



Ingeniería y Desarrollo

ISSN: 0122-3461

ISSN: 2145-9371

Fundación Universidad del Norte

SIERRA GIL, EDUARDO; BASULTO ESPINOSA, ALFREDO; ESCALONA AGUILAR, ARGELIS

A new statistical approach to customer classification and load profiling

Ingeniería y Desarrollo, vol. 38, no. 1, 2020, January-June, pp. 32-43

Fundación Universidad del Norte

DOI: <https://doi.org/10.14482/inde.38.1.519.5>

Available in: <https://www.redalyc.org/articulo.oa?id=85270674003>

- How to cite
- Complete issue
- More information about this article
- Journal's webpage in redalyc.org

UNEN  redalyc.org

Scientific Information System Redalyc

Network of Scientific Journals from Latin America and the Caribbean, Spain and Portugal

Project academic non-profit, developed under the open access initiative



ARTÍCULO DE INVESTIGACIÓN / RESEARCH ARTICLE

<http://dx.doi.org/10.14482/inde.38.1.519.5>

A new statistical approach to customer classification and load profiling

*Nueva aproximación estadística a
la clasificación de consumidores y la
construcción de curvas de carga*

EDUARDO SIERRA GIL *

ALFREDO BASULTO ESPINOSA **

ARGELIS ESCALONA AGUILAR ***

* Universidad de Camagüey “Ignacio Agramonte Loynaz. Departamento de Ingeniería Eléctrica, Facultad de Electromecánica (Cuba).

Ingeniero Electricista, Doctor en Ciencias Técnicas en la especialidad de Ingeniería Eléctrica, Profesor Titular de la disciplina de Sistemas Eléctricos del Departamento de Ingeniería Eléctrica de la Universidad de Camagüey, eduardo.sierra@reduc.edu.cu

** Empresa Eléctrica Provincial Camagüey. Dirección Técnica (Cuba). Ingeniero Electricista, Master en Ciencias en la especialidad de Ingeniería Eléctrica, Profesor Instructor Adjunto del Departamento de Ingeniería Eléctrica de la Universidad de Camagüey, basulto@eleccmg.une.cu

*** Empresa Inmobiliaria ALMEST. Inversiones (Cuba). Ingeniero Electricista, Especialista de inversiones, argelis@cxs.co.cu

Correspondencia: Eduardo Sierra Gil, teléfono: 53-32261667, eduardo.sierra@reduc.edu.cu

Origen de subvenciones o apoyos recibidos: Los resultados forman parte del proyecto de I+D+i “Gestión Integrada de Redes Eléctricas”, código: E82oCM900031, Objetivo: Desarrollar un sistema integrado de gestión de redes eléctricas de distribución y subtransmisión, financiado por la Empresa Eléctrica de Camagüey y el Ministerio de Educación Superior,

Fecha de ejecución: 2016-2018.



Resumen

En un sistema eléctrico de distribución es de suma importancia tener un conocimiento detallado de las características de las cargas que este alimenta y que determinan en última instancia el comportamiento de los parámetros en los diferentes regímenes de operación. Existen métodos diversos para la clasificación de los consumidores y la construcción de las curvas de carga diaria típicas sin embargo, estos métodos en su mayoría no consideran que estas curvas están sujetas a la conducta de cada tipo de consumidor. Este trabajo propone una nueva aproximación al problema a partir de un método basado en dos herramientas estadísticas, el coeficiente de concordancia de Kendall y el coeficiente de correlación por rangos de Spearman, y se comprueba la efectividad del mismo mediante su aplicación en dos circuitos de distribución, demostrándose que existe coincidencia entre los perfiles de carga obtenidos mediante el método propuesto, y los que se obtuvieron mediante mediciones realizadas en la subestación.

Palabras clave: Curvas de carga, demanda eléctrica, redes de distribución, transformadores de distribución.

