Improved Fuzzy Load Models by Clustering Techniques in Distribution Network Control

Gheorghe Grigoras, Ph.D.[*1] and Gheorghe Cartina, Ph.D. [1]
[1]"Gheorghe Asachi" Technical University, Iasi, Romania
[*]ghgrigoras@yahoo.com

Abstract: In operation and planning of the power systems, the analysis of the consumption trends depends on the evolution of economic activities and competition among several sources of energy, which affect forecasts. The loads estimation represents the basis of the system state estimation and influences various aspects of power system planning such as: transformer and conductor sizing, capacitor bank placement and so on. The main difficulties in modeling of the nodal loads result from the random nature of loads, the deficiency of measured data and the fragmentary and uncertain character of information on loads and customers. Thus, a modern method for expressing the uncertainty in load models is fuzzy technique. A fuzzy set is a set containing elements that have varying degrees of membership in the set. There are different ways to derive membership functions. Subjective judgment, intuition and expert knowledge are commonly used in constructing membership function. Because in many situations the choices of the membership functions are subjective, in the paper the clustering techniques are proposed for the improved of the defining of membership functions corresponding to the load profiles and customers consumption categories. Results obtained demonstrate the ability of the fuzzy load models to overcome difficult aspects encountered in process control and operation problems.

Keywords: clustering techniques, fuzzy load models, distribution networks, load profiles.

1. Introduction

The electric load in distribution system varies with time and place. Therefore electric companies need by the accurate load data of the supply customers for distribution network planning and operation, load management, customer service and billing. There are several factors that influencing the customer's load [1] - [3]:

- customer factor: type of consumption, type of electric heating, size of building etc;
- time factor: time of day, day of week, time of year;
- climate factor: temperature, humidity etc;
- other electric loads correlated to the target load;
- previous load values;
- load curve patterns and so on.

For an electric customer, the behavior is represented by a load profile figuring the electric power consumption for every period of time. Availability of such data depends on the type of customer. Generally, the small customers (like residential ones) are poorly described since a communicating meter is too expensive regarding to their consumption: for these customers there are only a few points of the curve every year. For larger customers, a communicating meter is often available for many reasons: the billing is done every month, the consumption is high and justifies the communicating meter investment, a detailed record of consumption is necessary because prices depend on the period. The load profiles/curves may correspond to individual customer curves or to aggregates over an electric substation.

2. Clustering techniques

Cluster analysis is the organization of a collection of objects (usually represented as a vector of

measurements) into clusters based on similarity. It is a wonderful exploratory technique to help us understand the clumping structure of the data.

Clustering is useful in several exploratory pattern-analysis, grouping, decision-making, and machine-learning situations, including data mining, documents retrieval, images segmentation, and patterns (objects) classification [1] - [3], [6], [7], [8].

A pattern (object) (or feature vector, observation, or datum) x is a single data item used by the clustering algorithm. It consists, typically, of a vector of d measurements: $x = (x_1, x_2, \ldots x_d)$. The individual scalar components x_i of a pattern x are called features (or attributes). A distance measure (a specialization of a proximity measure) is a metric (or quasi-metric) on the feature space, used to quantify the similarity of patterns.

There are two major methods of clustering: hierarchical clustering and k-means clustering.

Hierarchical clustering is subdivided into *agglomerative* methods, which proceed by series of fusions of the n objects into groups, and *divisive* methods, which separate n objects successively into finer groupings. Agglomerative techniques are more commonly used. Hierarchical clustering may be represented by a two dimensional diagram, known as dendogram, which illustrates the fusions or divisions made at each successive stage of analysis. Hierarchical clustering is appropriate for small tables, up to several hundred rows. An example of such a dendogram is given below in Figure 1.

