Big Data-based Harmonic Problem Research in Wind Farms

Song-Tao Yu, Da Xie, Yu-Pu Lu, Zu-Yi Zhao
Department of Electrical Engineering, Shanghai Jiao Tong University
Shanghai, China
Email: profxzg@hotmail.com

Yan-Chi Zhang
Department of Electrical Engineering, Shanghai Dian Ji University,
Shanghai, China
Email: yanchizhang@hotmail.com

Abstract **- More and more attention on the power quality of wind farms has been paid recent years, in which harmonic problem is one of the most concerned. On the other hand, power big data in wind farm generated all the time, and the volume is increasing continuously, which can be mined to extract some special or new information to solve current operating problem and adjust the running conditions of wind turbines. In this paper, a novel algorithm for power big data analysis has been put forward by a combined application of conventional harmonic analyzing method and typical clustering algorithm, which can be used to deal with the big data in wind farm to study harmonic problem. The measured big data of a 2MW DFIG wind turbine in operation have been used to verify the new algorithm, and some interesting conclusions of harmonics have been found at last.**

Keywords-Big Data; Wind Farms; Harmonics; Data Mining; Clustering.

I INTRODUCTION

The interest in the power quality of wind farms has increased as renewable energies become more important to face global environmental challenges. One of the most concerned aspects for power quality of wind farms is harmonics [1-2]. On the other hand, big data has been widely studied and applied in many sectors [3-5]. However, the research and application of power big data, especially big data in wind farms, are still in their infancy. Traditional methods of data analysis are based on simple small dataset or some simulation data in small amount [6-8], which are not suitable to analyze big data. As current software and hardware conditions can not meet the demand of big data analysis, there are only two basic ways to settle this matter: one is to build hardware and software supports of high performance, which undoubtedly will increase the investment; the other one is to put forward new algorithms for big data analysis under current software and hardware conditions by reducing the dimension of data. Large amount of research have been done for big data algorithms on high performance facilities [9-11], such as Hadoop and Storm etc. However, little research has been done on the algorithms of big data, which can be applied on the ordinary computing facilities that we use daily.

The contribution of this paper is to put forward a new algorithm for the harmonic problem in wind farms by the combined application of normal FFT method and typical clustering algorithm, which is suitable to deal with power big data generated in wind farms. First, conventional analyzing method of harmonic problem has been studied. Subsequently, basic concepts of big data, typical data mining methods, and the applied analysis of data mining in power big data in wind farm have been presented. Finally, actual running power big data of a typical wind turbine have been analyzed by the new algorithm.

The structure of the paper is as follows: Definition and traditional harmonic analyzing method have been described in Section2. Typical Data mining methods, characteristics pf power big data in wind farms, and the application of data mining on big data in wind farms have been discussed in Section 3. A new algorithm for the harmonic problem in wind energy generation system has been put forward and tested by actual running data of a 2MW doubly fed induction generator (DFIG) in Section4. A conclusion is provided in Section 5.

II CONVENTIONAL ANALYSIS OF HARMONIC PROBLEM

A. *Definition of Harmonics*

Harmonics refers to the sinusoidal oscillation whose frequency is an integer multiple of the fundamental frequency (50 or 60 Hz) [12]. Even harmonics are those whose frequencies are even integer multiples of fundamental frequency. Odd harmonics are those whose frequencies are odd integer multiples of fundamental frequency. Inter-harmonics are sinusoidal oscillations whose frequencies are not integer multiples of the fundamental frequency [12].

1) *Total harmonic distortion*

The degree of deviation from sine waveform in distorted waveforms, which is caused by harmonics, is called Total Harmonic Distortion (THD). THD is equal to the percentage of the total harmonics and the fundamental component in RMS [13]. This index is widely used to represent the distortion rate of the voltages or currents. Notably, the currents' THD is a relative concept, as the fundamental component of the currents will considerably vary with the change of the loads.

$$THD = \frac{\sum_{h=2}^{50} U_h^2}{U_1} \quad (1)$$

2) *Total demand distortion*

According to the international standard IEEE 519-1992, we can replace the RMS of fundamental current with that of maximum load current in equation (2-1) to avoid the variation of the fundamental current, which is called Total Demand Distortion (TDD) [13].

$$TDD = \frac{\sum_{h=2}^{50} I_h^2}{I_{rated}} \qquad (2)$$

B. *Calculation of Harmonics*

The calculation of harmonics is based on the concept of Fourier series, which can used to decompose a non-sinusoidal but periodic wave into a series of sinusoidal waves. The Fast Fourier Transform (FFT) algorithm is one of the most widely used techniques to calculate the harmonics. The accuracy of FFT is very high provided that [14]:

(a) The signal is strictly periodic with the fundamental frequency;

(b) The sampling frequency is integer multiples of the fundamental frequency;

(c) The sampling frequency is at least two times of the highest frequency of the signal to be analyzed.

