Electrical Power and Energy Systems 104 (2019) 21–28

Contents lists available at ScienceDirect

Electrical Power and Energy Systems

journal homepage: www.elsevier.com/locate/ijepes

Identification of the electrical load by C-means from non-intrusive monitoring of electrical signals in non-residential buildings

Emmanuel Guillén-García[a], Luis Morales-Velazquez[a,*], Angel L. Zorita-Lamadrid[b], Oscar Duque-Perez[b], Roque Alfredo Osornio-Rios[a], Rene de Jesús Romero-Troncoso[a]

[a] CA Mecatronica, Facultad de Ingenieria, Universidad Autonoma de Queretaro, Campus San Juan del Rio, Rio Moctezuma 249, Colonia San Cayetano, 76807 San Juan del Rio, QRO, Mexico

[b] Department of Electrical Engineering, University of Valladolid (UVa), 47011 Valladolid, Spain

ARTICLE INFO

Keywords:
C-means
Load classification
Energy signature
Energy power quantities

ABSTRACT

The combined action of different equipment connected to an electrical installation is capable of causing unexpected changes in the load kinds inside the installation; these load variations are responsible for some electrical failures. In this paper a methodology to classify and to identify the load kinds in industrial environments is presented. Energy power quantities (EPQ) and current values are used to establish indexes in order to use them as features for a C-means algorithm and perform the load classification. The experimentation is done in a healthcare facility gathering electrical data in different electric distribution boards. The results obtained from the classification method show variations in the load behavior along the day. Furthermore, some classes can be used to recognize equipment in the electrical installation for further inspection or failure detection.

1. Introduction

Electrical installations in residential and non-residential buildings supply the energy to electrical loads that cause events in the electrical signal [1–3] producing interference among elements that eventually force disconnections from the grid. These events are directly related to harmonic content, reactive loads and load distribution [4–6]; causing signal waveform distortion, power factor reduction and imbalanced lines. Consequently, electric bills can be increased and the production lines could be affected. Therefore, the data analysis of electrical signals is fundamental when the electrical installation performance is affected by the different elements connected to it, and it is necessary to develop methodologies for the electrical signals analysis in order to monitor the load performance and take actions accordingly.

Some published works are focused on the analysis of electrical signals from residential buildings to forecast the power consumption or to detect events in the electrical installation. For instance, some works are centered in the use of non-intrusive methods to monitor events in current signals, as in Cominola et al. [7] that developed an energy disaggregation hybrid algorithm for load monitoring in order to implement customized energy demand management strategies. Factorial hidden markov models and iterative subsequence dynamic time warping are used on real power consumption data sets for the energy disaggregation. On the same path in [8], Taha et al. presented a study of two-dimensional load signature for non-intrusive identification. They use power consumption information to cluster switching events and to implement wave-shape features for improving the predictions in non-intrusive load monitoring. Whereas Huang et al. [9] proposed a new method to improve the current waveform discrimination for non-intrusive appliance load monitoring. Their purpose is to decompose the current waveform into active and nonactive current to discriminate loads with similar indexes. The study cases in their work validate the method reducing the likeness of similar appliances. In another example focused on home applications, Corinne et al. [10] developed a non-intrusive expert system for assistance in smart homes based on electrical device identification. This work uses algorithmic approach to analyze load signatures represented by active power, reactive power and line-to-neutral, this way the proposed system can recognize erratic behaviors. Nguyen et al. [11] proposed a real-time NIALM system, integrating a Field-programmable gate array aimed to give consumers pertinent information about their residential power consumption. In this work a Cumulative Sum is presented for the event detection and a K-means algorithm for the event clustering. Abubakar et al. [12] perform a review of techniques and tools for the energy monitoring, including intrusive load monitoring and non-intrusive load monitoring, furthermore, they include analysis in measuring, communication and recognition devices, besides optimization tools and control devices. In their study, they reveal some issues to be attended as the load

* Corresponding author.
E-mail address: lmorales@hspdigital.org (L. Morales-Velazquez).

https://doi.org/10.1016/j.ijepes.2018.06.040
Received 26 February 2018; Received in revised form 9 May 2018; Accepted 18 June 2018
Available online 28 June 2018

monitoring and load management with the aim of being able to recognize as many kinds of loads as possible. Gillis and Morsi [13] presented a semi-supervised and wavelet design technique for NILM to improve the process of learning the load pattern, they used co-training in two machine learning classifiers to reduce contaminant signals and automate the process of learning the load pattern. Bonfigli et al. [14] proposed a NILM algorithm for home appliances using factorial hidden markov models with reactive an active power measures in addition of additive factorial approximate maximum a posteriori for the load disaggregation, their algorithm can output the disaggregated profiles in the active and reactive power components. Buddhahai et al. [15] presented a NILM for energy disaggregation implementing label classification, they used basic electrical parameters as current, active power and reactive power from a load distribution in a house to be analyzed through machine learning process and finally deliver a predictive performance for appliance loads.

On the other hand, some researches are centered on computing the energy power quantities (EPQ) as the active power aimed to forecast the power consumption in residential and non-residential buildings in order to reduce power consumption. For instance, Roger and Ian [16] developed energy consumptions benchmarks using energy signatures derived from daily energy consumption. The aim in this work is to produce control charts and diagnostic information for air-conditioned buildings; the results show that the daily signatures can generate robust energy consumption benchmarks compared to monthly or weekly energy signatures. For non-residential buildings, Vaghefi et al. [17] developed a data driven to forecast the power consumption with one day of anticipation. Machine learning is used to analyze historical power consumption data to make predictions about future patterns. They conclude that the controllable loads such as lighting, heating, ventilation and air conditioning can be used to lower a fraction of total power consumption. [18], Yulkseltan et al. analyzed the electrical consumption of three years from Turkish power market using a linear regression model in order to predict the demand over daily and weekly horizons.

