A New Proposal of Typification of Load Profiles to Support the Decision-Making in the Sector of Electric Energy Distribution

A. M.S.Ferreira[1], C. A. Cavalcante[1], C. H. Fontes[1], J. Marambio[2]

[1] Industrial Engineering Program, Polytechnic School – Federal University of Bahia, Professor Aristides Novis street, Federação, Salvador, Bahia Brazil. Zip code: 40210-630
[2] Norsul Engineering LTD. Tancredo Neves Avenue 1632 S 1802, Salvador Trade Center Building, South Tower, Caminho das Árvores Salvador – BA Zip code 41820 020

Email: adoniasmagdiel@ufba.br; arthurtc@ufba.br; cfontes@ufba.br

Abstract

This works presents a method of selection, classification and clustering load curves (SCCL) able to identify a greater diversity of consumption patterns existing in the distribution sector. The method was developed to estimate the features of a sample of load curves aiming to infer the behavior of consumption required by the population of consumers. The algorithm comprises four steps that extract essential features of a load curve of residential users with emphasis on seasonal and temporal profile, among others. The method was successfully implemented and tested in the context of an energy efficiency program developed by a company associated to the sector of electricity distribution (Electric Company of Maranhão, Brazil). This program comprised, among others, the analysis of the impact of replacing refrigerators in a universe of low-income consumers distributed by some cities in the state of Maranhão (Brazil), where it was possible to recognize patterns of load profiles using the typing method developed. The results were compared with a well known method of time series clustering already established in the literature, the Fuzzy C-Means (FCM). Based on the main features of a load profile, the analysis confirmed that the SCCL method was capable to identify a greater diversity of patterns, demonstrating the potential of this method in better characterization of types of demand which represents an important aspect to the process of decision making in the energy distribution sector. Furthermore, a well known index (Silhouette index) was also adopted to quantify the level of uniformity within and between clusters.

Keywords: typing load profiles; clustering; electricity sector.

1 Introduction

A decision-making process in an organization can be analyzed and improved through several methods or strategies. The development of technologies of storage and data mining has given opportunities for the implementation of efficiency strategies of pattern recognition through historical series. In this context, analysis of records of power measurements in substations and customers allows to identify opportunities for improvement of the load factor and energy efficiency of the distribution system, through relationship to the client, contributing also as support for the decision-making (Marambio *et al*, 2003, Cavalcante *et al*, 2008, Cavalcante *et al* 2009, Ferreira *et al*, 2010 and Ferreira *et al*, 2011).

Some works are related to the typification of load curves based on clustering techniques. Bitu and Born (1993) used the non-hierarchical method, through the dynamic clouds technique, which emphasizes the minimization of variance within each group, maximizing the distance between groups. Gerbec *et al* (2002) performed the typification of load curves through the hierarchical clustering method using the *Ward* technique (Hair, 2005), highlighting its convenience in the quantification of groups (and patterns). Gemignani *et al* (2009) combined the hierarchical and non-hierarchical methods to improve the efficiency of the clustering in the recognition of different demand patterns from the same level of tension,.

Paula and Tahan (2006) applied the auto-organizable neural net, or, *Self Organizing Map* (*SOM)*, for the recognition of typical load curves in different classes of consumers. Zalewski (2006) used the *fuzzy* logic for clustering and typification of load curves. The author performed the clustering of load profiles in order

to classify the substations in homogeneous groups according to the peak of consumption. Nizar's, Dong et al (2006) combined two methods, namely, the *Feature Selection* and *Knowledge Discover in Database* (*KDD*) (Han, 2000 and Goldschmidt, 2005) to get better patterns of load demand in a distribution system. Nizar, Dong and Zhang (2008) also combined the *Feature Selection* and *KDD*. Nizar, Dong and Wang (2008) presented the *Extreme Learning Machine (ELM)* method which improves the performance of the clustering methods based on neural nets and incorporates concepts of *KDD* and *Data mining*.

Considering the methods of typification of load curves, already consolidated in the literature, Pessanha (2002) compared the development of some of them and highlighted that the *Fuzzy C-Means* (*FCM*) presents a better level of quality of cohesion and distinction in problems of load curve clustering. Gerbec et al (2004), Zakaria & Hadi (2006), Prahastono et al (2008) e Anuar & Zakaria (2010) used the *FCM* method for the typification of load curves.

