

Short-term wind speed prediction based on the wavelet transformation and Adaboost neural network

Zhou Hai^{1,2}, Zhu Xiang¹, Shao Haijian³ and Wu Ji¹

¹ China Electric Power Research Institute, No.8 NARI Road, Nanjing, China

² State Key Laboratory of Operation and Control of Renewable Energy & Storage Systems, China Electric Power Research Institute, No.8 NARI Road, Nanjing, China

³ Southeast University, Nanjing, China

zhouhai@epri.sgcc.com.cn

Abstract. The operation of the power grid will be affected inevitably with the increasing scale of wind farm due to the inherent randomness and uncertainty, so the accurate wind speed forecasting is critical for the stability of the grid operation. Typically, the traditional forecasting method does not take into account the frequency characteristics of wind speed, which cannot reflect the nature of the wind speed signal changes result from the low generality ability of the model structure. AdaBoost neural network in combination with the multi-resolution and multi-scale decomposition of wind speed is proposed to design the model structure in order to improve the forecasting accuracy and generality ability. The experimental evaluation using the data from a real wind farm in Jiangsu province is given to demonstrate the proposed strategy can improve the robust and accuracy of the forecasted variable.

1. Introduction

Wind energy as a renewable energy and green energy, compared with the traditional thermal power generation, can reduce the operation cost of power system and environmental pollution, therefore all countries are encouraging large-scale development of wind power[1]. However, due to the randomness, intermittency and volatility of wind power, the power quality is seriously affected, which brings uncertainty to the optimal scheduling of power system. Therefore, accurate wind speed prediction can not only improve power quality, but also have positive significance for real-time power grid scheduling and wind power grid operation. Neural network is a commonly used prediction method of wind speed. In theory, a feedforward network with a single hidden layer can approximate any complex functions, but in the actual project, people's prior knowledge is needed to set a reasonable network model and appropriate parameters to achieve satisfactory results. It is difficult for common engineers to realize. So the neural network ensemble method came into being in 90s, the basic idea is to improve the generalization ability of the system and reduce the dependence on the user's experience through the training of multiple neural networks and the synthesis of the results. AdaBoost [2] algorithm is an important feature learning algorithm in machine learning. The basic idea is to design a strong learner using multiple weak learning devices to enhance the learning ability of the algorithm. For samples of larger training error, larger weights are given, back and forth, a strong learner consisting of classification function sequences can be obtained [3].

Wind speed is essentially a kind of time signal with different frequency components, and its frequency spectrum are composed of amplitude spectrum and phase spectrum. Wavelet transform is a



characterization method using the oscillation wave of mother wavelet to scale and shift to match the input signal [4]. Wavelet transform can accurately reflect the characteristics of the signal in time and frequency, and analyze the characteristics of the signal in time-frequency domain, and then use a variety of resolution for signal analysis. Wind velocity can be regarded as a cumulative superposition of different frequency components with volatility and periodicity, if multilevel decomposition is done, the resolution with similar frequency characteristics of each component can be found. Moreover, by increasing the resolution at different scales, we can find the appropriate analysis method, and then get the high precision model according to the characteristics of each frequency component [5]. This multi-resolution method of layer by layer analysis is an important direction of wavelet applications [6-9].

The main structure of this paper is as follows. The first part mainly discusses the wind farm data. The second part examines the methods of input variables selection (Input, variable, selection, IVS), model order estimation, etc. In the third and fourth parts, we discuss the neural network ensemble method and the AdaBoost method respectively. The fifth part mainly introduces the error index which is used to evaluate the performance of the model. The sixth part mainly uses the actual data of a wind farm in East China to design the experiment. The seventh part is a summary of this article.

2. Wind farm data

In this paper, the actual data of a wind farm in East China is adopted for analysis. Data of this wind farm are collected by a weak wind FD-77 anemometer tower with a diameter of 77 meters. The sampling frequency is 5 minutes per point, and the data contain all 18 variables. There are 183212*18 data between 2011/06/01 and 2013/04/19 (some dates do not have the corresponding data), among them, there were 135 groups that lasted more than 24 zero points (2 hours), and the longest group has more than 196 continuous zero points. The observation height of the anemometer tower includes 10 meters, 50 meters and 70 meters. Variables include: the average wind speed and wind direction of 10 meters, 50 meters and 70 meters height, the standard deviation of data acquisition within 5 minutes, real-time wind speed, real-time wind direction and the temperature, the humidity and the atmospheric pressure of 10 meters height.

