

Accuracy and Data Compression Trade-Offs for Power Quality Disturbance Representation with DWT and PCA techniques

L.B. Soares¹, R. E. Dapper³, C. Crovato², S. Bampi¹ and A. A. Susin^{1,3}

¹ Informatics Institute – Microelectronics Program (PGMicro)
Federal University of Rio Grande do Sul (UFRGS)

Campus do Vale – Av. Bento Gonçalves, 9500, Porto Alegre (Brazil)

Phone/Fax number:+55 (51) 3308-6156 / +55 (51) 3308-7308, e-mail: [lbsoares](mailto:lbsoares@inf.ufrgs.br), [bampi](mailto:bampi@inf.ufrgs.br) @inf.ufrgs.br,

² Feevale University - ICET

Campus II ERS239, 2755, Novo Hamburgo – CEP: 93352-000 (Brazil)

Phone number:+55 (51) 3586-8800, e-mail: cesarc@feevale.br

³ Electrical Engineering department

Federal University of Rio Grande do Sul (UFRGS)

Av. Osvaldo Aranha, 103, Porto Alegre (Brazil)

Phone number:+55 (51) 3308-3515, e-mail: [roque.dapper](mailto:roque.dapper@ufrgs.br), [altamiro](mailto:altamiro@ufrgs.br) @ufrgs.br

Abstract. This paper analyses and compares different digital signal processing techniques to capture with fidelity the power lines disturbances. The benefits and shortcomings of both Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) techniques are assessed. Their implementation, application and test-cases with recorded power signals are described in the paper. We define the data compression ratio (or efficiency) and the signal reconstruction accuracy after compression as the main criteria for comparing these techniques. In this paper the trade-offs of accuracy in the disturbances representation and compression efficiency were exploited. After systematic comparisons, our experimental results have shown shortcomings of the DWT and PCA techniques not yet addressed in previous works.

Key words

Power Quality, Data Compression, Discrete Wavelet Transform, Principal Component Analysis

1. Introduction

In Power Quality (PQ) scope, disturbances to AC power lines can play a harmful role to electrical equipment. Data acquisition and digital monitoring systems are applied to detect these undesired electrical disturbances. Since the acquired data may be stored in a digital memory, then compression techniques are mandatory. Researchers in fact have made an effort to achieve a good balance between signal reconstruction accuracy and data compression efficiency. Therefore, many proposals have been developed to deal with this scenario.

DWT has been applied in Power Quality disturbances for a long time [1]-[14]. Many authors have developed disturbance detectors and data compressors based on it. In compression terms, the most usual process is to discard detail coefficients according to a threshold value (e.g. [4], [7], [8], [11] and [14]). This type of approach

inserts irreversible loss of information. On the other hand, that is counterbalanced by data compression efficiency. In [4] a DWT technique was presented to compress power quality disturbance data. The author assumed that Daubechies 4 filter can produce good results for both compression efficiency and reconstruction accuracy. Based on that, examples of PQ disturbances were tested and analysed to provide quantitative results. Further, in [7] a more complex solution was proposed focusing in Wavelet Packet Transform (WPT) and Minimum Description Length (MDL). The main goal was to analyse the minimum quantity of nonzero coefficients needed to produce satisfactory quality in the reconstructed signal. The developed technique was tested for a set of different filters. According to the author, Symlets 7 provides the best result. Both works, in fact, showed best case results of specific examples and did not mention a DWT drawback for Power Quality: the large variation in results given by a certain filter. In this paper we show the impact of this disadvantage.

PCA has also been proposed for recognition of Power Quality disturbances. Reference [20] showed a solution based on PCA and BP Neural Network for AC power line disturbances classification. The author presented results of recognition accuracy, but does not address the compression efficiency neither the reconstruction accuracy. We show the PCA implementation for PQ signals regarding both data compression ratio and reconstruction accuracy.

This paper is organized as follows: Section II presents an overview of DWT and PCA techniques. Section III explains the experimental setup for each technique under test. Then, in section IV we analyse the results and compare them with previous works. Finally, conclusions are drawn in section V.