This work proposes a new method of selection, classification and clustering load curves (*SCCL*), based on a systematic extraction of features, that is able to identify a greater diversity of demand patterns and also represents a potential tool for the improvement of decision-making process through a better classification of heterogeneous profiles of consume in the electricity sector. The case study analyzed comprised an energetic efficiency program, accomplished by the Electric Company of Maranhão (CEMAR) (Brazil) that considered, among others, the analysis of the impact of replacing refrigerators in low-income consumers. Section 2 presents the *SCCL* method and Section 3 presents the case study and the results achieved through the application of Fuzzy C-Means and *SCCL*, attesting the potential of the latter for the analyzed problem.

2 *SCCL* - A New Method of Typification of Load Curves

The *SCCL* method (Figure 1) consists of two phases and uses consumption data sampled throughout the day. The first phase performs the classification of data (time series) and pattern recognition by successive iterations. The second phase performs the clustering of load curves according to the patterns recognized in the first phase.

The first phase is divided in four stages that apply, each one, one concept (or feature) important for the typification (pattern recognition), in accordance to the requirements and indicators practiced by the electricity sector. The first three stages perform the quantification of similarity between the load curves based on the curve with higher peak demand (reference curve). The last stage performs the quantification of similarity between curves according to a criterion of seasonality.

The four stages inserted into the first phase of typification are described below:

- 1st stage: Typification by probability distribution of hourly demand.

Each curve is normalized in the interval [0; 1] dividing the hourly measurements by the peak demand curve. The consume is quantified in the dimensionless value, named power per unit (*pu*) (Chicco et al, 2002 and Gerbec et al, 2005). The formation of groups is accomplished through the chi-square goodness of fit test that quantify the level of similarity between the distribution of the peaked curve and the distribution of other curves (Hair et al, 2005).

- 2nd stage: Typification by variation of hourly demand.

This step comprises the analysis of correlation between the load curves belonging to the same group (groups obtained in the previous stage) and its reference curve of the group (peaked curve or curve with the highest peak demand). The formation of groups is accomplished through the t-test of correlation coefficient (Gujarati, 2000), obtained for each load curve in relation to the reference curve.

