Compression of Power System Signals with Wavelets

Chun Sing Lai, *Student Member, IEEE*
Energy and Power Group
University of Oxford
Oxfordshire, UK
chun.lai@eng.ox.ac.uk

***Abstract*— This paper reviews some technical issues in investigating real-time data analytics for future power grid systems. Wavelet transform will be used to demonstrate how current and voltage signals are compressed as a possible solution to data compression, and limitations of the technique will be highlighted.**

***Index Terms*—Big Data, Data Compression, Wavelet, Smart Grid Analytics**

I. INTRODUCTION

Smart meters and smart devices are progressively entering the market, causing immense amounts of data to be transmitted over the communication system as illustrated in Fig. 1. Data are received by devices, and then transmitted with wireless technology such as ZigBee, WiFi [1]. This method ensures simplistic and inexpensive transmission of data. However, bandwidth is limited and transmission congestion could occur. Device data should therefore be compressed before transmission. Decision making with be a major issue, massive amounts of data are arriving at utility companies that have no idea what to do with them [2-3].

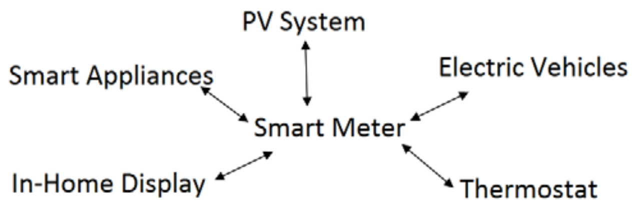

Figure 1: Typical Smart Home Network

Big data will be a major challenge in the future power grid, as it determines the fate of Smart Grid as defined in Section II later. As the power grid moves from the slow and static modern grid to the smart and intelligent grid, the system needs to manage huge amounts of data from millions of devices in real-time and able to incorporate various sets of control functions and data analytics. It is worth noting that the power grid is a critical piece of infrastructure especially in the 21st century. The system must be extremely reliable, highly scalable, environmental friendly and very secure. Today's outdated power systems fail to sustain the quality and scalability challenge of service guarantees at the same time [4].

This paper reviews some technical issues in analyzing immense quantity of real-time data analytics for future power grid systems.

II. BIG DATA

Big Data is recognized as the driver for competition [9]. The characteristics of Big Data are high volume, high variety and high velocity as results are expected to be transmitted rapidly [10]. Practically, Big Data is mainly motivated by the need to analyze substantial volumes of data to find information and relationships for decision making [11-12].

To give an idea how much data smart meters could create, a calculation was performed as follows:

By assuming that the sampling rate is 10 kHz and with each sample has 16 bits of information. In one second, the amount of information will be 10,000 samples/second * 16 bits/sample, that is 160,000 bits/second. Naturally, for each minute it will require 160,000 * 60 bytes that is 9.6 Mbytes of storage space. For 6 hours, there is a need of 9.6 * 6 * 60 Mbytes, that is, 3.456 Giga bytes. By assuming that there will be 50 million smart meters installed in the UK [26], 50*10^6 *3.456*10^9, that is 172.8 Petabyte of data will be created in 6 hours.

A good approach to study big data is to identify the various types of Big Data sources, apply the correct mining techniques to find the valuable information within each type, and then combine and report those new insights appropriately. This will ensure effective decisions can be made by the power utilities [13].

The large technology gap between industrial applications and decision makers remains a critical matter. Decision makers need to enhance their understanding in the technologies and data to be able to extract useful information to assist their strategic planning.

Processing data in a standalone manner is not enough. Methods are needed to reduce the difficulty to combine knowledge with Big Data analytics. As a good practice, this can only be done by combining both quantitative and qualitative analysis. Data sources together with the underlying systems need to be considered for integrating the information to achieve the speed and amount of data. This means there is a requirement to build a closer relationship between domain experts, mathematicians and computer engineers [14].

The data analytics system for future power grid face many unique challenges, which include real-time guarantees, scalability in term of the amount of data and the diverse of application to support high reliability, and security in a reasonable cost [4]. But it is not easy to find a system that can fulfill all the requirements. For example, the most commonly used system in power grid system, the OSIsoft PI system [20], can provide real-time guarantees but is not scalable to millions of sensors to be deployed in future power grids. The data systems developed in other domain can handle big data but cannot meet the real-time, reliability, and security requirements of power grid systems. It is an area for attention.

