Deriving LV load diagrams for market purposes using commercial information

M. A. Matos, *Member, IEEE*, J. N. Fidalgo, *Member, IEEE*, L.F. Ribeiro

***Abstract*—Classifying consumers, namely LV consumers, in order to assign them typical load diagrams, was always a concern of the electric utilities, which used this kind of information to better manage their distribution networks. Now, with the transition to a completely open market, the need for settlement between distribution operators and traders requires hourly consumption records that are not generally available, so deriving load diagrams for LV consumers is a mandatory task. This paper presents a new methodology for this purpose that uses typical diagrams obtained in measurement campaigns to create classes defined in the commercial information space that maximize the compactness of the diagrams in each class. The methodology was developed in a project with EDP (the Portuguese distribution operator) and the result will probably be adopted by the regulatory authority.**

***Index Terms*-- Clustering methods, Load modeling, Electricity Markets, Energy Management, Optimization methods, Simulated annealing**

I. INTRODUCTION

CLASSIFYING consumers, namely LV consumers, in order to assign them typical load diagrams, was always a concern of the electric utilities, which used this kind of information to better manage their distribution networks. Now, with the transition to a completely open market, the need for settlement between distribution operators and traders requires hourly consumption records that are not generally available, so deriving load diagrams for LV consumers is a mandatory task (the expensive alternative is to install interval meters in all LV consumers).

The problem of assigning diagrams to consumers, or load profiling, includes a basic dilemma that comes after a first phase of measurement campaigns to obtain a sample of diagrams: either you pre-define the classes, on the basis of commercial information you have for all customers (hired power, energy consumption, activity, etc.) or you cluster a sample of diagrams.

In the first option, the diversity of diagrams in each class may be excessive, so the average diagrams are far from being the prototypes even of the sampled diagrams. In the second hypothesis, you get good prototypes, but you are unable to relate them to the commercial information, so you can not classify the consumers outside the sample.

The scientific literature on power systems includes a large variety of clustering methodologies applied to the load profiling question [1-14]. Most of the reported techniques fall into one of the two precedent categories, relatively to both the definition of typical load diagrams and the classification process.

However, to the knowledge of authors, there's no standard established method, of relating a class profile obtained through clustering with the consumers' commercial characteristics. That's why most electrical distribution utilities used an *a priori* classification of consumers based on its commercial characteristics.

This paper presents a new methodology that uses typical diagrams obtained in measurement campaigns to create classes (defined in the commercial information space) that maximize the compactness of the diagrams in each class. The basic idea is to create first small classification cells in the commercial information space that are afterwards aggregated by an optimization procedure that minimizes a measure of compactness of the diagrams in each class.

The methodology was developed in a project with EDP (the Portuguese distribution operator) and the result will probably be adopted by the regulatory authority.

The structure of the paper is the following. In the next section, the initial step of creating the basic classification cells is described. Then in Section III, we present the clustering methodology and the optimization process behind it, followed (section IV) by the explanation of the process of applying the class diagram to a specific costumer. Finally, in section V, we show partial results of the project carried out in Portugal. The conclusions and references complete the paper.

II. DEFINING THE BASIC CLASSIFICATION CELLS

In order to apply the methodology, two blocks of data must be available:

- Commercial data of all the LV customers (hired power, energy consumption, etc);
- A sample of load diagrams for the period under analysis (e.g. one year), linked with the same kind of commercial information.

The first block of data is generally included in the utility's databases and raises no problems, except for the need of processing a large number of records.

On the other hand, gathering the sample diagrams requires the implementation of a measurement campaign that may be designed in different ways, as described in the literature [4,13-16]. Since it is a well-known topic, we will not go into details here, but in section V we'll give some information about our study for the Portuguese distribution company.

The first step of the basic classification cells definition consists of choosing the commercial variables to be used in the classification of customers (the commercial information space). Hired power and energy consumption are natural candidates, but other variables may be used, including information on the use of energy (see a comment on this in section V).