Without loss of generality, the data are divided into four groups as follows, according to observation height, and the data variables are numbered 1-18 respectively. Group 1(10m): x1 AWS, x2 AWD, x3 SSD, x4 RTWS, x5 RTWD; Group 2(50m): x6 AWS, x7 AWD, x8 SSD, x9 RTWS, x10 RTWD; Group 3(70m): x11 AWS, x12 AWD, x13 SSD, x14 RTWS, x15 RTWD; Group 4(Other): x16 temperature (10m), x17 humidity (10m), x18 atmospheric pressure (10m). Where, AWS represents average wind speed, AWD represents average wind direction, SSD represents sample standard deviation, RTWS is real-time wind speed, and RTWD is real-time wind speed.

3. Input variable selection

3.1. IVS strategy design

Different from other variables, an ideal variable input is a collection that is highly informationalized, independent of each other, has an appropriate number and is used to interpret the output. Therefore, the optimal set of input variables will use the least input variables to describe the characteristics of the output variables, and the optimal input variables are good for the structural design and generalization ability of the neural networks. Generally, for linear independent variable selection, methods include forward backward method and stepwise regression method and so on [10]. In fact, there are still many difficulties in the selection of nonlinear independent variables. In 1995, Dombi et al proposed the use of nonlinear models to filter the independent variable mean value (Mean impact value, MIV) method, which is recognized as the best independent variable screening method so far. MIV can be used to reflect the change of network weight matrix in neural networks, and determine the influence of the output neuron on the input neuron. The specific calculation process is as follows, after training of the neural network, we increase or decrease the independent variable feature in the training sample input index ISI (impact selection index) by 10% of its original value, then the two new training sample sets ISI1 and ISI2 are formed, they will be put into the network test respectively as input variables, and we

will get the test results S1 and S2, then we calculate the D-value and get the impact value of independent variables on output variables, after that, MIV will be calculated which has an impact on the network output variables. In this paper, MIV-ANN method is used to select reasonable independent variables as network input variables, which is used to evaluate the influence of each independent variable on the output variables, and then to improve the accuracy of the network model.

3.2. Model order estimation

In general, model order is often related to the current output variables, and model order or maximum time delay reflects the dynamic sustainability of the system. The greater the dynamic persistence of a system, the higher the model order, and the higher the dimension of the corresponding input variables. There are many methods for model order determination, such as correlation function, residual error method, F-test, criterion function method and so on. The correlation function can be used to measure the change of the correlation between any two sequences according to the data change over time, and the sub correlation function reflects the correlation between the current and the adjacent variables of its own sequence, and the partial correlation function can be used to describe the correlation between two sequences.

4. Adaboost algorithm

Sollich et al [11] define neural network ensemble as “neural network ensemble, which use finite neural networks to learn the same problems, while the corresponding outputs are determined by the outputs of the individual neural networks that make up the ensemble”. Taking the classification problem as an example, if the polynomial in the finite interval is used for sample identification and the accuracy is high, then the sample can be strongly learned [12]. If the accuracy is just better than the random guess, then the sample is weakly learned. We only need to find an algorithm better than the random guess, then we can upgrade the original algorithm to a strong learning algorithm, which will directly improve the generalization ability of the learning algorithm. This construction process is called the Boosting method. According to the AdaBoost algorithm, we adopt different weak learning algorithm for the training set, and then combine them into a strong learner to improve the overall learning generalization ability. The core idea is to focus on the samples with large prediction errors, and combine the weak learners of good performance to improve the sample weights and learning ability of the samples with poor training set. Boosting and Bagging proposed by Schapire[13] and Breiman[14] respectively were early used for the production of various networks, the existing network was used for the judgment of the error probability in the previous training set and improving it, so that it make the neural network ensemble able to deal with complex process. The AdaBoost (Adaptive Boost) algorithm proposed by Freund et al [15] later does not require the lower bound of the learning accuracy rate of the weak learning algorithm, so this method is very popular in practical applications. The training error of the final prediction function generated by the AdaBoosting method satisfies:

$$H_{\text{prediction}} = \prod_t \left[2(\varepsilon_t(1-\varepsilon_t))^{1/2} \right] = \prod_t \left[(1-4\gamma_t^2)^{1/2} \right] \leq e^{-2\sum \gamma_t^2} \quad (1)$$

Where ε_t is the training error of the prediction function, learned from the above formula, if the learning algorithm is slightly better than the random algorithm, the training error will descend with the t exponentially.