III. Wavelet Compression Example

"Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale [22]." The wavelets have a lot of benefits which are better than traditional Fourier method in analyzing physical situations where the signal contains discontinues. The wavelet applications are very useful in many fields such as image compression, turbulence, human vision, radar, and earthquake prediction. In a word, "a wavelet is a waveform of effectively limited duration that has an average value of zero [22]." Wavelet transform is widely used in image processing, data compression and transmission.

When considering data compression there are mainly two possible approaches, namely:

- Lossless algorithms are capable of a perfect reconstruction of the original data but have inferior compression ratios;
- Lossy algorithms have higher compression ratios but introduce some error when reconstructing the uncompressed data.

This is true for compression of sampled electrical signals.

In [23], it is noted that the wavelet transform can bring remarkable advantages to all fields in which large amounts of data are processed for systems such as fault recording and SCADA/EMS/DMS systems.

A. *Wavelet Compression Algorithm for Data Preservation and Power Quality Analysis*

As suggested by IEC 61000-4-30 [24], voltage and current waveforms are always preferred in dynamic performance analysis and trouble-shooting. Nevertheless, recording waveforms require large storage for data preservation, and also huge bandwidth for monitoring. Their equipment is usually expensive and unsuitable for visualization.

Waveform compression algorithm is proposed to reduce the storage and bandwidth requirements. It is designed for compressing voltage and current waveforms in the power grid effectively, and which can be executed in a low cost digital signal processor for continuous waveform recording.

Voltage and current waveforms are often used for trouble-shooting. However, recording and monitoring of these power waveforms require huge digital storage and network bandwidth.

The Discrete Wavelet Transform transforms the discrete data from time domain into time-frequency domain. The values of the transformed data in time-frequency domain are called the coefficients. The coefficients with small absolute values are dominated by noise, while the coefficients with large absolute values carry more data information than noise. In the second step the wavelet coefficients are set to zero (hard threshold rule) or shrink (soft threshold rule), if they are not crossing certain threshold level. The last step is to reconstruct the signal from the resultant coefficients using Inverse Discrete Wavelet Transform (IDWT).

The Discrete Wavelet Transform (DWT) is a powerful mathematic tool for time-frequency analysis of non-stationary signals. It uses multi-resolution filter banks for the signal analysis. The general form of DWT at L-level is written in terms of L detail coefficients $d_j(k)$, and the Lth Level approximation coefficients $c_L(k)$ can be expressed as [7]:

$$x(t) = \sum_{j=1}^{L} \sum_{k} d_j(k)\psi_j(t) + \sum_{k} c_L(k)\varphi_L(t) \qquad (1)$$

Where, the functions $\varphi_L(t)$ and $\psi_j(t)$ are respectively known as the scaling function and the mother wavelet.

A compression algorithm is proposed to reduce both storage and bandwidth requirements. Fig. 2 explains how the sampled data are compressed with wavelets. The mother wavelet used in the algorithm is db20 and the number of decomposition level is assumed to be 5.

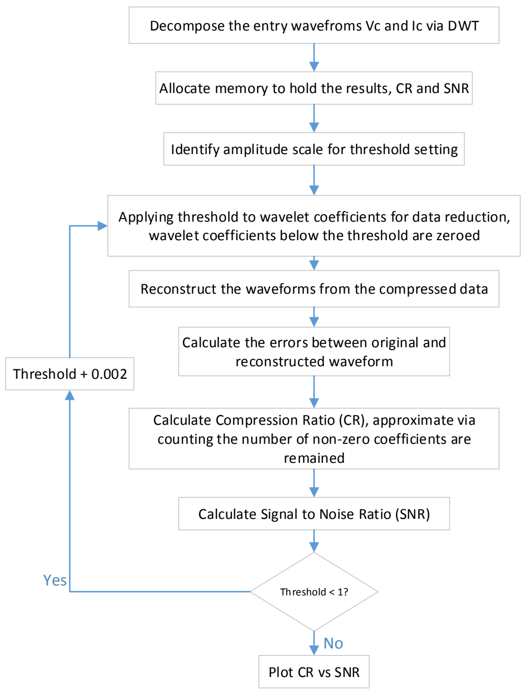

Figure 2: Flowchart of DWT used to compress voltage and current signals

Real data collected from a data logger were used to test the effectiveness of the compression algorithm. The data was obtained by sampling the voltage and current from a power network at Christ Church College, Oxford, UK on 2011-10-29 at 00-07-50. The sampling frequency was at 10 kHz for the duration of 10 minutes. The data processing and analysis were done offline in MATLAB.