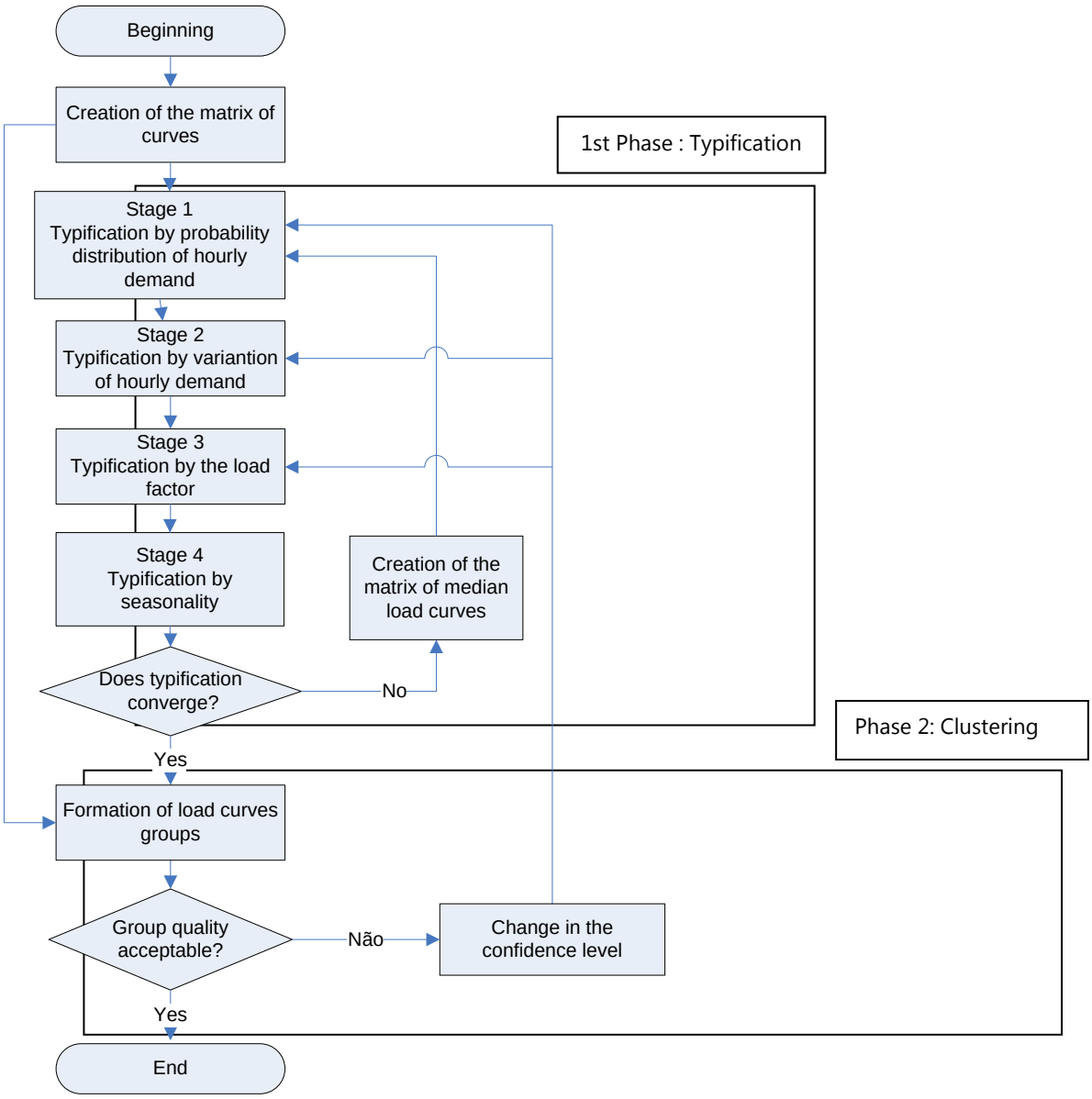

Figure 1: The *SCCL* method.

- 3rd stage: Typification by Load Factor (LF)

This stage comprises the analysis of similarity of LF between the curves of each group obtained at the end of the second stage and the respective reference curve. The formation of groups is accomplished through t-test of differences in mean demands (Gujarati, 2000), obtained for each load curve in relation to the reference curve. The LF is a feature of the load curve which is calculated by dividing the mean demand by the maximum demand. According to Filho (2001), LF is an indicator able to evaluate the rational use of electric energy by the consumer, provided by the concessionaire. In relation to the system of distribution, based on the consumer expectation, the higher the LF, the energy use will be more rational and economic. In relation to the equipment, the higher the LF, more efficient is its operation.

- 4th stage Typification by seasonality at peak hours and off-peak

The curves of each set generated in the third stage are submitted to a process of clustering by seasonal affinity which consists in calculating the median of the energy consumption at half-time edge (9 am to 18

pm) (m_A), at the edge time (18 pm to 21 pm) (m_B) and at the empty time (without loading) (0 am to 9 am and 21 pm to 24 pm) (m_C). Next, the curves are classified according to the Table 1.

Table 1 – Classification of patterns based on seasonal similarity.

Conditions	$m_B \leq m_C$	$m_B > m_C$
$m_A \leq m_B$	Type 1	Type 2
$m_A > m_B$	Type 3	Type 4

The first phase is composed by four successive stages involving different criteria. This phase is repeated several times (iterative process) to check if some prototypes (patterns) can be reassembled, and it's finished when there is convergence in the number of patterns obtained (prototypes). Thus, the number of prototypes is a result of the method itself and an initial estimate is not necessary.

The *SCCL*'s second phase (Figure 1) performs the clustering of the load curves of the initial sample using the smallest Euclidian distance (selection criteria) from the patterns obtained. The groups are submitted to a qualification process that comprises two steps. In the first, prototypes associated to few load curves are eliminated. The second step verifies if changes in the confidence level, used in the three first stages of the first phase, provides better quality of clustering. The metric used to qualify the clustering is the silhouette index (ROUSSEEUW, 1987) which measures the level of cohesion and separation among the groups. The median of the silhouette index of each load curve represents the General Index of Silhouette (GIS).

3 Case Study and Results

The *SCCL* method was applied to analyze possible changes in the consumption patterns in the context of an energy efficiency program implemented by the Electricity Company of Maranhão (CEMAR) (Brazil) and developed during the period November 2008 to July 2009. This program comprised the exchange of five thousand two hundred and fifty (5,250) old refrigerators by new ones in low-income communities. One sample with eighty load curves (old refrigerators), presenting a high consumption of electric energy and another sample of 80 load curves after the exchange of refrigerators. This sample size represents an error level of 11% variation in sample means and a confidence level of 95% in the prediction of the population parameter.