Hardware induced sampling errors are related to the analogue-to-digital Converter and the non-linear problem of voltage or current transformer. On the other hand, since the voltage and current wave change with time, windowed FFTs are used to deal with the time-varying characteristics of harmonics. As a consequence, software-induced sample errors may occur, such as aliasing, leakage and picket-fence effect [14]. Aliasing occurs when a signal includes the components with the frequencies higher than the Nyquist frequency $f_N = f_s/2$, where f_s is sampling frequency. The frequency component f_x higher than f_N will appear a low frequency component $f_{low} = f_s - f_x$, where $f_N < f_x < f_s$ [15]. Leakage refers to the phenomenon that energy obviously extent from one frequency to another frequency caused by non-integer periodic sampling. Picket-fence effect is basically a resolution bias error produced by inter-harmonics. Advanced signal processing techniques have been used to mitigate leakage and Picket-fence effect [14].

III DATA MINING VS. BIG DATA IN WIND FARMS

A. *Typical data mining method*

Data Mining is to discover the implicit, effective, valuable and apprehensible modes in large volume of unordered data, then to get useful knowledge, and to acquire trend and correlation, which can provide decision support ability on the level of solution for consumers [16]. The most popular methods of data mining can be classified into association analysis, classification analysis and clustering analysis. More specifically, typical clustering algorithms can be divided into five parts: partition methods [17] (k-means, k-medoids etc.), hierarchical method [17] (BIRCH, CURE etc.), density-based methods [17] (DBSCAN, OPTICS etc.), grid-based methods [18] (STING, CLIQE etc.) and model-based methods [19-20] (COBWEB, SOFM neural network etc.)

B. *Application Analysis of Data Mining Algorithms With Big Data in Wind Farms*

The dimension and amount of traditional data sets are generally small. For analyzing this kind of small datasets, many mature algorithms or analyzing method have been proposed and work well. Generally speaking, we can analyze these small datasets to achieve desired results by applying traditional data analyzing methods or typical data mining algorithms.

With the characteristics named 4Vs, big data surpasses traditional datasets both on dimension and amount, for which traditional data analyzing methods or typical data mining algorithms are not suitable to complete relevant analysis. It’s necessary to improve those methods or algorithms to adapt to the characteristics of big data, which needs much stronger software and hardware technologies on data processing.

With the characteristics named 3V3E, power big data in wind farms is a specific embodiment of big data on power industry, especially on the wind energy generating sector, which can be analyzed in the following strategy: (1) based on the characteristics of big data produced in the wind energy generating systems, manipulating and filtering these power big data by typical digital signal processing methods on power system to obtain a much smaller dataset on specific problem or phenomenon; (2) then mining the new dataset by traditional data mining methods. The strategy can be used to deal with the analyzing task of big data under normal software and hardware facilities. Therefore, banded application of normal data processing methods and typical data mining algorithms is the best choice to adapt to the analysis of big data on the condition that the existing software and hardware technology failed to meet the requirements of big data analysis.

IV BIG DATA ANALYSIS ON THE BASIS OF FFT AND CLUSTERING METHODS

A. *Big Data Measurement in a Wind Farm*

My research group has done the data measuring on the running conditions of a certain wind turbine in GuangRao wind farm in Shandong province. The wind turbine is a W2000 series DFIG produced by Shanghai electric group. The rated power is 2MW, the rated grid-connected voltage is 690V. The designed running range of wind speed is as follows: rated wind speed is 10.4 m/s, cut-in wind speed is 3 m/s, and cut-out wind speed is 25 m/s. The real-time monitoring point is located at the terminal of the wind turbine, where 10-minutes data has been recorded for every power or wind range. The measured data types include phase voltage, phase current and wind speed.

B. *FFT Analysis of Big Data in Wind Farm*

1) *Basic analytical methods of harmonic problem in wind farms*

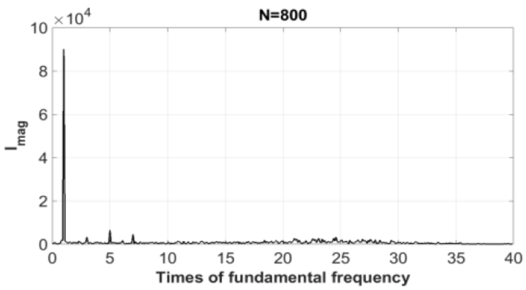
When we doing the power system analysis, we generally do the calculations of power system after transforming the original data value into per-unit value. As there’re different types of data in wind farm, it’s necessary to transform the measured data to per-unit value for objective analysis. Let $S_N = 2MW$, $U_N = 690V$, $V_N = 10.4m/s$ as the reference values of power capacity, voltage and wind speed, then we can get the reference values of current, cut-in and cut-out wind speed.