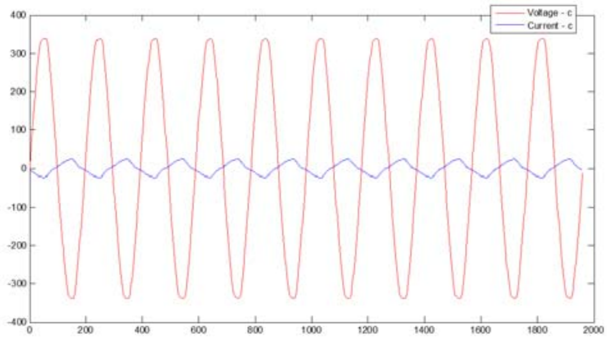

Figure 3: Original instantaneous sampled voltage (red) and current (blue) signals for the first 10 cycles

B. *Case Study 1: Change of Threshold*

There are two types of thresholds, hard and soft threshold. Hard threshold zeroes out small coefficients, resulting in an efficient representation. Soft threshold softens the coefficients exceeding the threshold by lowering them by the threshold value. When threshold is applied, no perfect reconstruction of the original signal is possible. [27]

The hard-threshold value is given by,

$$T_{hard}(w;t) = wI(|w| > t)$$

And the soft-threshold value is given by

$$T_{soft}(w;t) = sgn(w)(|w| - t)I(|w| > t)$$

Where w is the wavelet coefficients, t is the threshold and I is the usual indicator function.

The DWT of a signal is calculated by passing the signal though a series of filter. The signals are passed through a lower pass filter (LPF) and a high pass filter (HPF) simultaneously. The two filters are related to each other and they are known as quadrature mirror filter. [8]

Since half the frequencies of the signal have now been removed, half the samples can be removed according to Nyquist's rule. The filter outputs are then subsampled by 2.

