Paper accepted for presentation at 2009 IEEE Bucharest Power Tech Conference, June 28th - July 2nd, Bucharest, Romania

Improved Fuzzy Load Models by Clustering Techniques in Optimal Planning of Distribution Networks

G. Cartina, *Member, IEEE*, G. Grigoras, E.C. Bobric, *Member, IEEE*, and D. Comanescu

***Abstract*--The notion of modeling is essential to modern techniques of control and operation process. Developing a control process in fact means developing a model that allows us to predict the action and reduce the amount of feedback required. Recently the fuzzy modeling has become driving force that today is reflected in many different software and hardware products. The paper presents improvements of the fuzzy load models by clustering techniques in distribution networks planning. The hierarchic clustering techniques, conjunctively with fuzzy modeling, are proposed in this paper for determination of the typical load profiles, customers' categories, and so on. Obtained results demonstrate the ability of the fuzzy load models to overcome difficult aspects encountered in process control and operation problems.**

***Index Terms*--Clustering techniques, distribution networks, fuzzy load models, load profiles, optimal planning.**

I. Introduction

THE electric load in distribution system varies with time and place. Therefore electric companies need accurate load data of customers for distribution network planning and operation, load management, customer service and billing. There are several factors that influence the customers' load [1], [2], [5]:

- customer factor: type of consumption, type of electric heating, size of building;
- time factor: time of day, day of week, time of year;
- climate factor: temperature, humidity;
- 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 reflecting the electric power consumption for every period of time. Availability of such data depends on the type of customer. Generally, small customers (like residential ones) are poorly described since a communicating meter is too expensive with respect 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 investments, and a detailed record of consumption is necessary because prices depend on the period.

G. Cartina is with the Department of Power Systems, "Gh. Asachi" Technical University, Iasi, Romania (e-mail:gcartina@ee.tuaiasi.ro).

G. Grigoras is with the Department of Power Systems, "Gh. Asachi" Technical University, Iasi, Romania (e-mail:ggrigor@ee.tuaiasi.ro).

E.C. Bobric is with the Department of Electrical Engineering, "Stefan cel Mare" University, Suceava, Romania (e-mail:crengutab@eed.usv.ro).

D. Comanescu is with the Hidroelectrica S.A., Piatra Neamt, Romania (e-mail: daniela.comanescu@hidroelectrica.ro).

978-1-4244-2235-7/09/$25.00 ©2009 IEEE

II. 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, document retrieval, images segmentation, and patterns (objects) classification, [1], [3], [6]-[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 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, Fig. 1.

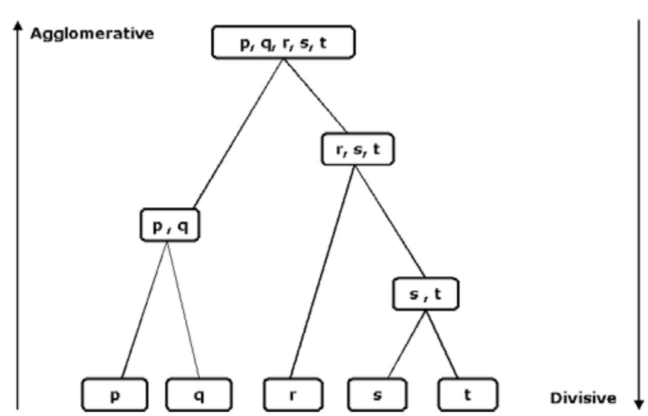

Fig. 1. Example of dendogram

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.

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

Single linkage clustering (connectedness or minimum method). The defining feature of the method is the distance between groups:

$$D(r,s) = \min\{d(i,j)\} \tag{1}$$

where object i is in cluster r and object j is in cluster s.