2. DWT and PCA techniques

In this section overviews about DWT and PCA are presented.

A. DWT

Wavelet Transform (WT) has been studied and applied in various scientific fields for a long time. The basic concept is the correlation between a given analysed function and an orthogonal basis [1]-[14]. The orthogonal basis are scaled and shifted in time. Then, after this process, the correlation is taken with the function under analysis. The WT's continuous form is shown in (1) [1].

$$CWT(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} x(t)g\left(\frac{t-b}{a}\right) dt \quad (1)$$

Where b and a are the shifting and scaling coefficients respectively. The term $a^{-1/2}$ is a normalization parameter and $g(\cdot)$ is the Wavelet basis. In fact, there is an additional term $h(\cdot)$ that represents a scale basis which is orthogonal to $g(\cdot)$ and works as an initial scale for the further decompositions.

The discrete form or Discrete Wavelet Transform (DWT) can be designed, in practice, by filter banks. Therefore, the Wavelet and Scale basis work as high and low pass filters respectively. The scaling coefficient a is implemented by dyadic decimation in time. Figure 1 presents the block diagram of the filters.

Notice that the DWT can be decomposed in various levels. This is called the multiresolution decomposition [17].

Reference [1] has motivated the use of Wavelets in Power Quality because its main advantage is the low and high resolutions in time for high and low resolutions of frequency spectrum respectively.

The data compression may be achieved by discarding least significant detail coefficients according to threshold heuristics and preset. Here, we present two ways to discard coefficients that were used in previous works (e.g. [4], [7], [8], [11] and [14]).

$$Th_1(n) = p \times \max(|d_n[t]|) \quad (2)$$

$$Th_2(n) = p \times \sum_{t=1}^T |d_n(t)|^2 \quad (3)$$

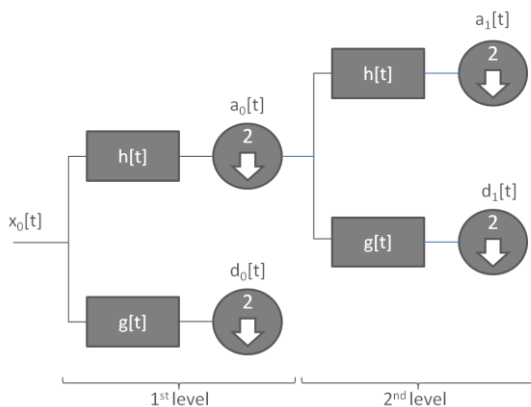


Fig. 1. DWT designed by filter banks.

Where, in (2) and (3), $0 \leq p \leq 1$ represents the factor of discarding coefficients and, for the latter equation, T is the length of detail coefficients in the n^{th} decomposition level. Notice that in (3) the coefficients are discarded by detail energy evaluation, as the squared norm is used. The decision is, then, performed according to (4).

$$\begin{cases} d_n(n), & |d_n(n)| \geq Th_1(n) \text{ or } Th_2(n) \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

B. PCA

Principal Component Analysis takes the correlation between multivariate data in order to reduce the data set redundancy. Hence, second-order statistics are used to transform multivariate data in PCA domain. First, the mean is subtracted from the data [21], as in (5):

$$B = x - E\{x\} \quad (5)$$

Where $E\{x\}$ is the expectation of x and, in practice, the sample mean is performed.

After that, the covariance matrix, C_x , is taken, as in (6):

$$C_x = E\{B \cdot B^T\} \quad (6)$$

Then, the PCA solution is given by the computation of matrix C_x 's eigenvalues and unit-length eigenvectors. The eigenvectors' matrix has its columns rearranged in respective eigenvalues descending order. This brings to the principal components y_n (7):

$$y_n = e_n^T x \quad (7)$$

Where, y_n and e_n is the n^{th} principal component and unit-length eigenvector respectively.

In theory, the benefit of this technique is that only a small number of principal components are needed to best represent the original data. Hence, this technique provides data compression based on the multivariate correlation.