The variables' ranges are then divided into discrete intervals, taking into account their distributions in the sample. The number of intervals must be sufficient to provide a good separation but limited to avoid the generation of too many cells.

Figure 1 shows a possible division of the (Pc, E) space in basic cells. The variables are hired power (Pc - kVA) and energy consumption (E - kWh). For instance, in cell (1;1) E≤1560 kWh and Pc is 1.1 or 3.5 or 4.6 kVA (three tariff categories). Dealing with both discrete (Pc) and continuous (E) variables is not a problem is this approach.

Note that some of the basic cells in the right-upper or left-lower corners may be empty, since they correspond to situations with low Pc/ high E, and vice-versa.

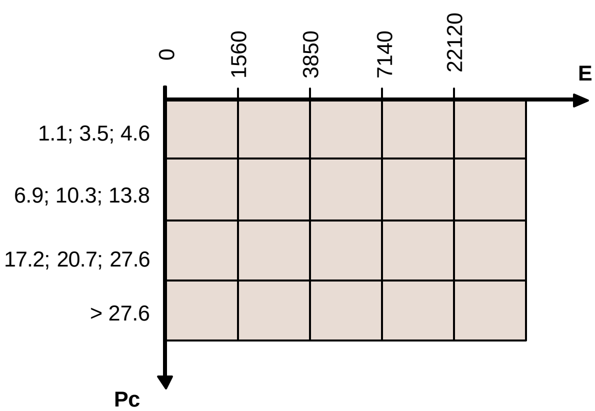

Figure 1 – Example of the definition of the basic classification cell matrix

The organization in cells may be done for all the customers, or separately according to the type of meter:

- single energy measure (type 1);
- peak and out of peak hours (type 2);
- peak, medium and low charged hours (type 3).

Decisions on this topic, and also on the scales granularity, depend, of course, of the trade-off between accuracy and the complexity and cost of the study. The optimization procedure that follows is able to deal with more intervals in each axis (or with more axis) without difficulty.

Once defined the cells, the sampled diagrams are easily assigned to a specific cell (since we have the link between the diagram and the commercial information of the corresponding customer). Each cell will then be characterized by an average diagram and the corresponding variance.

III. Clustering Methodology

A. Problem formulation

Having defined the basic classification cells, with the corresponding assigned diagrams, the next step is to cluster the cells (in the present formulation of the approach, the number N of classes is predefined).

Each possible solution X_i is thus a partition of the set of *n* cells c_1, c_2, ..., c_n into *N* classes C_1, C_2, …, C_N. The usual constraints associated to a partition apply:

- Each cell belongs to only one class;
- There are no empty classes;
- The union of all the classes recomposes the original set.

Moreover, due to the objectives of the exercise, we'll only accept as valid the partitions where all the cells in every class are contiguous.

Therefore, we seek for the valid partition X* that minimizes a measure of quality related to the compactness of the load diagrams. We used the intraclass variance (other options would be possible)

$$C_i = \frac{\sum_{j=1}^{m} (x_{ji} - \overline{x}_i)^2 \bullet NCQ_j}{NCC}, \; i = 1..24$$

C_i - Class load variance in hour i
NCQ_j - Number of consumers in the cell j of the class
NCC- Number of consumers in the class
X_{ji} - Value of the load diagram of cell j in hour i
$\overline{X}$ - mean value of X
m – number of cells in the class

The total variance for each class is given by:

$$CV = \frac{\sum_{i=1}^{24} C_i}{24}$$

where CV represents the class variance.

The fitness function is described by the following.

$$Fitness = \sum_{k=1}^{n} \frac{CV_k * NCC_k}{NCT}$$

where:
N – number of classes
CV_k - class k variance
NCC – number of consumers in the class
NCT – total number of consumers

B. Optimization procedure

The optimization procedure is based on the Simulated Annealing (SA) algorithm [19], a technique derived from statistical mechanics to address large combinatorial problems.