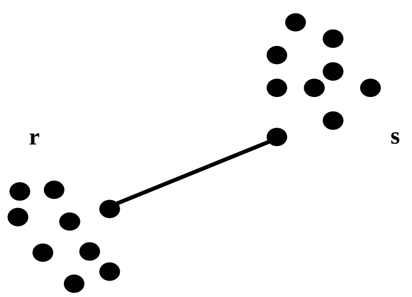

Fig. 2. Single linkage clustering

Complete linkage clustering (diameter or maximum method). The distance between groups is now defined as the distance between the most distant pair of objects, one from each group, Fig. 3:

$$D(r,s) = \max\{d(i,j)\} \tag{2}$$

where object i is in cluster r and object j is in cluster s.

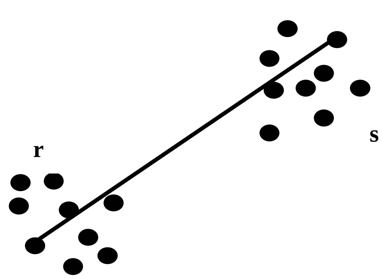

Fig. 3. Complete linkage clustering

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, Fig. 4:

$$D(r,s) = \frac{T_{rs}}{N_r \cdot N_s} \tag{3}$$

where T_{rs} is the sum of all pair wise distances between cluster r and cluster s, N_r and N_s are the sizes of the clusters r and s respectively.

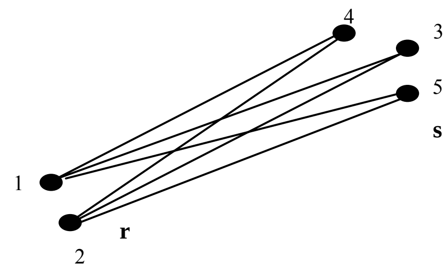

Fig. 4. Average linkage clustering

Centroid 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.

$$D_{KL} = \left\| \overline{X}_K - \overline{X}_L \right\|^2 \tag{4}$$

At each stage of hierarchical clustering, the clusters r and s, for which $D(r, s)$ is the minimum, are merged, Figs. 2-4.

K-*means clustering*. The 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 as to minimize an objective function, in this case a squared error function:

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2 \tag{5}$$

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, [2], [6]-[8].

III. 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, [2].

The uncertainty of the load level, reliability indices, the length of the feeders and so on will be represented as fuzzy numbers, with membership functions over the real domain $\Re$. *A* fuzzy number $\tilde{A}$ can have different forms but, generally, this is represented in triangular or trapezoidal form, usually represented by its breaking points:

$$\widetilde{A} \Leftrightarrow (x_1, x_2, x_3) = [m, a, b] \quad (6)$$

$$\widetilde{A} \Leftrightarrow (x_1, x_2, x_3, x_4) = [m, n, a, b] \quad (7)$$

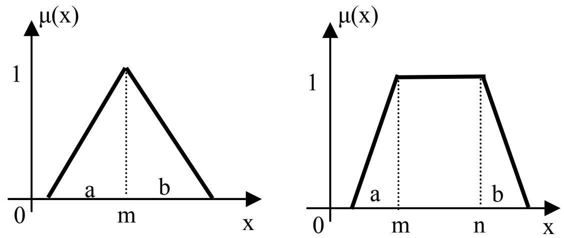

Fig. 5. Triangular and trapezoidal membership function

Using clustering techniques, the steps for defining some trapezoidal membership function in the case of a set of two dimensional objects, are presented in Fig. 6, Fig. 7, and Fig. 8. This is an improved fuzzy model by clustering techniques because the breaking points are calculated with statistical characteristics, (7), [1]-[3]:

$$\begin{cases} x_1 = m - k_1\sigma; \quad x_2 = m - \sigma \\ x_3 = n + \sigma; \quad x_4 = n + k_4\sigma \end{cases} \quad (8)$$

where m is average value and σ is standard deviation value.

