# Customer Classification and Load Profiling using Data from Smart Meters

**Conference Paper** · November 2014

DOI: 10.1109/NEUREL.2014.7011464



3 authors:

Grigoras Gheorghe
Gheorghe Asachi Technical University of Iasi


Ovidiu Ivanov
Gheorghe Asachi Technical University of Iasi


Mihai Gavrilas
Gheorghe Asachi Technical University of Iasi

# Customer Classification and Load Profiling using Data from Smart Meters

G. Grigoras, O. Ivanov, and M. Gavrilas, *Senior Member, IEEE*

***Abstract*—This paper presents a self-organization based integrated model for customer classification and load profiling in distribution systems. The consumer classification in consumption classes characterized by typical load profiles is made using information provided by Smart Meters. For determination of the consumption classes, every customer is characterized by the following primary information: daily (monthly) energy consumption, minimum and maximum loads. The proposed model was tested using household consumers from a rural area. The results demonstrate the ability of the methodology to efficiently used in distribution systems when information about the supplied customers is very poor (based only the data provided by classic meters).**

***Index Terms*—customer classification, load profiling, self-organization, smart meters, distribution systems.**

## I. Introduction

IN the last years, many electric companies are deploying Smart Metering Systems. The deployment of Smart Meter Systems begins with selection of the technology and the planning for installation, operation and maintenance. These systems have emerged as a breakthrough in the relationship between the consumer and the distribution company. Coupled with an Advanced Metering Infrastructure (AMI), Smart Metering Systems could allow electric distribution companies monitoring the electrical power conditions practically in every point of the network [1]-[4].

Generally, the categories like residential or small and medium non-residential ones are poorly described since a communicating smart. At present there two alternate solutions for this problem. First, suppliers can provide to every customer smart meters; this is mostly costly, but the most precise solution. Second, customers use traditional meters, and suppliers apply estimated load profiles to distributed monthly energy consumption over days and hours; this is the cheapest but less precise solution. Even if the second alternative is less precise, it is simple and cheap enough to be the most used approach applied to small customers. 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 [5]-[10].

The implementation of smart metering in EU countries for households is in different stages. Some countries are in implementation stage of legislative provisions, while others are in early stages. Spain, Finland and the UK are in the phase of large-scale adoption and France and Portugal are now set to join these. Italy or Sweden has not waited changes to the legislative and regulatory framework and in the past years a large number of smart meters have been rolled out [11].

A principal aspect of implementing smart metering technology is represented by a flow of data several magnitudes greater than any previous traditional metering schemes. For managing of this data, the electrical distribution companies implementing smart metering companies undergo changes. Thus, the meter data management systems (MDMS) provide utilities with a critical solution for storing, validating, aggregating and processing large volumes of data, in preparation for billing, settlements and other reports [12]. Power professionals believe the integration of Smart metering technology in electric networks might support the derelict power system through minimizing line losses, reducing power pilferages, clamping peak power demands, increasing generation efficiency, economic power system operation, and real time state estimation. Exploration of such data can also emerge as a research direction both in academic and industry sector, such as load aggregation and disaggregation (LD), load forecasting (LF) and demand response (DR) support [6], [13]-[14].

The paper presents how a classification and load profiling methodology can be implemented to small customers using the information provided by smart meters. This methodology uses self-organization techniques to process data, identify existing load classes or patterns, classify customers according to classes, and generate typical load profiles (TLPs). For determination of the consumers' classes, every customer must be characterized by the following primary information: daily (monthly) energy consumption, minimum and maximum loads.

## II. Description of Customer Classification and Load Profiling Methodology

An important research topic used in optimal operation and planning of distribution networks by electric companies refers at representative consumption categories of customers. This is corresponding particularly to small residential consumers, whose energy consumption is not usually measured with smart

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

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

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

meters. Because of the large number of small customers, sampling is the only possible way to collect data. The problem is how the selection and the analysis of the sample of customers should be made to finally get the most accurate load estimates for practical network calculations [6]-[8].