$$I_N = \frac{S_N}{\sqrt{3}U_N} = \frac{2MW}{\sqrt{3}*0.69kV} = 1673.5A \quad (3)$$

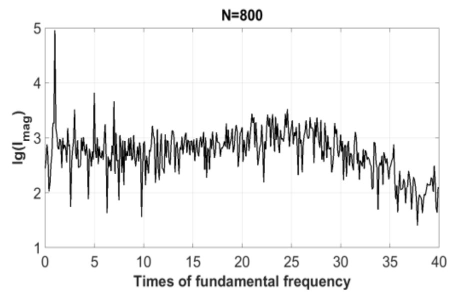
$$V_{ci} = \frac{V_{ci}}{V_n} = \frac{3m/s}{10.4m/s} = 0.2885 \quad (4)$$

$$V_{co} = \frac{V_{co}}{V_n} = \frac{25m/s}{10.4m/s} = 2.4038 \quad (5)$$

After doing data pre-processing by transform the original data into per-unit value, we can use the per-unit dataset to do harmonic analysis by FFT methods. The FFT output of the first time-window with 800 points as shown in Fig. 1(a). Generally speaking, the magnitude of fundamental component is much higher than the magnitude of every harmonic component, which will make it hard to observe the magnitude differences of all harmonic components. There're two common methods to deal with this problem. The first method is to take logarithm for the magnitude of every harmonic component, as shown in Fig. 1(b). This method can be used to observe the variation trend of every harmonic component. However, it's hard to identify the differences of all harmonic components precisely. The second method is to filter out the fundamental component before we do the harmonic analysis, as shown in Fig. 2(b). Notice the concerned object is the magnitude and frequency of every harmonic component, instead of the inter-harmonics between adjacent harmonics. Therefore, based on the basic characteristics of the FFT spectrum, we can screen out all harmonic components, as shown in Fig. 2(c) and Fig. 2(d).

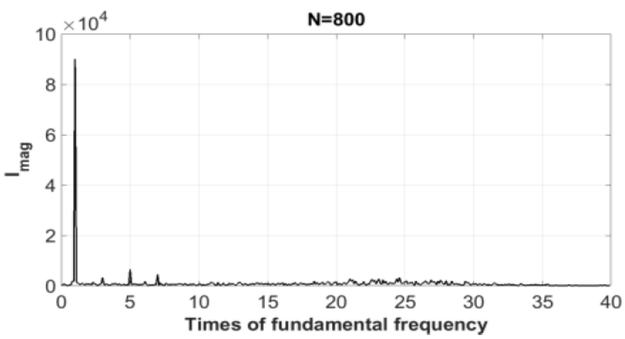


(a) Original drawing

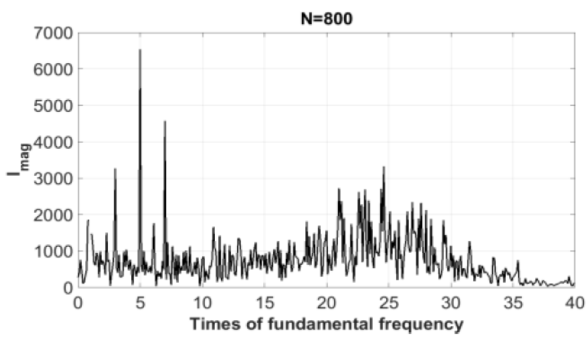


(b) Take logarithm

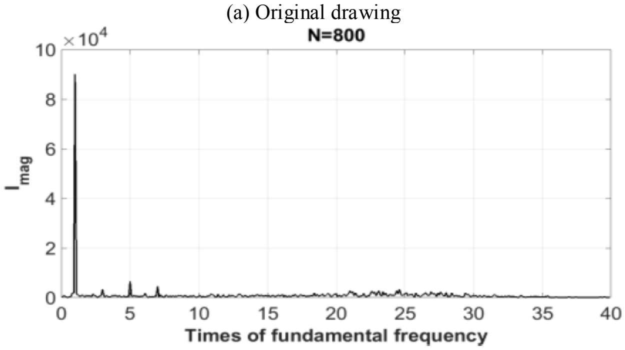
Figure 1. FFT output of the first time-window



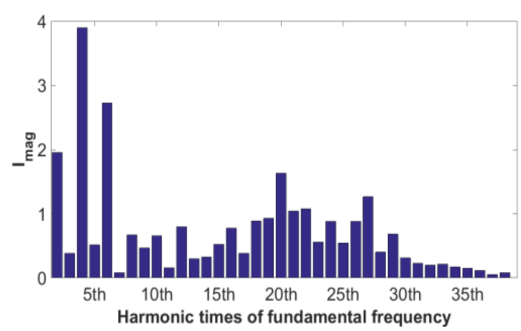
(a) Original drawing



(b) Remove fundamental component



(c) Include fundamental component



(d) Remove fundamental component

Figure 2. FFT output of the first time-window

2) *Big data analysis on the harmonic problem in wind farms*

(1) Data analysis methods

We can analyze power big data in wind farms by moving time-windowed FFT method, and do FFT analysis for dataset in every timed window. The basic analysis procedure can be listed as follows:

