



PQD classifier based on higher-order statistics and total harmonic distortion

Jesús-Manuel González-Bueno¹, José-Carlos Palomares-Salas¹, Juan-José González-de-la-Rosa¹, Olivia Florencias-Oliveros¹, José-María Sierra-Fernández¹, Manuel-Jesús Espinosa-Gavira¹ and Agustín Agüera-Pérez¹

¹University of Cádiz. Research Group PAIDI-TIC 168 – Computational Instrumentation and Industrial Electronics Dept. of Automation Engineering, Electronics, Architecture and Computer Networks Engineering School of Algeciras Ramón Puyol Av. s/n, E11202 Algeciras (Spain) Phone: +0034 956 028069, e-mail: juanjose.delarosa@uca.es

Abstract. Higher-Order Statistics (HOS) have been frequently applied in Power Quality Disturbance (PQD) analysis as a reliable tool for event detection. This paper outlines a technique based on mean, variance and zero-lag third and fourth cumulants - skewness and kurtosis - along with the Total Harmonic Distortion (THD) index for PQD detection. These statistics are obtained in order to characterize a waveform by a feature vector. A two-layer feed-forward neural network is then used to classify inputs (feature vectors) into a set of PQD categories. The impact of frame duration and number of hidden neurons is analyzed. The network is trained, validated and tested with synthetically-generated PQD waveforms obtained from parameter-controlled equations. As a first approach, five PQD categories are considered: sag, swell, interruption, impulsive transient and oscillatory transient. A promising overall classification rate of 99.7 % is achieved which allows future analysis with more PQD categories and/or a noisy context.

Key words

Higher-Order Statistics (HOS); Total Harmonic Distortion (THD); Feed-Forward Neural Network; Power Quality Disturbance (PQD).

1. Introduction

The increasing pollution of power lines and its impact on the quality of power delivery by electrical utilities to the users [1] make power quality (PQ) event detection and classification a leading issue for the power industry and so a glowing topic for researchers. Myriads of papers have proposed different methods, techniques and solutions to this problem. Most of them consist of feature extraction based on signal processing and artificial intelligence based classifiers [2]. This paper follows this research issue, using a reduced set of parameters that include both second (e.g. THD) and Higher-Order Statistics (HOS) in the feature extraction stage over the signal under test. The extracted features are presented to a multi-layer perceptron neural network to deal with the classification of the signal in a suitable category. The goal is to make the neural network as simple as possible. Recently use of HOS can be found in [3] and [4].

PQ problems faced by facilities operations include sags, swell, surges, outages, harmonics and flickers [5]. Related terms and definitions can be found in [6]. However, since a large and complete set of real disturbances is difficult to capture, a data generation stage is firstly required to provide a significant number of signals to work with.

The paper is organized as follows. Section 2 reviews the definitions of HOS and THD and the potential use in PQ event detection. Section 3 presents the parametercontrolled equations to model PQD. Section 4 presents the proposed technique for PQD detection and classification. Section 5 presents some numerical results. Finally, section 6 summarizes the conclusions.

2. HOS and THD in PQ context

Certain higher-order statistics, namely cumulants have become increasingly popular in several areas of system identification and signal processing [7], especially in problems which deal with non-Gaussian or non-linear processes [8]. A number of valuable properties of cumulants which make them an interesting choice can be found in [7] and [8].

Hereinafter some important definitions and theoretical developments are exposed in order to get a comprehensive approach to the mathematical guts of the problem. The starting point is a random variable, in a random vector, which models a stationary stochastic process.

The moment generating associated with a random variable *X* is given by the following integral:

$$\phi(s) = \int_{-\infty}^{\infty} f_X(x) e^{sx} dx = E\left[e^{sX}\right]$$
(1)

and the cumulant generating function can be obtained as the natural logarithm of the moment generating function:

$$\psi(s) = \ln \phi(s) \tag{2}$$

Moment theorem [9] states that the derivatives of (1) at the origin (s=0) equal the moments of X. On the other hand, cumulants (or cumulative moments) of a random variable X are by definition the derivatives at the origin of the cumulant generating function. That is, the *nth*-order cumulant of a random variable is defined as the *nth*coefficient in the McLaurin series expansion (provided it exists [8]) of the cumulant generating function.