The large quantity of data derived from the monitoring of different processes can be confusing and difficult to analyze. For this reason, artificial intelligence is recently used to cluster elements with similarities in data sets [19–21]. According to the aforementioned works, several studies are aimed at classifying electrical events and to recognize load patterns with artificial intelligence methodologies. For instance, Rafferty et al. [22] presented a method for detection and classification of multiple events in an electrical power system in real-time. Moving window is used to provide thresholds for event detection in real events recorder from the U.K. power system and a principal component analysis is used to discriminate between the events found. [23], Ming et al. developed a method to classify partial discharge events in gas insulated load break switches using probabilistic neural networks and fuzzy C-means as a sorter. This work is available to diagnose different defect models. Likewise, but improving the aforementioned method, Abubakar et al. [24] presented a technique for classifying partial discharge patterns using ensemble neural network learning with six neural network models. This technique improves the classification performance in comparison with a single neural network using the same testing and data sets. Similarly, Ali et al. [25] proposed a hybrid algorithm for power quality (PQ) disturbances detection using variational mode decomposition and S-transform to improve the accuracy in feature extraction. Finally, their proposed method uses support vector machines to classify the PQ events.

The aforementioned works are mostly focused onto data analysis in residential electrical installations for monitoring events in current signals to forecast the power consumption and to recognize patterns in the power consumption information. Moreover, the data window for the data processing is too large and the low power events are ignored, especially in non-residential buildings where the connection to plug loads is uncontrollable. At the same time, the relationship of the different load kinds with the equipment in an electrical installation remains unexplored.

This paper proposes a methodology based on a high resolution machine learning sorter, capable of processing data with a ten-second data window to identify low-power events and to classify the load kind present in non-residential electrical installations, no matter the number of elements connected to the installation. The events can be related to the loads presented in the line through the power consumption indexes to assist the management decisions based on the load performance of the electrical installation. The proposed methodology consists in computing EPQ from acquired data and to find events in the signal through statistical processes; afterwards, a C-means algorithm classifies the events accordingly to their load kind and EPQ. Finally, the classified events are located in a daily timeline with a ten-second resolution in order to show the load performance along the day. The methodology is validated using a test bench with resistive, capacitive and inductive loads. The experimentation is carried out in a healthcare facility in the region of Castilla y Leon (Spain).

2. Theoretical background

The recommended definitions for the measurement of electric power quantities (EPQ) are defined in the international standard [26], in order to quantify the flow of electrical energy under sinusoidal, nonsinusoidal, balanced and unbalanced conditions. The cluster analysis (C-means algorithm) allows classifying groups of data with similar characteristics according to the given features.

2.1. Electric power quantities

The active power (P) of any electrical installation in a time period t is given by

$$P = \frac{1}{kT}\int_{\tau}^{\tau+kT} p(t)dt \tag{1}$$

where T is the cycle time, k is a positive integer number, τ is the moment when the measurement starts and $p(t)$ is the product of voltage $v(t)$ and current $i(t)$.

The apparent power (S) is the amount of active power that can be supplied to a load under ideal conditions and is given by

$$S = V{\cdot}I \tag{2}$$

where V is the root mean square (RMS) value of the voltage and I is the RMS value of the current.

The reactive power (Q) is given by

$$Q = \sqrt{S^2-P^2} \tag{3}$$

The total harmonic distortion (THD) is an indicator of non-linear loads connected to the electrical installation and is given by

$$THD = \frac{I_H}{I_1}\sqrt{\left(\frac{I}{I_1}\right)^2-1} \tag{4}$$

where I_H is the current RMS value of the harmonic frequency, I_1 is the RMS value of the fundamental frequency

Power factor (PF) is defined as the relation between P and S powers as follows

$$PF = \frac{P}{S} \tag{5}$$

2.2. C-means algorithm

C-means is a cluster technique for data classification based on the minimization of the objective function (J_m) as follows

$$J_m = \sum_{i=1}^{D}\sum_{j=1}^{N}\mu_{ij}^{m}\ \|x_i-c_j\|^2 \tag{6}$$

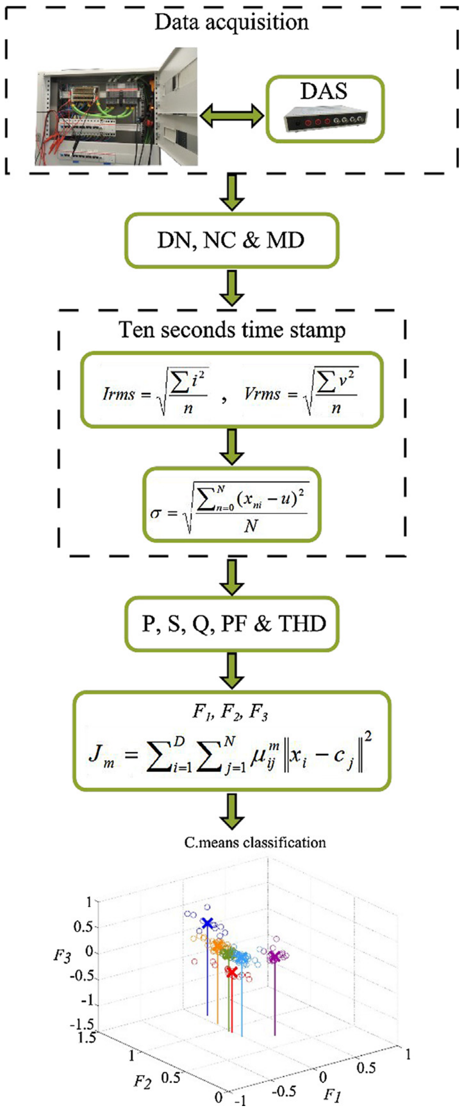


Fig. 1. Methodology.

where D is the number of data points, N is the number of clusters, m is the fuzzy partition matrix exponent, x_i is the ith data point, c_j the center of cluster, and μ_{ij} is the degree of membership for the jth given data point, this value is stored in a fuzzy partition matrix (μ) as follows

$$\mu = \begin{vmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1j} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{i1} & \mu_{i2} & \cdots & \mu_{ij} \end{vmatrix} \tag{7}$$

where the sum of the membership values for all clusters is one.

3. Methodology

The main purpose of this work is to classify the different load kinds present in non-residential electrical installations and to relate them to the elements connected to the installation in a given time. Fig. 1 depicts the proposed methodology. The voltage and current signals used in this work are acquired in a continuous way during 24 h at 8 Ksps. The signals are processed daily in time stamps of ten seconds to compute Irms, Vrms and EPQ as P, Q, THD and PF. In this work C-means is used as a cluster method because it is an unsupervised method for exclusive clustering, where each element in a group of data is grouped with the elements of its own cluster according to a membership degree, this membership degree can be used as a feedback to make sure that the elements in a group are highly related with the rest in the same group and adjust the clustering parameters in consequence.