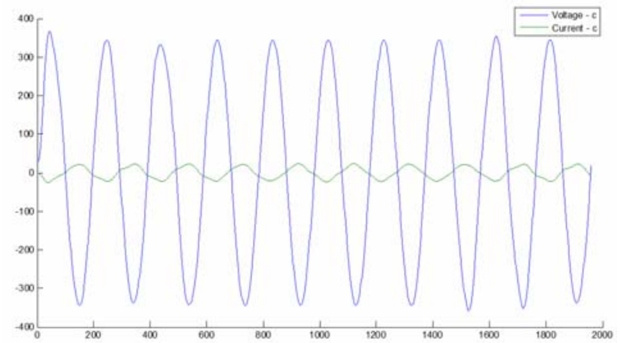

Figure 4: Reconstructed waveforms with threshold at 20%

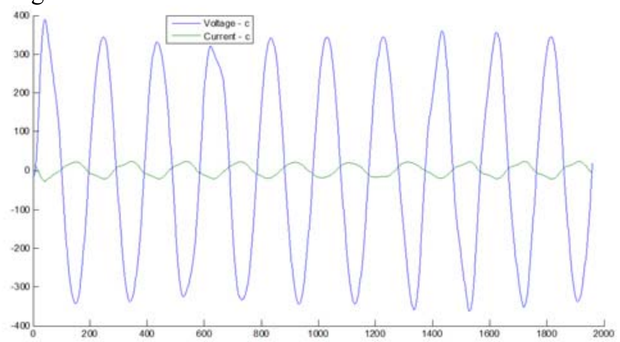

Figure 5: Reconstructed waveforms with threshold at 30%

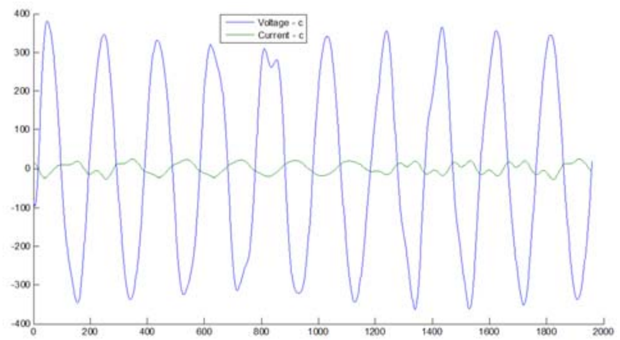

Figure 6: Reconstructed waveforms with threshold at 40%

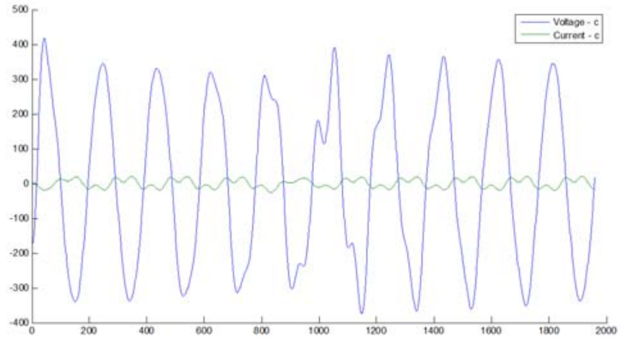

Figure 7: Reconstructed waveforms with threshold at 50%

The Compression Ratio (CR) and Signal-to-Noise Ratio (SNR) defined by equation (2) below were used to evaluate the waveform compression performance. A high SNR represents good preservation of waveform information and a high CR represents effective compression. The performance of the proposed algorithm was hence evaluated for achieving an optimal compromise between SNR and CR.

$$\text{Compression Ratio (CR)} = \frac{\text{original file size}}{\text{compressed file size}}, \qquad (2)$$

$$\text{SNR(dB)} = 10\log_{10}\left(\frac{\sum_{n=0}^{N-1}\|x[n]\|^2}{\sum_{n=0}^{N-1}\|x[n]-\hat{x}[n]\|^2}\right), \qquad (3)$$

where $x[n]$ and $\hat{x}[n]$, n=0,1,…, N-1 are the original and the reconstructed signal data respectively.

Table 1: Summary of CR and SNR with different threshold

	Compression Ratio		Signal to Noise Ratio	
Threshold	Vc	Ic	Vc	Ic
20	35.89	42.13	24.66	13.04
30	39.25	57.02	18.72	8.45
40	43.61	83.47	15.35	4.82
50	49.06	276.72	11.97	1.38

As the threshold increases, more sampling points will be zeroed, therefore the CR will be increased. However, less information will be retained and the SNR will be decreased.

C. *Case Study 2: Change of DWT Level*

After each level of decomposition, the input data is branched into two outputs, one associated to the upper half-band of the input signal and the other one to the lower half-band. In each branch, the sampling rate is half of that of the input data to that branch. Therefore, the overall sampling rate remains constant. Since decomposition of the DWT proceeds only on one branch, the lower half-band branch, it will result with different data. [21]

For many signals, the low-frequency content carries the most important information about the signal. In general, the low-frequency content carries the signal identity. The high-frequency content, on the other hand, gives details of the signal.

The approximation will be the high scale or low-frequency components of a signal.

The details are the low-scale, high frequency components.

A single level decomposition puts a signal through 2 complementary low-pass and high-pass filters as illustrated below:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} h_0 x[k][2n-k] \qquad (4)$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} h_1 x[k][2n-k] \qquad (5)$$

where h_0 is the Low Pass Filter and h_1 is the High Pass Filter.

Levels	Level 0 $x(k)$	Level 1 $d_1(k)$	Level 2 $d_2(k)$	Level 3 $d_3(k)$	Level 4 $d_4(k)$	Level 5 $d_5(k)$	Level 5 $c_5(k)$
Rates	10 kHz	5 kHz	2.5 kHz	1.25 kHz	625 Hz	312.5 Hz	312.5 Hz