Cumulants are related to the moments by the following recursive formula:

$$c_{n} = m_{n} - \sum_{k=1}^{n-1} {\binom{n-1}{k-1}} c_{k} m_{n-k}$$
 (3)

Similarly, the moment generation function of a random vector or collection of random variables $\{X_1, X_2, ..., X_n\}$ is defined by:

$$\phi(s_1, s_2, \dots, s_n) = E\left[e^{s_1 X_1 + s_2 X_2 + \dots + s_n X_n}\right]$$
(4)

and the cumulant generation function by:

$$\Psi(s_1, s_2, \dots, s_n) = \ln \phi(s_1, s_2, \dots, s_n)$$
(5)

Thus, the joint moment of *n* random variables follows:

$$\operatorname{mom}(X_1, X_2, \dots, X_n) = \mathbb{E}\left[X_1 \cdot X_2 \cdot \dots \cdot X_n\right]$$
(6)

and the joint cumulant by the Leonov – Shiryaev's formula:

$$\operatorname{cum}(X_{1}, X_{2}, ..., X_{n}) = \\ = \sum_{\pi} (-1)^{|\pi|-1} (|\pi|-1)! \prod_{B \in \pi} E \left[\prod_{i \in B} X_{i} \right]$$
(7)

where π runs through the list of all partitions of $\{1, 2, ..., n\}$, *B* runs through the list of all blocks of the partition π , and $|\pi|$ is the number of parts or blocks in the partition. That is, the nth-order cumulant of a collection of random variables is therefore defined in terms of its joint moments of orders up to *n* [8].

Finally, in the case of a stationary random process X(k), the nth-order moment and cumulant are given by:

$$\begin{split} m_{n}^{x}(k_{1},...,k_{n-1}) &= \\ &= \operatorname{mom}(X(k),X(k+k_{1}),...,X(k+k_{n-1})) = \\ &= E\Big[X(k)\cdot X(k+k_{1})\cdot...\cdot X(k+k_{n-1})\Big] \\ c_{n}^{x}(k_{1},...,k_{n-1}) &= \\ &= \operatorname{cum}(X(k),X(k+k_{1}),...,X(k+k_{n-1})) = \\ &= \sum_{\pi}(-1)^{|\pi|-1}(|\pi|-1)!\prod_{B\in\pi}E\bigg[\prod_{i\in B}X(k+k_{i})\bigg] \end{split}$$
(9)

Under zero-lag condition the collection or set of random variables $\{X(k), X(k+k_1), ..., X(k+k_{n-1})\}$ is reduced to $\{X(k), X(k), ..., X(k)\}$ and then joint moments and joint cumulants of an nth-order random vector become moments and cumulants of a one-dimensional random variable.

$$m_{n}^{x}(\underbrace{0,...,0}_{n-1}) = m_{n} = E\left[X^{n}(k)\right]$$
 (10)

$$\mathbf{c}_{n}^{x}(\underbrace{0,\dots,0}_{n-1}) = \mathbf{c}_{n} = \mathbf{m}_{n} - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \mathbf{c}_{k} \mathbf{m}_{n-k}$$
(11)

In practice, *nth*-order zero-lag sample moments and sample cumulants of a stationary and zero-mean stochastic process shall be given by:

$$\hat{m}_{n}^{x}(\underbrace{0,...,0}_{n-1}) = \frac{1}{N} \sum_{k=1}^{N} X^{n}(k)$$
(12)

$$\hat{c}_{n}^{x}(\underbrace{0,\ldots,0}_{n-1}) = \hat{m}_{n} - \sum_{k=1}^{n-1} \binom{n-1}{k-1} \hat{c}_{n} \hat{m}_{n-k}$$
(13)

Sample moments and cumulants so defined are demonstrated to give unbiased estimates [10], however cumulants are preferred [8].