Once the data are acquired, the first part of the methodology consists in defining three different configuration parameters: the number of standard deviations for the event detection (DN); the maximum number of classes to be created (NC); and the membership degree of the elements in the μ matrix (MD). Afterwards, it is necessary to process the current and voltage data to compute their RMS values and store them in two vectors; then, a rolling window method is used to find events in the current signal according to the standard deviation (σ) as follows

$$\sigma = \sqrt{\frac{\sum_{n=0}^{N}(x_{ni}-u)^2}{N}} \quad i = 1, \ldots, M-N \tag{8}$$

where u is the mean of the window data, N and M are the window size and the data size respectively.

A current RMS (*Irms*) value is recognized as an event when the difference between the value of one sample and another next to it is more than DN times the sample standard deviation. Once the current events are located, the electric power quantities (EPQ) as *P*, *Q*, *S*, *PF* and *THD* are computed for each event and stored in data vectors.

The EPQ values are used to define the features by implementing the C-means algorithm for the load classification. For this work three different features are used, first one (F_1) is a proposed EPQ weighing as a load kind indicator as shown in (9)

$$F_1 = \frac{Q}{S_m}\cdot\frac{1}{PF}\cdot S_\theta \tag{9}$$

where S_θ is the angle sign between the voltage and current signal and Q/S_m is a ratio to normalize the reactive power Q as an indicator of the presence of inductive and reactive loads and the maximum value that the apparent power can take S_m; this feature is used to identify the load kind at the event analyzed. Second feature (F_2) is the *THD* computed until the 25th harmonic from the acquired current data according to (4) from each found event; according to the aim of the methodology, it is important to mention that *THD* gives information about the presence of non-linear loads in the signal, for this reason is chosen as a feature for the C-means algorithm. The last feature (*F3*) is an *Irms* index used due to the currents signal can be affected for the load kind and is given by

$$F_3 = \frac{\Delta Irms}{Irms_{\max} - Irms_{\min}} \tag{10}$$

where $Irms_{max}$ is the maximum value of the *Irms* vector and $Irms_{min}$ is the minimum value of the *Irms* values in the time stamp, $\Delta Irms$ indicates the change in the *Irms* magnitude between one element and the previous one. About the used features, F_1 provides information about the load kind in the time when the installation is analyzed, F_2 gives information of the presence of non-linear loads in the signal, and in F_3 the current is the signal that can change owing to the load kind.

Besides the extracted features, in order to implement the C-means algorithm it is necessary to define the number of classes to be created. For this propose the C-means algorithm is run several times incrementing the number of classes to be originated until the degree of membership of the elements in the μ matrix be higher than MD. The cases in which this requirement is not accomplished are stored as incidences. When the membership degree is reached, the number of classes resulted are chosen for the classification. However, if the C-means algorithm reaches NC and MD is not achieved, the number of classes with fewer incidences is chosen to be used for the event classification.

Finally, the classification is executed and the classes obtained by the C-means algorithm are shown in a 3D graph and in a day time line. Furthermore, the classes are shown according to their behavior features to relate them to the equipment connected to the installation, besides that, a time line shows the time percentage taken for each class in each hour of the 24 h monitoring period.

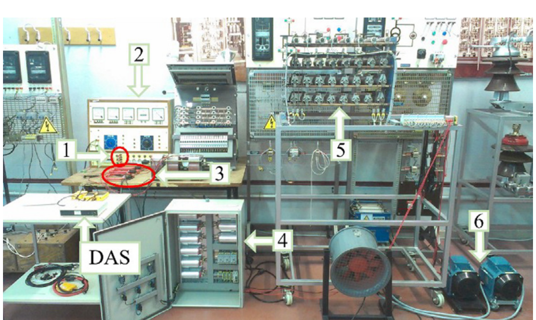
Resuming the proposed methodologies, rolling window can search values of interest in a data vector, for this work, it is used to find events in the current signal in concordance to the standard deviation of the signal. The energy power quantities give information about the electrical installation condition; in this work the power quantities provide the information of the load behavior in a given time. C-means is an unsupervised cluster method to divide a data set in similar groups of data and in this case is used to group different types of loads according to the proposed features.

4. Experimentation

The experimentation is carried out in two different stages: the first in a laboratory test using one test bench with known loads in order to validate the methodology and the C-means classification. The second one is made in a healthcare facility to prove the methodology in a real non-residential environment. For the data acquisition task, a proprietary equipment able to acquire signals from seven simultaneous channels at 8000 samples per second with a 16-bit resolution is used. Four channels are dedicated to 1A–1000A current measurements in several available ranges and three to 100 V–600 V voltage measurements. The requirements to apply this method are any system with 2.3 GHz two-core processor and at lease 4 Gb of RAM memory. DN is used in the events searching process, as a gain for the standard deviation value to store the maximum number of found events, in this case, a DN = 0.5 is proposed, but this value can be changed according to the current value in the installation. NC is the highest number of classes can be created, and accordingly to the experimentation the creation of maximum eight classes increment the membership degree in the elements of each class, for this reason, a NC = 8 is proposed. MD is the minimum wanted membership degree for the elements in each class, in this case, MD = 0.8 is proposed which represents a membership degree of 80%. The aforementioned values maximize the number of found events and classes in the C-means algorithm. The current and voltage data acquisition is done in a ten-minute time stamp with an 8 kilo samples per second of sampling frequency.

4.1. Laboratory test for methodology validation

For the methodology validation, a laboratory test bench shown in Fig. 2 consisting of three balanced star-connected to a 50 Hz, 230Vac is used. The loads are purely resistive of 71 Ω, purely capacitive of 75 μF, and mostly inductive of 560 mH. The test is done using current and voltage clamps for the data acquisition in the test bench to process it according to the aforementioned methodology in order to prove if the classification made by the C-means sorter is congruent with the loads presented in the test bench.



1. Voltage clamps 3. Current clamps 5. Resistive load
2. 230Vac power supply 4. Capacitive load 6. Inductive load

Fig. 2. Test bench for the validation test.

