

# An Implementation of Synthetic Generation of Wind Data Series

Liang Liang, and Jin Zhong

Department of Electrical and Electronic Engineering  
The University of Hong Kong  
Hong Kong SAR  
E-mail: lliang@eee.hku.hk

Jianing Liu, Puming Li, Cailiang Zhan, and Zijie Meng

Guangdong Electric Power Dispatch Center  
Guangdong Power Grid Corporation  
Guang Zhou, China  
E-mail: lipuming@gddd.csg.cn

**Abstract**—Wind power fluctuation is a major concern of large scale wind power grid integration. To test methods proposed for wind power grid integration, a large amount of wind data with time series are necessary and will be helpful to improve the methods. Meanwhile, due to the short operation history of most wind farms as well as limitations of data collections, the data obtained from wind farms could not satisfy the needs of data analysis. Consequently, synthetic generation of wind data series could be one of the effective solutions for this issue.

In this paper, a method is presented for generating wind data series using Markov chain. Due to the high order Markov chain, the possibility matrix designed for a wind farm could cost a lot of memory, which is a problem with current computer technologies. Dynamic list will be introduced in this paper to reduce the memory required. Communication errors are un-avoidable on long way signal transmission between the control centre and wind farms. Missing of data always happens in the historical wind data series. Using these data to generate wind data series may result in some mistakes when searching related elements in the probability matrix. An adaptive method will be applied in this paper to solve the problem.

The proposed method will be verified using a set of one-year historical data. The results show that the method could generate wind data series in an effective way.

**Index Terms**—Synthetic generation, wind power, renewable energy integration, power fluctuations, Markov chain

## I. INTRODUCTION

Renewable energy integration has been regarded as a key target for the future development of Smart Grid [1]–[6]. With more wind farms connected to the power system, the impacts of wind energy on the system can not be neglected. To follow the distribution of wind resources, one wind farm could have hundreds wind turbines. The power fluctuation at the accessing point of a wind farm to the power grid could significantly affect the frequency of the grid [7]–[10]. Some solutions have been proposed to mitigate the negative effects caused by power fluctuations, such as wind and energy storage hybrid systems [11], demand side management [12], and new control strategies for wind turbines [13]. To test these solutions,

various mathematic models of wind power generation have been developed to simulate the behavior of wind in power systems. These models all require time-varying wind data as input data. As a result, how to provide sufficient wind data time series becomes a key issue in wind power study and testing.

A long-period historical data obtained continuously from a wind farm is certainly the best data sources. However, obtaining continuous historical data for a long time is difficult for any wind farm. Methods have been proposed to generate wind data time series. In [14], wind data is generated by a proposed model, which is trained using historical data. The method can also be used to predict the output of a wind farm. A method with random variance is proposed in [15] to generate wind data time series for multiple wind farms using historical data from a reference site as an input. In these two methods, wind data are generated by training historical data instead of directly from historical data. The statistic characteristics of generated data might be different from the original historical data.

Different from above methods, Markov chain based stochastic models can generate wind data time series with the same statistical parameters, especially for the difference of wind power time series between two continuous points. The statistical results of Markov chain model with different state sizes were discussed in [16]. Markov chain models with different dimensions of possibility matrix have been studied. The statistical characteristics of these models were discussed. It was found that the dimension of the possibility matrix is a key factor for Markov chain model to generating data. The factor of time intervals to Markov chain models have been studied in [17]. It shows that Markov chain model with a time interval within a minute usually produces large error when generating wind data time series. Further more, it shows that increasing the order of Markov chain model is helpful in obtaining a more accurate result.

The paper is organized as following: The model of Markov Chain will be introduced in Section II. The methods of dynamic list and dynamic state size will be described in Section III. The results of case study will be presented and discussed in Section IV and Conclusions in Section V.

This work was partly supported by grants from the Research Grant Council of Hong Kong (Project No. HKU10/CRF/10), HKU Strategic Research Theme and the University Development Funding (Initiative on Clean Energy & Environment), and the State Key Laboratory of Power Systems of Tsinghua University, Beijing, China.