Step1: do data pre-processing by transform the original data into per-unit value;

Step2: analyze per-unit dataset by time-windowed FFT method, where time window width is 800 points (that is 10 circles);

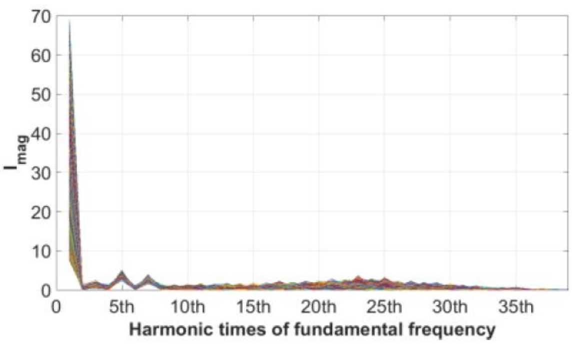
Step3: store the magnitude of different harmonic component into a new dataset;

Step4: analyze the harmonics based on the harmonic spectral dataset.

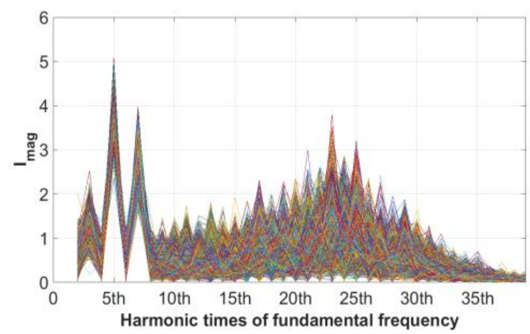
As the original data is stored according to different range of wind speed and every data segment is a set of sequential data in 10 minutes, it's necessary to loop over different data segment according to the procedure listed above, when we deal with the power big data of a whole day or longer time.

(2) Simulation analysis

First of all, we can put the FFT outputs of every time window for the first data segment onto a plane graph, as shown in Fig. 3. The magnitudes of 5th and 7th harmonics are significantly larger than the component with higher harmonic frequency in Fig. 3(b). However, the magnitudes of the components with higher frequency fluctuates obviously and are overlapped, which is unobservable to be analyzed.



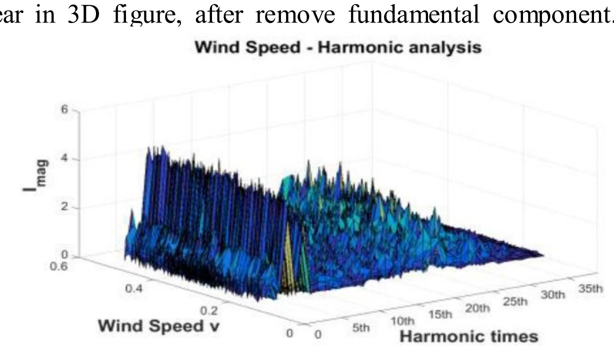
(a) Original drawing



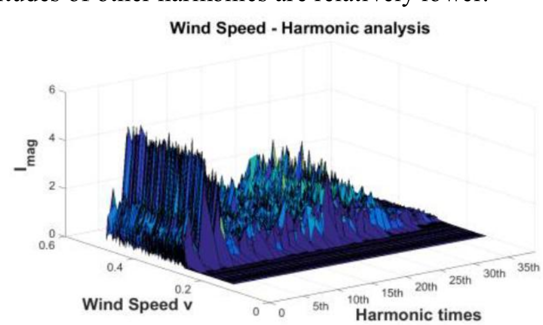
(b) Remove fundamental component

Figure 3. FFT-analyzing plane figures of the first data segment

As wind speed has volatility with time, the average wind speed or active power in different time window will fluctuate over time. In a power system whose fundamental frequency is 50Hz, the time span of 10 periods is $20\mu s$, during which the wind speed or actve power in every time window can be appximately considered as constant. As shown in Fig. 4(b), the magntides of harmonics at different wind speed vary a lot, it's hard to depict the variation of harmonic magnitude by plane figures. Therefore, it's necessary to depict the harmonc distribution at different wind speed, as shown in Fig. 4. As shown in Fig. 4(a), the variation of harmonic magnitudes becomee much more clear in 3D figure, after remove fundamental component. Notice that the wind turbine is running at the mode of off-line when wind speed is lower than 0.2885 p.u. (Eq. 11), as it dosen't fulfill the parallelled operating requirements, and the corresponding harmonic magnitudes are rather small. In order to observe the details of the harmonic conditions in the on-line operating period, it's necessary to filter out the wind speed region lower than cut-in wind speed, as shown in Fig. 4(b). Notice that the magnitudes of 5th , 7th ,23th harmonics are relatively larger than others, in which the 5th harmonics has the highest magnitude; for even harminics and the harminics whose times is larger than 31, the magnitudes are extremely low, which is negligible; the magnitudes of other harmonics are relatively lower.