5. Wavelet signal decomposition and reconstruction

Since wind speed can be considered as a cumulative superposition of many different frequency components with fluctuation and periodicity, multilevel decomposition is adopted, then, the resolution with similar frequency characteristics of each decomposed component will be found, and the appropriate wavelet scale is used to decompose the relevant input variables at an appropriate resolution level. The feature prediction based on analysis and prediction of wavelet time-frequency domain is used for generating the target sequence, then, a suitable analysis method is adopted to obtain a high accuracy model based on the characteristics of each frequency component. Take the classic Mallat algorithm [16] as an example. W_j is the high frequency component of the corresponding signal.

The new elements are generated in turn according to the decomposition of the previous elements. After that, through layer by layer recursion, the original signal is mapped into the 2^j wavelet space, and then form a structured two fork tree. Suppose each node represents the wavelet packet coefficients

$\{d_{level-n}^{j,n}\}$ multiplied by corresponding wavelet packet function, thus we will obtain the component of the original signal in the wavelet packet space U_j^n . Wavelet packet decomposition and reconstruction algorithms are as follows:

$$\begin{aligned} d_{level-n}^{j,2n} &= \sum_k h_{k-2l} d_k^{j-1,n}, d_{level-n}^{j,2n+1} = \sum_k g_{k-2l} d_k^{j-1,n} \\ d_{level-n}^{j-1,n} &= \sum_k (\tilde{h}_{k-2l} d_{level-n}^{j,2n} + \tilde{g}_{k-2l} d_{level-n}^{j,2n+1}) \end{aligned} \quad (2)$$

Where h and g are corresponding low frequency and high frequency filter coefficients, while \tilde{h}, \tilde{g} are the corresponding coefficients of the reconstructed filter.

6. Model evaluation index

The robustness of artificial neural networks (ANN) depends on their network parameter values and their specific shape near the sample error surface. The network robustness is better if the network parameters are distributed at the sample extreme point and the corresponding error surface distribution is flat, otherwise the robustness is considered to be poor. Therefore, the robustness of the network output can fundamentally solve the practical problems, and make the network's generalization ability and application prospects better. We uses three kinds of evaluation index to evaluate the performance of models, they are RMSE (root mean square error), MAE (average absolute error) and RMAE (relative absolute error),

$$\begin{aligned} RMSE &= \left(\frac{1}{n} \sum_{t=1}^n (y_{tf} - \hat{y}_{tr}) \right)^{1/2} \\ MAE &= \frac{1}{n} \times \sum_{t=1}^n |y_{tf} - \hat{y}_{tr}| \\ RMAE &= \frac{1}{\sum_{t=1}^n |y_{tf}|} \times \sum_{t=1}^n |y_{tf} - \hat{y}_{tr}| \end{aligned} \quad (3)$$

RMSE reflects the discrete or deviation degree of the forecast sample and actual sample. MAE reflects the absolute deviation degree of prediction samples and real samples, and RMAE reflects the absolute deviation between prediction samples and the proportion of real samples, it is usually used to measure the prediction accuracy.

7. Experimental evaluation

7.1. IVS and average influence value

The average wind speed of 70 meters height is less influenced by the ground friction, so it is usually used as the observation point in engineering, and this is also the closest observation point with the equipment weak wind type FD-77 anemometer tower, so the average wind speed of 70 meters height is predicted in this paper. The corresponding IVS_MIV values are shown in Table 1.