In this context, the electric companies require a set of load curves to represent all consumption classes. Deciding the optimal number of classes and the type of load curve for one class is a complicated problem. In order to form the consumption classes, a classification can be made taking into account the daily energy consumption, minimum and maximum loads of customers. After identification of consumption class, the typical load profiles corresponding to every class are determined.

The major innovation of the methodology described in this paper is based on the structure of the Data Mining in large databases using self-organization techniques to obtain useful knowledge from large amounts of data.

The proposed algorithm has the following steps, Fig. 1:

*Step 1*. Database: In this step, a consumer' representative sample will be select from the database. For this, the most important primary characteristics in terms of the consumption energy are identified. The customers have installed the smart meters. The consumers' load profiles, respectively variables that characterize the energy consumption will be recorded in a database. These variables are referred at: the daily energy, minimum and maximum active powers, and consumption category.

*Step 2*. Partitioning into macro-categories: The database is divided into consumption categories taking into account the type of consumer activity: residential, commercial, industrial.

*Step 3*. Data cleaning and Pre-processing: In this step, all records that contain missing data or outliers will be excluded. These situations are encountered in practice because the measurements are made for a large number of customers, spread over a wide geographical area, and problems with communication, interruption, failure of equipment or irregular atypical behaviour of consumers can appear. After being cleaned, pre-processed and reduced, the data are used to obtain the division in classes.

*Step 4*. Data Mining Model: a classification into more classes, in function by the daily energy consumption, minimum and maximum loads of the consumers is done. For customer classification is used the self-organization technique. This approach is presented in next paragraph. Finally, the typical load profile for each customer class is obtained by averaging the values for each hour.

*Step 5*. Assignation: Each customer class is attributed a typical load profile according to their consumption category (characterized by the daily energy consumption, minimum load, and maximum load).

## III. THE SELF-ORGANIZING TECHNIQUES

Self-organization provides a method of representation of multidimensional data in space with a smaller number of dimensions, usually one or two, maintaining topological relationships between data in the training set.

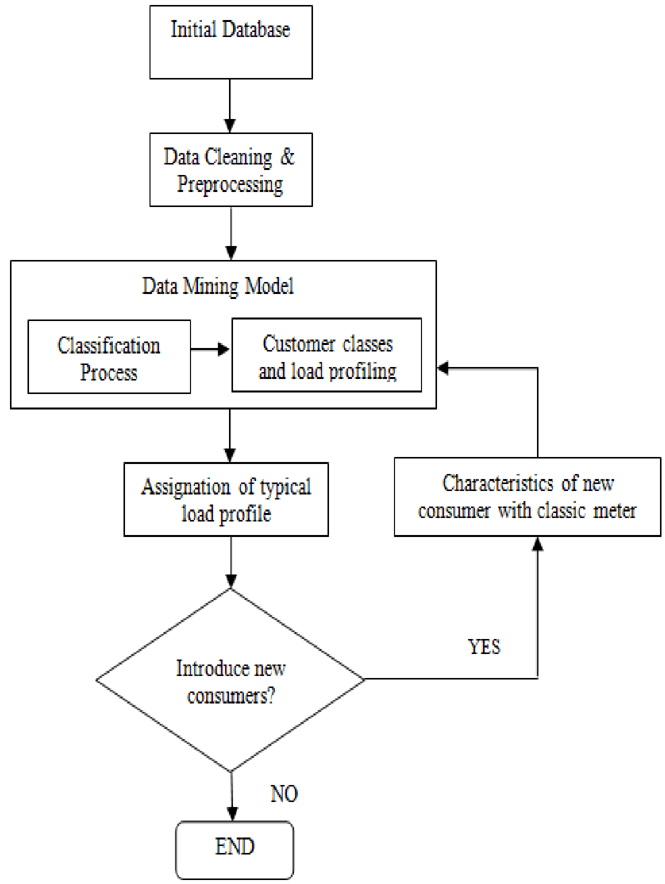



Fig. 1. Flow-chart of the proposed algorithm.