(a) Remove fundamental component



(b) Remove the part lower than cut-in wind speed

Figure 4. Wind speed vs. FFT-spectrum 3D figures of the first data segment

The first data segment has been analyzed by FFT method in Fig. 4, which just includes the data at the wind-speed range of 0.15~0.55 p.u. In order to do full-scale harmonic analysis from cut-in wind speed V_{ci} to cut-out wind speed V_{co}, it's necessary to process operating data for a whole day or longer, which will challenge the ability of basic hardwares and softwares to process big data and test the analyzing ability of analysts. Therefore, it become a feasible way to extract the typical harmonic modes of every data segment by data mining algorithm, then data analysis in full-scale can be carried out.

C. *Clustering Analysis of the FFT Outputs*

1) *Algorithm selection*

Clustering is one of the most common used analytical tools, and many clustering algorithms have been developed, in which k-means algorithm and k-medoids algorithm are typical partition clustering method. As based on the average values in different clusters, k-means algorithm is rather sensitive to abnormal data. Thus, the existing abnormal data will play a great influence on the selection of clustering center. However, k-medoids algorithm take medoids as clustering centers, which will be more robust than k-means algorithm on processing abnormal data and noises. Therefore, we should use k-medoids algorithm to reduce the influence of abnormal or noise data.

Similarity measurement in different samples are necessary when we do clustering analysis, and the general method is to calculate the distance of samples. Typical similarity measurement or distance includes Euclidean Distance, Manhattan Distance, Chebyshev Distance, and Minkowski Distance etc.

2) *Case study*

(1) Spectral clustering method

In this paper, we do the clustering analysis by k-medoids algorithm for the power big data of wind farm in the frequency domain, which is generated from the FFT analysis above. On the condition of filtering out the fundamental component, we take the magnitude of every harmonics as the object of clustering analysis. In this paper, we take Euclidean Distance as the similarity measurement, and the basic analyzing procedure is listed as follows:

Step1: extract all harmonic component from a data segment in 10 minutes;

Step2: do the clustering analysis of extracted FFT spectral data by k-medoids algorithm to get the k best center objects of the data segment;

Step3: Store the k best center objects into a new dataset in turn;

Step4: analyze the harmonics based on the new harmonic spectral dataset.

As mentioned above, the measured data is recorded in different ranges of wind speed and every data segment has 10-minutes long data. The generated spectral dataset by FFT method corresponds to the original data segment in 10 minutes. When do the spectral clustering analysis of power big data in wind farm for a whole day or longer time, we should loop over different data segment according to the procedure listed above (Step1~Step3). The whole program flow chart for FFT spectral clustering analysis of power big data in wind farm is shown in Fig. 5.