Table 1. IVS_MIV

MIV	6 months	9 months
x_1	0.0039	0.0050
$\cos(x_2)$	-0.0048	-0.0065
$\sin(x_2)$	-0.0050	-0.0059
x_3	0.0012	0.0015
x_4	0.0032	0.0045

$\cos(x_5)$	-0.0050	-0.0061
$\sin(x_5)$	-0.0048	-0.0061
x_6	0.0074	0.0060
$\cos(x_7)$	-0.0050	-0.0060
$\sin(x_7)$	-0.0048	-0.0062
x_8	7.6983e-04	0.0030
x_9	0.0068	0.0060
$\cos(x_{10})$	-0.0050	-0.0060
$\sin(x_{10})$	-0.0049	-0.0063
x_{11}	0.0080	0.0062
$\cos(x_{12})$	-0.0052	-0.0062
$\sin(x_{12})$	-0.0048	-0.0057
x_{13}	5.3146e-05	0.0033
x_{14}	0.0077	0.0062
$\cos(x_{15})$	-0.0051	-0.0062
$\sin(x_{15})$	-0.0049	-0.0060
x_{16}	-0.0031	-0.0054
x_{17}	-0.0029	-0.0057
x_{18}	-0.0044	-0.0056

Similarly, the change of weight matrix of RBF network is used to determine the influence size of output neurons on input neurons. The above table includes 6 months and 9 months IVS_MIV value of the sample. For the wind direction, in order to avoid the difference between different angles, the sine and cosine inputs are used simultaneously to fully reflect the characteristics of wind direction. For this wind farm in East China, MIV absolute values of 50m average wind speed, 70m average wind speed and 70m real-time wind speed were above 0.0074 bigger than the other variables. Without loss of generality, x_6 , x_{14} and x_{11} are used as the input variables for modelling. The absolute value of the IVS_MIV is the index that reflects the impact on the output, and its sign only represents the correlation direction of its influence.

7.2. Order estimation

By examining the autocorrelation and cross-correlation of the sequence, the correlation coefficient between the variables is greater than 0.95 when the order is 3 or 4, and the correlation between the wind speed variable x_{11} and its correlated variables standard deviation and power is higher. Table 2 shows the model order under different criterion functions.

Table 2. Model Order

Season	AIC	BIC	SBC	MDL	PHI
Spring(x_{11})	3	2	2	2	2
Summer(x_{11})	3	2	2	2	3
Autumn(x_{11})	3	2	2	2	3
Winter (x_{11})	3	3	3	3	3
Spring (x_6)	3	2	3	3	3
Summer (x_6)	3	2	3	3	3
Autumn (x_6)	3	2	2	2	3
Winter (x_6)	3	3	2	2	2
Spring(x_{14})	3	2	2	2	2
Summer(x_{14})	3	2	2	2	3
Autumn(x_{14})	3	2	2	2	3
Winter (x_{14})	3	2	3	3	3

AIC, Akaike Information Criterion (Marple 1987, Wei 1990, Priestley 1981, Akaike 1974); BIC, Bayesian Akaike Information Criterion (Wei 1990, Priestley 1981, Akaike 1978-1979); SBC, Schwartz's Bayesian Criterion (Wei 1994, Schwartz 1978); MDL, Minimal Description length Criterion (Marple 1987, Rissanen 1978-1983); PHI, Phi criterion (Pukkila et al. 1988, Hannan 1980, Hannan & Quinn, 1979).

From the above analysis results, 3 and 2 are chosen as the order of x_6 , x_{14} and x_{11} simultaneously. In order to increase the calculation efficiency, ensure the reasonable structure and forecast data analysis efficiency and other reasons, we can choose the order of significant correlation only, and in order to

ensure the rationality of algorithm performance test, we compare the results through cross validation of the four seasons, then we investigate the rationality of the model structure.

7.3. Wind velocity decomposition

Four different wavelets (Daubechies, Coiflet, Symlet and Biorthogonal) are tested. Range settings: Daubechies, N=1-10, Coiflet, N=1-5, Symlet, N=5-10, and Biorthogonal, M=1-6, N= 1, 3, 5, 2, 4, 6, 8. Based on the 12 sub sets and based on the error evaluation index, the following conclusions are obtained: (1) Wavelet decomposition resolution. Daubechies(4): The overall performance of the subset is good; Coiflet(5): The local effect of signal feature is better, but the overall effect is worse than Daubechies; Biorthogonal Wavelet (4,4): Performance is better for refactoring on time series; Symlets(10): On the feature decomposition of time series, the dimension is larger, and the performance of reconfiguration is worse than that of Biorthogonal Wavelet; (2) Wavelet decomposition level. From the trend, wind speed fluctuation of level-1 and level-2 is more intense, instantaneity is relatively strong, while there is a good correlation between wind velocity variables and the input variables selected before on level-3, the low frequency signals obtained by wavelet decomposition can well track the change trend of the original wind speed, after level-4, the energy of signal decomposition is very weak, and even compared with the original wind speed, there is no better correlation and feature segmentation. We finally selects Daubechies (4, 2) as the final scale and resolution.

