 Data

To take a wind farm in Northwest China as example, the operational data and numerical weather prediction (NWP) data are obtained to train and test the proposed model. The available data includes wind speed, wind direction and wind power; cover a time period of 12 months. The example of time-series power output of one randomly selected wind turbine in this wind farm is shown in Fig.2. This figure is presenting the variability of wind in this wind farm. So, it is more convincing to indicate how the proposed model performs in such wind conditions with respect of its robustness ability.

The data is classified into two parts as training data that is used to build up the proposed forecasting model and testing data, which is used to verify the precision and reliability. To compare the performance of the proposed model, two counterparts – artificial neural network (ANN) and quantile regression (QR) are used as benchmarks.

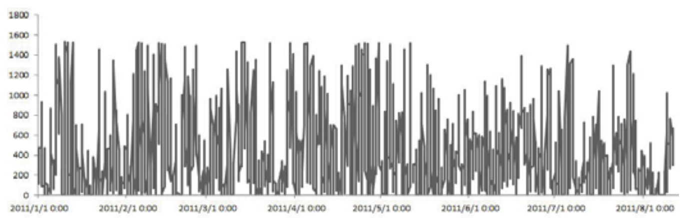

Fig. 2: Time series of power output from one wind turbine. X-axis is time series; Y-axis is power output (kW)

4.2 Criteria

Three commonly used statistical indices are employed to evaluate the error level and the reliability of the proposed RPWPF model, which are normalized root mean square error (NRMSE) in eq. (8), mean absolute error (NMAE) in eq.(9) [12] and reliability score in eq. (10), skill score in eq.(11) [13] .

$$NRMSE = \frac{\sqrt{\sum_{i=1}^{n}\left(P_{ai} \quad P_{pi}\right)^2}}{Cap \times \sqrt{n}} \tag{8}$$

$$NMAE = \frac{\sum_{i=1}^{n}\left|P_{ai} \quad P_{pi}\right|}{Cap \times n} \tag{9}$$

where, P_{ai} and P_{pi} respectively signify the actual and forecasting value of wind power output at i time; *Cap* indicates the installed capacity of a wind farm; *n* is the number of forcast samples involved.

$$r = \hat{\tau}^{(\tau)} - \tau \tag{10}$$

$$Sc = \frac{1}{N}\sum_{i=1}^{N}\left(\theta^{(\tau)} - \tau\right)\left(P_{m,i} - \hat{q}_i^{(\tau)}\right) \tag{11}$$

where, τ is the nominal confidence level; while $\hat{\tau}^{(\tau)}$ is the actual percentage validate results, where the measured wind power lies within the confidential range. N is the number of samples. $\theta^{(\tau)}$ is an indicator variable for the quantile forecasting $\hat{q}^{(\tau)}$. $P_{m,i}$ is the wind power observations at *i* time.

4.3 Results and analysis

To get an estimate of how well the proposed model performs, its results are compared to the ANN model and QR model which are shown in Table 3. The training data and testing data used for mentioned models are the same for the sake of fairness. The annual NRMSE of proposed model is 13.27%; while the ANN forecasting have 14.53% of NRMSE. For the NMAE, the model with moving model parameters has 10.79% of error level; while the NMAE of ANN model is 11.27%. Therefore, the NRMSE has decreased by up to 9.49% and the NMAE deceased by 4.44%. The improvement of annual forecasting accuracy validates the superiority of the proposed robust wind power forecasting method.

	NRMSE	NMAE	Reliability	Skill score
RPWPF	13.27%	10.79%	89.3%	-6.9
ANN	14.53%	11.27%	-	-
QR	-	-	88.9%	-7.8

Table 3: Comparison results of different models

In order to reduce the risk of forecasting, the uncertainty of forecasting results is provided with a nominal 90% confidential level. The results issued on four days are shown in Fig.3-6 to indicate the trend of actual power output, power forecasting and forecasting interval in different seasons. The reliability score is 89.3% and 88.9% respectively with robust probabilistic forecasting model and its benchmark QR method, which means that the percentage of the measured power output locating within the uncertain range is very close to the nominal confidence level. And the RWPPF method is slightly more reliable than QR. As for the skill score, an indicator for resolution or sharpness, a higher score value represents a better performance. So the proposed method has a 16.38% better result comparing with QR results.