Fig. 6. Example of ungrouped objects

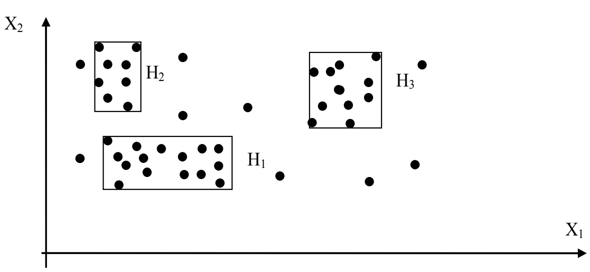

Fig. 7. Grouping of the objects

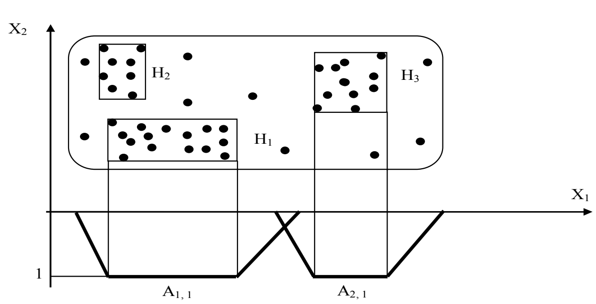

Fig. 8. Defining of the membership function

IV. Load Modeling of the Distribution Transformers

In distribution networks, except for the usual measurements from substations, there is little 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, [2].

Thus, the hourly loading factor of a particular distribution transformer can be employed to approximate the transformer load. And, because most utilities have no historical records of feeders, it is proposed to use linguistic terms, usually used by dispatchers, to describe the uncertain hourly loading factor.

Two primary fuzzy variables can be considered for modeling of the loads from distribution substations: the loading factor *kI* (%) and power factor $cos\varphi$, so that the fuzzy representation of the active and reactive powers result from relations, [2]:

$$P = \frac{kI}{100} \cdot S_n \cdot \cos\varphi; \quad Q = P \cdot \tan\varphi \quad (9)$$

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

The fuzzy variables, *kI* and $cos\varphi$, are associated to trapezoidal membership functions, (7), Fig. 5. The two fuzzy variables must be correlated, just like that fuzzy variables *P* and $cos\varphi$, [2].

In the Table I, the loading factor *kI* and the power factor $cos\varphi$ were divided into five linguistic categories (L.C.) with the trapezoidal membership function.

TABLE I
Linguistic Categories of the kI and cos φ

Linguistic Categories		x: kI (%)	x: cos φ	Linguistic Categories		x: kI (%)	x: cos φ
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

The fuzzy load modeling (linguistic categories) can be employed for the steady-state calculation of the distribution networks under conditions of lack of information, especially for evaluation of the power losses, optimal reconfiguration, missing data treatment and so on.

V. Load Profile Modeling of the Distribution Substation

In this section, an approach to determine the daily load profile for the nodes of the electric distribution network (20 kV) is presented.

The shape of load profiles is influenced by the type of node, by the type of day or season of the year. Because a large number of load profiles regarding various nodes create unnecessary problems in handling them, they could be grouped into coherent groups, given that some similarities exist between load profiles. The typical load profile (TLP) for each cluster is obtained by averaging the values for each hour.

For this purpose the use of the hierarchic clustering method conjunctively with fuzzy models is applied to a 20 kV distribution system, containing the data for 39 substations, to classify load profiles into groups, representing typical load profiles.

The active power profiles corresponding to the considered distribution system were normalized relatively to the proper energy consumption (during the day when the load peak was recorded), using the following relation:

$$p_i^h = \frac{P_i^h}{W_i}; \quad h = 1,\ldots,\ 24; \quad i = 1,\ldots, N_R \qquad (10)$$

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

Thus, four groups were determined for the active power in this case. For every obtained group, C_{P1}, C_{P2}, C_{P3} and C_{P4}, the statistical values, average and standard deviation ($m_{P_k}^h$ and $\sigma_{P_k}^h$, k =1, 2, 3, 4), during the peak load day of the distribution system, were calculated, Table II:

$$m_{P_k}^h = \frac{\sum_{i=1}^{N_{C_{Pk}}} p_i^h}{N_{C_{Pk}}} \qquad (11)$$

$$\sigma_{P_k}^h = \sqrt{\frac{\sum_{i=1}^{N_{C_{Pk}}} (p_i^h - m_{P_k}^h)^2}{N_{C_{Pk}}}}; \quad h = 1,\ldots, 24; \quad k = 1,\ldots,\ N_G \qquad (12)$$

where: N_G represents the number of the resulted groups from the clustering process, and $N_{C_{Pk}}$ is the total number of nodes from every group.