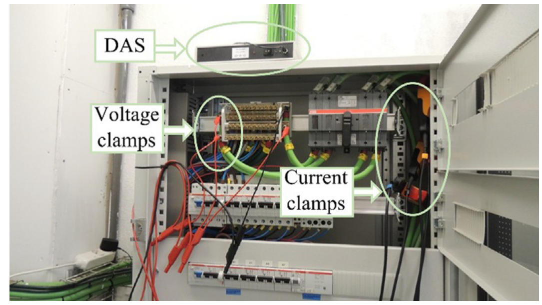


Fig. 3. DAS installation in board CS631.

4.2. Health care facilities

The data acquisition for the experimentation is done in a healthcare facility in the region of Castilla y Leon (Spain); this facility is a modern hospital construction with a round of 600 beds. The electrical installation has a three-phase configuration, but the lines feed different kinds of areas with different elements connected to the installation, for this reason the analysis is done with single phase measurements to obtain individual loads for each line. Data acquisition is done in a few load distribution boards in order to record the maximum number of load variations due to the action of switching power supplies, plug loads and medical equipment. A proprietary data acquisition system (DAS) is installed at the main board labeled as CG3. From this board, there are fourteen secondary boards connected from which the secondary boards CS301 and CS321 are chosen for monitoring. A second DAS system is connected to the secondary board CS301 that feeds some halls, electric plugs and warehouses; and a third DAS is connected to the secondary board CS321, which feeds commune zones and halls. The equipment installation is shown in the Fig. 3. For this work the input parameters are DN = 0.5, NC = 8 and MD = 80.

To process the acquired data, one day of samples is divided in ten-second time stamps to obtain approximately 8640 samples with 80,000 values each one. The rolling window method according to (8) and the EPQ are applied and computed from these samples, respectively. Furthermore, the THD is computed following (4) until the 25th signal harmonic.

5. Results

The obtained results are presented firstly with the methodology validation from the acquired data at the laboratory test bench. Finally the results from the data analysis in the healthcare facility in the general board CG3 and the secondary boards CS301 and CS321 are presented.

5.1. Test bench for methodology validation

Fig. 4 shows the classification results obtained for the validation process. Fig. 4(a) depicts the classes generated for the loads used in the test bench, where the centroids for the classification are represented with mark (x) in different colors (one color for each class), and the lines are the coordinates of the centroids in the (F_1, F_2) plane. In Fig. 4(a) the F_1 axis depicts the load kind where the negative side is for capacitive loads and the positive side for inductive loads. F_2 axis is the harmonic content percentage and the F3 axis is the *Irms* difference between one event and the other before it. The red elements in the negative F_1 side are capacitive loads, the green ones in the positive side are the inductive loads and the orange ones closer to the zero value are the resistive loads. Fig. 4(b) shows the signal waveform for the three different kinds of loads, where the current signals have the same color that the C-means classification in Fig. 4(a) and the voltage signals are presented in black. In Fig. 4(b) the current signal (red) shows the delay regarding the

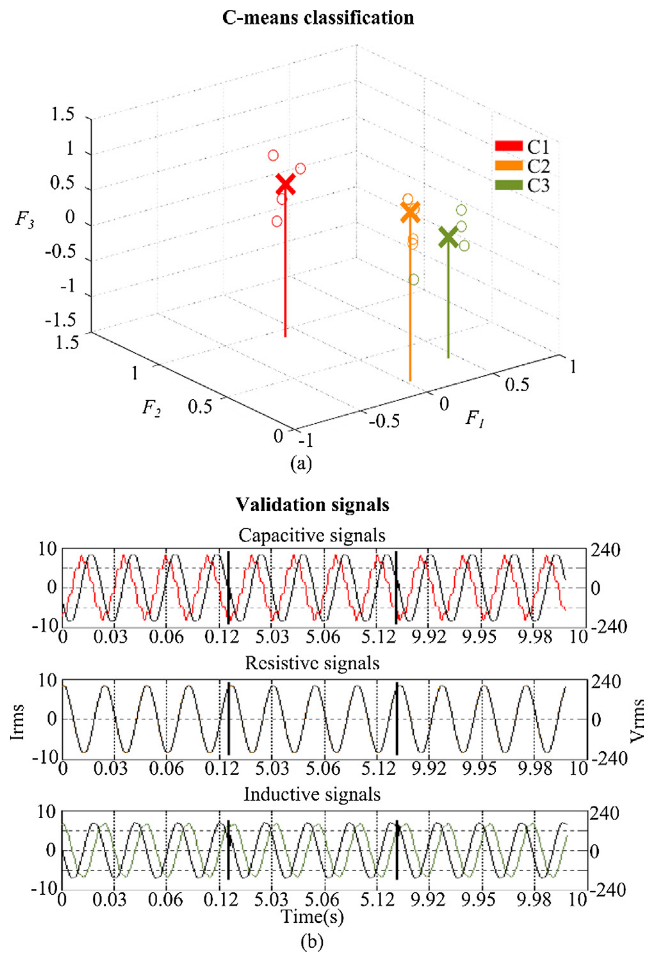


Fig. 4. Results for the validation process: (a) test bench generated classes, and (b) three different kind of loads, capacitive, resistive and inductive.

voltage signal (black) corresponding to the capacitive load, whereas the current signal (green) is lagging the voltage (black) corresponding to the inductive load, and the current signal for the resistive load (orange) is overlapping the voltage and cannot be viewed.

5.2. Healthcare facility

The results obtained in the healthcare facility are divided into two sections, the first section shows the results obtained by the C-means classification and the second one the relationship between the generated classes and the equipment connected to the electrical installation in the analyzed boards.

5.3. Data analysis

Fig. 5 depicts the events located in the current signal according to (8) in the 24 h lapse of data acquisition. The red line in Fig. 5(a)–(c) is the threshold value of the standard deviation median value of the all signals and the blue lines are the standard deviation values for each ten-second time stamp. Fig. 5(a) shows the main board CG3, Fig. 5(b) and (c) show the secondary board SB301 and the secondary board CS321, respectively.