## II. MARKOV CHAIN MODEL

Markov chain is usually used for describing stochastic process, especially for the cases that current states are decided by previous states [18]. Wind energy could be considered as a random process. As wind is a continuous process in natural, wind speed at any time is strongly related to previous wind speeds. Consequently, wind speed time series can be modeled using Markov chain model. Based on historical wind data time series, a transition possibility matrix of wind speed transition can be set up as followings:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & \cdots & \cdots & a_{1,j-1} & a_{1,j} \\ a_{2,1} & a_{2,2} & \cdots & \cdots & \cdots & a_{2,j-1} & a_{2,j} \\ \vdots & \vdots & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & & \vdots & \vdots \\ \vdots & \vdots & & & \ddots & \vdots & \vdots \\ a_{i-1,1} & a_{i-1,2} & \cdots & \cdots & \cdots & a_{i-1,j-1} & a_{i-1,j} \\ a_{i,1} & a_{i,2} & \cdots & \cdots & \cdots & a_{i,j-1} & a_{i,j} \end{bmatrix} \quad (1)$$

where,  $a_{i,j}$  represents the number of times wind speed changing from state  $i$  to state  $j$  in the historical data.

The summation of each row in (1) represents the total number of times that previous wind speed states belong to state  $i$ . Elements of possibility matrix can be obtained using (2) to represents wind speed transition.

$$b_{i,j} = a_{i,j} / \sum_{k=1}^n a_{i,k} \quad (2)$$

where,  $b_{i,j}$  represents the transition possibility from state  $i$  to state  $j$ .

Then a uniform distributed random number  $R$ , which is between 0 and 1, will be created to generate wind speed time series base on (2). The reformed possibility matrix can be obtained using (3).

$$c_{i,j} = \sum_{k=1}^j b_{i,k} \quad (3)$$

where,  $c_{i,j}$  represents the summation of transition possibility from state 1 to state  $j$  in row  $i$  in (2).

Once the matrix  $C$  is generated using historical wind data time series, random number  $R$  is used to generate wind speed. Comparing  $R$  with the elements  $c_{i,j}$  ( $j = 1, 2, \dots, n$ ), if  $R$  is smaller than  $c_{i,j}$ , the next wind speed belongs to state  $j$ . In this way, wind speed time series with unlimited length could be generated from historical data.

The proposed method could be extended to a higher order Markov Chain model to improve data generating accuracy. A fifth order Markov Chain model will be implemented in the following section as an example. Elements  $a_{i,j}$ ,  $b_{i,j}$ ,  $c_{i,j}$  will be converted to  $a_{i,j,k,l,m}$ ,  $b_{i,j,k,l,m}$ ,  $c_{i,j,k,l,m}$  to save the memory used for transition possibility matrix.

## III. IMPLEMENTATION OF SYNTHETIC GENERATION

### A. Dynamic List

In this paper, we use a fourth order Markov Chain model with a state size of 350. It requires  $350^5 * 8$ , which is 42T bytes memory. This memory requirement is not an easy task to be implemented in a PC system. However, lots of elements in the matrix are equal to zeros as lots combinations of wind speed time series are not exist in the historical data set. As a solution, dynamic list is applied in this paper to save the possibility matrix as a sparse matrix.

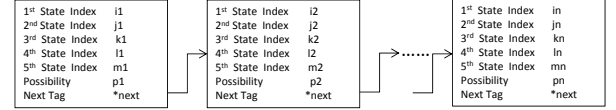


Fig. 1. The Structure of Dynamic List for Possibility Matrix

The structure of proposed dynamic list is shown in Fig.1. As the search direction of possibility matrix is one direction, the single-direction dynamic list is applied to obtain simple structure. The list follows decreasing sequence of  $m, l, k, j, i$ .

### B. Dynamic State Size

Data discontinuation is another issue in generating data using historical data. This is unavoidable with existing data transmission technologies due to the long distance between wind turbines and control centre. The discontinuity of historical data will result in failures of generating wind data time series, when searching in the transition possibility matrix with random number  $R$ .

Using a wind data series as an example. Assume the data time series is 10-23-45-24-34 in the historical wind data. The sequence terminates at state 34 due to communication errors, and the sequence 23-45-24-34 only appears once in the whole historical data. In the case that generated wind speed is state 23, following the historical data, state 45, 23, 34 should be generated as a sequence. However, because of the interruption after state 34, the next wind speed could not be generated. In order to solve the issue, a transition possibility matrix with smaller state size will be selected to reduce the discontinuity in the transition possibility. As a trade off, the accuracy of data generated by the smaller-sized transition possibility matrix is reduced accordingly.