3. Experimental Setup

In this section, we show the proposed experimental setup we developed to test the accuracy and compression for each technique under evaluation.

A. DWT

In order to verify the DWT technique, we analysed sixty measured test-cases from IEEE Working Group P1159.3 database [15]. The dataset has 60 test-cases sampled at 15.36 kHz with time duration of 100 ms. This brings 1536 samples for each test-case. Then, twenty-five different Wavelet filters were selected to decompose the test-cases from one to four levels. The Wavelab 850 was used to decompose and reconstruct the signals. This toolbox is available for MatLab in [16]. The data compression is taken by discarding the detail coefficients based on the theory explained in the last section. The factor p in (2) and (3) are set in four distinct values (0.05, 0.1, 0.15 and 0.2). This brings four hundred ($25 \times 4 \times 4$)

different combinations. The Table I summarizes the general settings.

Table I. – DWT Experimental Setup

Database source	IEEE working group P1159.3
# of waveforms	60
Sampling Frequency	15.36 kHz
# of filters	25
Range of decomposition levels	1-4
Threshold factors (p)	0.05, 0.1, 0.15, 0.2
# of resultant combinations	400
Toolbox	Wavelab 850

B. PCA

In order to evaluate the PCA technique performance (*i.e.* CR efficiency vs. signal reconstruction accuracy terms), we use the same IEEE working group P1159.3 test-cases previously mentioned.

The signal decomposition was performed by using the statistics MatLab toolbox. The test-case is analysed by using 4, 6, and 8 principal components. Table II summarizes the PCA Experimental Setup.

Table II. – PCA Experimental Setup

Database source	IEEE working group P1159.3
# of waveforms	60
Sampling Frequency	15.36 kHz
Principal Components	4, 6, 8
Toolbox	MatLab – Statistics Toolbox

Each test-case is decomposed in a 64 rows by 24 columns matrix. The statistics MatLab toolbox treats each matrix column and row as a different variable and its realization respectively. The matrix's variables (*i.e.* columns) are created by decimating in time a given test-case with a factor of 24 times. In decimation process, the first sample is picked up as a function of the respective column index. The remainder of the decimation procedure is to count 24 samples to pick the next sample up. These actions are performed for all 24 different variables.

4. Results

In this section we present the metrics and the results for both techniques that were evaluated.

A. DWT Results

First, we introduce three metrics that are evaluated to quantify the DWT technique performance for both data compression efficiency and signal reconstruction accuracy.

- 1) **Compression Ratio (CR):** We consider that data samples are acquired with 2 bytes resolution and the arithmetic is calculated in floating point single precision. Hence, the stored data must spend 4 bytes per retained detail coefficient. Since all discarded detail coefficients are set to zero and not stored, then the position of retained coefficients must be recorded for further reconstruction of the zeros. We assume a data

window containing a maximum value of 2^{16} samples. This brings 2 bytes per position. The coarser approximation coefficients are entirely stored. Putting all these information together, we show in (8) the CR's equation.

$$CR = \frac{2 \times \text{original input samples}}{4 \times (\text{rdc} + \text{appc}) + 2 \times \text{rdc}} \quad (8)$$

Where, *rdc* and *appc* means retained detail and coarser approximation coefficients' length respectively.

- 2) **Entropy (H):** The entropy defines the average bit length to code a data set without loss of information. Hence, we decided to evaluate this metric to know how good can be lossless coding based on entropy. Equation (9) shows the metric.

$$H = - \sum_{i=1}^N P_i \times \log_2 P_i \quad (9)$$

In (9), P_i defines the probability of occurrence for a given symbol. N is the total quantity of different symbols in the data set.

- 3) **Normalized Mean Squared Error (NMSE):** This metric relates the squared error in signal reconstruction and the squared original data [4], [7] (see (10) for more details).

$$NMSE = \frac{\|x_n - \hat{x}_n\|^2}{\|x_n\|^2} \times 100 \quad (10)$$

Notice from (10) that the output is given in percentage. Further, x_n is the original data and \hat{x}_n is the reconstructed one.