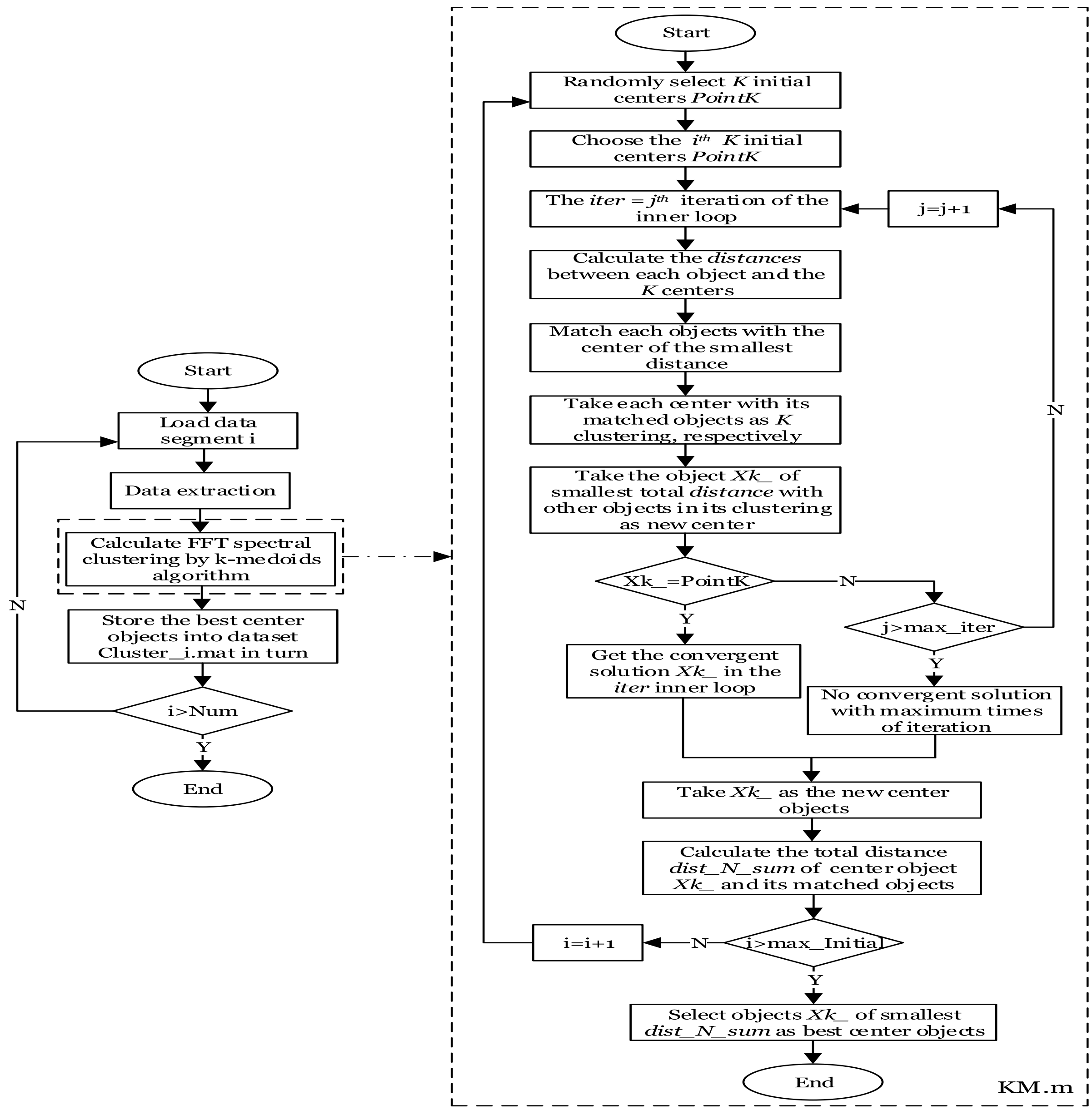
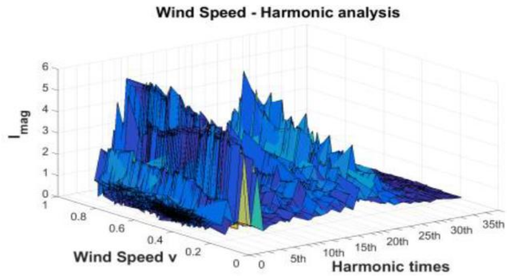


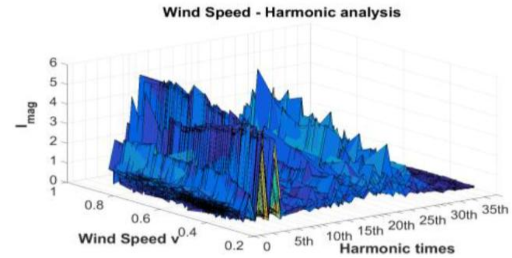
Figure5. Program flowchart of cl_main.m

(2) Simulation & Analysis

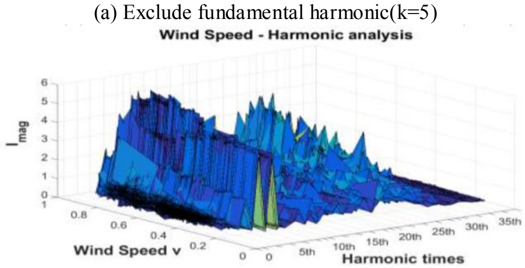
Let the cluster number to be k=5, k=10 and k=20, we can get the corresponding wind speed vs. harmonic spectral diagram, in Fig. 6(a), Fig. 6(c) and Fig. 6(e), by the spectral clustering algorithm above. Notice that the outcome of spectral clustering analysis includes the harmonic spectrum at the wind speed ranging from 0.15 p.u. to 1.0 p.u. However, we generally pay more attention to the harmonic generated in the on-line operating period of wind turbines, that is, the spectrum at the wind speed ranging from cut-in speed to cut-off speed(0.2885~2.4038). Thus, we can get the new wind speed vs. harmonic spectral diagram by filtering out those spectrum generated in the off-line period, as shown in Fig. 6(b), Fig. 6(d) and Fig. 6(f). Notice that the characteristics of the wind speed vs. harmonic spectrum change little when the spectral data is classified into different number k of clusters, which all can be used to do the full-scale wind speed vs. harmonics analysis. In addition, the k value is larger, the diagram of wind speed vs. harmonic spectrum will be more subtle, which will be beneficial to precise analysis. Therefore, it’s feasible to select different k values in different analyzing goals for big data in wind farms. First of all, extract k typical modes of wind speed vs. harmonics in every data segment and store them into a new dataset. Then, we can do the full-scale wind speed vs. harmonics analysis of the new dataset, which have ideal precision of data analysis.