Differences between agglomerative methods arise because of the different ways of defining distance (or similarity) between clusters. Several agglomerative techniques will now be described in the following [4], [6] - [8].
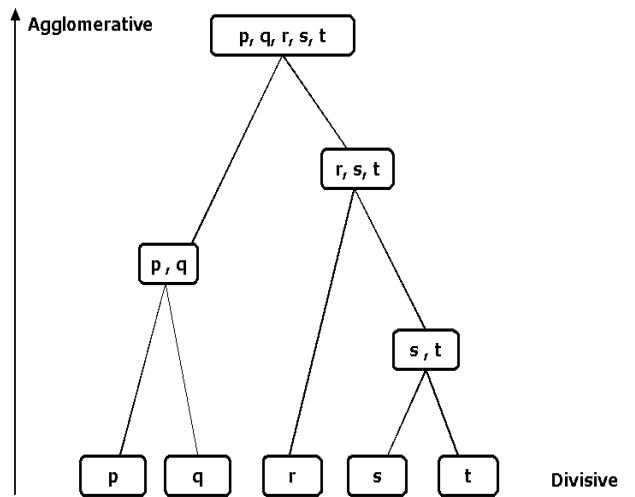

Figure 1. Example of dendogram

Single linkage clustering (*connectedness or minimum method*). The defining feature of the method is that distance between groups is defined as the distance between the closest pair of objects, where only pairs consisting of one object from each group are considered:

$$D(r,s) = \min \{d(i,j) : \text{where object } i \text{ is in cluster } r \text{ and object } j \text{ is in cluster } s\}$$

Complete linkage clustering (*diameter or maximum method*). Distance between groups is now defined as the distance between the most distant pair of objects, one from each group:

$$D(r,s) = \max \{d(i,j) : \text{where object } i \text{ is in cluster } r \text{ and object } j \text{ is in cluster } s\}$$

Average linkage clustering. The distance between two clusters is defined as the average of distances between all pairs of objects, where each pair is made up of one object from each group:

$$D(r,s) = \frac{T_{rs}}{N_r \cdot N_s} \tag{1}$$

where: T_{rs} - the sum of all pair wise distances between cluster r and cluster s;

$\qquad$ N_r, N_s - the sizes of the clusters r and s.

C*entroid Method*. In the centroid method the distance between two clusters is defined as the squared Euclidean distance between their means. The centroid method is more robust to outliers than most other hierarchical methods.

$$D(r,s) = \left\| \overline{X}_r - \overline{X}_s \right\|^2 \tag{2}$$

K-means clustering. K-means is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The main idea is to define k centroids, one for each cluster so minimize an objective function, in this case a squared error function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{3}$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j.

3. Membership functions

Mathematical models and algorithms in electric power system theory aim to be as close to reality as possible. Modeling can be performed in numerous ways, inclusively using the Fuzzy Techniques (FT). The basic idea of FT is to model and to be able to calculate with uncertainty. Uncertainty in fuzzy logic is a measure of nonspecifically that is characterized by possibility distributions. Linguistic terms used in our daily conversation can be easily captured by fuzzy sets for computer implementations. A fuzzy set is a set containing elements that have varying degrees of membership in the set. The membership values of each function are normalized between 0 and 1. There are different ways to derive membership functions. Subjective judgment, intuition and expert knowledge are commonly used in constructing membership function. Even though the choices of membership function are subjective, there are some rules for membership function selection that can produce well the results [4].

A fuzzy number A can have different forms but, generally, this is represented as trapezoidal or triangular form, Figure 2, usually represented by its breaking points:

$$A \Leftrightarrow (x_1, x_2, x_3) = [m, a, b] \tag{4}$$

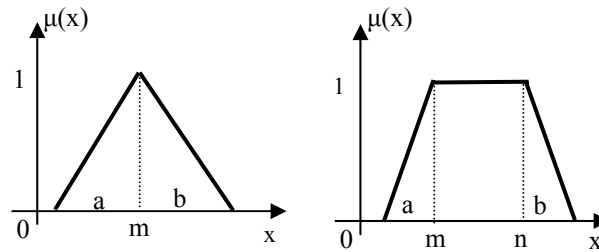$$A \Leftrightarrow (x_1, x_2, x_3, x_4) = [m, n, a, b] \tag{5}$$