The resulting classes from the C-means classification are shown with different colors in a 3D graph, where, the F_1 axis depicts the load kind in which the negative side is for capacitive loads and the positive side for inductive loads. F_2 axis is the harmonic content percentage and the F3 axis is the *Irms* difference between one event and the other before it. Moreover the centroid for each class is represented with a mark (x) and

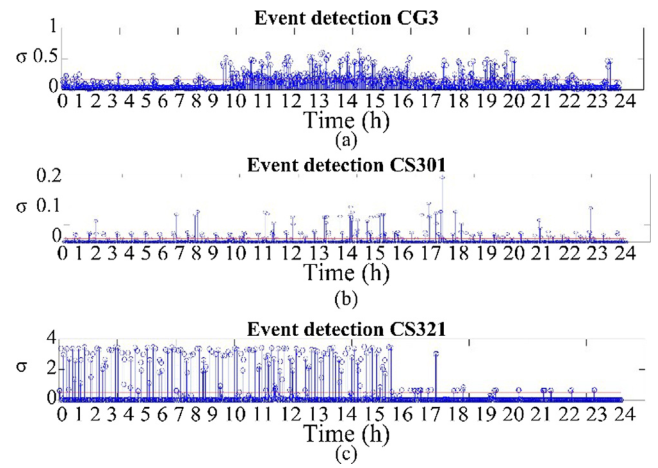


Fig. 5. Events detected, the blue lines are the RMS values for each time stamp and the red line is the all signal standard deviation value: (a) main board CG3, (b) secondary board SB301, and (c) secondary board CS321. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

projected in the (F_1, F_2) plane with a vertical line. Besides the 3D graph, the results are shown in time lines where the classes are represented with the same color as the 3D graph and the *Irms* value are shown in black, and one chart where the class behavior is illustrated according to the used features.

The main board CG3, Fig. 6(a), has four classes with high percentage of resistive loads. However, these classes are differentiated from

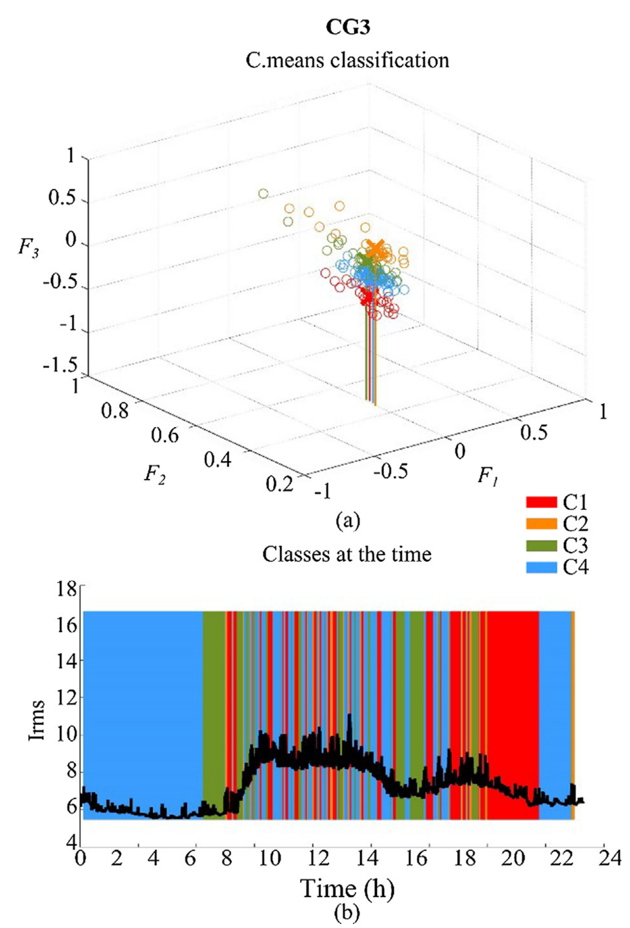


Fig. 6. Results for the CG3: (a) C-means classification, and (b) classes along the time.

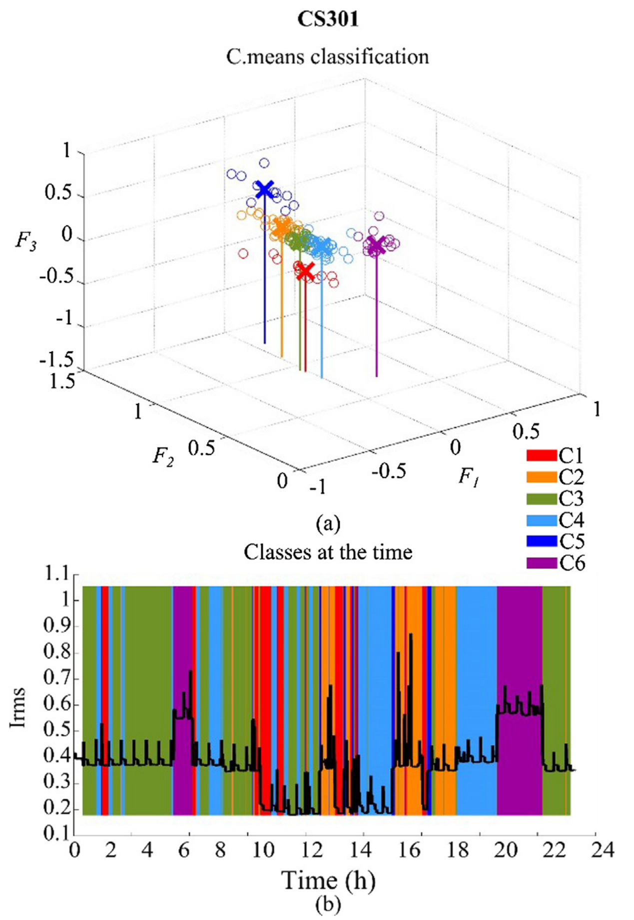


Fig. 7. Results for the CS301: (a) C-means classification, and (b) classes along the time.