Abstract

It is of utmost importance in an electrical distribution system to have a detailed knowledge of the characteristics of the loads it feeds and that they determine, in a final extent, the behavior of parameters in the different regimes of operation, there are various methods for the classification of consumers and the construction of the typical daily load curves, however these methods do not mainly consider that these curves are subjected to the behavior of each kind of consumers. This work proposes a new approximation to this problem based on a method sustained by two statistical tools, Kendall matching coefficient and the correlation coefficient for ranges stated by Spearman and its effectiveness is checked by means of its application in two distribution circuits, demonstrating that there is a coincidence between the load profiles obtained through the method proposed and the load profiles obtained through measurements accomplished at the substation.

Keywords: distribution network, distribution transformers, electrical demand, Load profiles.

1. INTRODUCTION

Due to the inability to perform measurements on each node of a network of primary distribution, to characterize the loads it feeds during the analysis of these networks different modeling methods have been developed to get as close as possible to their real values, as being the load the most influential elements in the results of a study in these circuits. The methods developed are generally based on the determination of the characteristic behavior of groups of customers, classifying them into layers, starting from the sum of the load curves of the clients associated with a distribution transformer to obtain the characteristic of the load on the node primary network [1], [2]. Such an approach is presented in reference [3] to a study of the behavior of residential, industrial and commercial customers by field measurements and through a methodology for the addition of load curves to determine the expected load on a part of the distribution network. A similar analysis is presented [4] specifically for industrial customers, where statistical methods are also used to determine the variations of the parameters of the curves of daily and monthly load stations.

Another approach used is to consider that residential loads can be made from any possible and relevant equipment, a given power demand, the frequency of use and the operation time, as a potential correlation equipment through a synthetic probabilistic time series [5]. An element to consider in the evolution of techniques of customer classification and load profiling is the incorporation of the infrastructure of automatic measurement AMI that permits obtaining measurements in real time [6]-[8]. One of the most developed directions is the use of clusters techniques for customer classification [7], [9], [10]

The use of artificial intelligence has also been treated as a tool for the stratification of load curves using fuzzy cluster techniques (Fuzzy K-Means), Self-Organizing Maps and Artificial Neural Networks [11]-[13]. However, all these methods only allow the establishment of a standard curve, usually per unit, by customer type. To determine the load curve of the distribution transformer, it is not only necessary to know the characteristic of each type of customer but also the number of customers in each layer associated with it and their maximum demands.

In the articles consulted load curves are treated as a perfect Gaussian and parametric statistics is used, which is based on the assumption that random samples were selected from populations that follow a normal distribution, but this is not the case of the load curves because customers do not have a normal distribution and their behavior is given according to each client's customs; for this reason a new approach from nonparametric statistics for the characterization of customers is proposed, since this is responsible for the study of samples that often do not pose any knowledge about the distributions of key populations, except if they are maybe continuous.

2. METHODOLOGY

Based on the fact that most of the loads analyzed in the study are of the residential type, there is no mistake considering the high probability of coincidence of the peaks of customers (85-90%) thus the sample size can be determined by equations (1) and (2):

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad (1)$$

$$n = \frac{n'}{1 + \frac{n'}{N}} \quad (2)$$

Where:

p: Probability of coincidence of the peaks.

se: Standard error.

N: Population size.

Considering the probability of coincidence of the peaks equals 0,9 (p=0,9) with a standard error of 0,05 (se=0,05) a statistically and significant sample is obtained from the total study population, which will be 36 customers, a value lower than the value samples taken from this study (52 customers).

A. Kendall matching coefficient (W)

Within nonparametric statistics there is a very important factor that measures the degree of association between various sets (k) of N entities. It is useful to determine the association between three or more variables utilized in several areas of knowledge like the economy, the evaluation of risks or medical studies [14]-[16]. This factor is called Kendall concordance coefficient and it is expressed by equation (3).

$$W = \frac{S}{\frac{1}{12} \times K^2 \times (N^3 - N) - K \times \sum Li} \quad (3)$$

Where:

W: Kendal matching coefficient.

S: Sum of the squares of the differences observed with respect to an average.

N: Sample size.

K: Number of variables included.

Li: Sum of links or ties between the ranges.