A dynamic state size method has been proposed to solve the issue in this paper. The principle of the method is that once the data interruption is detected during the search process, the state size automatically reduces to the half of the original size to maintain the continuity in possibility matrix, and the matrix size will be restored to the original size in the next data generation process. This method requires multiple transition possibility matrixes, which could cost more memory. The possible increased memory requirement can be mitigated by applying dynamic list technology.

### C. Wind Speed Time Series Generation Process

The process of wind data time series generation could be divided into 3 steps. First, historical wind speed time series data is input and converted into discrete states. Based on the converted discrete data states, transition possibility matrix  $a_{i,j}$  is created and transferred to matrix  $b_{i,j}$  and  $c_{i,j}$  for wind speed time series generating. Second step, uniform distributed random number  $R$  is then generated according to matrix  $c_{i,j}$ . In the situation of discontinued data, matrix  $C$  can not generate corresponding wind speed time series with number  $R$ , the size of matrix  $C$  will be reduced dynamically to provide a solution. The data generating process is described as the flow chart shown in Fig.2. Computer program based on the flow chart can be programmed to generate required wind data time series.

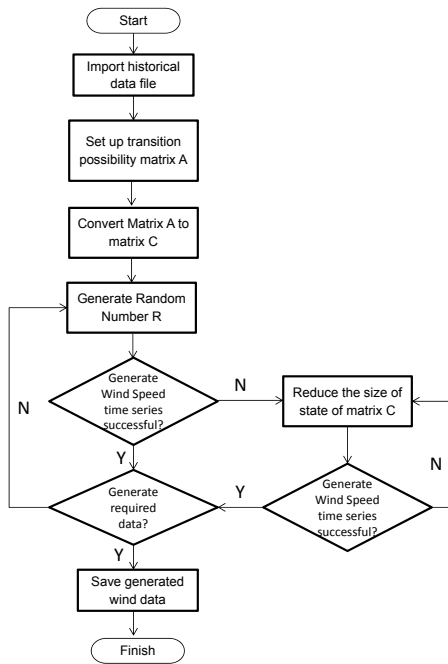


Fig. 2. The Flow Chart of Wind Speed Time Series Generation

## IV. CASE STUDY

### A. Historical Wind Data Time Series

A three-year historical wind data time series is used for case study. The data is recorded in a wind farm located at the U.S. with ID 18 in the data base of NREL.

Fig.3 shows the historical wind speed time series of the wind farm. Fig.4 shows the histogram for the same farm. The time interval of the historical data is 10 minutes. There are 157,824 data points in the historical data. The average wind speed is 7.95 m/s.

Fig.5 shows the historical for differential wind speed time series. Fig.6 shows the histogram for the differential wind speed time series. The average value of differential wind speed is  $-3.74 \times 10^{-4}$  m/s, which can be regarded as zero. The standard deviation for the differential data is 0.534.