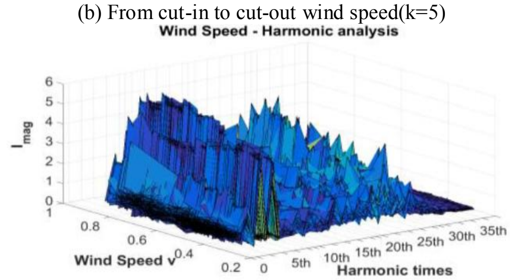
(a) Exclude fundamental harmonic(k=5)



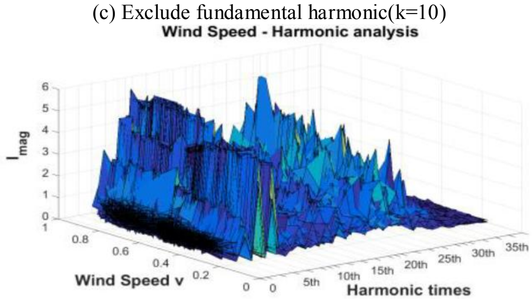
(b) From cut-in to cut-out wind speed(k=5)



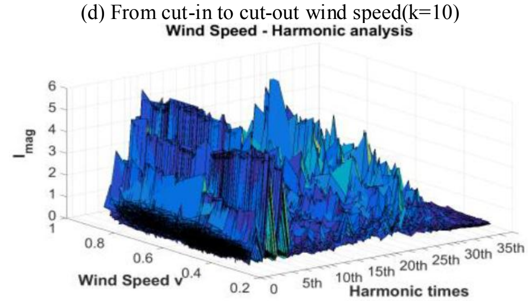
(c) Exclude fundamental harmonic(k=10)



(d) From cut-in to cut-out wind speed(k=10)



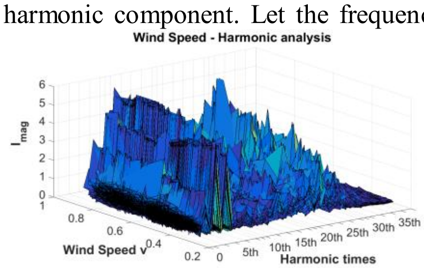
(e) Exclude fundamental harmonic(k=20)



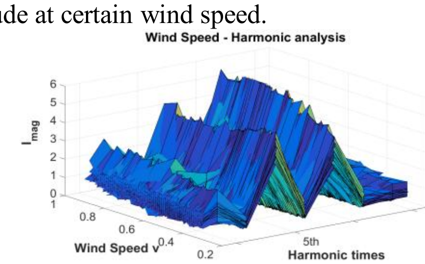
(f) From cut-in to cut-out wind speed(k=20)

Figure 6. Wind speed vs. FFT-spectrum 3D figures of a certain day (24h)

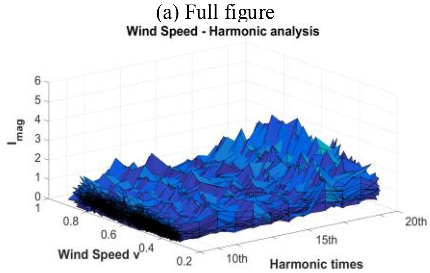
In order to acquire better harmonic spectrum, let the cluster number k=20 and observe the wind speed vs. harmonic spectrum ranging from cut-in wind speed to cut-out wind speed in detail, as shown in Fig. 7(a). Furthermore, we can enlarge the plot of the wind speed vs. harmonic spectrum in different frequency bands to observe the detail of every harmonic component. Let the frequency intervals [2.5 9.5], [9 20], [19.5 31.5] as observed objects, then plot corresponding relationship of wind speed vs. harmonic spectrum in Fig. 7(b), Fig. 7(c), Fig. 7(d). Notice that the magnitudes of 5th, 7th, 23th, 25th harmonics is the largest; that of 3th, 20th, 22th, 24th, 26th harmonics is relatively large; in addition, 27th harmonics have relatively large magnitude at certain wind speed.



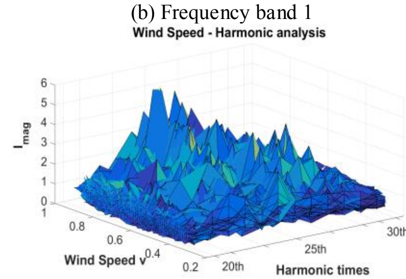
(a) Full figure



(b) Frequency band 1



(c) Frequency band 2



(d) Frequency band 3

Figure 7. Wind speed vs. FFT-spectrum 3D figures of a certain day (24h): k=20

V CONCLUSION

In this paper, with the characteristics of 3V3E, big data in wind farms has been analyzed by the combined application of normal FFT method and typical data mining algorithm to systematically research the harmonic problem in wind energy generating system.