The signification of the coefficients m is following: these coefficients transform the active energy consumed by the average member of the group in average active power demanded by it. These coefficients lead us to TLPs corresponding to the active power variation, for the distribution substations, Fig. 9, Table 2.

Using these TLPs for the active power and the factor *cosφ*, the loading factor of every transformer from the distribution substations can be calculated.

In addition, the fuzzy load models can be determined from hourly loads, using the average values (Table II, Table III and (8)).

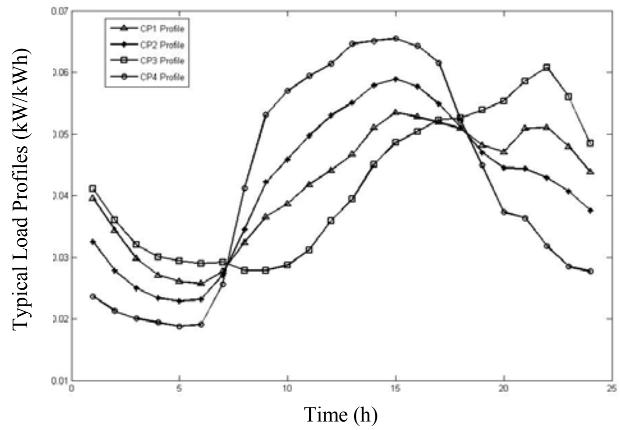

Fig. 9. TLPs for the obtained groups

TABLE II
AVERAGE VALUES FOR THE GROUPS $C_{P1} - C_{P4}$

Hour	$m_{C_{P1}}$ (kW / kWh)	$m_{C_{P2}}$	$m_{C_{P3}}$	$m_{C_{P4}}$
1	0.0396	0.0325	0.0411	0.0237
2	0.0344	0.0278	0.036	0.0213
3	0.0298	0.0250	0.032	0.0201
4	0.0271	0.0234	0.0301	0.0194
5	0.0261	0.0229	0.0294	0.0188
6	0.0257	0.0232	0.0290	0.0191
7	0.0277	0.0271	0.0292	0.0256
8	0.0324	0.0345	0.0279	0.0412
9	0.0366	0.0422	0.0279	0.0531
10	0.0387	0.0459	0.0287	0.0570
11	0.0418	0.0497	0.0312	0.0594
12	0.0441	0.0530	0.0359	0.0614
13	0.0467	0.0551	0.0395	0.0646
14	0.0510	0.0579	0.0451	0.0651
15	0.0535	0.0589	0.0486	0.0655
16	0.0528	0.0577	0.0504	0.0643
17	0.0519	0.0549	0.0523	0.0616
18	0.0509	0.0511	0.0526	0.0522
19	0.0482	0.0470	0.0539	0.0450
20	0.0471	0.0445	0.0554	0.0373
21	0.0509	0.0444	0.0586	0.0364
22	0.0511	0.0429	0.0608	0.0318
23	0.0480	0.0407	0.0560	0.0285
24	0.0439	0.0376	0.0485	0.0277