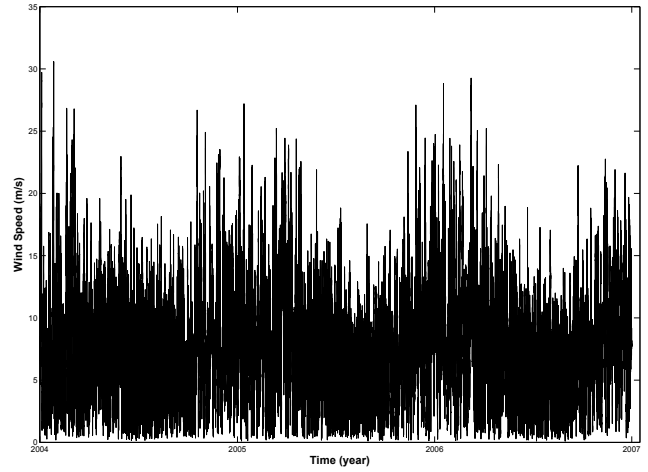


Fig. 3. The Historical Wind Speed Time Series from 2004 to 2006

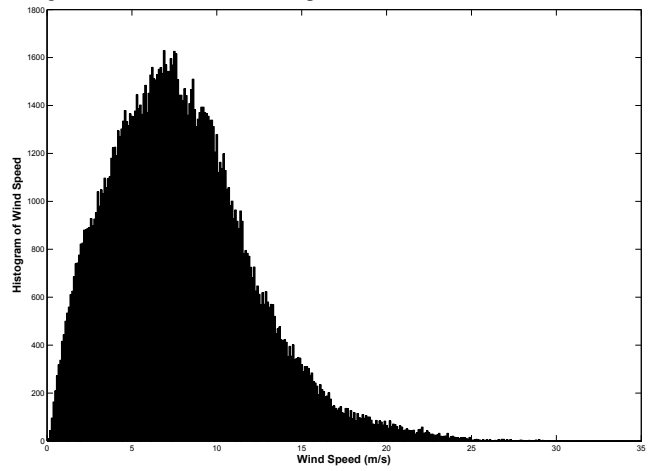


Fig. 4. The Histogram of Wind Speed Time Series from 2004 to 2006

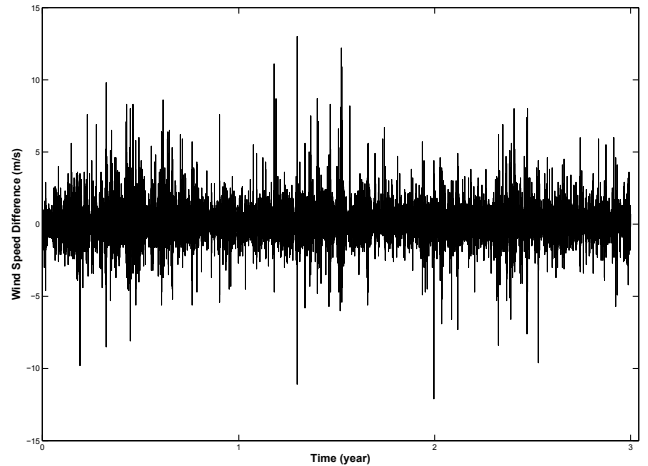


Fig. 5. The Historical Differential Wind Speed Time Series in 3 Years

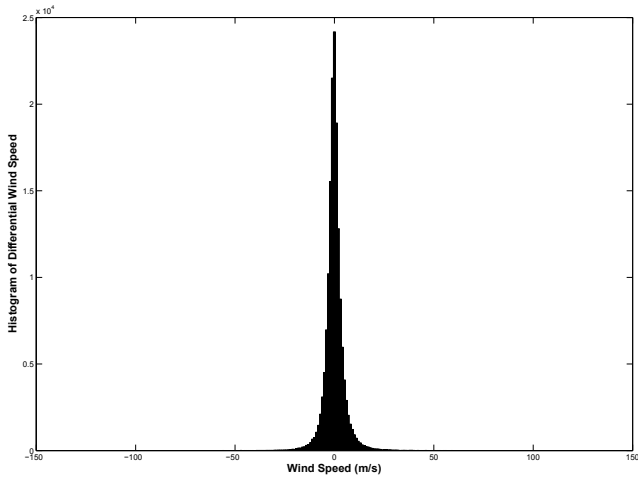


Fig. 6. The Histogram of Historical Differential Wind Speed Time Series in 3 Years

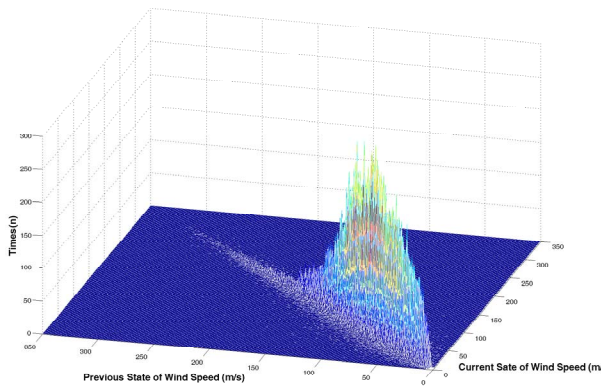


Fig. 7. The Distribution of Wind Speed States for two conjoint time points

Fig.7 shows the distribution of wind speed states with two conjoint time points. The figure shows that the wind speed in most-of-time keeps the same as previous state and most probably will not jump to a far away state. The centre of statistical result is around the average wind speed.