Figure 2. Triangular and trapezoidal membership functions

Using of the clustering techniques, the steps for defining some trapezoidal membership function in the case of a set of two dimensional objects, are presented in Figures 3, 4, and 5. This is an improved fuzzy models method by clustering techniques because the breaking points are calculated with statistical characteristics [3] - [5]:

$$\begin{cases} x_1 = m - k_1\sigma; \quad x_2 = m - k_2\sigma \\ x_3 = m + k_3\sigma; \quad x_4 = m + k_4\sigma \end{cases} \tag{6}$$

where: m - average value, σ - standard deviation value, and k_i ($i = 1, \dots, 4$) - coefficients whose value is determined by experts.
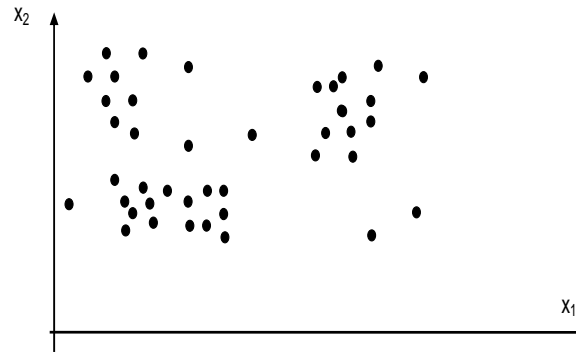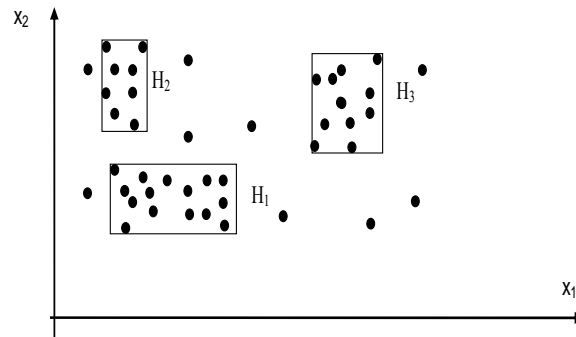
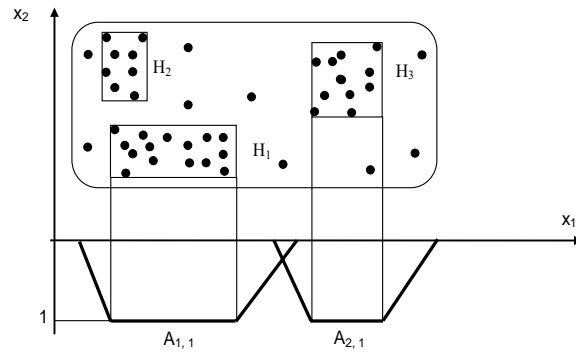Figure 3. Example of ungrouped objects

Figure 4. Grouping of the objects

Figure 5. Defining of the membership function

4. Fuzzy modeling of load profiles

For an electric customer, the behavior is represented by a load profile figuring the electric power consumption for every period of time. Availability of such data depends on the type of customer. Generally, the small customers (like residential ones) are poorly described since a communicating

meter is too expensive regarding to their consumption: for these customers there are only a few points of the curve every year. For larger customers, a communicating meter is often available for many reasons: the billing is done every month, the consumption is high and justifies the communicating meter investment, a detailed record of consumption is necessary because prices depend on the period. Analyzed load curves may correspond to individual customer curves or to aggregates over an electric substation [1].

A. Load modeling

In distribution networks, except the usual measurements from substations, there is few information about the network state. The feeders and the loads are not usually monitored. As a result, there is a high degree of uncertainty about the power demand and, consequently, about the network loading, voltage level and power losses. Therefore, the fuzzy approach may reflect better the real behavior of a distribution network under various loading conditions [4].

Thus, the hourly loading factor of a particular distribution transformer can be employed to approximate the nodal load. And, because the most utilities have not historical records of feeders, it is proposed to use linguistic terms, usually used by dispatchers, to describe the uncertain hourly loading factor. For modeling of the loads from distribution substations, two primary fuzzy variables can be considered: the loading factor kI (%) and power factor $cos\varphi$, so that the fuzzy representation of the active and reactive powers result from relations [4]:

$$P = \frac{kI}{100} \cdot S_n \cdot cos\,\varphi\,; \quad Q = P \cdot tan\,\varphi \quad\quad (7)$$

where S_n (kVA) is the nominal power of the transformer from the distribution substations.

