

# The Study of GRNN for Wind Speed Forecasting Based on Markov Chain

Shujie Gao, Jianyan Tian\*, Fang Wang and Yang Bai  
College of information engineering  
Taiyuan University of technology  
Taiyuan 030024, China  
\*Corresponding author

Wei Gao, Shengqiang Yang  
College of Mechanical Engineering  
Taiyuan University of Technology  
Taiyuan 030024, China

**Abstract**—In view of modeling accuracy problems of General Regression Neural Network (GRNN), the improved GRNN has been used to forecast wind speed. Firstly, K-fold cross validation was used to select smooth parameter of GRNN, and the influence of K value was analyzed. Then, Markov Chain (MC) was introduced to correct the results of GRNN to improve the accuracy of GRNN. And C-average clustering algorithm was used to state division, the number of state and correction results were emphatically discussed. Finally, the improved GRNN was applied to wind speed forecasting by use of the actual data gathered from a wind farm in Shanxi province to further test the proposed method. The experimental results show that the superiority of the proposed method compared with GRNN.

**Keywords**—wind speed forecasting; general regression neural network; error correction; markov chain

## I. INTRODUCTION

Wind energy has been recognized as a kind of ample, secure, clean energy. People from all over the world paid more attention to it. The principle of wind power generation is to convert wind energy to power through wind turbines. The random, intermittence and uncontrollability of wind speed causes the volatility and intermittence of wind power [1]. Therefore, wind speed forecasting is the foundation of wind power prediction, and it is also an effective method of improving the power quality.

At present, there are three different approaches to forecast wind speed. The first one is physical approaches such as Numerical Weather Prediction model [2]. The second one is statistical techniques such as times series method [3] and Kalman filters [2]. The last one is intelligent method such as BP neural network [4], and wavelet neural network [5].

General Regression Neural Network (GRNN) is a remarkable method based on statistical theory and kernel regression. Its advantages include less initialization parameter, efficient convergence, robustness to noise, precise forecasting compared with other types of neural networks. It has been successfully applied to short-term wind power prediction, surface deformation prediction and traffic volume [6]. In this paper, the improved GRNN is used to forecast wind speed based on the above characteristics. The smooth parameter has a significant influence on the predictable performance. It is optimized by K-fold cross validation based on the advantages such as obtaining more information from learned metrics, avoiding the local minimum point effectively and over-fitting

problem to some degree [7]. It is essential to improve the predictable accuracy of GRNN to enhance the generalization ability of network since predictable errors of GRNN increase in the condition of volatile wind speed. According to relevant literatures of the generation of synthetic wind speed and wind power time series, Markov Chain (MC) is proposed to revise the results of GRNN.

The paper is organized as follow: Section II describes the GRNN model based on wind speed forecasting. The optimal parameter and wind speed forecasting are given in Section III. Section IV demonstrates the Markov Chain and discusses its revised results. The conclusion is in the section V.

## II. GRNN MODEL BASED ON WIND SPEED FORECASTING

GRNN is a kind of feed forward network based on the nonparametric kernel regression, which takes sample data as a posterior condition and exploits the maximum a posterior criterion to calculate its output. GRNN for prediction can be achieved as follow [8]:

$$\hat{Y}(X_0) = E[Y | X_0] = \frac{\sum_{s=1}^S Y_s \exp\left[-\frac{(X_0 - X_s)^T (X_0 - X_s)}{2\delta^2}\right]}{\sum_{s=1}^S \exp\left[-\frac{(X_0 - X_s)^T (X_0 - X_s)}{2\delta^2}\right]} \quad (1)$$

where  $\{X_s, Y_s\}_{s=1}^S$  ( $X \in R^p$  is the p-dimensional input vector and  $Y$  is corresponding output) is the sample set composed of random variable  $X$  and  $Y$ . The number of sample set is expressed in  $S$ .  $\delta$  is the smooth parameter.  $\hat{Y}(X_0)$  represents the condition mean of  $Y$  given  $X_0$ , which is obtained in the utilization of Parzen window estimation and integral transform. It is obvious that the predicted value  $\hat{Y}(X_0)$  is weighted average of all of observed values  $Y_s$ , where the weight of each observed value is exponentiation of its Euclidean distance from  $X_0$ .