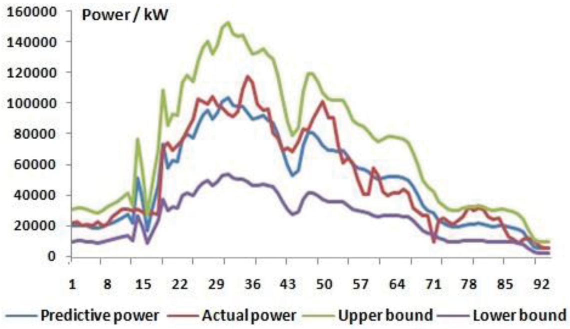

Fig. 3: Results issued on 25th May

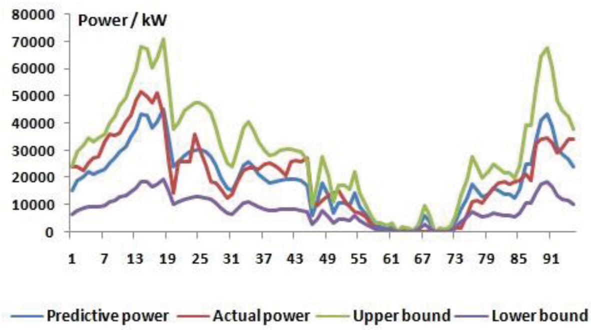

Fig. 4: Results issued on 29th July

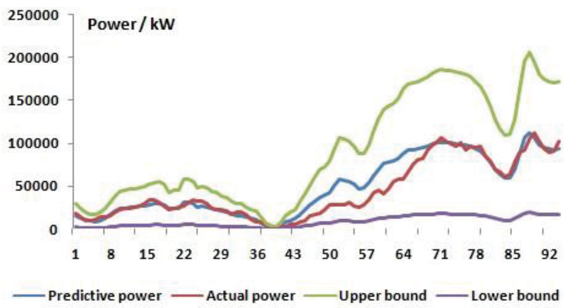

Fig. 5: Results issued on 25th Sep

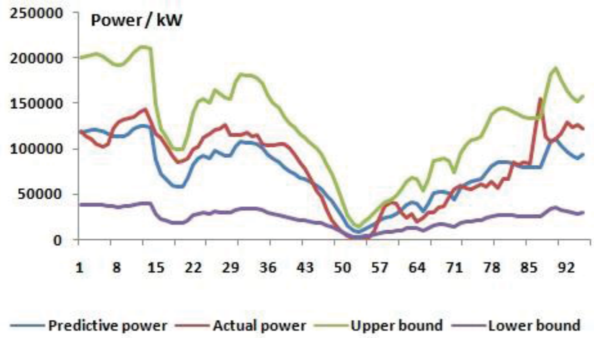

Fig. 6: Results issued on 26th Dec

Fig.7 shows the frequency distribution for the absolute error of point prediction; and Fig.8 is the distribution of uncertain fluctuation range (the range between the upper and lower bound) for 90% confidence. 80% of all point predictions have absolute errors in the range zero to 20MW whereas less than 1% has error over 60MW. To an extent, this validates the point prediction accuracy of the proposed model. The shapes of two distributions are similar except for the 20-60MW range. When the absolute error of point prediction is between 0 and 20MW, 54.1% of the uncertain ranges lie between 0 and 20MW, and only 0.2% is over 60MW. At the same time, when the absolute error is between 20MW and 60MW, 71.8% of the uncertain ranges lie between 20MW and 60MW, and 6.1% over 60MW.

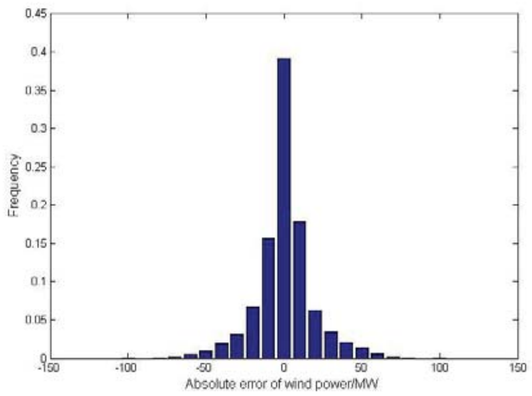

Fig. 7: Distribution of the absolute error of point prediction

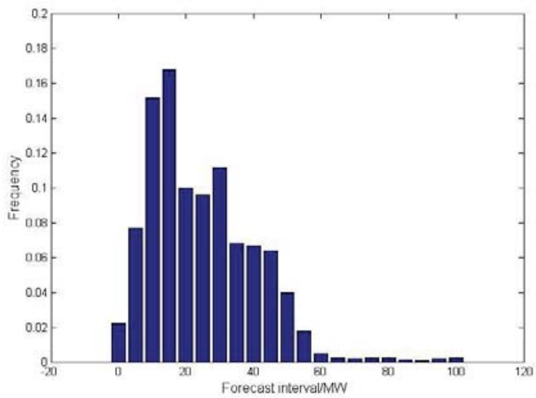

Fig. 8: Distribution of the uncertain range of wind power

4.4 Sensitivity analysis of individual model input

The proposed models is proved to be competent in terms of point prediction accuracy and probabilistic skill with all the available meteorological parameters as model inputs, containing wind speed, wind direction, pressure, temperature and humidity. However, it is important to learn the individual contribution of these 5 factors to the prediction accuracy. Knowing this could guide the researchers to balance the model complexity or computational burden and prediction quality, for instance to delete the less sensitive variable for the sake of time-effectiveness. Moreover, it could also help target efforts specifically on improving the accuracy of more sensitive NWP variables. Fig.9 depicts the NRMSE when individual NWP data streams were respectively inputted into the TVRVM model month by month. From this figure, it is observed that wind speed is a clear dominating factor for the prediction model since it could give an acceptable NRMSE around 19%, with wind direction a far second around 23%, and the others only have similar and minor effects almost 30%.