TABLE III
BREAKING POINTS FOR GROUPS $C_{P1} - C_{P4}$

	C_{P1}	C_{P2}	C_{P3}	C_{P4}
x_1	$m_{G_{P1}} - 1.1\sigma_{G_{P1}}$	$m_{G_{P2}} - 1.12\sigma_{G_{P2}}$	$m_{G_{P3}} - 1{,}11\sigma_{G_{P3}}$	$m_{G_{P4}} - 1.12\sigma_{G_{P4}}$
x_2	$m_{G_{P1}} - \sigma_{G_{P1}}$	$m_{G_{P2}} - \sigma_{G_{P2}}$	$m_{G_{P3}} - \sigma_{G_{P3}}$	$m_{G_{P4}} - \sigma_{G_{P4}}$
x_3	$m_{G_{P1}} + \sigma_{G_{P1}}$	$m_{G_{P2}} + \sigma_{G_{P2}}$	$m_{G_{P3}} + \sigma_{G_{P3}}$	$m_{G_{P4}} + \sigma_{G_{P4}}$
x_4	$m_{G_{P1}} + 1.1\sigma_{G_{P1}}$	$m_{G_{P2}} + 1.12\sigma_{G_{P2}}$	$m_{G_{P3}} + 1.11\sigma_{G_{P3}}$	$m_{G_{P4}} + 1.12\sigma_{G_{P4}}$

As an example, Fig. 10 presents TLP for C_{P1} versus its breaking points.

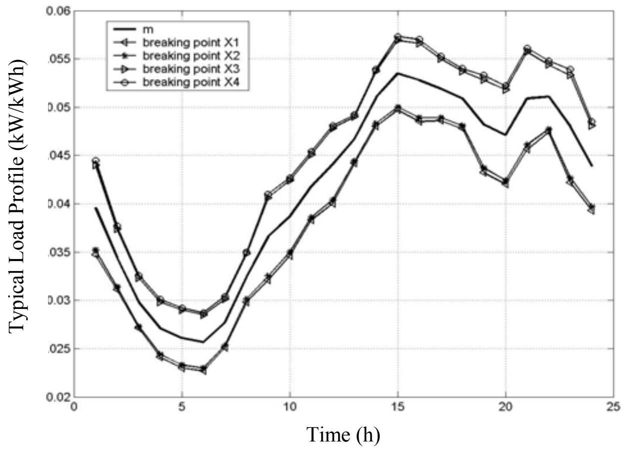

Fig. 10. Typical load profile C_{P1} versus its breaking points

For each load profile, C_{P1} – C_{P4}, considering VH load at the peak, Table I, the fuzzy hourly loads are presented in Table IV and Fig. 11.

TABLE IV
LOADING FACTORS FOR THE GROUPS C_{P1} – C_{P4}

h	kI (Table I)							
	C_{P1}		C_{P2}		C_{P3}		C_{P4}	
	(%)	(L.C.)	(%)	(L.C.)	(%)	(L.C.)	(%)	(L.C.)
1	0.600	M	0.497	M	0.608	H	0.326	S
2	0.521	M	0.425	M	0.533	M	0.293	S
3	0.451	M	0.382	S	0.474	M	0.276	S
4	0.410	M	0.358	S	0.445	M	0.267	S
5	0.395	S	0.350	S	0.435	M	0.258	S
6	0.389	S	0.355	S	0.429	M	0.262	S
7	0.419	M	0.414	M	0.432	M	0.352	S
8	0.491	M	0.527	M	0.413	M	0.566	M
9	0.554	M	0.645	H	0.413	M	0.730	H
10	0.586	M	0.701	H	0.425	M	0.783	H
11	0.633	H	0.759	H	0.462	M	0.816	VH
12	0.668	H	0.810	VH	0.531	M	0.844	VH
13	0.707	H	0.842	VH	0.585	M	0.888	VH
14	0.772	H	0.885	VH	0.667	H	0.894	VH
15	0.810	VH	0.900	VH	0.719	H	0.900	VH
16	0.799	H	0.882	VH	0.746	H	0.883	VH
17	0.786	H	0.839	VH	0.774	H	0.846	VH
18	0.771	H	0.781	H	0.778	H	0.717	H
19	0.730	H	0.718	H	0.798	H	0.618	H
20	0.713	H	0.680	H	0.820	VH	0.513	M
21	0.771	H	0.678	H	0.867	VH	0.500	M
22	0.774	H	0.656	H	0.900	VH	0.437	M
23	0.727	H	0.622	H	0.829	VH	0.392	S
24	0.665	H	0.575	M	0.718	H	0.381	S

In the Table IV the (L.C.), shading columns, represent linguistic categories of the loading factors in accord with Table I.