The SA algorithm is a modified local search procedure (we change to a neighbor solution if it better than the present solution) where sometimes a worse solution is accepted, to avoid being trapped in local optima. The probability of accepting a worse solution is controlled by a constant K and a variable parameter called temperature that decreases during the process.

In order to design the SA procedure for this particular application, we must define:

a) Solution – as stated before, a solution is a valid partition X_i.
b) Neighborhood – we get a neighbor of the present partition by selecting a cell at random and then changing its assignment from the present class to a contiguous one. For instance, in Figure 2 the marked cell (now belonging to class A) can change to class B or C, but not to class D.

A	A	B	B	
A	$\boxed{A}$	B	B	B
C	C	C	D	D
	C	C	D	D

Figure 2 - Example of a cell changing movement possibilities.

c) Evaluation – a partition is evaluated by its interclass variance, as explained before.

Once established these fundamental issues, the procedure follows the standard SA algorithm:

Define the initial temperature T_o and the constant K;
Define an initial partition X_o;
Calculate the quality index Q_o for X_o;
While (stopping criterion not satisfied):
   Choose randomly a neighbor X_1 of X_o and calculate its quality index Q_1;
   If $Q_1<Q_o$ or rand $\leq \exp(-(Q_1-Q_0)/KT)$ replace X_o by X_1 as the new reference partition;
   Decrease the temperature according to $T_{k+1} = a \times T_k$, where α is a constant close to, but smaller than, 1;

In order to set the value for K, we used the two first solutions and calculate the value of K that corresponds to a probability P (say 0.25) of accepting the worst solution in the initial phase of the algorithm:

$$K = \frac{\Delta Q}{T \cdot \log(P)}$$

On the other hand, the exponential cooling scheme we used was first proposed Kirkpatrick et al. [19] with α=0.95. Finally, the stopping criterion is the lack of new partitions accepted for a degree of temperature or temperature close to zero.

C. Practical issues

Besides the general arrangements described so far, there is a number of issues that we must decide in a practical study:

1. Using real diagrams, normalized or scaled;
2. Weighting each consumer type in each cell according to its number in the sample or in the universe;
3. Weighting each cell in the class proportionally to the number of consumers in the sample or in the universe;
4. Clustering all months or separately for each month.

According to our experience in the Portuguese study, the best solution has been achieved by the following combination:

1. Using real diagrams (no normalization);
2. Considering the relative weight of each consumer type within each cell proportionally to its amount in the universe;
3. Considering the weight of each cell proportional to the number of clients in the universe;
4. Considering all months in a single set.

IV. Applying typical diagrams to customers

A. Operational procedure

After the classes have been defined, the classification of a given consumer is straightforward, using the cell limits that apply in each class.

Figure 3 and Table 1 show an example for 3 classes, using the cells of Figure 1. As explained before, the blank corners in Figure 3 correspond to empty cells in the universe of consumers.

A	A	A	A	
A	A	A	A	A
B	B	B	C	C
	B	B	C	C

Figure 3 – Example of cell classification for LV consumers, simple meter, 3 classes, workday

Table 1 – Class definition for LV consumers, simple meter, 3 classes, workday

	Pc	**E**
Class A	<= 13.8	all values
Class B	>13.8	<= 7140
Class C	>13.8	>7140

The next step consists in applying the profile for billing purposes. This includes:

1. Building the week typical diagram for each class;
2. Reconstructing, for the particular application month, the corresponding typical diagram: repetition of

workdays, Saturdays and Sundays according to the calendar;
3. Scale the month diagram such a way that its energy matches the measured consumption.
4. Aggregate costs for each hour and compute the total charge for the considered month.

In the case of consumers with period meters (for instance, peak, medium and low charged hours), step 3 should be made in such a way that peak, medium and low charged hours of the month diagram equal the measured consumption in each period.