### B. Data Series Generation

The wind speed time series has been generated for an extended three years based on the proposed 4<sup>th</sup> order Markov chain method and programs.mentioned in Section III.

The generated extended 3 years wind speed time series is shown in Fig.8. The average wind speed for the generated data is 7.00 m/s, which is less than the average of the historical wind data but still in an acceptable region. The histogram of the generated wind data is shown in Fig.9. The distribution of generated wind speed time series generally follows the distribution of historical wind data. The centre has shifted to the left about 1m/s, which is similar to the difference of average wind speed between historical wind data and generated wind data.

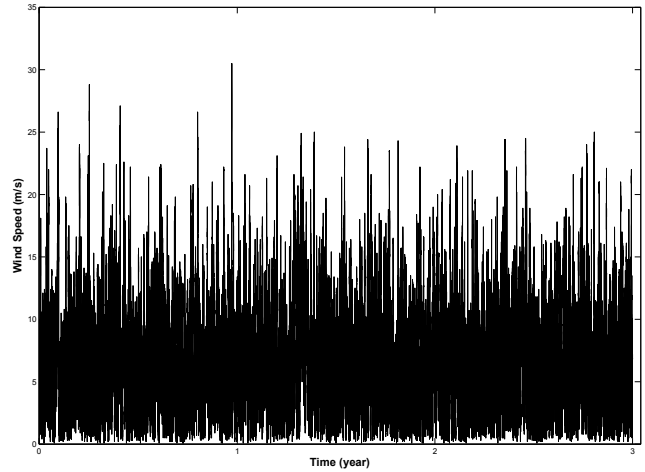


Fig. 8. The Generated 3 Years Wind Speed Time Series

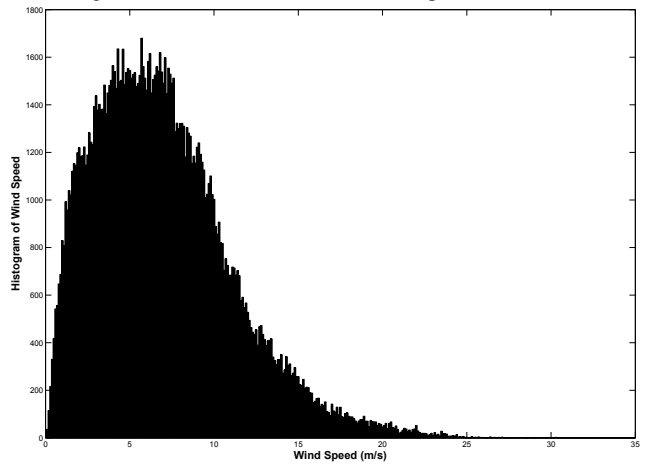


Fig. 9. The Generated 3 Years Wind Speed Time Series

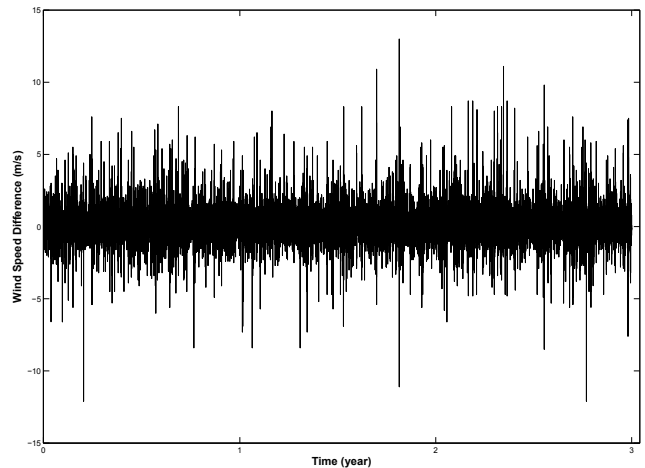


Fig. 10. The Difference of Generated Wind Speed Time Series in 3 Years

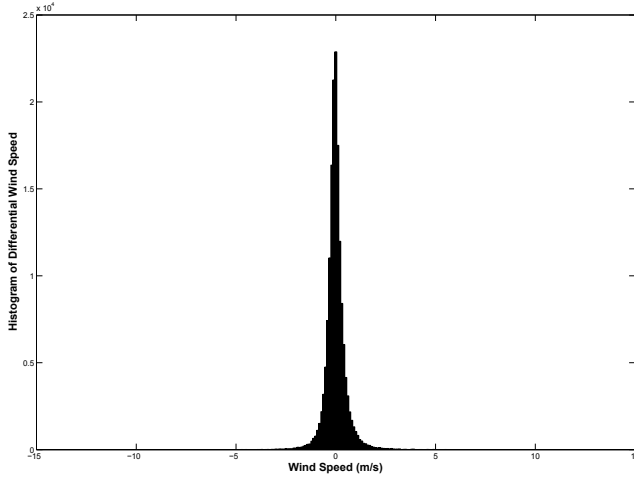


Fig. 11. The Histogram of Generated Difference Wind Speed Time Series in 3 Years

Fig.10 and Fig.11 show the results of generated wind speed time series in differential mode. The average wind speed for differential mode is  $-0.146 \times 10^{-4}$  and the standard deviation is 0.5585, which are very close to the historical wind data. As the differential wind speed data can directly represent the power fluctuation of wind farm, generated wind speed time series are mainly used to study the power fluctuations of wind farms. Consequently, the statistical characteristic of differential wind speed time series are more important than the original wind data. The proposed 4<sup>th</sup> order Markov chain model could provide a high accuracy in generating differential wind speed data.