The total harmonic distortion (THD) is a measure of the effective value of the harmonic components of a distorted voltage or current waveform defined as the ratio of harmonics in percentage of the fundamental component [12].

$$THD = \frac{\sqrt{\sum_{k=2}^{\infty} V_k^2}}{V_1}$$
(14)

where V_k is the amplitude of the kth harmonic component.

3. PQD Parameter-controlled equations

It is common to find in the literature PQD signal models based on mathematical equations [2], [12], [13] and [14]. Although real-time PQ disturbances signals are difficult to capture [12], synthetically-obtained waveforms can be an interesting alternative. They are quite similar to real signals and allow the generation of a large number of controlled samples to be used for research purposes. In fact, they are generated by power-signal generators. In this paper, six types of PQ disturbances were generated, including the healthy case (pure sine, sag, swell, interruption, impulsive transient, and oscillatory transient). Waveforms are produced using parametric equations in MATLABTM. The proposed expressions and parameter variation are given in Table I. Values are uniformly distributed in their suitable ranges.

Different types of PQ disturbances waveforms are shown in Figure 1. Non-disturbed signals have normalized amplitude of 1.0 p.u. and fundamental frequency of 50 Hz (20 ms cycle time). The sampling frequency and period are 20 kHz and 0.05 ms, respectively, that is 400 points for each cycle. Each waveform has an overall duration of 20 cycles or 0.4 s (8000 points).

Table I.- PQ disturbances equations and parameter values

Healthy waveform (pure sine) $v(t) = A \cdot \sin(2\pi \cdot f \cdot t + \phi)$ A = 1V f = 50Hz $\phi = U(-\pi,\pi)$

Sag/Swell

 $v(t) = [1 - (1 - \alpha) \cdot w(t)] \cdot \sin(2\pi \cdot f \cdot t + \phi)$ $\alpha = U(0.5, 0.8)$ $v(t) = [1 + \alpha \cdot w(t)] \cdot \sin(2\pi \cdot f \cdot t + \phi)$ $\alpha = U(0.2, 0.5)$ $\phi = U(-\pi, \pi)$ $w(t) = u(t - t_1) - u(t - t_2)$ $t_2 - t_1 = U(5T, 9T)$

Interruption

It is considered a sag with $\alpha = U(0, 0.1)$

Impulsive transient

$$\begin{split} v(t) &= \sin(2\pi \cdot f \cdot t + \phi) + \alpha \cdot e^{-(t-t_1)/\tau} \cdot w(t) \\ \phi &= U(-\pi,\pi) \\ \alpha &= U(3A,4A) \\ \tau &= U(100 \cdot 10^{-3} / \ln(10), 180 \cdot 10^{-3} / \ln(10)) \\ w(t) &= u(t-t_1) \end{split}$$

Oscillatory transient

 $\begin{aligned} v(t) &= \sin(2\pi \cdot f \cdot t + \phi) + \alpha \cdot e^{-(t - t_1)/\tau} \cdot \sin(2\pi \cdot f_t \cdot t) \cdot w(t) \\ \phi &= U(-\pi, \pi) \\ \alpha &= U(3A, 4A) \\ \tau &= U(100 \cdot 10^{-3} / \ln(10), 180 \cdot 10^{-3} / \ln(10)) \\ f_t &= U(0, 4 \text{KHz}) \\ w(t) &= u(t - t_1) \end{aligned}$



Fig. 1. Healthy and PQD waveforms: (a) Pure sine; (b) Sag; (c) Swell; (d) Interruption; (e) Impulsive transient; (f) Oscillatory transient.

4. PQD detection and classification

The signals generated by mathematical models can be easily used in the classification of PQ disturbances to extract their distinctive features [12]. However some preprocessing is required before the classification stage in order to get a zero-mean process and presume stationarity and ergodicity.