GRNN model based on wind speed forecasting is established on the foundation of (1), which composes of input layer, pattern layer, summation layer and output layer [7] as shown in Figure 1. Based on our previous study, the temperature, barometric pressure and humidity at the moment of  $t$ , wind speed at the moments of  $t, t-1, t-2, t-3$  are adopted as input vector of neurons of input layer, and the input layer has

the same numbers as the dimensions of vector. Gauss function is chosen as the transfer function of neurons in pattern layer which shares the same numbers with wind speed sample set  $S$ . Summation layer consists of two kinds of neurons, one is applied to calculate the weighted summation of each output in the second layer, where the weights is replaced by the values of wind speed sample  $Y_s$ . The other is employed to calculate the sum of pattern layer. The wind speed at the moment of  $t+1$  is adopted as output.

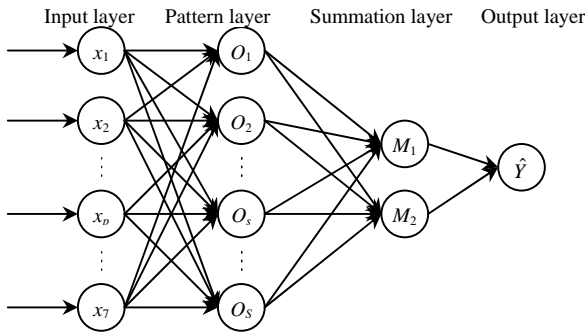


FIGURE I. GRNN MODEL BASED ON WIND SPEED FORECASTING

### III. K-FOLD CROSS VALIDATION FOR GRNN

#### A. Wind Farm Data

The real data is obtained from a wind farm in Shanxi province. It refers to average values collected every five minutes from wind towers. Data sieving, abnormal data erasing and interpolation are the prior steps of data processing. Then, the processed data is converted into information collected every 30 minutes to reduce the random influence of wind speed. The processed half-hour wind speed series of June in 2014 (including 1400 data) at a height of 70 meters (the installation height of wind turbines is usually higher than 50 meters in china) are applied in this part because of the limitation of space. The proposed method is used to forecast wind speed for half-hour ahead, where the first 1300 values of these data are used for training and the last 100 values are used for validation.

#### B. Parameter of GRNN

Smooth parameter is a critical parameter which has a significant influence on predictable performance of GRNN. During the process of optimization, the parameter is made large, making the curve of output more smooth and the network training complicated, and small smooth parameter will result in inaccuracy of forecasting. In this paper K-fold cross validation is chosen to optimize the smooth parameter. The optimizing principals are: the learning samples are divided into K groups firstly, then K-1 groups of samples are used to establish GRNN model. In the meanwhile, the remaining group is used to test the performance of model with Mean Absolute Error (MAE) inaccuracy of forecasting. In this paper, Analysis of influence of different K values until all the groups have the same time for testing. Finally, the optimum parameter is determined by the one corresponding to the

minimum value of MAE [9]. However, K value is made certain, the smooth parameter is determined. Therefore, how to determine K value is the key problem to solved in next part.

TABLE I. ANALYSIS OF INFLUENCE OF DIFFERENT K VALUES

K	$\delta$	Time(s)	MAE (m/s)	MAPE (%)	RMSE (m/s)
2	0.27	14.24	0.7738	19.16	1.0036
3	0.25	19.61	0.7701	18.85	0.9976
4	0.23	23.32	0.7632	18.78	0.9889
5	0.23	27.14	0.7601	18.62	0.9820
6	0.23	29.96	0.7631	18.74	0.9878
7	0.22	35.59	0.7610	18.80	0.9857
8	0.23	38.36	0.7617	18.90	0.9839
9	0.22	41.37	0.7590	18.64	0.9832
10	0.21	44.54	0.7615	18.64	0.9859

#### C. Evaluation and the Selection of K Value

For all performed experiments, mean absolute error (MAE), mean absolute percentage error (MAPE) and root-mean-square error (RMSE) are utilized as evaluation criteria for evaluating the predictable performance [5]. Considering the volatility of neural network, all the results are the average values of series of simulation experiments. The influence of different K values on forecasting results and simulation times are shown in Table 1. It shows that there is a progressive increment in simulation time. However, the forecasting results have no progressive decrement, and smooth parameter is comparatively stable as K value increases. K=5 is an appropriate value for forecasting when evaluation criteria and simulation times are considered together. In comparison of the results, it shows that the MAPE of GRNN is all greater than 18%, which is not accurate enough for application. Therefore, it is necessary to take necessary measures to improve the precision.