If in the case of supervised learning using multilayer perceptron neural networks (Multilayer Perceptron - MLP) training set contains known input-output pairs and neural network must achieve mapping the inputs toward the outputs, Kohonen networks use self-organizing learning where the training set contains inputs and neural network tries to find the characteristic features in the data presented. These characteristics are stored in synaptic weights for connections between neurons in the form of so-called prototypes [13], [14].

From the viewpoint of the structure, a Kohonen network consists two layers (input and output) of class-neurons, one or two-dimensional. Fig. 4 presents a one-dimensional network. Each neuron-class is connected with one weight by each neuron from the input layer, whose size corresponds to the number of elements from a model of learning set.

The prototypes of class (weights of neurons) are randomly initialized and all class-neurons compete for the right to learn. Each model from the training set ($x^{(m)}$) is compared with prototypes $w_c$ associated the class-neurons and winning neuron will be find at the shortest distance (Best-Matching Unit). For distance calculation is usually used Euclidean norm, $||x^{(m)}-w_c||$, but other distances can be used. Class-neurons are placed in symmetrical grid, Fig. 2.

After determining the winning neuron $c$, this adapts the prototype alongside all the class-neurons from its vicinity, ($V_c$), while the weights of others neurons remain unchanged.

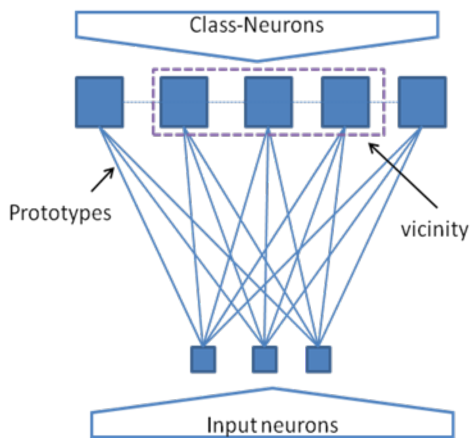



Fig. 2. Kohonen neural network

The adaptation of the weighs is made according to the equations (1), initially in a large vicinity, which narrows progressively in order to avoid blocking in a local minimum.

$$\begin{aligned} w_c &= w_c + \eta \cdot \left(x^{(m)} - w_c\right) \quad c \in V_c \\ w_c &= w_c \quad c \notin V_c \end{aligned} \tag{1}$$

where η is the learning rate.

Finally, in weights of class-neurons will be stored the features of found classes in training set. If the number of class-neurons is too large, some will remain unused. Therefore, sometimes it prefers initializing of Kohonen network with a single class, creating some new ones when the distance between the existing classes and current input model exceeds a threshold.

## IV. CASE STUDY

A database described by 180 load curves corresponding to the small residential consumers from a rural distribution network from Romania was considered. The measurements of load curves were performed using smart meters. Every measurement is represented by a curve of 24 hourly points describing the behaviour of a consumer during a day. For each consumer, the following consumption data are also known: the daily energy consumption (*W*), minimum (*Pmin*) and maximum (*Pmax*) loads. The consumers with missing values, outliers or energy consumption equal with zero were excluded. In the pre-processing, these consumers were excluded. Finally, only 149 consumers were eligible.

The self-organization approach used a feature map with 5 class-units. After the training phase was completed, the self-organizing feature map produced five customer classes whose characteristics are those from Table I. The customer classes obtained are represented in Fig. 3.

From analysis of Table I, it can observed that the most representative class is C5 (51 % from all consumers) and the least representative class is C2 (4 % from all consumers). These results are normal because in Romania the most consumers from rural area have daily energy consumption between 1.5 and 2 kWh. Very few consumers have a high consumption more than 4 kWh.

The values of these characteristics are the medium values obtained through averaging inside each cluster. In the next step, the load profiles corresponding to the consumers from each category were normalized relatively to the daily energy consumption.

Then, the average values corresponding for each consumer class were calculated. These values transform the energy consumed by the medium member (consumer) of the class in average active power demanded by it. The typical load profile for each class from the consumers' categories is obtained by representation of these coefficients, Figs. 4 – 8.