The value of W ranges between 0 and 1. The value of 1 means a concordance of all agreements and zero total disagreement. This coefficient is calculated to see the degree of similarity between all residential customers (110 and 220 V) which were most of the measurements performed. If the calculus of the coefficient might provide a value above 0,5 ($\geq 0,5 W$), it is assumed that there is an acceptable similarity between customers of that layer, hence the average curve is taken as the characteristic of this layer, otherwise ($W < 0,5$) does not have a desired similarity and then customers are separated in two groups, by voltage levels such as residential consumers can be separated into two groups by voltage levels, one for which their average consumption is above the ensemble mean and another for those who consumed below the average.

The corresponding calculations yielded the following results shown in Table 1:

TABLE 1

Voltage level(V)	W
110	0,61
220	0,32

As seen in the case of customers 220V Kendall matching criterion is below the target, therefore as indicated above customers are separated based on their average consumption. After this is done, the calculation of W starts, resulting in a level even lower than that of the whole ($W_{\text{xupward}}=0,271$ and $W_{\text{xdownward}}=0,269$) match, which leads to the conclusion that the set curve is much more accurate that the ones separated by average consumption.

B. Spearman Correlation Coefficient Rank (r_s)

This ratio is used to compare the curve resulting from the sum of the characteristic curves obtained in the circuit under study measurement on a time period for the already-mentioned circuit.

Spearman correlation coefficient rank is used which is a measurement of association that requires both variables to be measured, at least, on an ordinal scale, so that the objects or individuals being studied can be placed in two ordered series [17], [18].

To calculate r_s , a list of N subjects is made. After each subject is registered, its range is entered in the variable X and variable Y

Now different values are defined below, the difference between the two ranges ($d_i = X_i - Y_i$).

Each is squared and all its values are added.

The value of the sum of the previous step and N (number of subjects) is replaced directly in equation (4).

$$r_s = 1 - \frac{6 \times \sum_{i=1}^n d_i^2}{(N^3 - N)} \quad (4)$$

Similar to Kendall matching coefficient r_s value ranges between 0 and 1. The closer it is to one, the bigger the similarity between the variables compared will be, the opposite will occur when the value is close to zero.

The ranges are determined by organizing joint variables ascending and assigning value to each of its place within the total set.

3. RESULTS AND DISCUSSION

For the database to be used, a network analyzer was placed in different points of the circuit C-41 of the city of Camagüey, taking the consumption of various types of electricity customers (mainly residential) over a period of 11 days. The instrument took samples every second of different electrical parameters.

To obtain data curves the measurements made for each type of client were processed, the active (P), reactive (Q) and apparent powers (S) were averaged at intervals of 15 minutes.

The characteristic curves are determined by adding all the curves of the measured customers [19] of the same layer and it is divided by the number of customers, as shown in equation (5).

$$\overline{\text{Curve}} = \frac{\sum_{n=1}^k CI_n}{N} \text{ (KW)} \quad (5)$$

Where:

CI: Individual clients

N: Number of clients

After this curve is obtained, it is divided by its maximum value to obtain the curve per unit (p.u) that will be the load curve characteristic of each consumer. This characteristic load curve will remain in the database of the electric company so they can make predictions and studies related to the loads they provide service with.

$$CV \text{ p. u.} = \frac{\overline{\text{Curve}}}{\text{MaxValue}} \quad (6)$$

Bearing load curves characteristics of each client, if you want to know the actual behavior of any client, just multiply the maximum value of that customer demand for the characteristic curve (in p.u) that corresponds.

A. Comparison of the result

Using Spearman correlation coefficient rank, the curve resulting from adding up all the characteristic curves obtained was compared with the load curve measured in the C-41 circuit, yielding satisfactory results considering the criteria explained about this coefficient. The level of correlation between these two curves was 0,855.

For greater certainty that the results obtained were correct, it was decided to make the same comparison but in another circuit of the same city, the C-31. For this aim, the characteristic curves of customers from C-41 that coincided with this one were taken as reference.

For the types of consumers that did not match the sample taken for the study on circuit C-41, existing curves in other databases of the utility were taken and the results were equally satisfactory. The Spearman correlation coefficient rank in this case was 0,973.