After the metrics explanation, the results extraction was performed as follows: First, the sixty test-cases were mapped in DWT domain and then the detail coefficients were handled regarding the two discarding heuristics (Th_1 and Th_2) presented in Section II. The discarding procedure is taken as a function of the four threshold factor values established in Table I. After that, the proposed metrics were performed for each test-case. Finally, we calculated and stored the sample mean, standard deviation, best and worst cases for each criterion as a function of the combinations previously mentioned. For instance, we have the CR, Entropy (H) and NMSE information for the following combination: Daubechies 4 filter decomposed in two levels and with a threshold factor p of 0.1. Notice that this process is done for all four hundred combinations. The heuristics Th_1 and Th_2 are analysed separately.

To extract a feasible combination, we need to balance the trade-off between average CR or Entropy and average NMSE. Figure 2 illustrates NMSE (in %) and CR analysis as a function of the combinations for Th_1 case. The combinations (x axis) are sorted in NMSE ascending order.

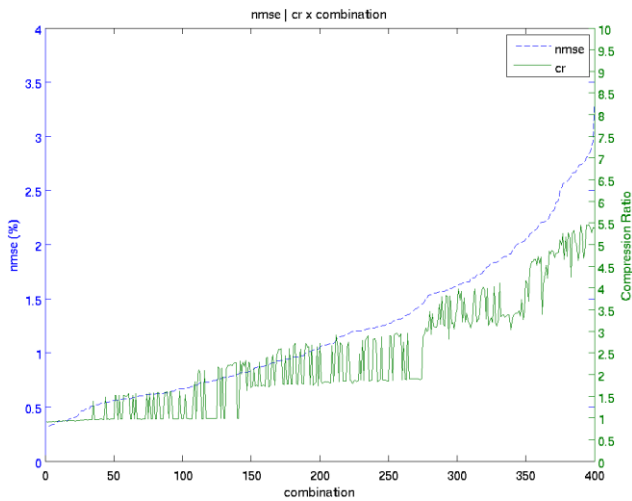


Fig. 2. CR and NMSE analysis for Th_1 case

This brings the resultant values of CR for each combination as a function of $NMSE$ criterion. Therefore, we first search for an acceptable average $NMSE$ and then look for the best average values of CR or H . In practice, the priority is attributed to accuracy, because, according to IEC standards (i.e. [18] and [19]), we always need to guarantee good quality in signal reconstruction. Otherwise, the loss of information can degrade disturbances compromising further analysis in Power Quality. The results can be seen in Table III. For each metric we show the average, standard deviation and coefficient of variation (11). In addition, the last two rows present the maximum and minimum metrics' values for a given combination between the sixty test-cases. For example, in a certain combination it is possible to detect which test-case has the lowest $NMSE$ and so forth. To extract a feasible combination, we first imposed the constraint of an average $NMSE$ less than 1%. Then, we searched, in the resultant combinations, for the best average CR and Entropy (H) respectively. Since we are looking for the most regular (or low variability) combination, then we calculate the coefficients of variation (CV) for each metric (11):

$$CV = \frac{\sigma}{\mu} \times 100 \quad (11)$$

Where σ is the standard deviation and μ is the sample mean. The most regular combination for $NMSE$ is the Coiflets 3 with four levels of decomposition and p of

0.05. In fact, Table III reveals that there is no selected combination providing $NMSE$'s CV less than 75.84%. For this particular solution, the best and worst $NMSE$ test-cases are ranging from 0.23% to 5.12%. That loss in $NMSE$ is unacceptable for a general-purpose PQ disturbance compressor. Further, notice that for Th_2 and a specific test-case, the minimum CR produces more stored data than original one.

The previous works published similar results. In [4] three disturbance examples are tested and the best case is 5.74 times for CR and 0.003% for $NMSE$ (with Daubechies 4). In [7] the WPT technique is applied to six test-cases and the proposed solution showed 36.36 times for CR and 10.01% for $NMSE$ (with Symlets 7). Notice that if we consider just the best case for each criterion, then our work reaches 6.02 times for CR (with Coiflets 3) and 0.04% for $NMSE$ (with Coiflets 2). Our results provide a more systematic evaluation of four hundred filter combinations, since we applied the DWT and thresholding technique to a much larger set, i.e. up to sixty power line disturbance test-cases. Our experiment, covering this large data set, showed that the variability turns out to be an important drawback for the DWT.