The *SCCL* method was also used to define the number of groups to be considered on the *FCM* method. The confidence level adopted at the 3 first stages of *SCCL*'s first phase were 97% in both cases 1 (before the exchange of refrigerators) and 2 (after the exchange of refrigerators).

The application of *SCCL* method in case 1 was capable to recognize the existence of two groups or demand profiles. The *FCM* method recognized different patterns of demand, with an inferior quality of clustering in relation to those identified by the *SCCL* method. The GIS obtained by *FCM* was 0.31 and by the *SCCL* method was 0.52 (Figure 2).

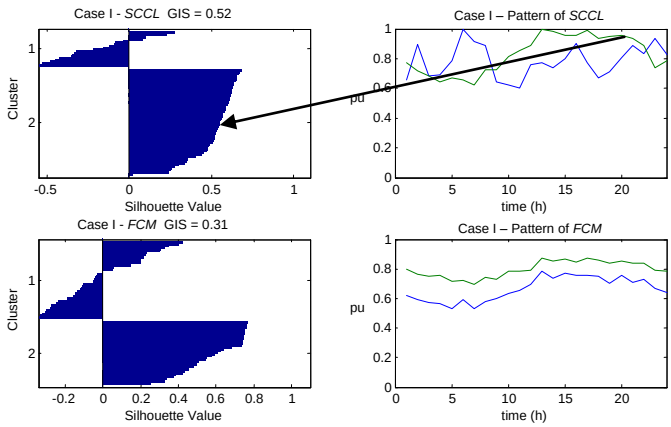

Figure 2: Indices silhouette and patterns recognized by *SCCL* and *FCM* methods (case I).

For the case 2, the *SCCL* method also identified two groups with different consumption patterns. The *FCM* method recognized two patterns with similar profiles of demand what implies that the *FCM* was capable to recognize only one pattern of consumption in the homes after the exchange of refrigerators. The GIS obtained by *FCM* presented a value slightly inferior (0.39) compared to *SCCL* method which was 0.40 (Figure 3).

The value of GIS associated to the *SCCL* method was lower in case 2, attesting that after the exchange of refrigerators the profiles of electric energy became more similar.

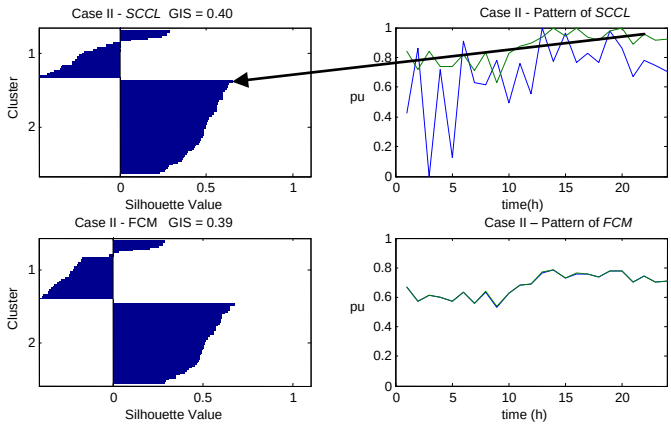

Figure 3: Indices silhouette and patterns by *SCCL* and *FCM* methods (case II)

In both cases the *SCCL* recognized demand patterns representative of more heterogeneous groups among themselves and with a higher level of homogeneity internally. Despite the replacement of refrigerators have contributed to the load curves become more homogeneous, the *SCCL* was still able to recognize different demand patterns, revealing the existence of two distinct types of consumers present in case 2.

An additional analysis was made to verify the homogeneity between load curves of the same group and its adherence, comparing to the pattern recognized. In this situation only the groups associated to the highest number of load curves (generators groups) were considered both cases 1 and 2. The analysis comprised the calculus of shape parameters (mean, variance, asymmetry and kurtosis) and one tendency parameter (slope) for each one of the curves belonging to the group. Next, all these parameters were normalized and compared (using Euclidian Distance) with the same parameters of pattern curve. The Table 2 shows the coefficient of variation of distribution of these dissimilarity measures for each of the methods (*SCCL* and *FCM*). The lower the value of the coefficient of variation more homogeneous is the group. The results show that the *SCCL* got a better pattern of homogeneity than *FCM* in both cases 1 and 2.