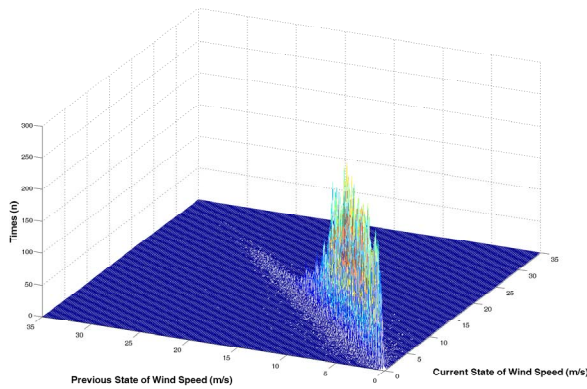


Fig. 12. The Distribution of Wind Speed States for two conjoint time points

The distribution of generated wind data states for two conjoint time points is shown in Fig.12. The 4<sup>th</sup> order Markov chain model can keep the characteristics of two conjoint steps as the historical wind speed time series, which is important for the study of power fluctuations of wind farms.

The comparison of historical and generated wind speed time series can be summarized in Table I.

TABLE I  
THE COMPARISON OF HISTORICAL AND GENERATED WIND SPEED TIME SERIES

Item	Historical Data	Generated Data
Length	3 years	3 years
Average Wind Speed	7.95 (m/s)	7.00 (m/s)
Standard Deviation	4.21	4.13
Average Wind Speed in Diff	$-3.74 \times 10^{-4}$	$-1.46 \times 10^{-4}$
standard Deviation in Diff	0.534	0.5585

## V. CONCLUSIONS

A wind data generation method with 4<sup>th</sup> order Markov chain model has been proposed in this paper. Several technologies have been introduced to implement such model with a computer program. Based on a 3 year historical operation data of a wind farm, an extended 3 year wind speed time series data has been generated. Results show that the generated wind speed time series could follow the statistical characteristic of historical data, especially in the differential mode.

Attributing to the integration of dynamic list method for recording the transition possibility matrix, a higher order Markov chain model could be implemented in the future when needed. The generated wind data according to historical wind data with high accuracy will be contributed to the study of renewable energy integration.