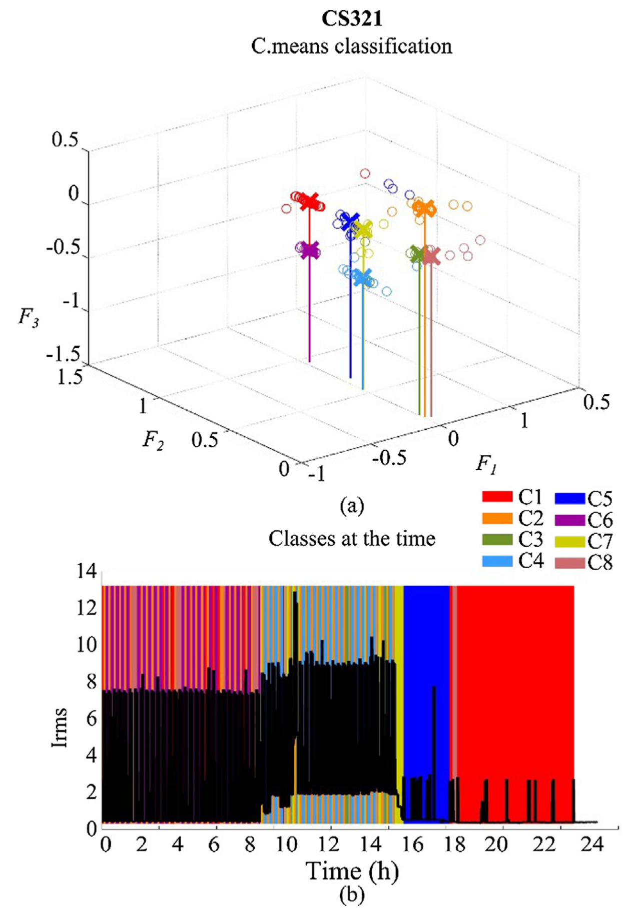


Fig. 8. Results for the CS321: (a) C-means classification, and (b) classes along the time.

each other due to their THD and *Irms* values, respectively. The high THD in class C2 and C3 are due to the action of switching power supplies in the commune zones and the medical bay. The class C1 and C4 are differentiated by the *Irms* value. The time chart for the CG3 Fig. 6(b) shows that from the 7 to the 19 h the classes are more variable at the time.

For the secondary board CS301, Fig. 7(a), the elements are more dispersed and the number of generated classes is higher than the main board CG3. The class C6 shows elements with the presence of inductive loads and the rest of classes show elements with resistive loads but with different percentage of harmonic content. The time chart for the secondary board CS301 in Fig. 7(b) shows the change from resistive to inductive loads at the 6:00 and 21:00 h approximately; besides 8:00 and 23:00 h the loads return to be resistive. According to the electric blueprints this behavior is surely caused by the action of some AC motors connected in the warehouse areas.

The secondary board CS321, Fig. 8(a), is one with the most dispersed elements for the classification and in consequence the created classes are larger. The inductive loads present in C5 and C8 can be related to the activity of cleaning and maintenance equipment used in the common zones. C1 and C6 are the classes with more presence of THD, and C3 is a class with balanced loads and very low harmonic content. The time chart in the secondary board CS321 from Fig. 8(b) is more homogenous due to the data dispersion and the generated classes. However, it is possible to identify four principal time groups.

5.3.1. Classes and equipment relation

Fig. 9 shows the behavior of each generated class according to the features used in the C-means classification. An ID number over the bars shown in Fig. 9(a)–(c) is generated as electric signature for each class through the feature median values summation. Furthermore, Fig. 9 shows the time percentage for each hour in the 24 h lapse of data acquisition.

The classes in the main board CG3 from Fig. 9(a) show a similar level of F_2 and F_1 with different variations in F_3. The positive increases in F_3 indicate the rise in the use of devices connected to the main board, and the negative increase is in accordance with the devices disconnection. Fig. 9(d) shows high activity from 7 h to 20 h; this is in concordance with the working hours in cardiology, medical consultants, and common zones. According to the behavior shown in Fig. 9(a) and the activity in Fig. 9(d), the classes C2 and C3 must be associated with the use of specialized medical equipment, C1 with computer equipment and C4 with vending machines and lights.

For the secondary board CS301 in Fig. 9(b), the classes from C1 to C5 show very high levels in F_2 and values close to zero for F_1; this behavior is narrowly related to the usage of switching power supplies. The F_1 increment in the class C6 suggest the presence of inductive loads connected to the use of AC motors. In Fig. 9(e) the classes C1 and C5 show the connection and disconnection of equipment from 6 h to 18 h. Furthermore the inductive loads presence are divided in two periods one from 5 h to 6 h and another from 21 h to 23 h; this can be correlated to the functions of the maintenance personal in the warehouses.

Fig. 9(c) shows high variations in F_2 and F_3 between some classes. For the C1 and C6 classes the increment in F_2 must be linked to the action of the electric ballast in halls and hallways. The classes C5 and C8 can be associated with the equipment used for the cleaning task in the halls and common zones. Furthermore, C2 and C4 classes show the on-off function of vending machines in the halls. In Fig. 9(f) it is possible to identify the working hours for the personal from 8 h to 18 h