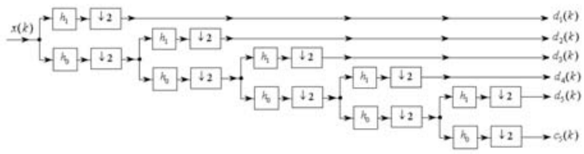

Figure 8: Five level wavelet decomposition tree

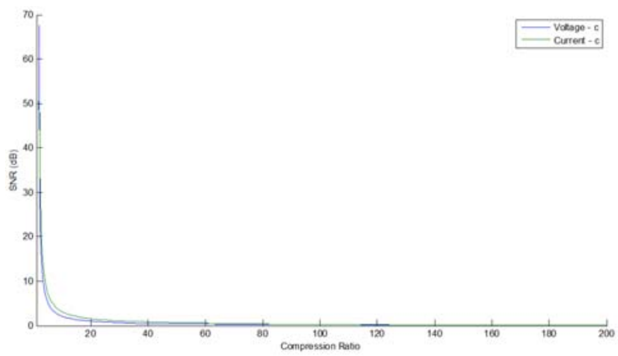

Figure 9: DWT at Level 1

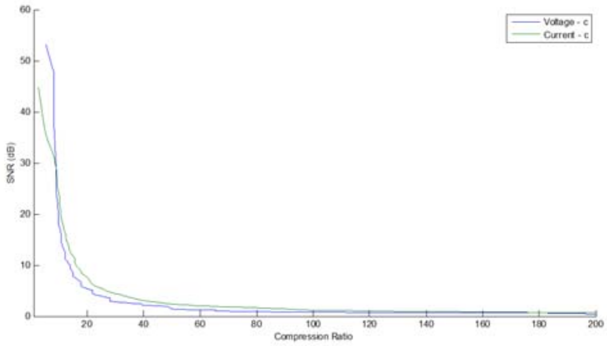

Figure 10: DWT at Level 3

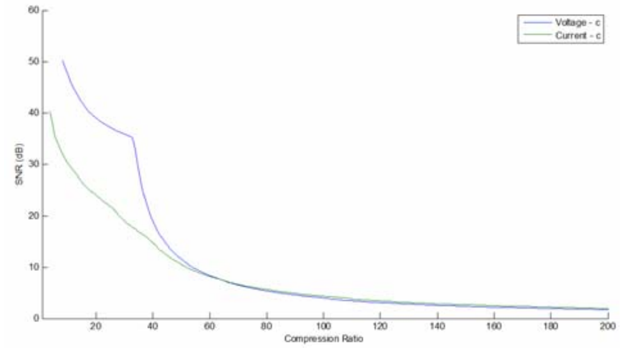

Figure 11: DWT at Level 5

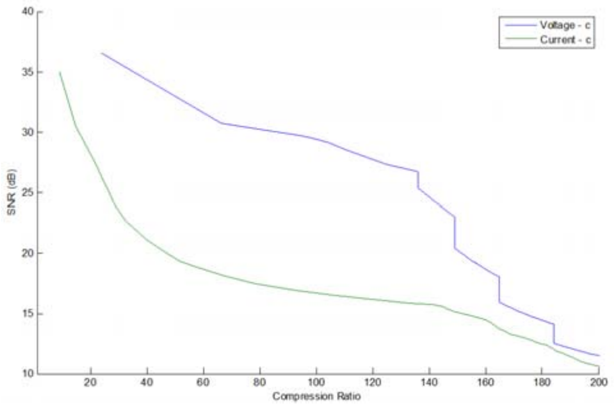

Figure 12: DWT at Level 10

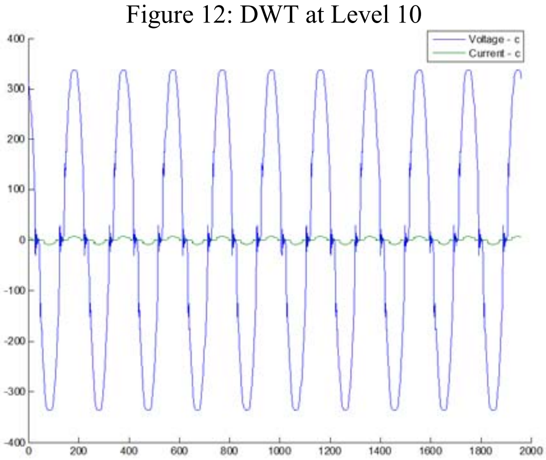

Figure 13: Threshold 30% Level 1

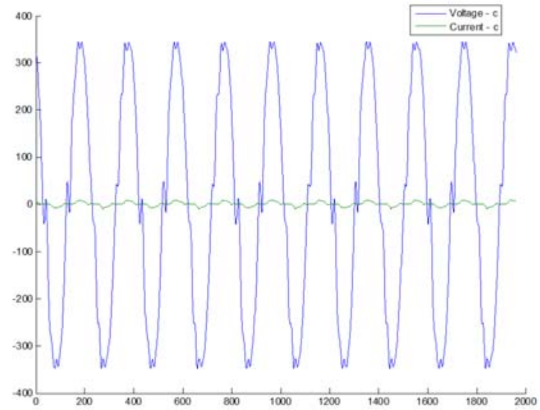

Figure 14: Threshold 30% Level 3

After performing compression on the current and voltage waveforms, the power waveforms can be reconstructed with these results. Figure 16 shows the power waveform from the compression results. The errors are in general reasonable small. For practical cases, it will be acceptable in many of the situations.

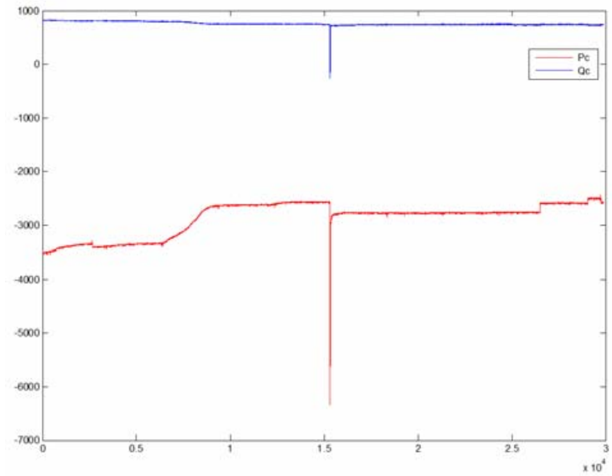

Figure 15: Original power waveforms P: Active Power; Q: Reactive Power

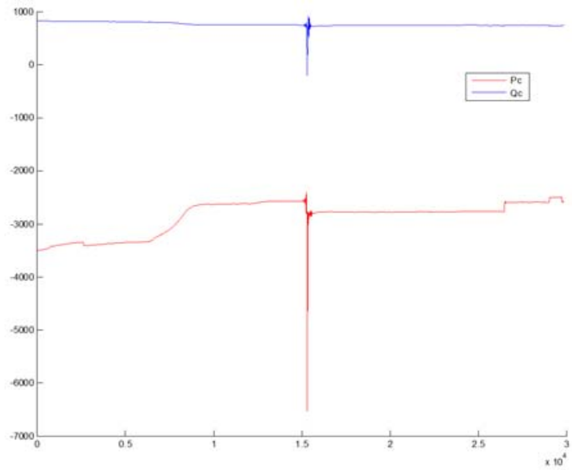

Figure 16: Reconstructed power waveforms

IV. CONCLUSIONS

This paper has given a critical review on the impact of big data on smart grid. It could be seen that due to smart grid deployment, there is a need to deal with 'fast' data, huge volume of data and different types of data sources in real-time. The paper has pointed out some essential requirements for the utilities to work for in order to maximum the benefit of smart grid. The paper has tutorial value for smart grid with big data. Although presently there are algorithms and techniques to relieve the big data crisis, but present methods need improvements and novel methods are desired in order to cope with the future big data issues. It is believed that Decision Support Systems need to incorporate with data compression mechanism to deal with big data situations effectively.

From the simulation results, it can be seen that it is possible to achieve a high compression rate by applying a higher threshold level and an increase in the sampling levels with wavelets. With this method, the waveforms are highly distorted and may not be applicable in some power system applications and future research is needed.

V. ACKNOWLEGEMENT

Sampled site data is provided by Christ Church, Oxford. The help is very much appreciated.