Furthermore, features estimates are to be normalized by the nth-power of the standard deviation to make them scale-invariant.

$$\hat{\gamma}_{n}^{x}(0,...,0) = \frac{\hat{c}_{n}^{x}(0,...,0)}{\left[\hat{c}_{2}^{x}(0)\right]^{\frac{n}{2}}}$$
(15)

Figure 2 shows the ability of cumulants 2, 3 and 4 to discriminate among different types of disturbance.



Fig. 2. Cumulants 2, 3 and 4 for different PQD categories (black: sag; green: swell; magenta: interruption; cyan: impulsive transient; red: oscillatory transient).

The feature vector presented to the classifier is composed by the mean, the variance, the third- and fourth-order standardized cumulants and the THD of each 20 cycles waveform.

This time window is used to ensure an adequate frequency discrimination for the calculation of the THD.



Fig. 3. Power spectrum of a 0.1 s., 0.2 s, and 0.4 s. (20 cycles), respectively, 50 Hz, 1 V pure tone.

Once the feature vector is obtained it should be classified in its appropriate category. Pattern recognition techniques are automated tools for decision-making processes [15]. Among them neural networks have been applied extensively in PQ [16]. Important features in the design of neural networks are the study of the necessary input pattern, the neural network structure, transfer functions, and learning algorithms [17].

This paper proposes a two-layer feed-forward network, with sigmoid hidden neurons, able to classify vectors arbitrarily well given enough neurons in its hidden layer. The performance of such a network is obtained from one to a hundred of neurons for different feature vectors. Figure 4 shows a better performance from 20 neurons and a feature vector consisting of HOS-1 to HOS-4 and THD of zero-mean waveforms.

A set of 600 waveforms – 100 per each PQD category – is randomly divided as 70% for training, 15% for validation, and 15% for testing the network. The number of inputs is set to 5, which is equal to the elements of the feature vector, and the number of output neurons is set to 6, corresponding to the number of elements in the target vector, that is the number of PQD categories. The network is trained with scaled conjugated gradient backpropagation.



Fig. 4. 1-hidden-layer neural network performance for different number of hidden neurons and feature vectors.

5. Results

Figure 5 shows the confusion matrices for training, validation, and testing, and the three kinds of data combined. The network outputs are very accurate as seen from the overall 99.7% accuracy. The target classes or categories can be identified as: 1) Pure sine; 2) Sag; 3) Swell; 4) Interruption; 5) Impulsive transient; and 6) Oscillatory transient.



Fig. 5. Training, validation, test, and overall confusion matrices. Gray cells show true positive rates (green figures) and false negative rates (red figures).

6. Conclusions

PQ disturbances can be synthetically generated through parameterized models that conform to standard definitions. It is thus possible to have a sufficient number of test waveforms to work with. Although they are not real, their use allows to reach early conclusions about the performance of systems.

The use of neural networks on a predefined five-element feature vector (up to fourth-order cumulants and THD index) makes possible the classification and recognition of patterns associated to a number of PQ disturbances (pure sine, sag, swell, interruption, impulsive transient, and oscillatory transient) with an acceptable level of success.

Data segmentation in 20-cycles frames seems to be appropriate to ensure a suitable frequency discrimination for THD computation. However this first approach should be used as a basis to study the effect of frame duration on the time features and the performance of the PQD classifier.

Future work will focus on determining the optimal frame duration in terms of classification accuracy. The expansion of the feature vector through the addition of more cumulants (above fourth order) will also be considered in order to study their capability to make a better classification and unambiguously characterize a wider and complex set of PQ disturbances with and without a noisy background.