TABLE I
THE CHARACTERISTICS OF RESULTED CLASSES

| Class | No. of customers | Pmax [kW] | | Pmin [kW] | | W [kWh] | |
|---|---|---|---|---|---|---|---|
| | | m | σ | m | σ | m | σ |
| C1 | 13 | 0.239 | 0.050 | 0.095 | 0.027 | 3.522 | 0.911 |
| C2 | 6 | 0.703 | 0.088 | 0.032 | 0.031 | 4.445 | 1.306 |
| C3 | 19 | 0.366 | 0.119 | 0.046 | 0.017 | 3.879 | 0.745 |
| C4 | 35 | 0.060 | 0.052 | 0.001 | 0.003 | 0.436 | 0.454 |
| C5 | 76 | 0.183 | 0.051 | 0.021 | 0.017 | 1.825 | 0.531 |

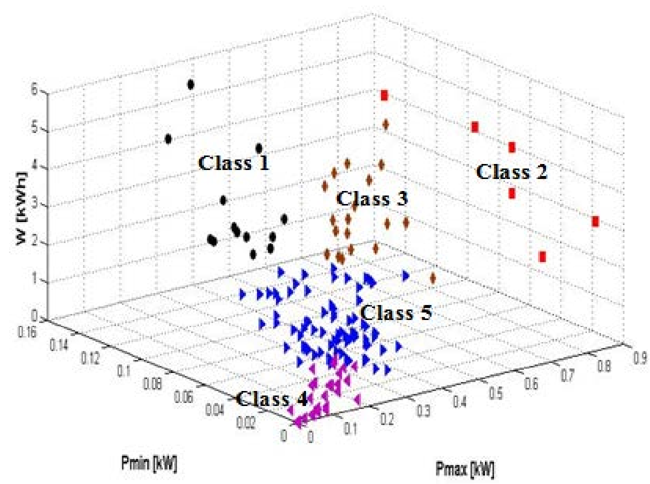



Fig. 3. The representation of the classes.

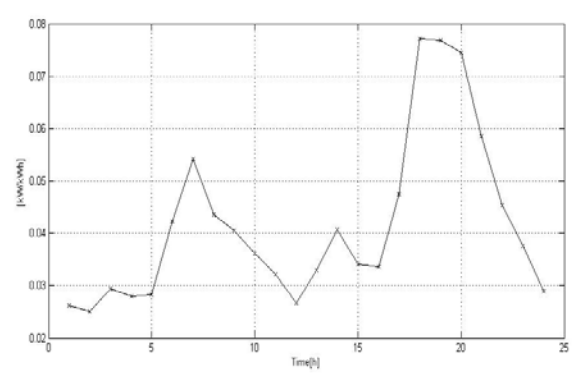



Fig. 4. Typical load profile for class C2.

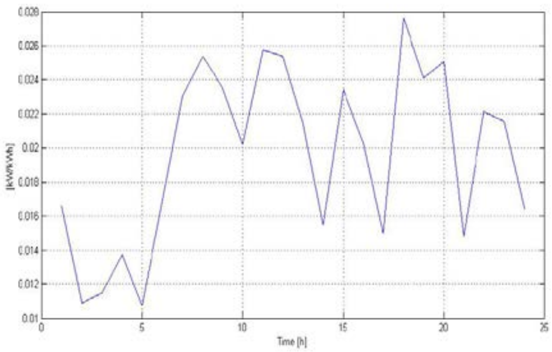



Fig. 5. Typical load profile for class C2.

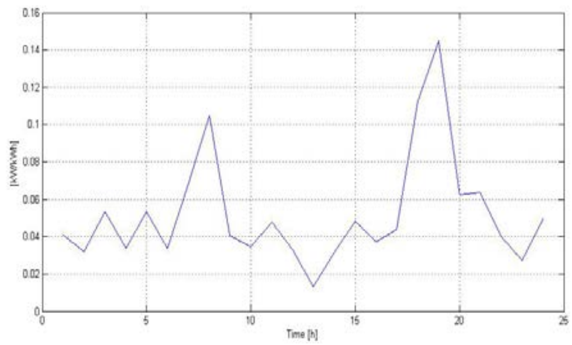



Fig. 6. Typical load profile for class C3.