Table 2 – Statics of groups generated through the most significant typical curves found by the STAC and *FCM* methods.

STATISTICS	**CASE 1**		**CASE 2**	
	SCCL	*FCM*	*SCCL*	*FCM*
Load factor of the pattern curve (*pu*)	0.83	0.80	0.86	0.68
Percent of curves belonging to generators groups (%)	75	55	66	58
Coefficient of variation (%)	53	71	35	68

According to the Table 2, the *SCCL* attracted a greater number of load curves to the generator group whose the pattern curve had a higher load factor in relation to the *FCM* in both cases 1 and 2. Thus, the *SCCL* shows that the exchange of refrigerators also contributed to a more uniform use of the energy load of the refrigerator. However, the *FCM* shows an opposite situation with load factor reduction after the refrigerators' replacement, what would represent a contradiction, considering that newer devices would contribute for the most efficient use of energy. This result obtained by the *FCM* could imply in the application of investments not required. Therefore, the STAC was able to generate, from the group with the highest number of load curves, typical curves (patterns) most representative and with load factors more consistent with the reality.

According to *SCCL*, the consumption averages associated to the generator group was 80.62 kWh and 57.79 kWh in cases 1 and 2, respectively, showing a reduction of 35.76% in the energy consumption and attesting the success of energy efficiency program.

The patterns recognized by *SCCL* associated to the minority of the refrigerators presented a less uniform profile in both cases 1 and 2. It reveals the existence of consumers with habit of opening the refrigerator more often throughout the day (improper use), reducing the load factor. Even considering the reduction in the energy consumption, this information provides for the efficiency program managers the possibility of more specific and focused actions for improving the consumption habits.

4 Conclusion

This work introduces a new method for the typification of load curves that uses criteria and characteristics inherent to the electricity sector and represents a potential tool for the recognition of patterns of consumption in a given population sample. Among other applications, the method of selection, classification and clustering load curves (*SCCL*) is capable to evaluate the impact of energetic efficiency programs, promoted by the sector, representing, therefore, an important support tool for decision-making at the management level.

The real case analyzed in this work comprised an energetic efficiency program accomplished by the Electricity Company of Maranhão (CEMAR) (Brazil) that analyzed the impact of replacing 5,250 refrigerators in low-income consumers. The results obtained through *SCCL*, compared to a well known method of clustering (Fuzzy C-Means, *FCM*), reveal the viability and potentiality of the first to recognize patterns and generate conclusions consistent with the electric sector, supporting the implementation of efficient management actions, based on the real features of the consumer market.

References

ANUAR, N.; ZAKARIA, Z.. Cluster Validity Analysis for Electricity Load Profiling. IEEE International Conference on Power and Energy, Kuala Lumpur Malaysia, 2010.

BITU, R., BORN, P.Electric Energy Prices: conceptual and methodological aspects. MM Publisher. São Paulo. 1993.

CAVALCANTE, C. A. M. T; SA BARRETTO, S. T.; NASCIMENTO, R. S.; PACHECO, L. A.; FERREIRA, A. M. S.. Development of methodology for real-time parameterization of mathematical models of decision processes. XXVIII National Meeting of Production Engineering (, Rio de Janeiro. ENEGEP), 2008

CAVALCANTE, C. A. M. T.; SA BARRETTO, S. T.; PACHECO, L. A. ; NASCIMENTO, R. S. ; FERREIRA, A. M. S. ; LEITE, W. S . Adaptive Models for Decision Support in Production Management Unit Thermoelectric. In: Rio Oil & Gas Expo and Conference, 2009, Rio de Janeiro. Rio Oil & Gas Expo and Conference, 2009.

CORMACK, R. A review of classification. Journal of the Royal Statistical Society Series A), v.134, p.321-367, 1971.

FERREIRA, A. M. S.; CAVALCANTE, C.A.; FONTES, C.H.; MARAMBIO, J. E. S. A New Classification method of Electric Energy Hourly Demand. XXXI National Meeting of Production Engineering (ENEGEP), 2011, Minas Gerais.

FERREIRA, A. M. S.; CAVALCANTE, C.A.; FONTES, C.H.; MARAMBIO, J. E. S. The Brazilian Electricity Sector. The Polytechnic Institute of Bahia Magazine. Year 3 n.7E.Quarterly Issue, Jul/Sep-2010, 18-24.