The results of the comparison are shown in figures (1), (2), (3) and (4).

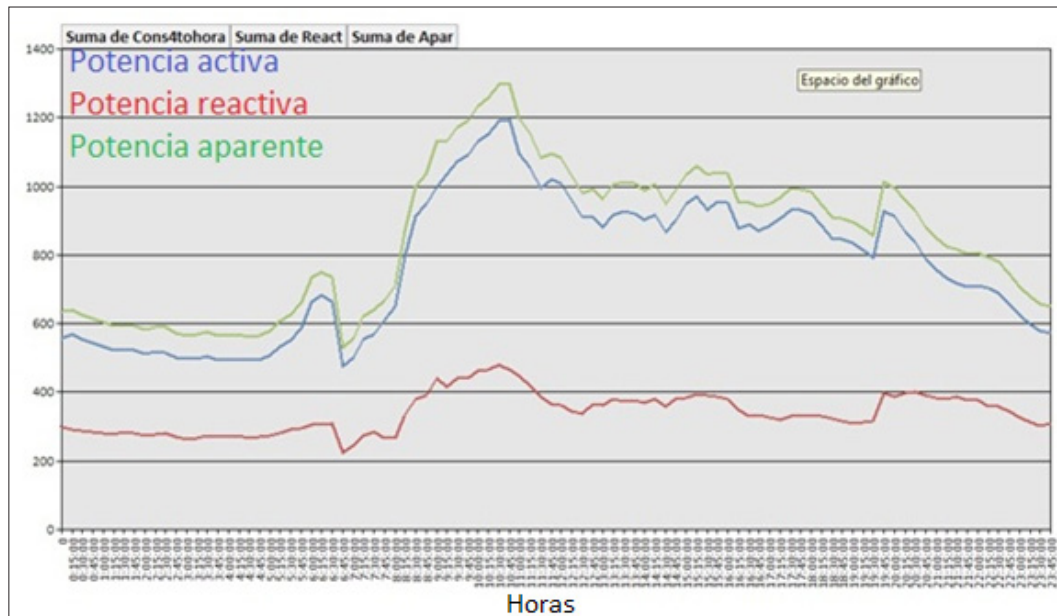


FIGURE 1. LOAD CURVE OBTAINED FROM THE SUM OF
THE CURVES OF CLIENTS IN CIRCUIT C-41

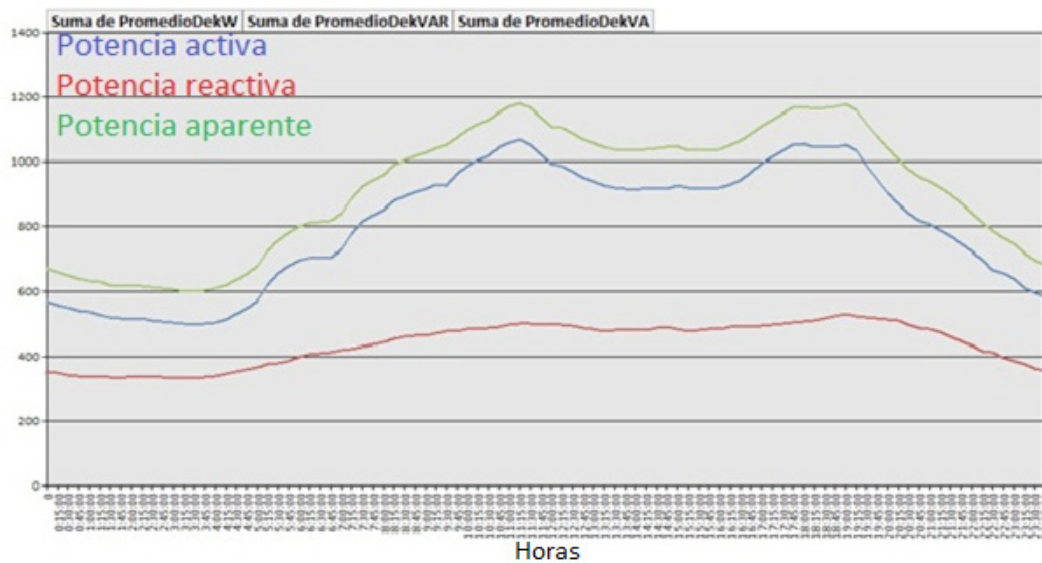


FIGURE 2. LOAD CURVE OBTAINED FROM THE MEASUREMENT
IN THE TOTALIZER FROM CIRCUIT C-41

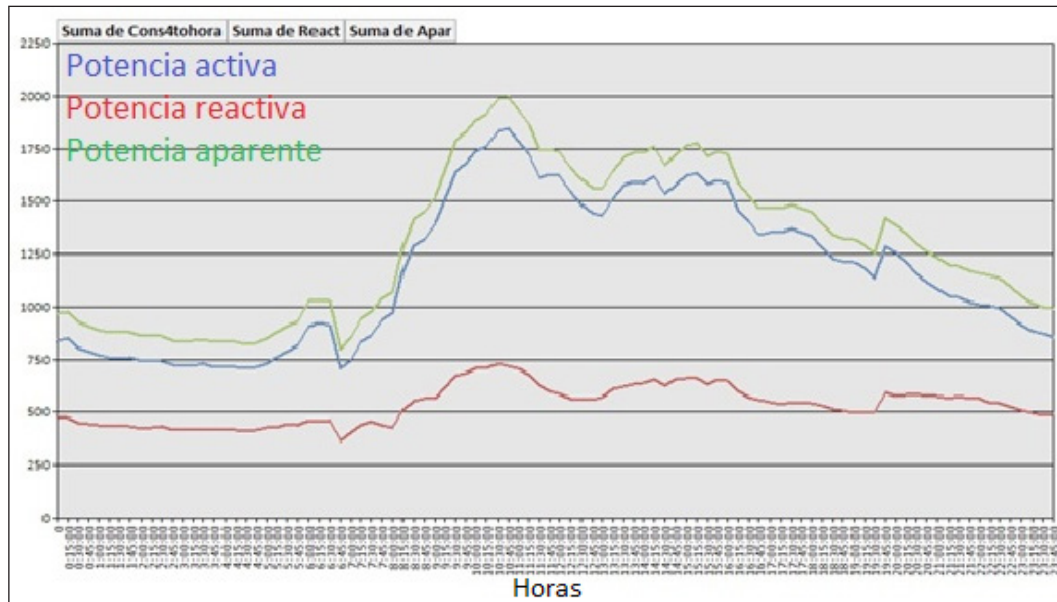


FIGURE 3. LOAD CURVE OBTAINED FROM THE SUM OF
THE CURVES OF CLIENTS IN CIRCUIT C-31

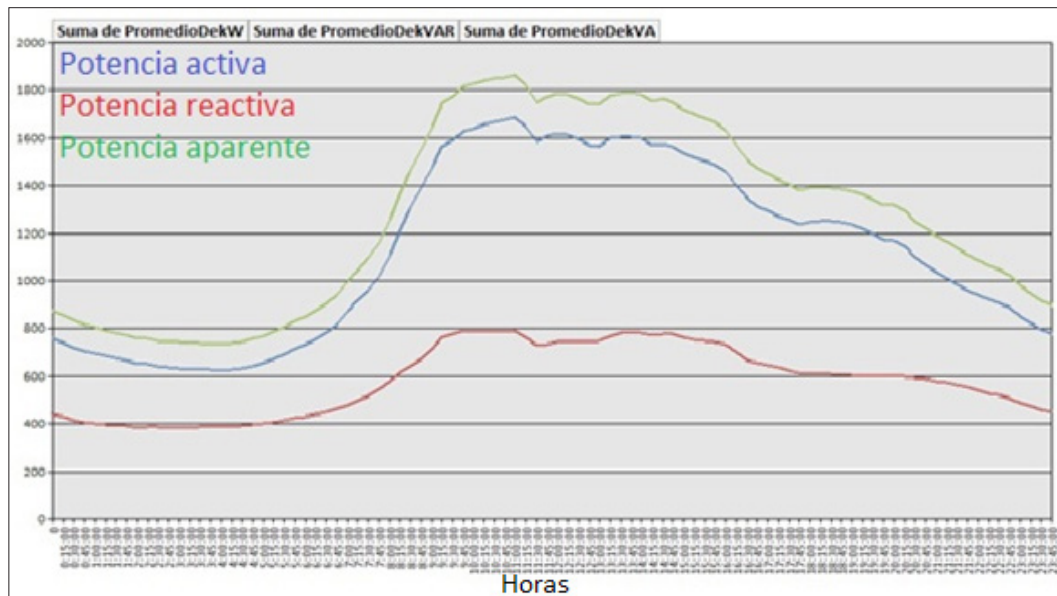


FIGURE 4. LOAD CURVE OBTAINED FROM THE MEASUREMENT
IN THE TOTALIZER FROM CIRCUIT C-31

In regards to circuit C-41 both curves show the same peaks, differing in the fact that the peak of the late afternoon/evening of the measurement curve is bigger. The behavior of this circuit is like this because although there are a considerable number of commercial customers with a high consumption mainly between 9:00 and 16:00 hours, they also have a significant number of residential customers, which make a well-defined peak consumption in the afternoon.