### IV. ERROR CORRECTION OF GRNN BASED ON MC

GRNN for wind speed forecasting is based on the historical sequences of wind speed and relevant factors with function of nonlinear mapping. It is inevitable to cause forecasting error. Therefore, it is necessary to make further revision for results of GRNN to improve wind speed forecasting accuracy. Markov Chain is a method of predicting the future state and 77development trend by state transition probability. It has been successfully applied to many fields such as hydrology, communications and geotechnical engineering [12].

#### A. Key Techniques and Process of Correction of MC

At present, there is no uniformed method for Markov Chain for state division, which has a significant effect on revision. Firstly, fuzzy C-average clustering algorithm is applied in state division to avoid the influence of artificial factors Then, the autocorrelation coefficient of error sequence is normalized as the weights of state probability to synthesize each kind of state probability Finally, the leveled characteristic value is used to judge the state of current wind speed in order to consider the influence of all state probability instead of maximum probability [12]. The flow chart of Markov chain prediction is shown in Figure 2.

It should be noted that the boundary between two states can be obtained by calculating the average of lower bound in previous state and upper bound in later state. The weight  $w$  is a  $k$ -dimension row vector. Transition probability matrix  $P^{(k)}$

consists of  $P_{ij}^{(k)}$ , which is equal to the ratio between the number of state  $i$  shifting to state  $j$  by jumping  $k$  steps and the number of state  $j$ .  $P(0)$  is a matrix of  $k$  rows and  $c$  columns.

