Short Term Power Forecasting of a Wind Farm Based on Atomic Sparse Decomposition Theory

Mingjian Cui, Xiaotao Peng, Junli Xia, Yuanzhang Sun, *Senior Member, IEEE/CSEE*, Ziping Wu

***Abstract*—The wind power data have very strong nonlinearity and non-stationarity, but the traditional method mainly focuses on the nonlinear problem of the wind power data and doesn't analysis the non-stationary problem. This paper proposed the combining method of atomic sparse decomposition and artificial neural network (ANN) to research the short-term forecasting of the wind power. Firstly, wind power data samples were decomposed into non-orthogonal atom sequences and residual sequences. Then ANN was used to model and predict the residual sequences, and the atom sequences adopt the adaptive prediction. Finally, the forecasting results were stacked and reconstructured. The generation power of an actual wind farm was forecasted by this method. The results show that the combining method of atomic sparse decomposition and ANN can reduce non-stationary behavior of the signal, produce sparser decomposition effect and better predict the variation tendency of the wind power.**

***Index Terms*-- wind generator; wind power forecasting; atomic sparse decomposition; ANN**

I. Introduction

Wind energy is a renewable clean energy. Due to the shortage of fossil fuel crisis and the urgent needs of environmental protection, wind power develops more and more rapidly in recent years. Because of the volatility and intermittency of the wind, when large capacity of wind power connects to the electricity grid, it will bring serious challenges to the power system security, stable operation and ensuring the power quality. Short-term forecasting of the wind farms' power output within 48h can adjust the scheduling plan timely, reduce the rotating spare capacity and cost of the power system, lay the foundation on participating in bidding for wind power generation. 15min earlier forecast of the wind power generation power has more important meanings. Firstly it can reasonably arrange the generating capacity of conventional energy, which will ensure the balance between power supply and user demand and make power system operate stably. Secondly extreme weather such as storm will make the energy output of wind farms change fast. 15min earlier forecast of the wind power can make timely response to the above events. Besides, it is also the foundation of longer forecasting[1-4].

At present, the short-term forecast for the output power of wind farms consists of two kinds of methods. One is the physical methods, which gets the weather data such as wind speed, wind direction, pressure and temperature through numerical weather prediction system. Then according to the physical information around the wind farms it gets the wind speed, wind direction and other information of the wind power generators hub height. Finally, it calculates the real output power of the fan using the fan power curve. The other is statistical method, which creates a mapping relationship between weather conditions and output power according to the historical data (wind speed or power) so as to carry on the forecasting. Modeling methods include time sequence method, Kalman filtering method, the neural network, the support vector machine (SVM) method, etc. The physical information around wind farm in physical method has a great influence on the forecast accuracy. But statistical method can modify the forecasting model at any time according to the characteristics and position of wind farms, and the accuracy is higher[5-7].

For wind power time series, a single step forecast has widespread obvious lag phenomenon comparing with the other existing modeling methods. In order to improve the problem, this paper proposes a forecasting method combining the atomic sparse decomposition and artificial neural network (ANN). Atomic sparse decomposition method, as the lead processing step of ANN, provides input characteristic vectors to ANN. The atoms decomposed signals are input to conventional neural network. This method achieved satisfactory results when applied to the actual power prediction of a certain wind farm in China.

II. Atomic Sparse Decomposition

A. Basic Concept

In recent years the atomic decomposition technique has become a hot spot in signal processing field. The method is derived from the thoughts of signal decomposition in the over-complete atomic library proposed by Mallat and Zhang[8]. This technology plays important roles in signal modeling, compression and feature extraction. In order to get the sparse expression of the signal, it is necessary to construct a super complete spread function set. The signal can choose the best spread functions from the set self-adaptively according to their own characteristic. Then the signal can be expressed more

This work was supported by a grant from the National High Technology Research and Development program of China (863 Program, 2012AA050218) and the National Natural Science Foundation of China (Key Program, 51190100).