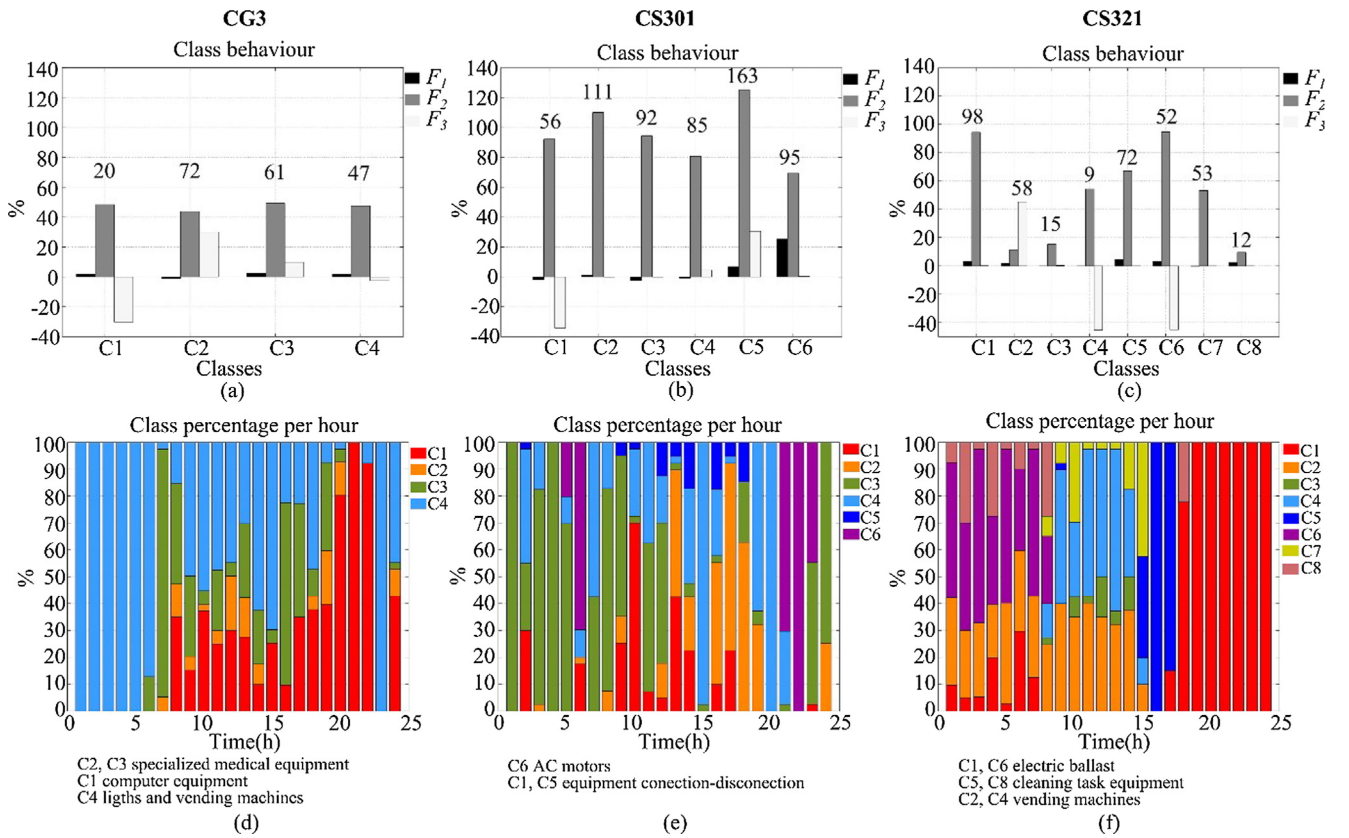


Fig. 9. Class behavior and class percentage: (a) CG3 class behavior, (b) CS301 class behavior, (c) CS321 class behavior, (d) CG3 class percentage per hour, (e) CS301 class percentage per hour and (f) class percentage per hour.

where the presence of different classes is evident. From 1 h to 7 h and from 18 to 24 h the increment in the usage of lights is also evident. Moreover, from 15 h to 17 h the very high presence of the class C5 suggests the closing in some areas for the maintenance or cleaning tasks.

The present method uses an unsupervised C-means algorithm for data clustering, as a difference of others using neural networks or support vector machines for data classification. In comparison with neural networks as the competitive learning, the clustering results are similar in the location of the centroids for the generated groups, however, competitive learning does not have a membership degree as a feedback about the similitude between the elements in one group, and the members for each centroid are not identified, the Fig. 10 shows the clustering results for the secondary board CS301 using neural networks. On the other hand, support vector machine is a supervised method and needs to be trained with a data set that contains the expected results. Furthermore, in the present methodology the continuous voltage and current signals acquisition, and the proposed features to classify the identified events, gives the method the possibility of identifying the load behavior along the day.

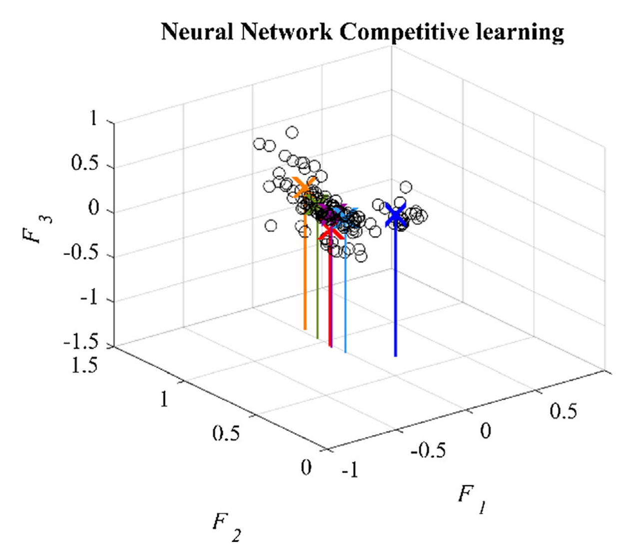


Fig. 10. Results for the CS301 with Neural Network Competitive Learning.

6. Conclusions

In this paper a novel non-intrusive methodology is presented to classify the loads found in an electrical installation. The proposed methodology allows identifying low power events due to the time stamp size. With the EPQ integration, it is possible to classify the load kinds in an electrical installation for a given period of time. Moreover, it is possible to relate the classes resulted from the C-means classification with equipment in the installation and assign them an electric signature. Additionally, due to the variations in the dispersion of the selected features in each secondary board, the C-means algorithm results an ideal tool to make the clustering according to the different degrees of membership.

For the cases analyzed in this work the main board presents a higher percentage of resistive loads. However, in the analysis made over the secondary boards the presence of inductive loads is evident, especially in the secondary boards connected to areas with plug loads such as common zones in medical bays and public access halls. Besides analyzing the behavior of the generated classes, it is possible to identify the on-off action of several devices connected to the installation.

In contrast to most recently published works, the proposed methodology can be applied in non-residential buildings. The event classification is focused on the load kinds in the installation and related with the equipment connected to it. The use of a proprietary data acquisition system allows the acquisition of a great deal of raw current and voltage data to be processed.

Some limitations presented in this work could be the manual adjust of the entry parameters DN and NC, for a future work, it could be interesting to implement an algorithm for the self-adjustment of these parameters.

Acknowledgments

The authors want to thank the Administration Board of Sanidad de Castilla y Leon (SACYL) for providing the access to the healthcare facilities. This project was supported in part by the National Council of Science and Technology (CONACYT), Mexico, under scholarship 487612; SEP-CONACyT 222453-2013 grant, FOMIX QRO-2014-C03-250269 grant, and FOFIUAQ-FIN201613 grant.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ijepes.2018.06.040.