B. Error analysis

Even if the partition obtained with the optimization process is the best according to the quality measure we used, it's inevitable that the application of profiles leads to deviations, given the diversity of individual consumer's behavior within each class. The portrayal of these errors must be characterized either in energetic terms as well as in billing terms, in order to have a global picture of the adequacy of the entire profiling scheme.

The energy deviation term is represented by the mean squared error (EQM):

$$EQM = \sqrt{\frac{\sum_{n=1}^{Nd}\sum_{h=1}^{96}\left[P_r(n,h) - P_p((n,h))\right]^2}{96\,Nd}}$$

where Nd represents the number of diagrams considered for each case, $P_r(n,h)$ the power of the real diagram *n* in the hour *h*, and $P_p(n,h)$ the power of the profiled diagram *n* in hour *h* (we assume the diagrams are on a 15 min base).

In what respect to economic deviations, one follows the procedure described previously for billing and compare this value with the one would be obtained if the consumer had an hourly (15 min) meter.

V. Illustrative results

In the first part of the project developed for the Portuguese distribution company, about 1000 hourly meters where installed in LV consumers' installations. The consumers where first divided into 5 categories (domestic, commercial, industry, hotels/restaurants and others), each category being afterwards divided into 4 strata, according to the global consumption. The sample definition was based on Neyman's stratified sampling methodology [17,18].

The development of the methodology was based on the following pre-requisites:

1. The classification should be clear: simple and not ambiguous;
2. The classification should not be based on the type of consumer (domestic, commercial, etc.);
3. The consumers should be grouped on the basis of diagrams proximity, having in mind the previous conditions.

Note that the second specification was imposed by the Portuguese regulatory authority, in order to preserve the principle of non-discrimination on the basis of energy use. So, this requirement is not a limitation of the methodology and could be dropped in other applications.

A. Class definition and profiles

Coming to the clustering process, we already showed some results for the 3 class exercise, in Figure 3 and Table 1. Now, Figure 4 depicts the best partition for 4 classes, the main difference consisting of a split of class 1 of Figure 3 into two classes in Figure 4.

A	A	B	B	
A	A	B	B	B
C	C	C	D	D
	C	C	D	D

Figure 4 – Example of cell classification for LV consumers, simple meter, 4 classes, workday

Table 2 depicts the quality measure (average variance) for the cases with 2, 3 and 4 classes, for workdays. The variance tends to be smaller as the number of classes grows. However, the variation from the 2-class to the 3-class case is considerably larger than the one resulting from changing from the 3-class to 4-class case, which may lead to the adoption of the 3-class option in this situation.

Table 2 – Quality measure for 2, 3 and 4 classes, workday

Nº de classes	Variance (x10^{-5})
2	4.7
3	4.2
4	4.0

To illustrate the results, Figure 5 shows typical class diagrams for weekdays, Saturdays and Sundays of LV consumers, single meter. These normalized diagrams will constitute the profiles that later can be applied to the costumers for billing purposes.

B. Error analysis

In order to evaluate the global adequacy of the profiles for different profiling strategies, energy and economic deviations can be computed, as explained in section IV. B. Typical studies could be:

1. Clustering the whole set of LV consumers;
2. Clustering separately the LV consumers according to their meter type;
3. Clustering separately domestic and non-domestic consumers.