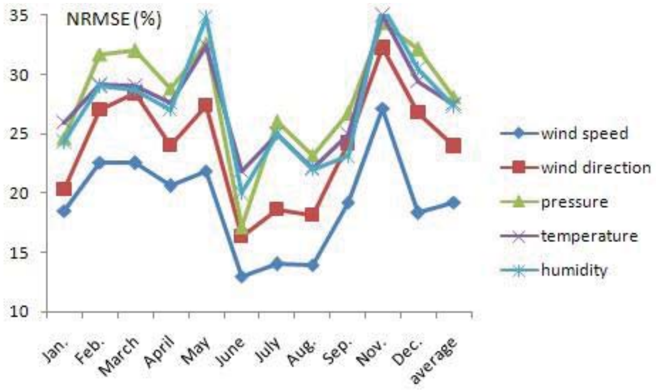

Fig. 9: NRMSE applying different individual weather variable

5 Conclusions

The proposed robust probabilistic forecasting model (RPWPF) has been shown capable of fitting on different wind scenarios thus providing improved forecasting performance in both deterministic and probabilistic way. The characteristics of the power generation and wind direction dominance are extracted to divide the wind into separate groups. The forecasting models are built up for each group in order to accurately simulate the power output with different external wind conditions. The results in a northwest wind farm show that the proposed model outperforms its counterpart ANN method with respect of NRMSE and NMAE. The accuracy improvement is up to 9.49% and 4.44%. As for the reliability and resolution ability of the uncertainty analysis, the proposed method has better performance comparing to mainstream method (QR) with 16.38% improvement of the overall score. Furthermore, sensitivity analysis is carried out to indicate the contribution of individual input variable to help optimize the dimension of forecasting model.

Acknowledgements

The paper is supported by the project of Physical Method Study for Wind Power Prediction based on CFD Numerical Simulation Database (Project No. 51206051) from National Nature Science Foundation of China.

References

[1] N. Bigdeli, K. Afshar, A. S. Gazafroudi. "A comparative study of optimal hybrid methods for wind power prediction in wind farm of Alberta, Canada", *Renewable and sustainable energy reviews*, 27, pp. 20-29, (2013).

[2] MC Mabel, E. Fernandez, "Analysis of wind power generation and prediction using ANN: a case study," *Renewable Energy*, 33(5): 986-92, (2008).

[3] E. Cadenas, W. Rivera, "Short term wind speed forecasting in La Venta, Oaxaca, México, using artificial neural networks," *Renewable Energy* 34 (1), pp. 274-8, (2009).

[4] Yongqian Liu, Jie Yan, Shuang Han, Yuhui Peng. Uncertainty analysis of wind power prediction based on quantile regression [C]. 2012 Asia-Pacific Power and Energy Engineering Conference, APPEEC 2012-Proceedings, March 27-29, 2012. Shanghai, China, (2006)

[5] H. A. Nielsen, H. Madsen, T. S. Nielsen, "Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts", Wind Energy, 9(1-2): 95-108, (2006)

[6] Drake B, Hubacek K, "What to expect from a greater geographic dispersion of wind farms-a risk portfolio approach", Energy Policy, 35:3999-4008, (2007).

[7] Yannick Degeilh Chanan Singh, "A quantitative approach to wind farm diversification and reliability", Electrical Power and Energy Systems. 33:303-314, (2011)

[8] J. Yan, Y. Q. Liu, S. Han, et al. "Wind power grouping forecasts and its uncertainty analysis using optimized relevance vector machine", *Renewable and sustainable energy reviews*, **27**, pp. 613-621, (2013).

[9] M.E. Tipping, "Sparse Bayesian Learning and the Relevance Vector Machine," Journal of Machine Learning Research, Vol.1, pp. 211-244, (2001).

[10] J. P. S. Catalao, H. M. I. Pousinho, V. M. F. Mendes. "Short-term wind power forecasting in Portugal by neural networks and wavelet transform", *Renewable energy*, **36**, pp. 1245-1251, (2011).

[11] Liu, W. G. 2005. MATLAB program design. Beijing: China Hydropower Publishing House.

[12] N. Amjady, F. Keynia, H. Zareipour. "Short-term wind power forecasting using ridgelet neural network", *Electric Power Systems Research Dec*, **81**, pp. 2099-2107, (2011).

[13] Pinson P, Nielsen HAa, Møller JK, Madsen H, Kariniotakis G, "Nonparametric probabilistic forecasts of wind power: required properties and evaluation," Wind Energy, 10: 497–516, (2007).