REFERENCES

[1] http://www.zigbee.org/About/UnderstandingZigBee.aspx [Assessed on 15 March 2014]

[2] http://www.forbes.com/sites/jeffmcmahon/2013/09/25/utilities-dumbstruck-by-big-data-from-smarter-grid/ [Assessed on 15 March 2014]

[3] http://www.smartgridnews.com/artman/publish/Business_Analytics/Forbes-Utilities-are-dumb-about-smart-grid-data-6084.html#.UpI5KOnuM5s [Assessed on 15 March 2014]

[4] Jian Yin, Poorva Sharma, Ian Gorton, Bora Akyol, "Large-Scale Data Challenges in Future Power Grids", 2013 IEEE Seventh International Symposium on Service-Oriented System Engineering

[5] http://www.mathworks.co.uk/help/wavelet/ug/data-compression.html

[6] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, "Wavelet Toolbox User's Guide", MATLAB, R2013b

[7] Yousef M. Hawwar, Ali M. Reza, Robert D. Turney, "Filtering (Denoising) in the Wavelet Transform Domain"

[8] Paulo Fernando Ribeiro, Carlos Augusto Duque, Paulo Marcio Ribeiro, Augusto Santiago Cerqueira, "Power Systems Signal Processing for Smart Grids", Wiley, 2013

[9] McGuire T., Manyika J., Chui M., July / August 2012, "Why Big Data is the New Competitive Advantage", Ivey Business Journal, www.iveybusinessjournal.com/topics/strategy/why-big-data-is-thenew-competitive-advantage [Assessed on 15 March 2014]

[10] Mark B., "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data". Gartner, June 27, 2011, http://www.gartner.com/newsroom/id/1731916 [Assessed on 15 March 2014]

[11] Johnson E. J., July/August 2012, "Big Data + Big Analytics = Big Opportunity", Journal of Financial Executive, pp. 1-4.

[12] Nichols, W., March 2013, "Advertising Analytics 2.0", Harvard Business Review, 91(3): 60-68.

[13] http://www.zdnet.com/top-10-categories-for-big-data-sources-and-mining-technologies-7000000926/ [Assessed on 15 March 2014]

[14] Davenport, Thomas H., and D. J. Patil., October 2012, "Data Scientist: The Sexiest Job of the 21st Century." Harvard Business Review 90(10):70-76.

[15] Firat Tekiner, John A. Keane, "Big Data Framework", 2013 IEEE International Conference on Systems, Man, and Cybernetics

[16] Allen B., Bresnahan J., Childers L., Foster I., Kandaswamy G., Kettimuthu R., Kordas J., Link M., Martin S., Pickett K., Tuecke S., February 2012, "Software as a Service for Data Scientists", Communications of the ACM, 55(2):81-88.

[17] Barton D. and Court D., October 2012, "Making Advanced Analytics Work for You", Harvard Business Review 89(10):78-83.

[18] McCallum J. S., March/April 2001, "Managing in the new economy: evolution or revolution?", Ivey Business Journal, http://www.iveybusinessjournal.com/topics/theorganization/managing-in-the-new-economy-evolution-or-revolution [Assessed on 15 March 2014]

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

[20] http://www.osisoft.com [Assessed on 15 March 2014]

[21] Stephane G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE, 1989

[22] Y. Meyer, "Wavelets: Algorithms and Applications", Society for Industrial and Applied Mathematics, Philadelphia, 1993, pp. 13-31, 101-105.

[23] L. Shang, J. Jaeger, R. Krebs, "Efficiency Analysis of Data Compression of Power System Transients using Wavelet Transform", 2003 IEEE Bologna PowerTech Conference, June 23-26, Bologna, Italy.

[24] Part 4-30: Testing and measurement techniques – Power quality measurement methods, International Standard, IEC 61000-4-30, 212.175.131.171/IEC/IEC61000-4-30%7Bed1.0%7Den_d.pdf [Assessed on 15 March 2014]

[25] Norman C. F. Tse, John Y.C. Chan, Wing-Hong Lau, T. Y. Poon, L. L. Lai, "Real-Time Power-Quality Monitoring With Hybrid Sinusoidal and Lifting Wavelet Compression Algorithm", IEEE Transactions on Power Delivery, Vol. 27, No. 4, October 2012.

[26] "Energy monitors: Smart meters and energy monitors explained", http://www.which.co.uk/energy/creating-an-energy-saving-home/guides/smart-meters-and-energy-monitors-explained/smart-meter-roll-out/ [Assessed on 15 March 2014]

[27] B. Ismail and Anjum Khan, Image De-noising with a New Threshold Value Using Wavelets, Journal of Data Science 10(2012), 259-270