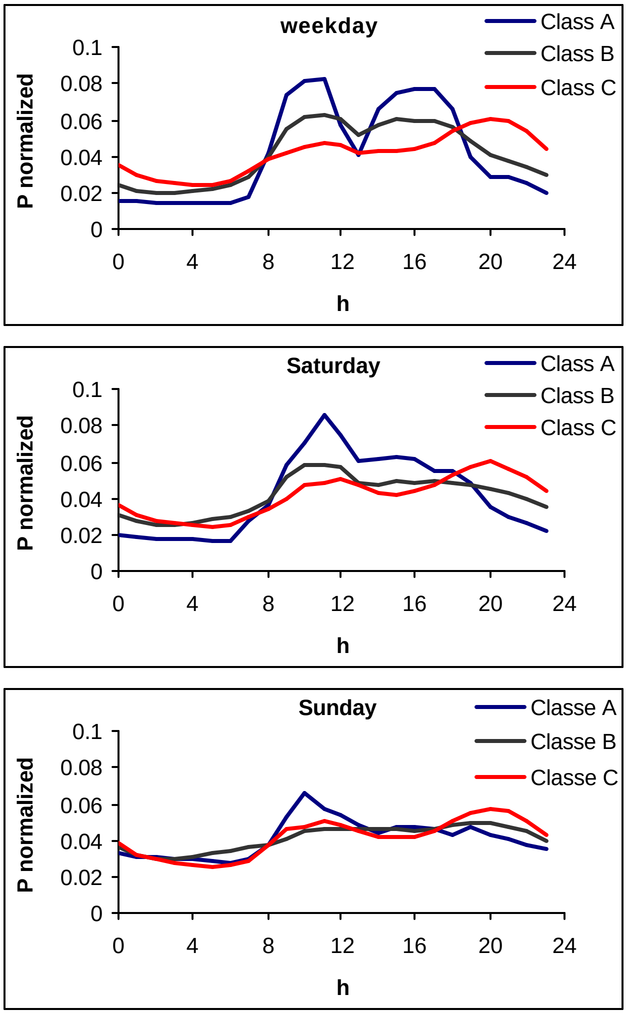

Figure 5 – Classes for LV, simple meter

First, Figure 6 and Figure 7 show typical results for the energy and economic deviations, for LV consumers with a simple meter, for case 1 (considering the set of LV consumers as a whole), for a 3-class partition. These quantities correspond to the average values of the sample for the referred situation. Figure 8 gives an idea of the distribution of the deviations for one of the classes.

In these figures, HP, HM, HL and HSL mean peak hours, medium charged hours, low charged hours and super-low charged hours, respectively.

Now, Figure 9 to Figure 10 show examples of the economic deviations for costumers with meter type 2 (peak and out of peak hours) in case 1 (when no meter discrimination is applied). The results are presented in terms of individual billing variations relatively to the real charge that would occur if the costumers had hourly meters.

The graphics show a fair distribution of deviations, some being positive and others negative, and the values may be considered generally small for the majority of the individuals.