Mingjian Cui, Xiaotao Peng, Junli Xia, Yuanzhang Sun are with the School of Elctrical Engineering, Wuhan University, Wuhan, P.R.China (e-mail:MJian.Choi@gmail.com)

Ziping Wu is with the department of Electrical and Computer Engineering, University of Denver, Denver, CO, 80210 USA.

978-1-4673-2868-5/12/$31.00 ©2012 IEEE

concisely and flexibly. This kind radical of the super complete set is called atom. Atomic decomposition method adopts a greedy self-adaptive decomposition strategy. The atomic library is highly redundant to ensure that any signal can self-adaptively choose a best group of atoms for the signal expression to make the decomposition results sparse, which is called sparse decomposition. The core problem of the atomic sparse decomposition is how to self-adaptively search for the best mapping atom and its parameters. The application of atomic decomposition method has just started in power system, which is mainly in power quality disturbance of signal characteristics research.

B. Structure of Dictionary Set

In signal processing, in order to get the sparse expression of the signal, an over-complete dictionary set is used to express the signal sparsely. In view of the different problems appropriate kernel functions can get the sparse expression. This paper uses the Gaussian kernel functions. Usually, a Gaussian function is defined by two parameters: Centre and Scale.

$$g = \exp(-\frac{(x-c)^2}{2\cdot\sigma^2}) \tag{1}$$

Where g is a Gaussian kernel functions, c and σ are the centre and scale, respectively. Different centre and scale parameters define disparate Gaussian kernel functions, which can structure a dictionary set by a series of Gaussian kernel functions. In Fig. 1, three different Gaussian kernel functions are enumerated. g1 shows a Gaussian kernel function with centre=0, scale=2. g2 shows a Gaussian kernel function with centre=2, scale=2. g3 shows a Gaussian kernel function with centre=0, scale=3.

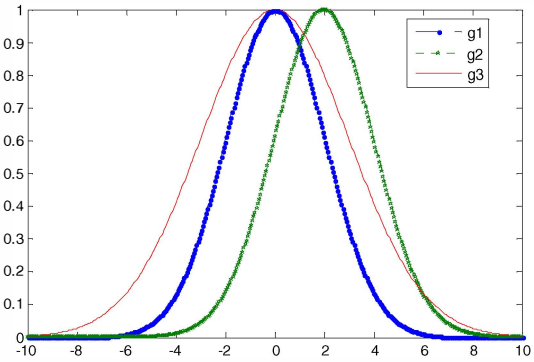

Fig. 1. Comparison of three different Gaussian kernel functions

C. Two Dictionaries Matching Pursuit (TDMP)

In MP decomposition process, the selected optimal atoms can be divided into two categories: old are previous selected atoms, new are not yet selected atoms. Corresponding to the two kinds of atoms, the over-complete dictionary set contains two separate dictionary sets, one consists of old atoms, and the other consists of new atoms. Therefore, at the very beginning, all the atoms are parts of the new dictionary set. In the several previous iterations of MP decomposition process, most of the chosen atoms belong to the new dictionary set. With the continuous iterations, the old dictionary set gradually extended. When the old dictionary set is large enough, the selected optimal atom is from either old or new. From the point of sparsity, the selection of the new atom goes against the sparse solution of the decomposition. To achieve the purpose of sparse decomposition the atoms should be selected from the old dictionary as far as possible. This paper puts forward a standard to choose atoms, specifically stated as follow:

The over-complete dictionary set is divided into two dictionary sets, ϕ^{old} and ϕ^{new}. We calculate the inner product of the residual signal and each atom in the two dictionary sets respectively and we can achieve the two local maximum coefficients of the inner product, c^{old} and c^{new}, in the two dictionary respectively. Accordingly the two local optimal atoms, ϕ^{old} and ϕ^{new} corresponding to c^{old} and c^{new}, can be determined in the old and new atom dictionary sets too.

If $\left|c^{old}\right| \geq \left|c^{new}\right|$, select the atom ϕ^{old} as the global optimal atom for the current iteration and update the searching variables with $c_{opt} = c^{old}$, $\phi_{opt} = \phi^{old}$ and $R = R - c_{opt}\phi_{opt}$.Finally accumulate the current coefficient c^{old} to the adder of the selected old atom ϕ^{old}.