Firstly, transform the big data from time domain to frequency domain by FFT method, and then extract the harmonic components to form a new harmonic dataset, which is an effective method to reduce the dimension of power big data in the frequency domain, as the volume of dataset in 10 minutes is 2400000*8, and the volume of dataset of corresponding extracted harmonic dataset is 3000*42*6. Secondly, analyze the harmonic dataset in different data segments by clustering algorithm (k-medoids) to extract typical harmonic modes at different wind speed. Finally, store all the typical harmonic modes of a whole day or longer time into a new spectral dataset, and then do the comprehensive analysis of wind speed vs. harmonics relationship.

As big data in wind farms continuously increases, typical modes of full-scale wind speed vs. harmonic spectrum from cut-in speed to cut-out speed can be acquired by the algorithm above. In addition, dataset of typical modes in different data granularity can be acquired by changing the clustering number k of every data segment, which can be used to analyzing the characteristics of harmonic magnitude in full-scale wind speed. Some interesting findings have been found by analyzing big data of wind farm for 24 hours: the magnitudes of 3th, 5th, 7th, 23th, 25th harmonics are relatively higher, in which 5th and 7th harmonics have the highest magnitudes; the magnitude of 5th harmonics decreases with the increase of wind speed, while that of 3th, 7th, 23th, 25th harmonics increases and that of even harmonics change little. In addition, the current TDD increases with the increase of wind speed, and fluctuates between 1.25% and 2.76%.

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (51277119), and Surface Project (51477098).

REFERENCES

[1] Coster E J, Myrzik J, Kruimer B, et al. Integration issues of distributed generation in distribution grids[J]. Proceedings of the IEEE, 2011, 99(1): 28-39.

[2] Ivry P M, Rawa M J, Thomas D W P, et al. Power quality of a voltage source converter in a smart grid[C]//PowerTech (POWERTECH), 2013 IEEE Grenoble. IEEE, 2013: 1-6.

[3] Swan M. The quantified self: Fundamental disruption in big data science and biological discovery[J]. Big Data, 2013, 1(2): 85-99.

[4] Chen H, Chiang R H L, Storey V C. Business Intelligence and Analytics: From Big Data to Big Impact[J]. MIS quarterly, 2012, 36(4): 1165-1188.

[5] LaValle S, Lesser E, Shockley R, et al. Big data, analytics and the path from insights to value[J]. MIT sloan management review, 2013, 21.

[6] Quinn G P, Keough M J. Experimental design and data analysis for biologists[M]. Cambridge University Press, 2002.

[7] Nelson W B. Accelerated testing: statistical models, test plans, and data analysis[M]. John Wiley & Sons, 2009.

[8] Brillinger D R. Time series: data analysis and theory[M]. Siam, 1981.

[9] Chandrasekhar U, Reddy A, Rath R. A comparative study of enterprise and open source big data analytical tools[C]//Information & Communication Technologies (ICT), 2013 IEEE Conference on. IEEE, 2013: 372-377.

[10] Mera D, Batko M, Zezula P. Towards Fast Multimedia Feature Extraction: Hadoop or Storm[C]//Multimedia (ISM), 2014 IEEE International Symposium on. IEEE, 2014: 106-109.

[11] Jovic A, Brkic K, Bogunovic N. An overview of free software tools for general data mining[C]//Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on. IEEE, 2014: 1112-1117.

[12] Schlabbach J, Blume D, Stephanblome T. Voltage quality in electrical power systems [M]. The Institution of Electrical Engineering, 2001.

[13] Arrillaga J, Watson N R, Chen S. Power system quality assessment [M]. John Wiley & Sons, 2000.

[14] Baghzouz Y. An overview on probabilistic aspects of harmonics in power systems[C]//Power Engineering Society General Meeting, 2005. IEEE. IEEE, 2005: 2394-2396.

[15] Bollen M H, Gu I. Signal processing of power quality disturbances [M]. John Wiley & Sons, 2006.

[16] Kantardzic M. Data mining: concepts, models, methods, and algorithms[M]. John Wiley & Sons, 2011.

[17] Gole S, Tidke B. A survey of big data in social media using data mining techniques[C]//Advanced Computing and Communication Systems, 2015 International Conference on. IEEE, 2015: 1-6.

[18] Wang W, Yang J, Muntz R. STING+: An approach to active spatial data mining[C]//Data Engineering, 1999. Proceedings., 15th International Conference on. IEEE, 1999: 116-125.

[19] Fisher D H. Knowledge acquisition via incremental conceptual clustering[J]. Machine learning, 1987, 2(2): 139-172.

[20] QingLan H, LongZhen D. Multi-level association rule mining based on clustering partition[C]//Intelligent System Design and Engineering Applications (ISDEA), 2013 Third International Conference on. IEEE, 2013: 982-985.