In the sense of entropy, results revealed a minimum average length of 1.54 bits (with Symlets 7). It seems that a lossless codification can be a good additional approach in the WT data sets. Although this benefit, the same filter produces a CV of 91.91 % for $NMSE$.

B. PCA Results

The PCA technique performance evaluation for both compression ratio and signal reconstruction accuracy is taken by Compression Ratio and $NMSE$ (see (10)) metrics. Since the PCA information are stored in a different manner, then the CR must be redefined.

1) *Compression Ratio (CR)*: We also consider that data samples are acquired with 2 bytes resolution and that the arithmetic is calculated in floating point single precision. Hence, the stored data must spend 4 bytes per observation.

$$CR = \frac{2 \times \text{original input samples}}{4 \times (24pc + 64pc + 24)} \quad (12)$$

In equation (12) pc denotes the number of principal components used to represent the signal.

Table III. – DWT Results

Threshold heuristic	Th_1				Th_2			
	$CR \times NMSE$		$H \times NMSE$		$CR \times NMSE$		$H \times NMSE$	
Analysed trade-off	Coiflets 3		Symlets 7		Coiflets 2		Coiflets 2	
Filter	Coiflets 3		Symlets 7		Coiflets 2		Coiflets 2	
Number of levels	4		4		3		3	
Threshold factor p	0.05		0.05		0.05		0.05	
Metric	CR	$NMSE$ (%)	H	$NMSE$ (%)	CR	$NMSE$ (%)	H	$NMSE$ (%)
μ	2.72	0.96	1.54	0.99	1.78	0.95	2.22	0.95
Σ	0.91	0.79	0.41	0.91	0.65	1.52	0.49	1.52
CV (%)	33.45%	75.84%	26.62%	91.91%	36.51%	160%	22.07%	160%
Minimum	1.08	0.23	0.8	0.21	0.61	0.04	1.37	0.04
Maximum	6.02	5.12	2.66	5.79	3.27	7.68	3.28	7.68