GEMIGNANI, M.M.F.; OLIVEIRA, C.C.B.; TAHAN, C.M.V. Proposition and Comparative Analysis of Alternative Selection and Classification of Load Curve for Defining Types for Higher Fare. Thirteenth Ibero-American Regional Meeting of Cigré– XIII ERIAC, 2009.

GERGBEC, D.,; GASPERIC, S.; SMON, I.; GUBINA, F. a Methodology to Classify distribution Load Profiles. Presented at the IEEE, 2002.

GERBEC, D.; S. GASPERIC, SMON, I.; GUBINA, F. Determining the load profiles of consumers based on fuzzy logic and probability neural networks. IEE Proc.-Gener. Transm. Distrib., Vol. 151, No. 3, May 2004.

GERBEC, D.; GASPERIC, S.; SMON, I.; GUBINA, F.. Allocation of the Load Profiles to Customers Using Probabilistic Neural Networks. IEEE Transactions On Power Systems, Vol. 20, No. 2, May 2005.

GUJARATI, D. N. Basic Econometrics, Makron Books, 2000.

HAIR, J. F., ANDERSON, R. E., TATHAM, R. L., BLACK, W. C. Multivariate Analysis of Data. 5. ed. Porto Alegre: Bookman, 2005.

JACCARD. P. The distribution of flora in the alpine zone. 1912. New Phytologist. 11, 37–50.

FILHO, J. MAMEDE .Project elements In:_____Industrial Electric Installations. 6. ed. Rio de Janeiro: LTC, 2001. p. 1-33.

HADDAD, JAMIL. Efficient use of energy: Recent regulatory incentives to the current energy efficiency law. The Brazilian Journal of Energy. Itajubá SBPE, v. 9, n. 1, 1st wk. 2002. 202 p.

MARAMBIO, J. E. S.; LUZ, A. D.; FERREIRA, A. M. S. ; CHAGAS, E. H. C.; C.MUCCINI, M. J.; SOARES JR, F. A. ; SANTOS, S. O. . Methodology for Planning and Accompaniment of GLD Programs in Non-Traditional Growing Market. In: II Congress of Technologic Innovation in Electric Energy. 2003, Salvador.

NIZAR, A. H., DONG, Z. Y.; WANG, Y., "Power Utility Nontechnical Loss Analysis With Extreme Learning Machine Model" IEEE Trans. on Power Systems, Vol. 23, No. 3, pp. 946-955, August 2008.

NIZAR, A. H.; DONG, Z. Y.; ZHAO, J. H., "Load profiling and data mining techniques in electricity deregulated market," presented at the IEEE Power Engineering Society (PES) General Meeting 2006, Montreal, Quebec, Canada, June 2006.

NIZAR, ANISAH HANIM; ZHAO, JUN HUA; DONG, ZHAO YANG. Customer Information System Data Pre-Processing with Feature Selection Techniques for Non-Technical Losses Prediction in an Electricity Market. International Conference on Power System Technology, 2006.

PAULA, G. M.F.; TAHAN, C.M.V. Typical Load Curves for the Operational Planning of the Distribution System. International Congress of Electric Distribution. – CIDEL Argentina, 2006.

PESSANHA, J. F. M.; Velásquez, R. M. C.; Melo, A. C. G.; et al. Cluster ANALYSIS Techniques in the Building of Load Curves Classification. XV NATIONAL SEMINARY OF ELECTRIC DISTRIBUTION (November, 2002, Salvador, Bahia). Anais. Bahia, 2002

Prahastono, Iswan; King, David J.; Ozveren, C.S.; Bradley, D. ELECTRICITY LOAD PROFILE CLASSIFICATION USING FUZZY C-MEANS METHOD .Proceedings of the Universities Power Engineering Conference, Proceedings of the 43rd International Universities Power Engineering Conference, UPEC 2008

ROUSSEEUW.P.J. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. 1987. Journal of Computational and Applied Mathematics. 20. 53-65.

ZAKARIA, ZUHAINA.; Lo, K L,;, SOHOD, HADI MOHAMAD. Application of Fuzzy Clustering to Determine Electricity Consumers' Load Profiles. First International Power and Energy Conference PECon 2006 99 November 28-29, 2006, Putrajaya, Malaysia

ZALEWSKI, W. Application of Fuzzy Inference to Electric Load Clustering. Presented at the IEEE, 2006.