The fuzzy variables, kI and $cos\varphi$, are associated to trapezoidal membership functions, (4), (5), Figure 2. The two fuzzy variables must be correlated, as indicated in Table 1.

Table 1. Linguistic categories of the kI and $cos\varphi$

Linguistic Category		x		Linguistic Category		x	
		kI (%)	cosφ			kI (%)	cosφ
VS (Very Small)	x_1	10	0.75	M (Medium)	x_3	55	0.87
	x_2	10	0.77		x_4	65	0.89
	x_3	15	0.79	H (High)	x_1	55	0.87
	x_4	25	0.81		x_2	65	0.89
S (Small)	x_1	15	0.79		x_3	75	0.91
	x_2	25	0.81		x_4	85	0.93
	x_3	35	0.83	VH (Very High)	x_1	75	0.91
	x_4	45	0.85		x_2	85	0.93
M (Medium)	x_1	35	0.83		x_3	95	0.95
	x_2	45	0.85		x_4	95	0.97

These linguistic terms are defined in function by the loading of the transformers at the peak load. Each loading level represented by a linguistic variable is described by a fuzzy variable and its associated membership function. The loading factor kI and the power factor $cos\varphi$ were divided into five linguistic categories with the trapezoidal membership function, Table 1.

B. Load profile modeling

In this section, an approach to daily load profile determination for the nodes of an electric distribution network (20 kV) is presented. For this purpose the use of hierarchic clustering method conjunctively with fuzzy models is applied to a distribution network, containing the data for 39 nodes, to classify load profiles of the 20 kV nodes into groups, representing typical load profiles.

Active power profiles corresponding to the considered nodes were normalized relatively to the proper energy consumption (during the day when the load peak was recording) using the following relation was used:

$$p_i^h = \frac{P_i^h}{W_i}; \quad h=1,..., 24; \quad i=1,..., N_R \tag{8}$$

where: P_i^h - the active power (kW), demanded by the i node at h hour, W_i - the active energy [kWh], consumed by the i node during the day when the peak load was recorded, and N_R = the total number nodes that were taken into consideration in the clustering process, for the active power.

Thus, four groups were determined for the active power. For every obtained group, C_{P1}, C_{P2}, C_{P3} and C_{P4}, the average and the variance values (m_{Pj}^h and σ_{Pj}^h, $j=1, 2, 3, 4$), during the peak load day of the distribution system, were calculated using relations:

$$m_{Pj}^h = \frac{\sum_{i=1}^{N_{C_{Pj}}} p_i^h}{N_{C_{Pj}}}; h=1,...,24; \quad j=1,..., N_G \tag{9}$$

$$\sigma_{Pj}^h = \sqrt{\frac{\sum_{i=1}^{N_{C_{Pj}}} (p_i^h - m_{Pj}^h)^2}{N_{C_{Pj}}}} \tag{10}$$

where: N_G - the number of the resulted groups from the clustering process and $N_{C_{Pj}}$ - the total number nodes from every group.

The signification of the coefficients m_{Pj}^h is following: these coefficients transform the energy consumed by the medium member of the group in average active power demanded by it. These coefficients lead us to the typical load profiles corresponding to the active power, corresponding to the 20 kV nodes, Figure 6.
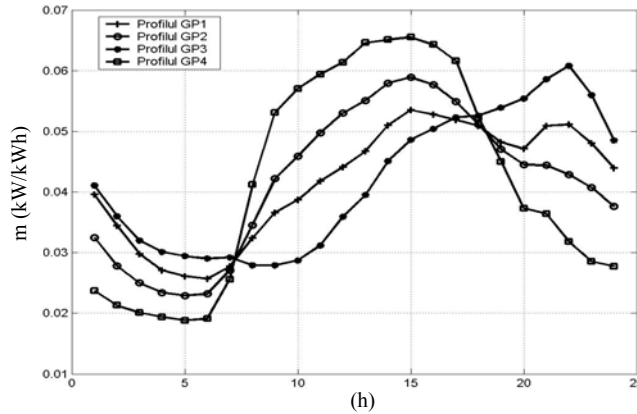

Figure 6. Typical Load Profiles for the obtained groups

Using these load profiles corresponding to the active power and the factor cosφ, the loading factor of the every transformer from the distribution substations can be determined. Also, using the statistical model (6) and values of the coefficients k_i ($i = 1,...,4$) from the a fuzzy load model is determined, Table 2 and Table 3.