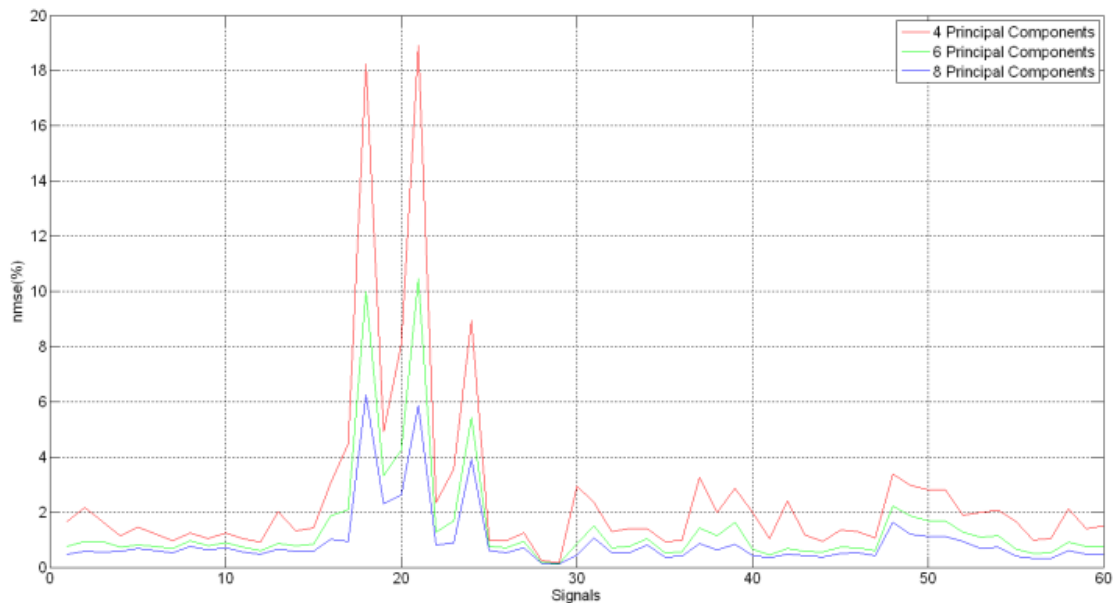


Fig. 3. *NMSE* analysis for PCA technique

Notice that in order to save the compressed data related to one test-case, it is needed to store a *pc* rows by 24 columns matrix, a *pc* rows by 64 columns matrix (*i.e.* the data represented in the PCA domain), and also the average of each principal component.

Since the *CR* depends only on the number of principal components used for signal's representation, then the *CR* is previously known before the data compression process is performed. On the other hand, the *NMSE* analysis is nondeterministic and varies in accordance with the number of principal components chosen. As previously depicted by Table II, the PCA experimental analysis takes into account 4, 6, and 8 principal components. Figure 3 shows the *NMSE* analysis vs. signals (or test-cases) for 4, 6, and 8 principal components. As can be seen, the *NMSE* analysis relies approximately at the same level for most of the test-cases. However, in a few cases (*e.g.* from signal index 15 to 25) the reconstruction error is significantly attenuated when 8 principal components are used to represent those signals. This is because these signals has more details and need more components to be reconstructed without significant loss of information. The Table IV shows the PCA results.

Table IV. – PCA Results

# Components	4	6	8
<i>CR</i>	2.04	1.39	1.05
μ (<i>NMSE</i>)	2.5648	1.4200	0.9261
σ (<i>NMSE</i>)	3.36	1.87	1.13
Maximum(<i>NMSE</i>)	18.9	10.46	6.26
Minimum (<i>NMSE</i>)	0.168	0.1363	0.1177

As described previously, it is possible to determine the *CR* as a function of principal components used. The *NMSE* variation is a function of the signal's characteristics. Therefore, the larger the number of principal components used, the better the signal reconstruction is. However, the *CR* decreases significantly.

5. Conclusion

DWT and thresholding technique is applied to Power Quality disturbance test-cases in order to assess data compression efficiency and accuracy of reconstructed signals. After a systematic evaluation process, we show that DWT produces variability in accuracy. That is a significant drawback for general-purpose Power Quality disturbance compressor, which has not been addressed in previous works that proposed the DWT with few specific filters. A given Power Quality disturbance can be highly correlated with a certain DWT basis, but that usually is not true for other types of electrical events. Further, least significant detail coefficients are not only related to Gaussian noise but, in fact, can be a relevant amount of information that is lost after the thresholding process. Therefore, the use of DWT and discarding process can be designed for a specific type of Power Quality disturbance and cannot provide good results for a general-purpose system that stores disturbance signals.

In comparison with the DWT technique, PCA results show a better average behaviour when considered a larger set of signals. On the other hand, for a specific group of signals (*i.e.* from signal 15 to 25 in Figure 3) the reconstruction error raised significantly even using more principal components to represent the signal. In this case, the number of principal components should be greater than the presented in the experimental analysis. However, the stored data increases significantly and the resultant size gets larger than the original one.

Based on that, the analysed techniques do not present a good performance for *CR* vs. *NMSE* considering all types of power line disturbances. For each disturbance type there is a technique that best compresses the data and reconstructs the signal with fidelity.

Acknowledgement

The authors would like to thank Feevale University, FINEP, FAURGS, and CAPES, Brazil, for financial support.