If $\left|c^{old}\right| < \left|c^{new}\right|$, the global optimal atom in the *k*th iteration will be decided with the following steps:

(1) Calculate the two residual signals in the old and new atom dictionaries as:

$$R^{old} = R^{(k-1)} - c^{old}\phi^{old},\ R^{new} = R^{(k-1)} - c^{new}\phi^{new} \tag{2}$$

with $R^{(k-1)}$ the residual signal in the prior iteration.

(2) Calculate the RE *r* between the residual signals in the old and new atom dictionaries as:

$$r = \frac{\left\|R^{old} - R^{new}\right\|}{\left\|R^{new}\right\|} \tag{3}$$

(3) Decide the global optimal atom by the given threshold function *T*.

If $r \leq T$, choose ϕ^{old} as the global optimal atoms of the current iteration, and then update the searching variables:

$$c_{opt} = c^{old},\ \phi_{opt} = \phi^{old} \tag{4}$$

If $r > T$, choose ϕ^{new} as the global optimal atoms of the current iteration, and then update the searching variables:

$c_{opt} = c^{new}$, $\phi_{opt} = \phi^{new}$, add the atomic to the old dictionary set, and delete it from the new dictionary set. When the global optimal atomic is selected, accumulate its coefficient to the corresponding optimal atoms' coefficient;

(4)Update the residual signal: $R^{(k)} = R^{(k-1)} - c_{opt}\phi_{opt}$

Through a given threshold function T, the global optimal atoms are chosen in the old and new dictionary sets. If $T = 0$, TDMP will degenerate into the MP algorithm. To ensure the convergence and stability, T is a function decreasing with iteration. This paper adopts the annealing function of the simulated annealing algorithm:

$$T(k) = T_0 \times \alpha^{k^{1/N}} \quad (5)$$

Where $0.7 \le \alpha < 1$, T_0 means the initial temperature and is set to less than 1. k is the current iteration number, N is the annealing speed factor. With the number of iterations increased, T tends to 0. Algorithm flow chart of TDMP is shown as Fig.2.

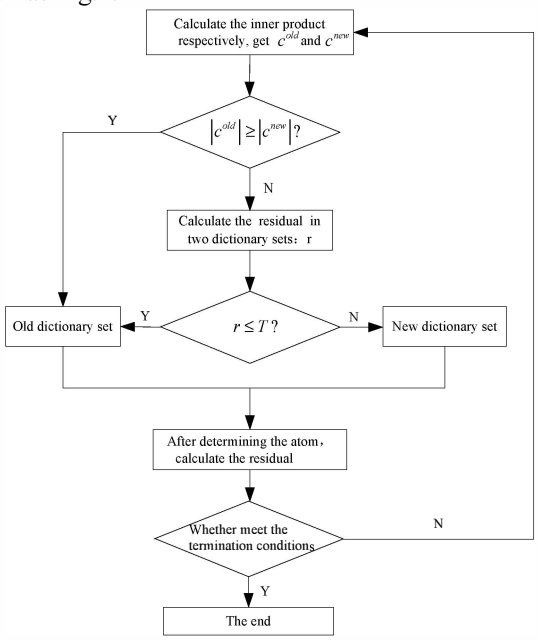

Fig. 2. Algorithm flow chart of TDMP

III. Forecast Modeling

ANN has very strong nonlinear mapping capability. Through the training of the samples, we can realize the mapping of the input n d space to the output m d space. This paper adopts BP neural network with single hidden layer. The input nodes number is m. The input data corresponds to each component of the input nodes. The input nodes number is 1. The output data corresponds to the power value of the next moment.