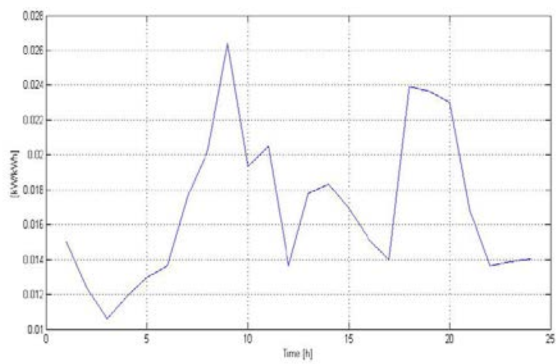



Fig. 7. Typical load profile for class C4.

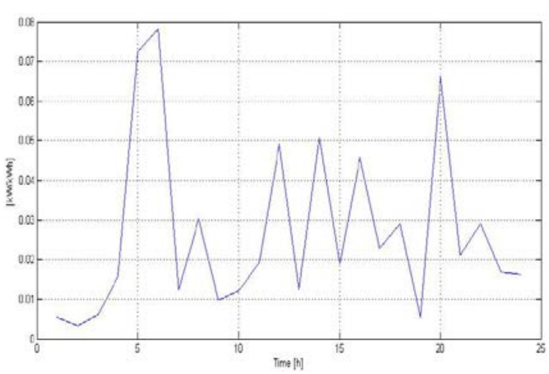



Fig. 8. Typical load profile for class C5.

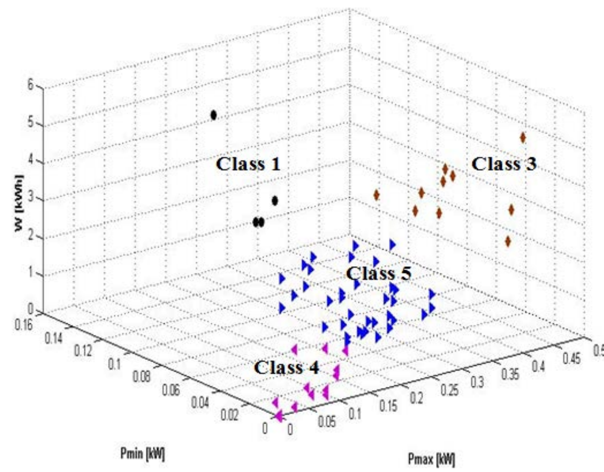



Fig. 9. Assignation of the consumers from testing database to classes

Further, the method is tested on a database formed by another 66 small residential consumers supplied from an electric distribution substation (20/0.4 kV) equipped with a power transformer with rated power by 40 kVA. Each consumer was assigned to a class, depending by his load characteristics (daily energy consumption, minimum load and peak load), Fig. 9.

From analysis of Fig. 9, it can observed that the most representative class is C5 (56 % from all consumers) and class C2 is not represented.

Using the typical load profiles corresponding to these consumers, the load of the distribution substation was obtained. In Fig. 10 and Table II, the real and estimated values of the load are presented.