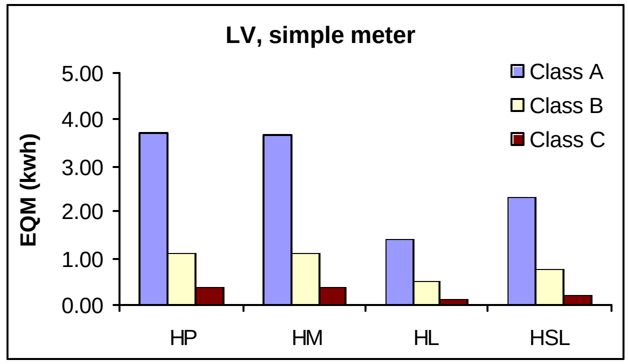

Figure 6 - Energy deviations for LV, simple meter

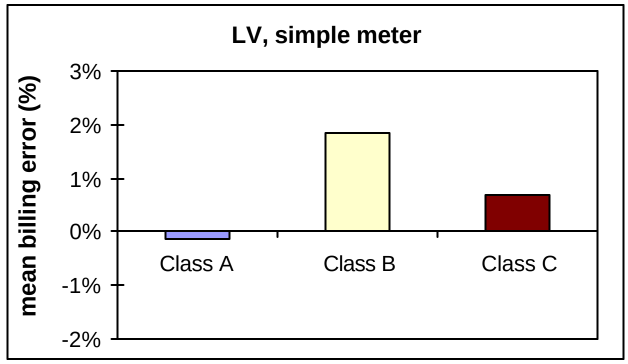

Figure 7 – Billing deviations for LV, simple meter

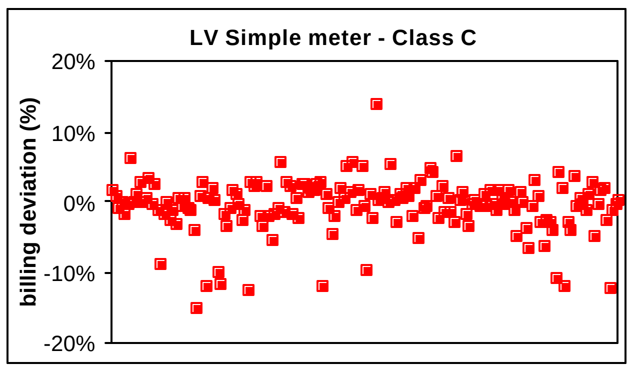

Figure 8 – Distribution of billing deviation for LV, simple meter.

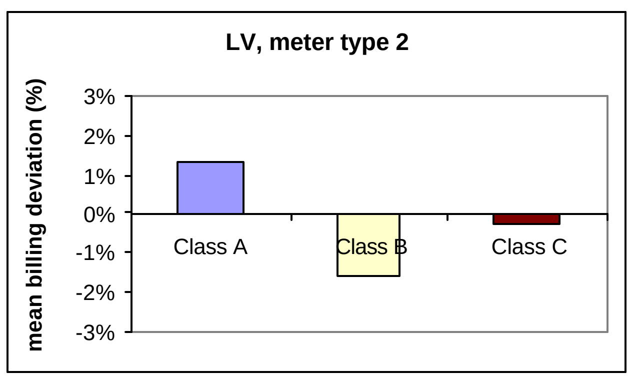

Figure 9 –Billing deviation (meter type 2)

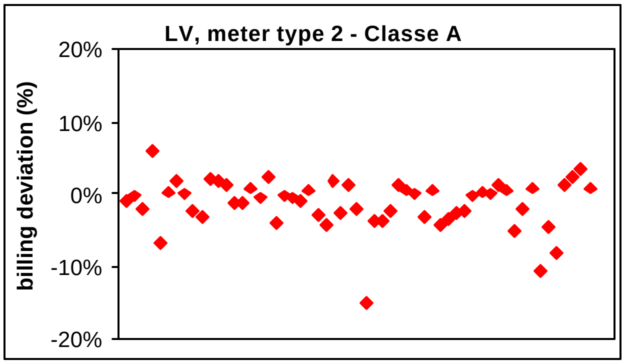

Figure 10 – Distribution of billing deviations (meter type 2)

VI. Conclusions

Deriving profiles for market purposes is a challenging exercise, due to the type of information we must deal with: on one hand we seek for similar load diagrams in each class, on the other hand the clusters must also be understandable in terms of commercial variables available for all the customers.

The new approach presented in this paper solves this dilemma through the use of an optimization process based on the meta-heuristic Simulated Annealing that globally is a clustering methodology relating two different spaces.

The illustrative results, similar to the ones produced in a project for the Portuguese distribution company (EDP), show the applicability of the approach.

VII. Acknowledgment

The authors are grateful to the EDP team for their collaboration in the project.

VIII. References

[1] S. V. Allera, A. G. Horsburgh, "Load Profiling for Energy Trading and Settlements in the UK Electricity Market", DistribuTECH Europe DA/DSM Conference, 1998.

[2] Maine Public Service Company, "Maine Public Service Company Settlement and Load Profiling Methodology, Chapter 321 Report #2", 2000.

[3] Kimiro Meguro, Michiyo Soejima, Fumio Yamazaki, Tsuneo Katayama, "Classification Of Urban Areas By Characteristics Of Power Load Curves", University of Tokyo, 1999.