Generating capacity of the wind farms is affected by the factors of wind speed, wind direction and so on. While these meteorological factors change rapidly, it makes the wind power data provided with strong non-stationary essentially. If the model is built based on the neural network simply, only nonlinearity of the system can be fitted out. But the data non-stationary makes the prediction result unsatisfactory. In order to improve the prediction accuracy, this paper makes use of atomic decomposition algorithm to decompose the signals, and then structures forecast model. The specific steps are as follows:

(1) Decompose the time series of power to n atoms. Get n atoms signal components and a residual signal component:

$$x(t) = \sum_{j=1}^{n} a_j(t) + r(t) \quad (6)$$

(2) Make the model of the residual signals component using BP neural network and forecast[9-10].

(3) Accumulate the forecasting results of each signal component and then get the final forecasting results.

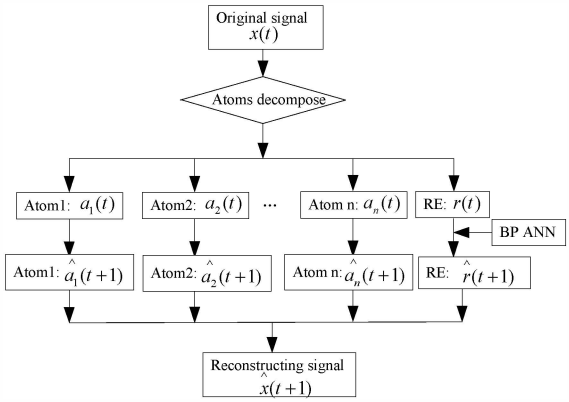

Fig. 3. Structure of the wind power forecasting model

IV. Results from an Actual Wind Farm

A. Test Data

Fig. 3 shows the output power data of one wind generator in a certain wind farm with 58 wind generators from May 10, 2006 to May 24, 2006. Time resolution is 15min. In this paper, we choose the first 450 points of the total 1440 sample points as the sample data to experiment, which is shown in Fig. 4 and Fig. 5.

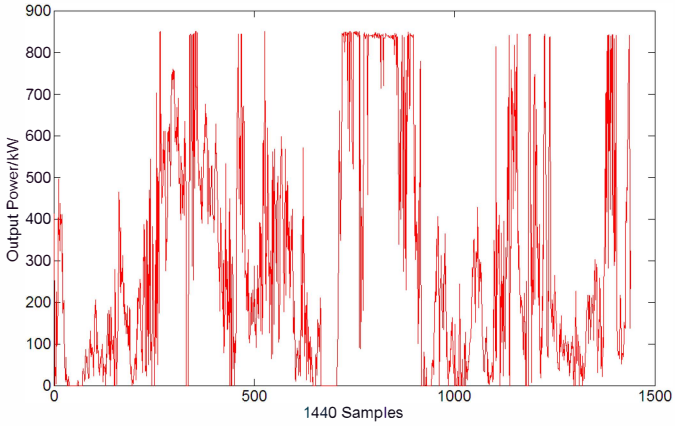

Fig. 4. The total 1440 sample points

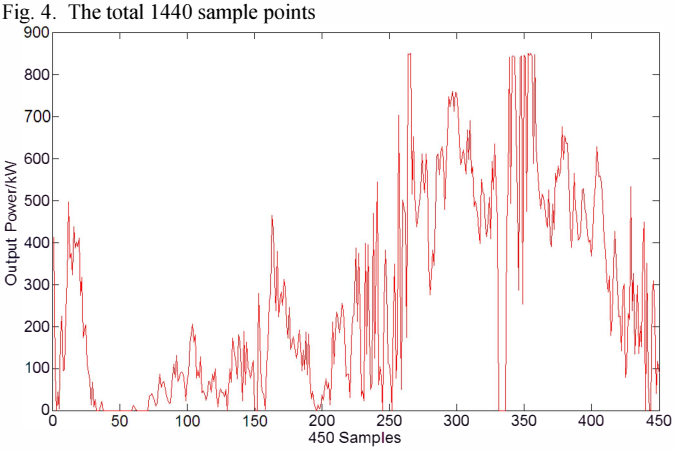

Fig. 5. The first 450 points of the total 1440 sample points

B. Experiment Results

This paper uses MATLAB 7.0 to program the algorithm. The selected kernel function is shown as follows:

$$\Phi(x) = \exp\left[-\frac{(x-c)^2}{2\sigma}\right] \quad (7)$$

Through sum-of-squares standards, we can get the optimal cluster number of the sample data $N_c = 5$ of the sample data. Scale parameter σ adopts the alternative set $\{2.0, 2.5, 3.0, 3.5, 4.0\}$, $\alpha = 0.935$, initial temperature $T_0 = 0.09$, annealing speed factor $N = 2.5$, terminating error 0.01[11-12]. The front 400 points of the residual signals RE in Fig.5 are selected as the training samples and the latter 50 points as test samples. BP neural network is used to model the residual signals RE, and then final prediction result is reconstructed through superposition. The decomposition results of the normalized 400 points ahead are shown in Fig.6, forecasting result in Fig.7, parameters in Table I.

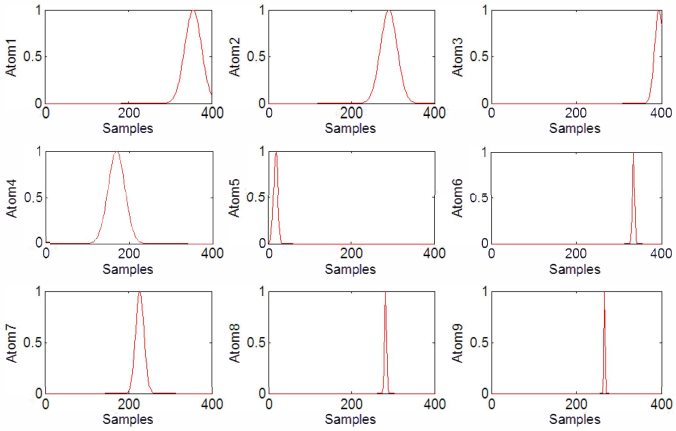

Fig. 6 Decomposition results of the normalized 400 points ahead

TABLE I
PARAMETERS OF THE NORMALIZED 400 POINTS AHEAD DECOMPOSITION RESULTS

Atoms	Centre	Scale	Weight
1	7.7444	1.0067	0.84456
2	4.4361	1.0067	0.8764
3	9.6491	0.50333	0.44304
4	-1.5789	1.0067	0.26101
5	-9.198	0.25166	0.4962
6	6.6416	0.12583	-0.58314
7	1.3283	0.50333	0.25308
8	4.0351	0.12583	-0.48269
9	3.2331	0.062916	0.47938

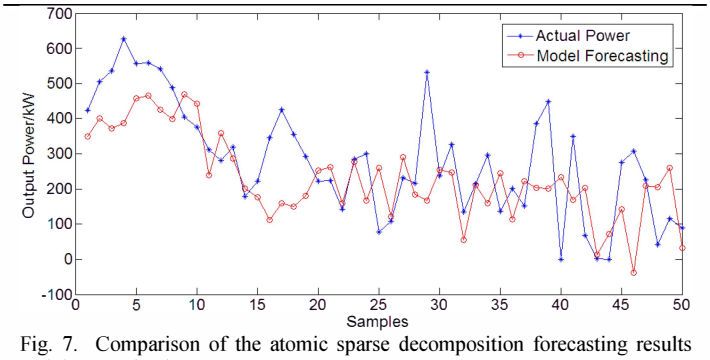

Fig. 7. Comparison of the atomic sparse decomposition forecasting results and the actual values

The size of the prediction error has immediate relationship with the fan capacity. In order to evaluate the forecasting effect quantificationally, this paper adopts the normalized mean absolute error e_{NMAE} and the normalized root mean square error e_{NRMSE} as the basis, which is generally international, defined as follows:

$$e_{NMAE} = \frac{1}{P_{cap.}} \frac{1}{N} \sum_{i=1}^{N} \left| \hat{x}(i) - x(i) \right| \tag{8}$$

$$e_{NRMSE} = \frac{1}{P_{cap.}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{x}(i) - x(i))^2} \tag{9}$$

Where $x(i)$ is the actual value, $\hat{x}(i)$ is the forecasting value, N is the number of the forecasting samples, $P_{cap.}$ is the nominal capacity of the fan. Through the calculation, he normalized mean absolute error is 13.10%, the normalized root mean square error is 16.53%.