From these results, it is evident that the obtained profiles can be useful in view of the estimation of power/energy losses, demand-side management, optimal operation and planning of distribution system, missing data treatment and so on.

The loading factors (*kI* (%) from Table IV) permit a more accurate determination of the energy losses in distribution systems.

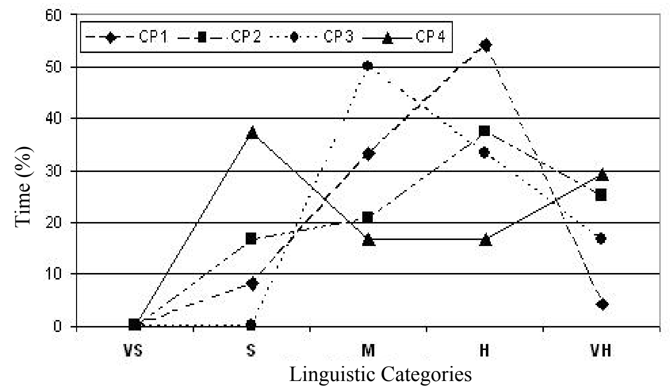

Fig. 11. The weight of the load linguistic categories for the profiles, reported to the installed power of transformers

An important simplification of the energy losses evaluation is obtained by using the weights of linguistic categories, [L.C.], Fig. 11, in conjunction with the Table I.

By extending the fuzzy typical load profile to high voltage networks it is possible to ensure the optimal control planning of these networks and prevent unnecessary tap operations, based on the near future voltage and reactive power. Most control actions are conditioned by the tendency of the load (to increase, decrease, or remain steady). The operators will not be willing to change the voltage set point if they take for granted that the opposite action will be necessary in the near future, in twenty minutes, for example, [9].

VI. CONCLUSIONS

In this paper, a methodology based on clustering techniques is proposed for improving fuzzy load models. These models can be extremely useful for assisting the distribution services providers in the process of electric customer classification based on load profile. The results obtained on a database demonstrate that the methodology can be used with success in the optimal operation and planning of distribution system.

VII. REFERENCES

[1] G Cartina, ″Improving of Fuzzy Models by Clustering Techniques Applications in Power Systems,″ in *Proc. 2005 International Conference on Fundamentals of Electrotechnics and Circuit Theory*, pp. 7a – 8h.

[2] G. Cartina, G. Grigoras, and V. Alexandrescu, ″Pilot pattern choosing by clustering and fuzzy techniques,″ in *Proc. 2005 International Conference Energy-Environment*, pp. 1.13 – 1.17.

[3] G. Cartina, G. Grigoras, E. C. Bobric, and C. Lupascu, ″Improving of Fuzzy Models by Clustering Techniques in Optimal Reconfiguration of the Distribution Networks,″ in *Proc. 2006 International World Energy System*, CD –ROM.

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

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

[6] ″Hierarchical Clustering″, [Online]. Available: http://www.resample. com/xlminer/ help/HClst/ HClst.intro.html.

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

Gheorghe Cartina (Member IEEE) received the M. SC. and Ph. D. degrees in Electrical Engineering from Technical University “Gh. Asachi” 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.

Gheorghe Grigoras received the M. SC. and Ph. D. degrees in Electrical Engineering from Technical University “Gh. Asachi” 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.

Elena-Crenguta Bobric (Member IEEE) received the M. SC. and Ph. D. degrees in Electrical Engineering from Technical University “Gh. Asachi” of Iasi, Romania, in 1995 and 2006 respectively. He is currently lecturer in the Electrical Engineering Faculty at the University “Stefan cel Mare” of Suceava. His research interests are the fuzzy technique applications in power systems.

Daniela Elena Comanescu received the M. SC. in Electrical Engineering from Technical University “Gh. Asachi” of Iasi, Romania, in 1988. She is a diplomat engineer working with Subsidiary “Bistrita”-Piatra Neamt, Hidroelectrica, Romania.