In C-31 circuit both curves show a sustained peak form 9:00 to 16:00 hours, tending to decrease in the curve resulting from the sum of the customers. Such a peak consumption is caused by the influence of connected commercial consumers to the circuit, which virtually lead their behavior since they have a higher weight in the load and also because on it there is no influential residential representation that can significantly raise the consumption in the peak hours in the afternoon.

4. CONCLUSIONS

The current statistic approaches for the characterization of consumers in distribution networks are based on parametric statistics, however the different layers of costumers do not correspond to a normal distribution and their behavior differ, at first, depending on the habits of each client, it makes nonparametric statistics to be suitable to use in this case.

The method shown using the Kendall matching coefficient is characterized by its simplicity and effectiveness to measure the degree of association of different sets, in this case, of the load curves obtained by sampling the layers from networks of distribution in which customers are classified.

The obtained results support the above statement with Spearman coefficient correlation ranks made possible to corroborate the similarity between the curve obtained by the method proposed and the one obtained by measurements in the distribution substation.

The robustness of the method was found when inducing the results of the curves obtained in C-41 to circuit C-31 where a high degree of similarity is found, as shown by Spearman correlation coefficient ranks between the sum of the load curves from the customers' circuit and the load curve measured at the distribution substation.

REFERENCES

1. Y. Wang, Q. Chen, C. Kang, M. Zhang, K. Wang, Y. Zhao. "Load profiling and its application to demand response: A review". *Tsinghua Science and Technology*. Vol. 20. 2015. pp. 117-129. Doi: 10.1109/TST.2015.7085625