Table 2. Values of coefficients k_{1-4} from the model (6) for each group C_{P1-P4}

k	C_{P1}	C_{P2}	C_{P3}	C_{P4}
k_1	1.10	1.12	1.11	1.12
k_2	1.00	1.00	1.00	1.00
k_3	1.00	1.00	1.00	1.00
k_4	1.10	1.12	1.11	1.12

Table 3. Fuzzy model for typical load profile C_{P1}

h	m (kW/kWh)	σ	x_1 (kW/kWh)	x_2	x_3	x_4	h	m (kW/kWh)	σ	x_1 (kW/kWh)	x_2	x_3	x_4
1	0,0396	0,0044	0,0347	0,0352	0,0440	0,0445	13	0,0467	0,0023	0,0442	0,0444	0,0490	0,0492
2	0,0344	0,0030	0,0311	0,0314	0,0374	0,0377	14	0,0510	0,0027	0,0480	0,0483	0,0537	0,0539
3	0,0298	0,0025	0,0271	0,0273	0,0323	0,0326	15	0,0535	0,0035	0,0497	0,0500	0,0569	0,0573
4	0,0271	0,0027	0,0241	0,0244	0,0298	0,0301	16	0,0528	0,0039	0,0485	0,0489	0,0566	0,0570
5	0,0261	0,0028	0,0230	0,0233	0,0290	0,0292	17	0,0519	0,0030	0,0486	0,0489	0,0550	0,0553
6	0,0257	0,0027	0,0227	0,0230	0,0285	0,0287	18	0,0509	0,0028	0,0478	0,0481	0,0537	0,0540
7	0,0277	0,0024	0,0251	0,0253	0,0301	0,0304	19	0,0482	0,0046	0,0432	0,0437	0,0528	0,0533
8	0,0324	0,0024	0,0298	0,0301	0,0348	0,0350	20	0,0471	0,0047	0,0420	0,0424	0,0518	0,0522
9	0,0366	0,0041	0,0321	0,0325	0,0406	0,0410	21	0,0509	0,0048	0,0456	0,0461	0,0557	0,0561
10	0,0387	0,0037	0,0346	0,0350	0,0424	0,0427	22	0,0511	0,0033	0,0474	0,0477	0,0544	0,0548
11	0,0418	0,0032	0,0383	0,0386	0,0451	0,0454	23	0,0480	0,0053	0,0422	0,0427	0,0533	0,0539
12	0,0441	0,0037	0,0400	0,0404	0,0478	0,0481	24	0,0439	0,0042	0,0393	0,0397	0,0481	0,0485

C. Determination of the Consumption Categories

In this paragraph it is presented an algorithm for identification of the consumption categories of the consumers and grouping them into the classes with the similar characteristics.

The customer classes must be obtained from historical data and must be updated to follow the changes on consumer's behavior. Due to the large amount of data predicted to be available in the future and the necessity of easy actualization, the algorithm provides a clear separation of different steps based on the application of clustering techniques. The algorithm is based on the load profiling process. The major steps identified are:

1. Load Research: In this step a representative sample of the set of consumers is identified, the most relevant attributes to be measured, the cadence for data collection is defined. Finally the collected data is gathered in a large database.

2. Data cleaning and pre-processing: In real problems, like this, involving a large number of measurements, spread over a large geographic area, collecting data during a considerable period of time different kind of problems will affect the quality of the database. The most relevant and frequent are communication problems, outages, failure of equipment and irregular atypical behavior of some consumers. The result will be a very large database with problems like noise, missing values and outliers. This data (after being cleaned, pre-processed and reduced) is used to obtain the division of the initial data set in classes.

3. Partitioned into macro-categories: The whole customer database is preliminary partitioned into consumption categories defined by the activity type: residential, commercial and industrial.