[4] E.Bompard, E.Carpaneto, G.Chicco, R.Napoli, F.Piglione, P.Postolache, M.Scutariu, "Stratified sampling of the electricity customers for setting up a load profile survey", Proc. RIMAPS 2000 Conference, 2000.

[5] B.D.Pitt and D.S.Kirschen, "Application of data mining techniques to load profiling", Proc. IEEE PICA'99 Conference, pp.131-136,1999.

[6] Wojciech Zalewski, "Combination of Fuzzy Regression Analysis and the Least Squares Regression Method to Modelling of Electrical Load", PMAPS 2000 Proceedings, 2000.

[7] S. Añó Villalba, "Bayesian Linear Model for Load Estimation in Distribution Electric Systems", PMAPS 2000 Proceedings, 2000.

[8] Eurelectric - System Tariff Issues Working Group, "Metering, Load Profiles and Settlement in Deregulated Markets", March 2000.

[9] Eurelectric - Market Research and Customer Needs Working Group, "End-use electricity consumption by the type of customer and type of utilisation", July 1999.

[10] Miriam L. Goldberg, "The Multiplier Method for Load Profiling", EPRI Study, May 1998

[11] Masashi Kitayama, Ryunosuke Matsubara, Yoshio Izui, "Application of Data Mining to Customer Profile Analysis in the Power Electric Industry", IEEE 2002 Winter Meeting Proceedings, 2002.

[12] J. Stuart McMenamin, "Hourly Load Analysis Using Neural Networks", AEIC Conference 1997.

[13] José Antonio Jardini *et. al.*, Daily Load Profiles for Residential, Commercial and Industrial Low Voltage Consumers", IEEE Trans. on Power Delivery, Vol.15, No. 1, Jan. 2000

[14] C.S. Chen et. al., "Determination of Costumer Load Characteristics By Load Survey System at Taipower", IEEE Trans. on Power Delivery, Vol.11, No. 3, July 1996

[15] D.M. Falcão, H.O.Henriques, "Load Estimation in Radial Distribution Systems Using Neural Networks and Fuzzy Set Techniques", 2001 IEEE PES summer Meeting, Vol. 2, 2001

[16] ADEME, e EDF, "Demand-Side Management End-Use Metering Campaign in the Residential Sector", Programa SAVE, Contrato nº 4.1031/93.58, ADEME, e EDF, 1998

[17] Anderson, T. W., and Sclove, Stanley L.. *Statistical Analysis of Data. 2nd ed.* Scientific Press, Palo Alto, Cal., 1986

[18] Neyman, J. (1934), On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection, Journal of the Royal Statistical Society, Ser. A, 97, 558-606.

[19] Kirpatric, S., C. D. Gelatt, and M. P. Vecchi, 1983, "Optimization with Simulated Annealing," Science, 220, p. 671-680.

IX. Biographies

Manuel A. Matos (El. Eng., Ph.D., Aggregation) was born in 1955 in Porto (Portugal). He is presently Full Professor at the Faculty of Engineering of the University of Porto, Portugal, and the Manager of the Power Systems Unit of INESC Porto. He also collaborates with the Management School of the University of Porto. His research interests include fuzzy modeling of power systems, optimization and decision-aid methods. He is a member of IEEE.

J. N. Fidalgo is an Assistant Professor in the Dept. of EE of the Fac. of Eng. of Porto and a Senior Researcher at Power Systems Unit of Institute of Engineering Systems and Computers (INESC Porto). He obtained a PhD degree in Electrical Engineering and Computers from the same University in 1995. His interests include pattern recognition, computational intelligence and neural networks and its applications to power systems.

Luís Ribeiro was born in 1978 in Aveiro (Portugal). He graduated in Applied Mathematics to Technology, in the Science Faculty of Porto University in 2001. He is currently with INESC Porto, as a junior researcher and as MSc Student at Engineering Faculty of Porto University. His research interests include GIS and optimization.