2. A. Mutanen, M. Ruska, S. Repo, P. Jarventausta. "Customer Classification and Load Profiling Method for Distribution Systems" *IEEE Transactions on Power Delivery*. Vol. 26.2011. pp. 1755-1763. Doi: 10.1109/TPWRD.2011.2142198
3. J. A. Jardini, M.V. Tahan, M.R. Gouvea, S. Un Ahn. "Daily load profiles for residential, commercial and industrial low voltage consumers". *IEEE Transactions on Power Delivery*. Vol. 15. 2000. pp. 375-380. Doi: 10.1109/61.847276
4. C. Mihai, I. Lepadat, E. Helerea, D. Călin. "Load curve analysis for an industrial consumer". in *IEEE 12th International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*. Bazov. 2010. pp. 1275-1280. Doi: 10.1109/OPTIM.2010.5510494
5. J. Dickert, P. Schegner. "A time series probabilistic synthetic load curve model for residential customers". in *IEEE PowerTech Conference*. 2011. pp. 1-6. Doi: 10.1109/PTC.2011.6019365
6. P. R. Harvey, B. Stephen, S. Galloway. "Classification of AMI Residential Load Profiles in the Presence of Missing Data". *IEEE Transactions on Smart Grid*. Vol. 7. 2016. pp. 1944-1945. Doi: 10.1109/TSG.2016.2558459
7. F. McLoughlin, A. Duffy, M. Conlon. "A clustering approach to domestic electricity load profile characterisation using smart metering data". *Applied Energy*. Vol. 141. 2015. pp. 190-199. Doi: 10.1016/j.apenergy.2014.12.039
8. J. L. Viegasa, S.M. Vieiraa, R. Melícioa, V.M.F. Mendesb, J.M.C. Sousaa. "Classification of new electricity customers based on surveys and smart metering data". *Energy*. Vol. 107. 2016. pp. 804-817. Doi: 10.1016/j.energy.2016.04.065
9. G. Chicco, R. Napoli, F. Piglion. "Comparisons among clustering techniques for electricity customer classification". *IEEE Transactions on Power Systems*. Vol. 21. 2006. pp. 933-940. Doi: 10.1109/TPWRS.2006.873122
10. D. Jang, J. Eom, M.J. Park, J.J. Rho. "Variability of electricity load patterns and its effect on demand response: A critical peak pricing experiment on Korean commercial and industrial customers". *Energy Policy*. Vol. 88. 2016. pp. 11-26. Doi: 10.1016/j.enpol.2015.09.029
11. P. T. Thanh Binh, N. Hong Ha, T. Cong Tuan, L. Dinh Khoa. "Determination of representative load curve based on Fuzzy K-Means" in *IEEE 4th International Power Engineering and Optimization Conference (PEOCO)*, 2010, pp. 281-286. Doi: 10.1109/PEOCO.2010.5559257
12. S. V. Verdu, M.O. Garcia, C. Senabre, A.G. Marin, F.J.G. Franco "Classification, Filtering, and Identification of Electrical Customer Load Patterns Through the Use of Self-Organizing Maps". *IEEE Transactions on Power Systems*. Vol. 21. 2006. pp. 1672-1682. Doi: 10.1109/TPWRS.2006.881133

13. E. D. Varga, S.F. Beretka, C. Noce, G. Sapienza. “Robust Real-Time Load Profile Encoding and Classification Framework for Efficient Power Systems Operation”. *IEEE Transactions on Power Systems*. Vol. 30. 2015. pp. 1897-1904. Doi: 10.1109/TPWRS.2014.2354552
14. J. Bulckaen, I. Keseru, C. Macharis. “Sustainability versus stakeholder preferences: Searching for synergies in urban and regional mobility measures”. *Research in Transportation Economics*. Vol. 55. 2016. pp. 40-49. Doi: 10.1016/j.retrec.2016.04.009
15. X. Qin, Y. Mo, L. Jing. “Risk perceptions of the life-cycle of green buildings in China”. *Journal of Cleaner Production*. Vol. 126. 2016. pp. 148-158. Doi: 10.1016/j.jclepro.2016.03.103
16. W. Xu, Z. Chen, W. Liu. “A new coefficient of concordance with applications to biosignal analysis” in *MIPPR 2015: Remote Sensing Image Processing, Geographic Information Systems, and Other Applications*, Enshi, China 2015. Doi: 10.1117/12.2210993
17. T. W. MacFarland, J. M. Yates. “Spearman’s Rank-Difference Coefficient of Correlation”. in *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, ed Cham: Springer International Publishing. 2016. pp. 249-297. Doi: 10.1007/978-3-319-30634-6_8
18. M. T. Puth, M. Neuhäuser, G.D. Ruxton. “Effective use of Spearman’s and Kendall’s correlation coefficients for association between two measured traits”. *Animal Behaviour*. Vol. 102. 2015. pp. 77-84. Doi: 10.1016/j.anbehav.2015.01.010
19. O. Urbano, et al., “Enfoque técnico-económico para el dimensionamiento de transformadores de distribución,” *Ingeniería y Desarrollo*, Vol. 34. 2016. pp. 267-285. Doi: 10.14482/inde.33.2.6368