4. Classification: Inside each consumption category a classification into 3 – 5 classes, in function by the daily energy consumption of the consumers is done. For realization of this classification, the k-means method is used. For every consumer from the class is determined the normalized load profile using a suitable normalizing factor (energy consumption by the period. Then, using a hierarchical clustering method, the normalized load profiles are refined so as to desist at the unrepresentative profiles. The typical load profile for each class is obtained by averaging the values for each hour.

5. Assignation: Finally, to the each customer class is assigned a typical load profile according to their consumption category.

In our study, we have considered a database described by 296 load curves of various consumers. Each measurement is a curve of 24 hourly points describing the behavior of a consumer during a day. The whole customer database was preliminary partitioned into consumption categories defined by the activity type: residential (147 consumers), commercial (97 consumers) and industrial (52 consumers).

In the next step the load profiles corresponding to the consumers from each category were normalized relatively to the daily energy consumption. For this purpose, the following relation was used:

$$p_{ij}^h = \frac{P_{ij}^h}{W_{ij}}; \quad h = 1,...,24; \quad i = 1,...,N_{Cj} \tag{11}$$

where: P_{ij}^h - the active power of the customer i from the class j, at the hour h, (kW), W_{ij} - the daily energy consumption of the customer from the class j, (kWh), and N_{Cj} - the total number of the customers from the class j that were taken into consideration in the clustering process.

Then, using the average linkage method from the hierarchical clustering methods [2], [3], the normalized load profiles of the customers are refined so as to desist at the unrepresentative profiles. Thus, using this refinement, the number of the load profiles of the customers from each class is less.

The average values corresponding for each consumer class were calculated using the relations:

$$m_{Pj}^h = \frac{\sum_{i=1}^{N_{ij}} p_{ij}^h}{N_{Cj}}; \quad h = 1,...,24; \quad i = 1,...,N_{Cj} \tag{12}$$
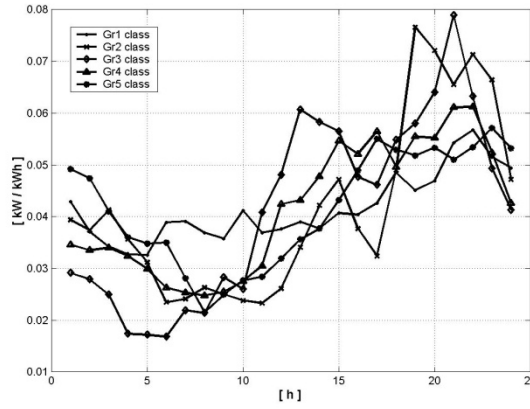

Figure 7. TLPs of the consumption classes from
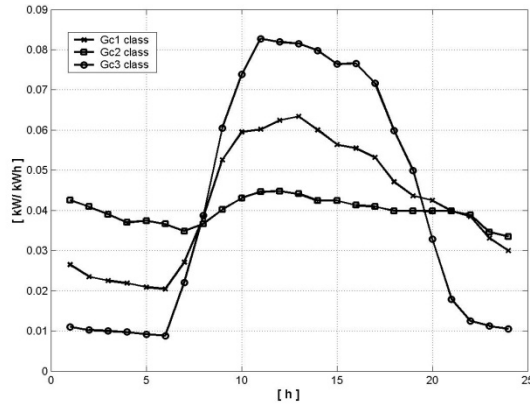the residential consumer's category

Figure 8. TLPs of the consumption classes from
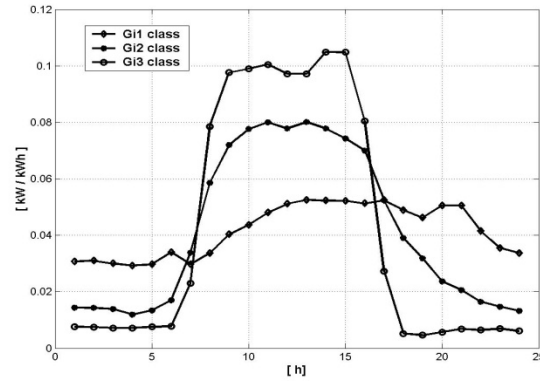the commercial consumer's category