7.4. Experimental evaluation

The 12 step (i.e. 1 hours) forecast is carried out in this paper. 60% of the data in the sample is used as a training set, and 20% of the data is used as a validation set to test the current learning performance. The remaining 20% are used as a test set to test the resulting model structure. At the same time, in order to compare the performance, we will compare with the traditional design without any structure and without the AdaBoost method. Table 3 shows the experimental results of different methods.

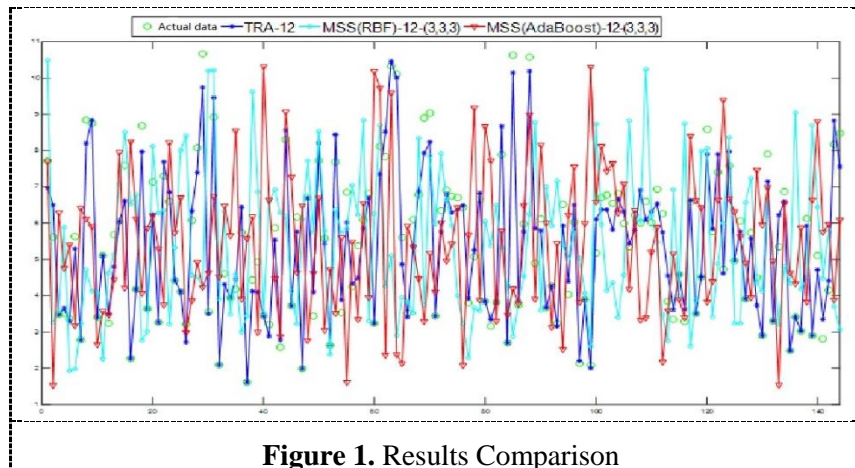
Table 3. Experimental results

	ERROR	TAR	MSS	MSS-Wav	MSS-Ada
Spring	RMSE	1.2953	1.5848	1.4688	1.5794
	MAE	1.4012	0.9022	0.9935	0.7446
	RMAE	0.1726	0.1219	0.1198	0.1212
	ET	698.68	269.09	171.81	152.62
Summer	RMSE	1.7303	1.3951	1.5113	1.6720
	MAE	0.3628	0.9016	0.9275	0.8442
	RMAE	0.1606	0.1218	0.1267	0.1136
	ET	768.99	269.09	191.09	176.25
Autumn	RMSE	1.2794	1.0343	1.1795	1.1295
	MAE	1.0202	0.7716	0.9376	0.9636
	RMAE	0.1676	0.1224	0.1259	0.1411
	ET	738.39	232.17	154.56	135.27
Winter	RMSE	1.1256	1.9306	1.4179	1.8573
	MAE	0.8505	0.7106	0.9214	0.8083
	RMAE	0.1513	0.1219	0.1217	0.1206
	ET	887.36	223.99	192.68	124.81

MSS-Wav refers to the combination of MSS and wavelet decomposition Daubechies(4, 2); MSS-Ada is a combination of MSS and AdaBoost and BP neural networks. Through the LM (Levenberg-Marquardt) learning algorithm, the learning of the sample is strengthened when the prediction error is more than 0.1, and the learning rate is 0.01.

The conclusions are as follows: (1) the stability and robustness of the prediction results are obviously increased through the MSS-Ada method. When the weak learning cycles is set to the value of 10 or around, the output remains basically unchanged, while the hidden nodes, after similar testing steps like SPREAD, are tested repeatedly to get the experience setting for each subset roughly (12,18). (2) The

frequency decomposition of wind speed can improve the prediction accuracy. After the wavelet transform, the accuracy of wind speed is obviously higher than that of TRA, and the time consumption is greatly reduced. Except the result of the summer and autumn corresponding to MSS-Wav and the autumn corresponding to MSS-Ada are slightly worse, the remaining accuracy is improved, and the computational time is reduced. Figure 1 shows the prediction results for the first subset.