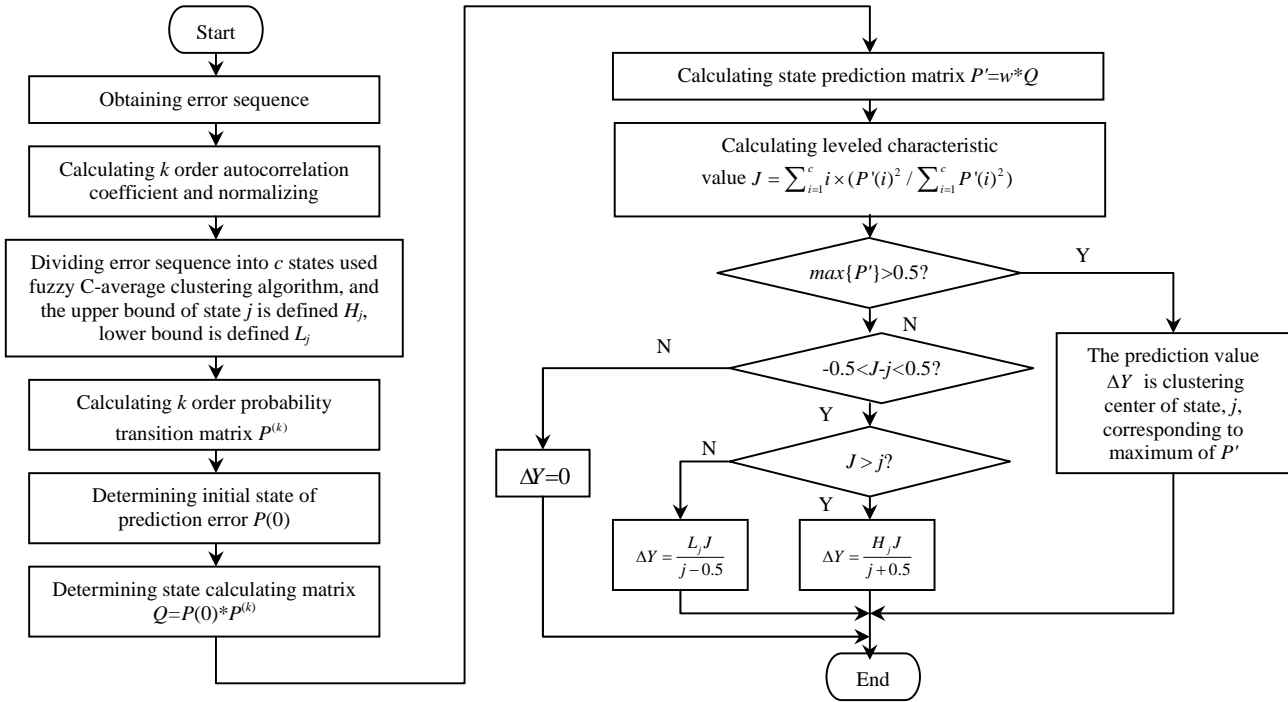


FIGURE II. THE FLOW CHART OF MARKOV CHAIN

### B. Discussion of the Vital Parameter of Markov Chain

Table 2 demonstrates the impact of  $c$  value on prediction at  $K=5$ . It is shown that  $c$  value causes the different errors. It has no effect or weakly negative effect on prediction of GRNN at  $c = 3$  and  $c=10$ . In the meantime, it has the strongest effect on prediction at  $c=8$  when comparing with the result at  $K=5$  in Table 1. Consequently, the historical errors of GRNN are divided into eight intervals. The results of state division using fuzzy C-average clustering algorithm are shown in Table 3.

According to Section II, previous four moments of errors are used to calculate the autocorrelation coefficient as previous four moments of speed.  $k$  ( $k=1,2,3,4$ ) order autocorrelation coefficient,  $r_k$ , and its normalized values  $w_k$  are calculated. Then, the corrections of wind speed are obtained from it. The comparative analysis of wind speed forecasted by GRNN and the modified GRNN (MC-GRNN) are listed in Table 4, and the corrections corresponding to wind speed are also listed in Table 4.

It is noticed that the results of GRNN are no obvious improvement after correction when compared Table 1 with Table 2. However, the results of simulation tests show that the current correction will cause a passive effect when the sign of next correction is changed as shown in Table 4. Therefore, the current correction is changed plus-minus to have a positive effect on the result of GRNN under this condition.

TABLE II. IMPACT OF C VALUE ON WIND SPEED FORECASTING

$c$	MAE (m/s)	MAPE (%)	RMSE(m/s)
3	0.7680	18.42	0.9866
4	0.7242	17.97	0.9298
5	0.7435	17.51	0.9496
6	0.7376	17.63	0.9448
7	0.7446	18.07	0.9466
8	0.7235	17.46	0.9438
9	0.7621	18.02	0.9812
10	0.7661	18.62	0.9688

TABLE III. THE RESULTS OF STATE DIVISION

State	Division of intervals (m/s)
state1	error $\leq -1.5479$
state2	$-1.5479 < \text{error} \leq -0.7982$
state3	$-0.7982 < \text{error} \leq -0.2723$
state4	$-0.2723 < \text{error} \leq 0.2026$
state5	$0.2026 < \text{error} < 0.7205$
state6	$0.7205 < \text{error} < 1.4147$
state7	$1.4147 < \text{error} < 2.8087$
state8	error $> 2.8087$

Table 5 shows the comparison of the wind speed forecasting improvements in MAE, MAPE and RMSE of MC-GRNN and GRNN. They are reduced by 26.28%, 27.77%, 20.41%, respectively. The result of comparison shows the superiority and effectiveness of MC-GRNN in the non-stationary wind speed forecasting.

TABLE IV. COMPARATIVE ANALYSIS OF WIND SPEED FORECASTED BY GRNN AND MC-GRNN

Real value (m/s)	Prediction by GRNN (m/s)	correction $\Delta Y$ (m/s)	Prediction by MC-GRNN(m/s)
2.4413	3.1053	-0.2867	2.8186
4.0665	3.1453	-0.2622	2.8831
4.8867	3.8923	0.2637	4.1560
4.0923	4.6172	0.2734	4.8906
4.9777	4.5089	-0.2707	4.2383
4.3403	4.5348	0.2374	4.7723
4.4908	4.4936	-0.2965	4.1971

TABLE V. ANALYSIS OF FORECASTING RESULTS OF GRNN AND MC-GRNN

Method	MAE (m/s)	MAPE (%)	RMSE (m/s)
GRNN	0.7771	19.30	0.9920
MC-GRNN	0.5729	13.94	0.7895

Figure 2 shows that forecasting values of GRNN become better in different degrees after correction. For example, there is a large result improvement of MC-GRNN around the point at 18. However, there is a small improvement around the point at 45. Figure 2 and Table 5 show that MC-GRNN method has better forecasting ability of forecasting wind speed series.