Figure 9. Typical load profiles of the consumption
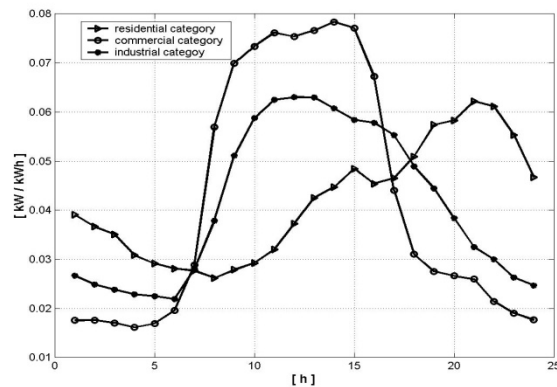classes from the industrial consumer's category

Figure 10. TLPs of the consumer categories
(residential, commercial and industrial)

These coefficients transform the energy consumed by the medium member (customer) of the class in average active power demanded by it. The typical load profile (TLP) for each class from the consumer categories is obtained by representation of these coefficients, Figures 7 – 10. For the residential consumer's category it obtained 5 consumption classes, the TLPs of these are presented in Figure 7. In the case of the commercial consumer's category, 3 consumption classes were obtained. In Figure 8, the TLPs of these classes are shown. Finally, three consumption classes were obtained for industrial consumer's category, Figure 9. In Figure 10, the TLPs corresponding to the case in which customers categories are not grouping in classes are represented.

From the comparison of the typical load profiles it results that a classification of the customer's categories (residential, commercial and industrial) in classes is useful in view of the building of the tariff structures, demand-side management, optimal operation and planning of distribution system and so on.

5. Conclusions

In this paper an algorithm, based on the clustering techniques, is proposed for determination of the consumption profiles and categories of the customers. Clustering techniques are extremely useful for assisting the distribution services providers in the process of electric customer

classification on the basis of load profile. The results obtained on a database of residential customers demonstrate that the methodology can be used with the success in building of the tariff structures for the customers or in the optimal operation and planning of distribution system.

6. References

[1] S.Gasperic, D. Gerbek, and F.Gubina, "Determination of the Consumers' Load Profiles". Available online: www.telmark.org /2002Sep/ 2-5_Gasperic.pdf.

[2] G. Chicco, , R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader, "A Review of Concepts and Techniques for Emergent Customer Categorization". Available online: www.telmark.org/ 2002Sep/ 2-4_Chicco.pdf.

[3] G. Cartina, G. Grigoras, E. C Bobric, and C. Lupaşcu, "Improving of Fuzzy Models by Clustering Techniques in Optimal Reconfiguration of the Distribution Networks", *International World Energy System*, Torino, Italy, 2006.

[4] G. Cartina, G. Grigoras, and E. C. Bobric. Clustering Techniques in Fuzzy Modeling. Power Systems Applications, Iasi, Romania: Venus Publishing House, 2005.

[5] G. Cartina, G. Grigoras, and V. Alexandrescu, "Pilot pattern choosing by clustering and fuzzy techniques", *Proc. of International Conference Energy-Environment, Bucharest, Romania*, 2005, pp. 1.13 – 1.17.

[6] "Hierarchical Clustering", Available online: http://www.resample.com/xlminer/help/HClst.

[7] A.K. Jain, M.N Murty, and P.J. Flynn. "Data Clustering: A Review". Available online: www.cermics.enpc.fr/~keriven/vision/ articles.

[8] JMP Statistics and Graphics Guide. Version 3, SAS Institute Inc., Cary, NC, USA, 1995.

Gheorghe Grigoras was born in Vatra Dornei, Romania, on February 28, 1976. He received the M. SC. and Ph. D. degrees in Electrical Engineering from "Gheorghe Asachi" Technical University of Iasi, Romania, in 2000 and 2005 respectively. He is currently lecturer in the Department of Power Systems of Electrical Engineering Faculty at the same university. His main areas of interest are analysis, planning, and optimization of power systems.

Gheorghe Cartina received the M. SC. and Ph. D. degrees in Electrical Engineering from "Gheorghe Asachi" Technical University of Iasi, Romania, in 1964 and 1972 respectively. He is currently professor in the Department of Power Systems of Electrical Engineering Faculty at the same university. His research interests include especially to monitoring and optimal control of power systems.