To sum up, the design of this method can effectively improve the accuracy of wind speed prediction modelling.

8. Conclusion

A hybrid modelling method based on neural network ensemble and multi-scale wavelet analysis is proposed in this paper, which realizes the prediction of wind speed series. First, we use the method of neural network ensemble based on AdaBoost to improve the defect of overfitting of neural networks and the defect that is easy to fall into the local minimum, and enhance the adaptability of different samples to improve the generalization ability of neural network; Then, the non-stationary wind speed sequence is transformed into sub sequences with suitable spectrum range through the multi-scale wavelet analysis method, which reduces the adverse impact of non-stationary input on predictive modelling. Tests show that compared with the single neural network model method, the proposed hybrid modelling method has higher prediction accuracy.

9. References

- [1] Ackermann T 2014 *Wind power in power systems, Second Edition. Wind Power in Power System.* 361-386
- [2] Schwenk, Holger and Y Bengio 1997 *AdaBoosting neural networks. International Conference on Artificial Neural Networks.* 967-972
- [3] Wu Junli, Zhang Buhan and Wang Kui 2012 *Wind speed prediction based on AdaBoost and BP neural net-works. Power System Technology.* **36(9)** 221-225
- [4] Meyer, Yves 1992 *Wavelets and operators* (Cambridge: Cambridge University Press)
- [5] Zhou Yufeng, Cheng Jingquan 2008 *Wavelet tranfor-mation and application. Physics.* **37(1)** 24-32
- [6] Tai Xiongling, Hou Zhilian, Li Tao, Jiang Chuanwen and Song Jiong 2003 *New principle based on wavelet transform for power system short-term load forecasting. Processing of the CSEE.* **23(1)** 45-50
- [7] Wang Xiaolan, Wang Mingwei 2010 *Short-term wind speed forecasting using the wavelet decomposition and least square support vector machine. Power Systems Technology.* **34(1)** 179-184
- [8] Wang Lijie, Dong Lei, Liao Xiaozhong and Gao Yang 2009 *Short-term power prediction of a wind farm based on wavelet analysis. Processing of the CSEE.* **29(28)** 30-33

- [9] Liu Hui, Tian Hongqi and Li Yanfei 2010 *Short-term forecasting optimization algorithm for wind speed based on wavelet analysis method and rolling time series method*. *Journal of Central South University (Science and Technology)*. **41(1)** 370-375
- [10] Cui Zhiquan, Fu Xuyun, Zhong Shisheng and Wang Tichun 2013 *Aero-engine arguments selection based on wavelet network mean impact value*. *Computer integrated manufacturing system*. **19(12)** 3062-3067
- [11] Touretzky, Edited D S, P Sollich and A Krogh 1996 *Learning with ensembles: How over-fitting can be useful* (Advances in Neural Information Processing Systems) p 190-196
- [12] Jiangyuan, Cheng Zhaoqian and Zhou Zhihua 2003 *A rule learning algorithm based on Neural Network Ensemble*. *Computer research and development*. **40(10)** 1419-1423
- [13] Schapire R E 2002 *The strength of weak learnability*. *Foundations of Computer Science*. **5** 197-227.
- [14] Breiman L 1996 *Bagging predictors*. *Machine Learning*. **24(2)** 123-140
- [15] Freund Y, Schapire R E 1997 *A decision-theoretic generalization of on-line learning and an application to boosting*. *Journal of Computer and System Sciences*. **55(1)** 119-139
- [16] Mallat S E P 1999 *A wavelet tour of signal processing* (Academic Press) p 83-85

Acknowledgments

This work was funded by NSFC-NRCT (National Natural Science Foundation of China - National Research Council of Thailand) Joint Research Project (51561145011), SGCC (State Grid Corporation of China) Science and Technology Project (NY71-17-009) and CEPRI (China Electric Power Research Institute) Science and Technology Project (Research on Risk Zoning and Assessment of Typhoon Disaster in Coastal Power Grid).