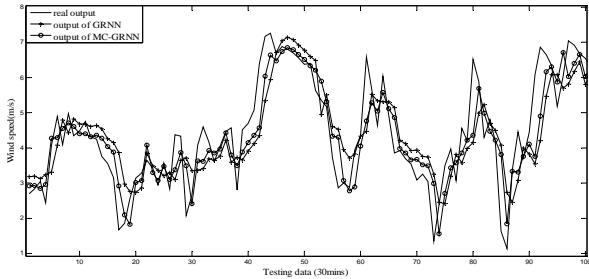


FIGURE III. COMPARISON BETWEEN REAL WIND SPEED AND FORECASTING VALUES CONCLUSIONS

## V. CONCLUSIONS

In this paper, the improved GRNN has been used to forecast wind speed of the wind tower based on real data. K value of K-fold Cross validation has a significant influence on the performance of GRNN. The qualified results can be obtained at K=5 according to our experiments. Markov chain is used to correct the results of GRNN to improve its performance, and the number of state division is discussed. On the basic of the results, dividing wind speed errors into eight states can be a reference for the future research. In order to compare the performance of the proposed method and GRNN, the real data collected from a wind farm in Shanxi province is used for evaluating the proposed method. The numerical results demonstrate the superiority of the proposed method.

## ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC) (51277127).

## REFERENCES

[1] L. Li and L.Ye, "Short-term wind power forecasting based on an improved persistence approach," Transactions of the CSAE, Vol. 78, pp. 182-187, December 2010.

[2] F. Cassola and M. Burlando, "Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output," Appl. Energy, Vol. 99, pp. 154-166, March 2012.

[3] W. N. Wang, W. Pedrycz, X. D. Liu, "Time series long-term forecasting model based on information granules and fuzzy clustering," Eng. Appl. Artif. Intel., Vol. 41, pp. 17-24, 2015.

[4] D. M. Wang, L. Wang, and G. M. Zhang, "Short-term wind speed forecast model for wind farms based on genetic BP neural network," J Zhejiang Univ. Sci. B., Vol. 46(5), pp. 837-841, May 2012.

[5] H. Liu, H. Q. Tian, D. F. Pan and Y. F. Li, "Forecasting models for wind speed using wavelet, wavelet packet, time series and Artificial Neural Network," Appl. Energy, Vol. 107, pp.191-208, March 2013.

[6] H. Zhang, L. Qu, Y. Chen G. Zhao and Z. Guo, "Support Vector Regression Based on Grid-Search Method for Short-Term Wind Power Forecasting," J. Appl. Math., June 2014.

[7] C. Y. Gao, X. M. Cui, N. Gao and X.Q. Hong, "Application of GRNN in the Prediction of Surface Deformation in Mining Areas," Metal Mine, Vol. 43(3), pp. 97-100, March 2014.

[8] E. E. Elattar, "Prediction of wind power based on evolutionary optimised local general regression neural network," IET Gener. Transm. Distrib., Vol. 8(5), pp. 916-923, May 2014.

[9] L. X. Yao, X. Q. Liu, L. Wu and M. J. Xue, "Mid-&long-term load forecast based on GRNN," Electr. Power. Autom. Equip., Vol. 27(8), pp. 26-29, August 2007.

[10] M. Han and J. Ding, "Improvement of BP algorithm based on cross-validation method and its implementation," Comput. Eng & Des., Vol. 29(14), pp. 3738-3739, July 2008.

[11] C. H. Du, X. Y. Huang, Z. Y. Yang, M. X. Tang and F. X Yang, "Real-time Rolling Traffic Flow Forecasting Based on Neural Networks and Markov Chains," J. Syst. Sim., Vol. 20(9), pp. 2464-2468, May 2008.

[12] Y. B. Wang, X. J. Liang, Y. Qiao, L. Wang, H. Y. Xu and W. Chen, "The Prediction of Rainfall Based on Superimposed Markov Chains and BP Neural Network Model," China Rural Water & Hydropower, Vol. (9), pp. 80-82, September 2014.