References

- M. V. Ribeiro, C. A. G. Marques, C. A. Duque, A. S. Cerqueira, and J. L. R. Pereira, "Detection of Disturbances in Voltage Signals for Power Quality Analysis Using HOS," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, no. 1, Dec. 2007.
- [2] S. Khokhar, A. A. B. Mohd Zin, A. S. B. Mokhtar, and M. Pesaran, "A comprehensive overview on signal processing and artificial intelligence techniques applications in classification of power quality disturbances," *Renewable and Sustainable Energy Reviews*, vol. 51, pp. 1650–1663, Nov. 2015.
- [3] J. Diego Silva Guedes, D. Diego Ferreira, B. Henrique Groenner Barbosa, C. Augusto Duque, and A. Santiago Cerqueira, "Non-Intrusive Appliance Load Identification Based on Higher-Order Statistics," *IEEE Latin America Transactions*, vol. 13, no. 10, pp. 3343–3349, Oct. 2015.
- [4] O. Florencias-Oliveros, J.-J. González-de-la-Rosa, A. Agüera-Pérez, and J. C. Palomares-Salas, "Discussion on Reliability and Power Quality in the Smart Grid: a prosumer approach of a time scalable index," *Renewable Energy and Power Quality Journal*, vol. 1, pp. 108–113, Apr. 2018.

- [5] J. J. González de la Rosa, A. M. Muñoz, A. Gallego, R. Piotrkowski, and E. Castro, "Higher-order characterization of power quality transients and their classification using competitive layers," *Measurement*, vol. 42, no. 3, pp. 478– 484, Apr. 2009.
- [6] Institute of Electrical and Electronics Engineers, *IEEE* recommended practice for monitoring electric power quality. New York: Institute of Electrical and Electronics Engineers, 2009.
- [7] A. Swami, "System Identification Using Cumulants," Ph.D. dissertation, USC SIPI Report 140, Dep. Elec. Eng.-Syst., Univ. Southern California, Los Angeles, CA, 1988.
- [8] J. M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications," *Proceedings of the IEEE*, vol. 79, no. 3, pp. 278–305, Mar. 1991.
- [9] A. Papoulis, Probability, random variables, and stochastic processes, 3rd ed. New York: McGraw-Hill, 1991.
- [10] J. A. R. Fonollosa, "Sample cumulants of stationary processes: asymptotic results," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 967–977, Apr. 1995.
- [11] J. F. Kenney and E. S. Keeping, *Mathematics of statistics*, vol. 2, 2 vols. Van Nostrand company, 1947.
- [12] S. Khokhar, A. A. M. Zin, A. S. Mokhtar, and N. Ismail, "MATLAB/Simulink based modeling and simulation of power quality disturbances," 2014, pp. 445–450.
- [13] K. Manimala, K. Selvi, and R. Ahila, "Hybrid soft computing techniques for feature selection and parameter optimization in power quality data mining," *Applied Soft Computing*, vol. 11, no. 8, pp. 5485–5497, Dec. 2011.
- [14] S. Khokhar, A. A. Mohd Zin, A. S. Mokhtar, and N. Zareen, "Automatic pattern recognition of single and multiple power quality disturbances," *Australian Journal of Electrical and Electronics Engineering*, vol. 13, no. 1, pp. 43–53, Jan. 2016.
- [15] A. M. Gaouda, S. H. Kanoun, M. M. A. Salama, and A. Y. Chikhani, "Pattern Recognition Applications for Power System Disturbance Classification," *IEEE Power Engineering Review*, vol. 22, no. 1, pp. 69–70, 2002.
- [16] W. R. A. Ibrahim and M. M. Morcos, "Artificial Intelligence and Advanced Mathematical Tools for Power Quality Applications: A Survey," *IEEE Transactions on Power Delivery*, vol. 17, no. 2, p. 6, 2002.
- [17] I. Monedero, C. Leon, J. Ropero, A. Garcia, J. M. Elena, and J. C. Montano, "Classification of Electrical Disturbances in Real Time Using Neural Networks," *IEEE Transactions on Power Delivery*, vol. 22, no. 3, pp. 1288– 1296, Jul. 2007.