V. CONCLUSION

In recent years, atomic sparse decomposition can be used for more flexible, simple and signal adaptive expression, but it has not yet been applied in the wind power forecasting field. In the light of the nonlinearity and non-stationarity of the wind power system, this paper proposed atomic sparse decomposition and ANN to the wind power time series forecasting model. Atomic sparse decomposition is able to reduce the signal non-stationarity better. ANN can well map out the nonlinear relationship between the system input and output. Firstly, this method used atomic sparse decomposition to decompose the original data. Then the residual sequence was input to the BP network to forecast. The generation power of an actual wind farm was forecasted by this method. The results show that the combined method of atomic sparse decomposition and ANN can reduce the non-stationary behavior of signals, produce sparser decomposition effect and predict the variation tendency of the wind power better. But forecast precision problem still needs to be further studied and analyzed as the next step of work.

VI. REFERENCES

[1] Yuanzhang Sun, Jin Lin, Guojie Li, Xiong Li, "A survey on the grid integration of wind farms with variable speed wind plant systems," *Automation of Electric Power System*, vol. 34, pp. 75-80, Feb. 2010.

[2] P. Sørensen, N. A. Cutululis, A. Vigueras-Rodríguez, L. E. Jensen, J. Hjerrild, M. H. Donovan, H. Madsen, "Power fluctuations from large wind farms," *IEEE Trans. Power Systems*, vol. 22, pp. 958-965, Aug. 2007.

[3] Sideratos G, Hatziargyriou N D, "An advanced statistical method for wind power forecasting," *IEEE Trans. Power Systems*, vol. 22, pp. 258-265, Feb. 2007.

[4] Lijie Wang, Lei Dong, Xiaozhong Liao, Yang Gao, "Short-term power prediction of a wind farm based on wavelet analysis," *Proceedings of the CSEE*, vol. 29, pp. 30-33, Oct. 2009.

[5] Xiuyuan Yang, Yang Xiao, and Shuyong Chen, "Wind speed forecast model for wind farms based on time series analysis," *Proceedings of the CSEE*, vol. 25, pp. 1-5, Jun. 2005.

[6] Hongtao Shi, Jingling Yang, Maosheng Ding, Jinmei Wang, "A short-term wind power prediction method based on wavelet decomposition and BP neural network," *Automation of Electric Power System*, vol. 35, pp. 44-48, Aug. 2011.

[7] Nengling Tai, Zhijian Hou, Tao Li, Chuanwen Jiang, Jiong Song, "New principle based on wavelet transform for power system short-term load forecasting," *Proceedings of the CSEE*, vol. 23, pp. 45-50, Jan. 2003.

[8] S. Mallat, Z. Zhang, "Matching pursuits with time frequency dictionaries," IEEE Trans. Singnal Processing, vol. 41, pp.3397-3415, Dec. 1993.

[9] G. Karniotakis, G. S. Stavrakakis, and E. F. Nogaret, "Wind power forecasting using advanced neural network models," *IEEE Trans. Energy Conversion*, vol. 11, pp. 762-767, Dec. 1996.

[10] Gaofeng Fan, Weisheng Wang, Chun Liu, Huizhu Dai, "Wind power prediction based on artificial neural network," *Proceedings of the CSEE*, vol. 28, pp. 118-123, Dec. 2008.
[11] Peng Xu, Dezhong Yao, "Two dictionaries matching pursuit for sparse decomposition of signals," *IEEE Trans. Signal Processing*, vol. 86, pp. 3472-3480, Nov. 2006.
[12] Ning Wang, Linchuan Li, Qingquan Jia, Haiyan Dong, "Classification of power quality disturbance signals using atomic decomposition method," *Proceedings of the CSEE*, vol. 31, pp. 51-58, Feb. 2011.