TABLE I

THE RESULTS OBTAINED WITH PROPOSED ALGORITHM

| Hour | Preal [kW] | Pest [kW] | Err [%] |
|---|---|---|---|
| 1 | 3.25 | 3.15 | 3.17 |
| 2 | 3.15 | 3.22 | 2.17 |
| 3 | 3.20 | 3.25 | 1.54 |
| 4 | 4.00 | 4.05 | 1.23 |
| 5 | 6.76 | 6.60 | 2.42 |
| 6 | 7.30 | 7.20 | 1.39 |
| 7 | 5.12 | 5.25 | 2.48 |
| 8 | 6.17 | 6.05 | 1.98 |
| 9 | 5.10 | 5.20 | 1.92 |
| 10 | 4.80 | 4.95 | 3.03 |
| 11 | 5.27 | 5.35 | 1.50 |
| 12 | 6.00 | 6.15 | 2.44 |
| 13 | 4.50 | 4.40 | 2.27 |
| 14 | 6.90 | 6.76 | 2.07 |
| 15 | 5.05 | 5.20 | 2.88 |
| 16 | 6.00 | 6.10 | 1.64 |
| 17 | 5.30 | 5.45 | 2.75 |
| 18 | 7.00 | 7.20 | 2.78 |
| 19 | 8.00 | 8.15 | 1.84 |
| 20 | 8.45 | 8.60 | 1.74 |
| 21 | 5.75 | 5.60 | 2.68 |
| 22 | 5.90 | 5.83 | 1.20 |
| 23 | 4.60 | 4.50 | 2.22 |
| 24 | 4.30 | 4.39 | 2.05 |

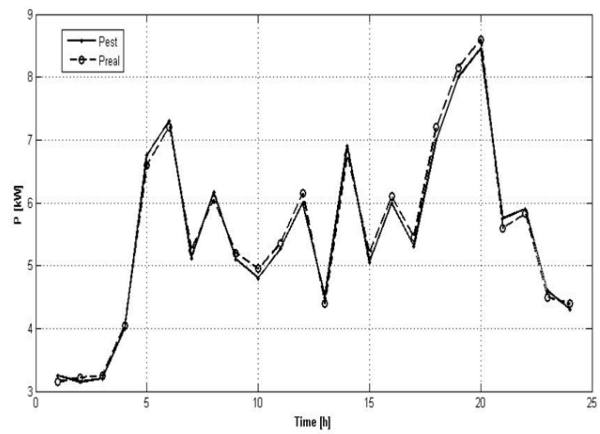



Fig. 10. Real and estimated load of the electric distribution substation

The shape of the load profile is similar with the typical load profiles of consumers from class C5, the deviations being reduced. In order to confirm this, the estimation errors are shown in Fig. 11.

The estimation errors for each hour $h$ = 1,…, 24 were calculated with the relation:

$$Err = \frac{\left|P_{real\ h} - P_{est\ h}\right|}{P_{real\ h}} \cdot 100, \quad (\%) \tag{2}$$

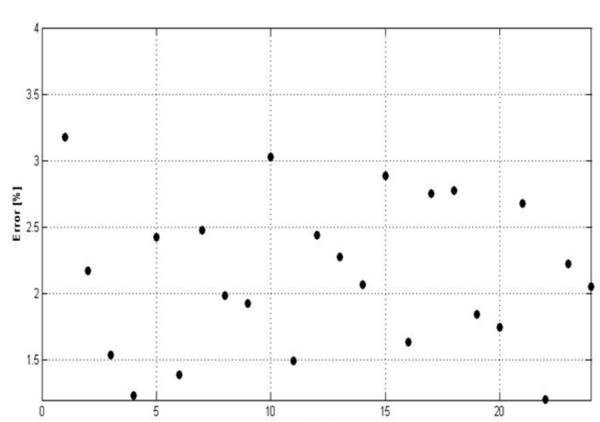
where: $P_{real\ h}$ is the real values for active load at hour $h$; $P_{est\ h}$ represent the estimated values for active load at hour $h$.

The highest value is recorded at hour 1 (3.17%) and the lowest value recorded at 22 (1.2%). The average value of estimation errors corresponding to hourly load of distribution substation is 2.14%.

The comparison with the results obtained with a clustering based approach given in [6], for same test database, indicated a decrease of average estimation error from 3.85 % to 2.14 % (obtained using self-organization based approach).

From the analysis of results, one can say that a classification of the customer classes based on data from smart meters is useful in view of the optimal operation and planning of distribution system, in the building of the tariff structures or demand-side management.

Fig. 11. Estimation errors obtained with proposed algorithm




## V. Conclusions

Self-organizing techniques are extremely useful for assisting the electric distribution services providers in the process of customers' classification on the basis of electrical characteristics (daily/monthly energy consumption, minimum load, and maximum load). The analysis of the algorithm allows assessing the best structure of the typical load profile set.