## REFERENCES

- [1] C. Cecati, C. Citro, and P. Siano, "Combined operations of renewable energy systems and responsive demand in a smart grid," *Sustainable Energy, IEEE Transactions on*, vol. 2, no. 4, pp. 468–476, oct. 2011.
- [2] M. Liserre, T. Sauter, and J. Hung, "Future energy systems: Integrating renewable energy sources into the smart power grid through industrial electronics," *Industrial Electronics Magazine, IEEE*, vol. 4, no. 1, pp. 18–37, march 2010.
- [3] P. Varaiya, F. Wu, and J. Bialek, "Smart operation of smart grid: Risk-limiting dispatch," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 40–57, jan. 2011.
- [4] A. Saber and G. Venayagamoorthy, "Resource scheduling under uncertainty in a smart grid with renewables and plug-in vehicles," *Systems Journal, IEEE*, vol. 6, no. 1, pp. 103–109, march 2012.
- [5] J. Byun, I. Hong, B. Kang, and S. Park, "A smart energy distribution and management system for renewable energy distribution and context-aware services based on user patterns and load forecasting," *Consumer Electronics, IEEE Transactions on*, vol. 57, no. 2, pp. 436–444, may 2011.
- [6] G. Venayagamoorthy, "Dynamic, stochastic, computational, and scalable technologies for smart grids," *Computational Intelligence Magazine, IEEE*, vol. 6, no. 3, pp. 22–35, aug. 2011.
- [7] P. Sorensen, N. Cutululis, A. Viguera-Rodriguez, L. Jensen, J. Hjerrild, M. Donovan, and H. Madsen, "Power fluctuations from large wind farms," *Power Systems, IEEE Transactions on*, vol. 22, no. 3, pp. 958–965, aug. 2007.
- [8] C. Luo, H. Banakar, B. Shen, and B.-T. Ooi, "Strategies to smooth wind power fluctuations of wind turbine generator," *Energy Conversion, IEEE Transactions on*, vol. 22, no. 2, pp. 341–349, june 2007.
- [9] L. Jin, S. Yuan-zhang, P. Sorensen, L. Guo-jie, and G. Weng-zhong, "Method for assessing grid frequency deviation due to wind power fluctuation based on time-frequency transformation"exi," *Sustainable Energy, IEEE Transactions on*, vol. 3, no. 1, pp. 65–73, jan. 2012.
- [10] S. Muyeen, H. Hasaniien, and J. Tamura, "Reduction of frequency fluctuation for wind farm connected power systems by an adaptive artificial neural network controlled energy capacitor system," *Renewable Power Generation, IET*, vol. 6, no. 4, pp. 226–235, july 2012.

- [11] H. Lee, B. Y. Shin, S. Han, S. Jung, B. Park, and G. Jang, "Compensation for the power fluctuation of the large scale wind farm using hybrid energy storage applications," *Applied Superconductivity, IEEE Transactions on*, vol. 22, no. 3, p. 5701904, june 2012.
- [12] D. Westermann and A. John, "Demand matching wind power generation with wide-area measurement and demand-side management," *Energy Conversion, IEEE Transactions on*, vol. 22, no. 1, pp. 145 –149, march 2007.
- [13] E. Muhando, T. Senjyu, K. Uchida, H. Kinjo, and T. Funabashi, "Stochastic inequality constrained closed-loop model-based predictive control of mw-class wind generating system in the electric power supply," *Renewable Power Generation, IET*, vol. 4, no. 1, pp. 23 –35, january 2010.
- [14] W. Yuan-Kang, L. Ching-Ying, T. Shao-Hong, and S.-N. Yu, "Actual experience on the short-term wind power forecasting at penghu - from an island perspective," in *Power System Technology (POWERCON), 2010 International Conference on*, oct. 2010, pp. 1 –8.
- [15] M. J. Woods, C. J. Russell, R. J. Davy, and P. A. Coppin, "Simulation of wind power at several locations using a measured time-series of wind speed," *Power Systems, IEEE Transactions on*, vol. PP, no. 99, p. 1, 2012.
- [16] F. Hocaoglu, O. Gerek, and M. Kurban, "The effect of markov chain state size for synthetic wind speed generation," in *Probabilistic Methods Applied to Power Systems, 2008. PMAPS '08. Proceedings of the 10th International Conference on*, may 2008, pp. 1 –4.
- [17] K. Brokish and J. Kirtley, "Pitfalls of modeling wind power using markov chains," in *Power Systems Conference and Exposition, 2009. PSCE '09. IEEE/PES*, march 2009, pp. 1 –6.
- [18] B. Everitt, *The Cambridge dictionary of statistics / B.S. Everitt*, 2nd ed. Cambridge University Press, Cambridge, U.K. ; New York :, 2002. [Online]. Available: <http://www.loc.gov/catdir/toc/cam031/2002514499.html>