References

[1] Lovett T, Lee JH, Gabe-Thomas E, Natarajan S, Brown M, Padget J, et al. Designing sensor sets for capturing energy events in buildings. Build Environ 2016;110:11–22. http://dx.doi.org/10.1016/j.buildenv.2016.09.004.
[2] Basu K, Debusschere V, Douzal-Chouakria A, Bacha S. Time series distance-based methods for non-intrusive load monitoring in residential buildings. Energy Build 2015;96:109–17. http://dx.doi.org/10.1016/j.enbuild.2015.03.021.
[3] Di Giorgio A, Pimpinella L. An event driven Smart Home Controller enabling consumer economic saving and automated Demand Side Management. Appl Energy 2012;96:92–103. http://dx.doi.org/10.1016/j.apenergy.2012.02.024.
[4] Senra R, Boaventura WC, Mendes EMAM. Assessment of the harmonic currents generated by single-phase nonlinear loads. Electr Power Syst Res 2017;147:272–9. http://dx.doi.org/10.1016/j.epsr.2017.02.028.
[5] Sainz L, Balcells J. Harmonic interaction influence due to current source shunt filters in networks supplying nonlinear loads. IEEE Trans Power Deliv 2012;27:1385–93. http://dx.doi.org/10.1109/TPWRD.2012.2187314.
[6] Gwon G-H, Kim C-H, Oh Y-S, Noh C-H, Jung T-H, Han J. Mitigation of voltage unbalance by using static load transfer switch in bipolar low voltage DC distribution system. Int J Electr Power Energy Syst 2017;90:158–67. http://dx.doi.org/10.1016/j.ijepes.2017.02.009.
[7] Cominola A, Giuliani M, Piga D, Castelletti A, Rizzoli AE. A hybrid signature-based iterative disaggregation algorithm for non-intrusive load monitoring. Appl Energy 2017;185:331–44. http://dx.doi.org/10.1016/j.apenergy.2016.10.040.
[8] Hassan T, Javed F, Arshad N. An empirical investigation of V-I trajectory based load signatures for non-intrusive load monitoring. IEEE Trans Smart Grid 2014;5:870–8. http://dx.doi.org/10.1109/TSG.2013.2271282.
[9] Huang TD, Wang WS, Lian KL. A new power signature for nonintrusive appliance load monitoring. IEEE Trans Smart Grid 2015;6:1994–5. http://dx.doi.org/10.1109/TSG.2015.2415456.
[10] Belley C, Gaboury S, Bouchard B, Bouzouane A. Nonintrusive system for assistance and guidance in smart homes based on electrical devices identification. Expert Syst Appl 2015;42:6552–77. http://dx.doi.org/10.1016/j.eswa.2015.04.024.
[11] Nguyen TK, Dekneuvel E, Jacquemod G, Nicolle B, Zammit O, Nguyen VC. Development of a real-time non-intrusive appliance load monitoring system: an application level model. Int J Electr Power Energy Syst 2017;90:168–80. http://dx.doi.org/10.1016/j.ijepes.2017.01.012.
[12] Abubakar I, Khalid SN, Mustafa MW, Shareef H, Mustapha M. Application of load monitoring in appliances' energy management – a review. Renew Sustain Energy Rev 2017;67:235–45. http://dx.doi.org/10.1016/j.rser.2016.09.064.
[13] Gillis JM, Member S, Morsi WG. Non-intrusive load monitoring using semi-supervised machine learning and wavelet design 2017;8:2648–55.
[14] Bonfigli R, Principi E, Fagiani M, Severini M, Squartini S, Piazza F. Non-intrusive load monitoring by using active and reactive power in additive Factorial Hidden Markov Models. Appl Energy 2017;208:1590–607. http://dx.doi.org/10.1016/j.apenergy.2017.08.203.
[15] Buddhahai B, Wongseree W, Rakkwamsuk P. A non-intrusive load monitoring system using multi-label classification approach. Sustain Cities Soc 2018;39:621–30. http://dx.doi.org/10.1016/j.scs.2018.02.002.
[16] Hitchin R, Knight I. Daily energy consumption signatures and control charts for air-conditioned buildings. Energy Build 2016;112:101–9. http://dx.doi.org/10.1016/j.enbuild.2015.11.059.
[17] Vaghefi A, Farzan F, Jafari MA. Modeling industrial loads in non-residential buildings. Appl Energy 2015;158:378–89. http://dx.doi.org/10.1016/j.apenergy.2015.08.077.
[18] Yukseltan E, Yucekaya A, Bilge AH. Forecasting electricity demand for Turkey: modeling periodic variations and demand segregation. Appl Energy 2017;193:287–96. http://dx.doi.org/10.1016/j.apenergy.2017.02.054.
[19] Du Z, Fan B, Jin X, Chi J. Fault detection and diagnosis for buildings and HVAC systems using combined neural networks and subtractive clustering analysis. Build Environ 2014;73:1–11. http://dx.doi.org/10.1016/j.buildenv.2013.11.021.
[20] Mostafavi N, Farzinmoghadam M, Hoque S. Urban residential energy consumption modeling in the Integrated Urban Metabolism Analysis Tool (IUMAT). Build Environ 2017;114:429–44. http://dx.doi.org/10.1016/j.buildenv.2016.12.035.
[21] Benmouiza K, Tadj M, Cheknane A. Classification of hourly solar radiation using fuzzy c-means algorithm for optimal stand-alone PV system sizing. Int J Electr Power Energy Syst 2016;82:233–41. http://dx.doi.org/10.1016/j.ijepes.2016.03.019.
[22] Rafferty M, Member S, Liu X, Laverty D, Mcloone IS. Real-time multiple event detection and classification using moving window PCA. IEEE Trans Smart Grid 2016;7:1–12. http://dx.doi.org/10.1109/TSG.2016.2559444.
[23] Su M-S, Chia C-C, Chen C-Y, Chen J-F. Classification of partial discharge events in GILBS using probabilistic neural networks and the fuzzy c-means clustering approach. Int J Electr Power Energy Syst 2014;61:173–9. http://dx.doi.org/10.1016/j.ijepes.2014.03.054.
[24] Abubakar A, Abubakar Mas'Ud A, Stewart BG, McMeekin SG. Application of an ensemble neural network for classifying partial discharge patterns. Electr Power Syst Res 2014;110:154–62. http://dx.doi.org/10.1016/j.epsr.2014.01.010Review.
[25] Abdoos AA, Khorshidian Mianaei P, Rayatpanah Ghadikolaei M. Combined VMD-SVM based feature selection method for classification of power quality events. Appl Soft Comput 2016;38:637–46. http://dx.doi.org/10.1016/j.asoc.2015.10.038.
[26] Fluke Corporation, Three-Phase Power Quality and Energy Analyzers, Fluke Co., Everret, WA, USA, 2012. doi: https://doi.org//10.1109/IEEESTD.2010.5439063.