The results obtained demonstrated that the proposed approach can be used with the success in the optimal operation and planning of distribution systems when information about the supplied customers is very poor (based only the data provided by classic meters).

## References

[1] D. R. V. Leite, H. Lamin, J. M. C. de Albuquerque, and I. M. T. Camargo, "Regulatory Impact Analysis of Smart Meters Implementation in Brazil," in *Proc. of Innovative Smart Grid Technologies (ISGT)*, Washington DC, USA, 2012.

[2] J. Pithart, "Integration and utilization of the smart grid," [Online]. Available: http://www.feec.vutbr.cz/EEICT/2009/sbornik/.

[3] P. Deng and L. Yang A., "Secure and Privacy-Preserving Communication Scheme for Advanced Metering Infrastructure," in *Proc. of Innovative Smart Grid Technologies (ISGT)*, Washington, DC, USA, 2012.

[4] R. Piacentini, "Modernizing Power Grids with Distributed Intelligence and Smart Grid-Ready Instrumentation," in *Proc. of Innovative Smart Grid Technologies (ISGT)*, Washington, DC, USA, 2012.

[5] G. Hebrail, "Practical Data Mining in Large Utility Company," [Online]. Available: https://upcommons.upc.edu/revistes/bitstream/2099/4160/4/ article. pdf

[6] G. Grigoras, Fl. Scarlatache, "Use of data from smart meters in optimal operation of distribution systems", in *14th International Conference on Optimization of Electrical and Electronic Equipment OPTIM 2014*, Brasov, Romania, May 22-24, 2014.

[7] G. Grigoras, G. Cartina, "Improved Fuzzy Load Models by Clustering Techniques in Distribution Network Control", *International Journal on Electrical Engineering and Informatics*, vol. 3, no. 2, 2011, pp. 205-2014.

[8] M. Gavrilas, G. Cartina, G. Grigoras, O. Ivanov and C. Lupascu, "Integrated Model for Electric Load Profiling" in *Proc. of 4th International Conference on Electrical and Power Engineering EPE 2006*, Iasi, Romania, October 12-14, 2006, pp. 1193-1198.

[9] Mutanen, M. Ruska, S. Repo, P. Järventausta, "Customer Classifcation and Load Profilling Method for Distribution Systems", *IEEE Trans. on Power Delivery*, vol. 26, no. 3, pp. 1755 – 1763, 2011.

[10] J. N. Fidalgoa, M. A. Matosb, L. Ribeiroc, "A new clustering algorithm for load profiling based on billing data," *Electric Power Systems Research*, vol. 82, pp. 27– 33, 2012.

[11] Communicating smart meters to customers – which role for DSOs? [Online]. Available: http:// www.eurelectric.org/ media/ 82228/eurelectric _comroledsos_final_to_ publish-for_website-2013-030-0457-01-e.pdf.

[12] Achieving high performance with smart meter data management systems. A leading practice approach for utilities, [Online]. Available: http://www.accenture.com/SiteCollectionDocuments/ PDF/ Resources/ Accenture_POV_Smart_Grid_Meter_Data_Management_Systems.pdf.

[13] Z. Guo, Z. J. Wang, A. Kashani, "Home Appliance Load Modeling from Aggregated Smart Meter Data," *IEEE Transactions on Power Systems*, 2014, in press.

[14] M. Z. Degefa, R. J. Millar, M. Koivisto, M. Humayun and M. Lehtonen, "Load Flow Analysis Framework for Active Distribution Networks Based on Smart Meter Reading System," *Engineering*, vol, 5, pp. 1-8, 2013.

[15] M. Gavrilas, G. Cartina, V. Alexandrescu, O. Ivanov, G. Grigoras, "Load Modeling in Electrical Networks (in romanian)", *PIM Publishing*, Iasi, Romania, 2007.

[16] F. Mc Loughlin, A. Duffy, M. Conlon, "Analyzing Domestic Electricity Smart Metering Data Using Self Organizing Maps", in *CIRED Workshop*, Lisbon, Portugal, 29-30 May, 2012.