References

- [1] D.C. Robertson, O.I. Camps, J.S. Mayer, and W.B. Gish, "Wavelets and Electromagnetic Power System Transients", in *IEEE Transactions on Power Delivery*, 1996, pp. 1050-1058.
- [2] S. Santoso, E.J. Powers, W.M. Grady, P. Hofmann, "Power Quality Assessment via Wavelet Transform Analysis", in *IEEE Transactions on Power Delivery*, 1996, pp. 924-930.
- [3] G.T. Heydt and A.W. Galli, "Transient Power Quality Problems Analyzed using Wavelets", in *IEEE Transactions on Power Delivery*, 1997, pp.908-915.
- [4] S. Santoso, E. J. Powers and W. M. Grady, "Power Quality Disturbance Data Compression using Wavelet Transform Methods", in *IEEE Transactions on Power Delivery*, 1997, pp. 1250-1257.
- [5] S. Santoso, W. M. Grady, E. J. Powers, J. Lamoree and S. C. Bhatt, "Characterization of Distribution Power Quality Events with Fourier and Wavelet Transforms", in *IEEE Transactions on Power Delivery*, 2000, pp. 247-254.
- [6] M. Karimi, H. Mokhtari, and M. R. Iravani, "Wavelet Based on-line Disturbance Detection for Power Quality Applications", in *IEEE Transactions on Power Delivery*, 2000, pp. 1212-1220.
- [7] E. Y. Hamid, Z. I. Kawasaki, "Wavelet-based Data Compression of Power System Disturbances using the Minimum Description Length Criterion", in *IEEE Transactions on Power Delivery*, 2002, pp. 460-466.
- [8] P.K. Dash, B.K. Panigrahi, D.K. Sahoo and G. Panda, "Power Quality Disturbance Data Compression, Detection, and Classification using Integrated Spline Wavelet and S-Transform", in *IEEE Transactions on Power Delivery*, 2003, pp. 595-600.
- [9] C-H. Lin and C-H. Wang, "Adaptive Wavelet Networks for Power-Quality Detection and Discrimination in a Power System", in *IEEE Transaction on Power Delivery*, 2006, pp. 1106-1113.
- [10] G. Liu, S. Su, P. Lu, H. Rong and L. Zheng, "Study on Wavelet Transformation Detection Arithmetic and its Actualization Technique of Transient Power Quality Disturbance based on DSP", in *Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies*, 2008, pp. 1641-1645.
- [11] L. Lin, N. Huang and W. Huang, "Review of Power Quality Signal Compression based on Wavelet Theory", in *International Conference on Test and Measurement*, 2009, pp. 235-238.
- [12] C. Chun-ling, X. Tong-yu, Y. Ye and J. Fengli, "Detection and Location of Power Quality Transient Disturbance Based on Lifting Wavelet" in *Asia-Pacific Power and Energy Engineering Conference*, 2010.
- [13] Z. Yan, X. Yonghai, X. Xiangning, Z. Yongqiang and G. Chunlin, "Power Quality Disturbances Identification Based on dq Conversion, Wavelet Transform and FFT, in *Asia-Pacific Power and Energy Engineering Conference*, 2010.
- [14] M. Zhang, K. Li and Y. Hu, "A High Efficient Compression Method for Power Quality Applications", in *IEEE Transactions on Instrumentation and Measurement*, 2011, pp. 1976-1985.
- [15] IEEE Working Group P1159.3, <http://grouper.ieee.org/groups1159/3/index.html> - Accessed in October 2012.
- [16] Wavelab 850, <http://www-stat.stanford.edu/~wavelab/> - Accessed in October 2012.
- [17] S. G. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation", in *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1989, pp. 674-693.
- [18] International Electrotechnical Commission, IEC 61000-4-30, "Testing and Measurement Techniques – Power Quality Measurement Methods". Geneva, 2008.
- [19] International Electrotechnical Commission, IEC61000-4-7, "Testing and Measurement Techniques – General Guide on Harmonics and Interharmonics Measurements and Instrumentation, for Power Supply Systems and Equipment Connected Thereto". Geneva, 2008.
- [20] N. Huang, "Power Quality Disturbances Recognition Based on PCA and BP Neural Network", in *Asia-Pacific Power and Energy Engineering Conference*, 2010.
- [21] A. Hyvärinen, J. Karhunen and E. Oja, "Independent Component Analysis", Espoo, 2001, John Wiley and Sons Inc.