A NOVEL DATA MINING PROCEDURE GENERATING A TYPICAL ELECTRICAL ENERGY DAILY SHAPE BASED ON VALUES FROM SMART METERING

Cătălin MIHAI[*], Marian PEARSICĂ[**], Laurian GHERMAN[**]

[*]SC FDEE Electrica, Brasov, Romania (catalin77@gmail.com)
[**]"Henri Coandă" Air Force Academy, Brasov, Romania
(marianpearsica@yahoo.com, lauriang@gmail.com)

DOI: 10.19062/1842-9238.2018.16.3.8

Abstract: *This paper presents a data mining procedure to generate a typical daily shape of energy consumption based on smart metering measurements. Such a typical daily shape describes an observed or estimated graphical representation of the variation of the electric load against time. This paper proposes two methods for the development of the novel data mining procedure: i) statistical analysis applied to a set of measured data; ii) cluster analysis using the machine-learning algorithm to determine a typical daily shape. The new procedure can be part of the Smart Metering and more precisely of the Automated Meter Reading (AMR). The results of the statistical analysis using the set of measured data show whether or not the average values are representative for the data set. The results of the cluster analysis extend the potential of forecasting a typical daily shape and of offering tools for potential applications to data mining procedure.*

Keywords: *Electrical energy load, statistical analysis, cluster analysis, data mining, typical daily shape, smart metering.*

1. INTRODUCTION

The implementation of smart grid concepts is inevitably linking distributed generation (DG), energy efficiency and storage. The element which forms the 'smart' core of the whole system is the internal capacity of the bidirectional flow management of both energy (due to the DG) and information, under the supervision and control of an intelligent control system.

Current Information and Communications Technology (ICT) methods to achieve the transmission, reception and processing of information enable the design of an intelligent network through which the exchange of information between all users of the electricity network is possible.

A key component of ICT-based smart grids is smart metering. The old-style meters are transformed into intelligent tools, interfaced with the data network, able to communicate with the data collection center of the distributor in order to transmit information on the detected energy consumption. Smart meters are divided into two categories:

i) Automated Meter Reading (AMR) systems, which implement simple functions of remote reading of energy consumption profiles, even in real time, from users and tariff profiles, thus leaving the human operation to cases of failure of the equipment and not to detection of consumption;

ii) Automated Meter Management (AMM) systems, which instead support bi-directional communication to the distribution utility and can thus improve and speed up the commercial services rendered to customers, perform self-diagnosis functions, signaling faults and power quality analysis in terms of interruptions, voltage variation and load measurement. In connection with the DG, some considerable synergies can also be obtained by using the smart meter communications infrastructure to remotely control the generators.

Several contributions to knowledge on energy load management and strategies for load profiling have recently been made. Al-Otaibi *et al*. [1] provide details about the construction and calibration for clustering of daily load curves from smart metering by applying a new method of a conditional filtering on meter resolution in order to obtain new consumption pattern recognition. Koivisto *et al.* perform an analysis with data from an AMR system.

Minchala-Avila *et al*. [2] present load profiles for consumers with renewable energy (wind, photovoltaic, energy storage) based on historical data, using the nonlinear model predictive control (NMPC) algorithm in islanding micro-grids. A variety of novel techniques, addressing and improving the problem of one-day-ahead load forecasting based on an interval type 2 fuzzy logic systems (IT2FLS), have been developed in paper [3]. Borges *et al*. [4] used aggregation for short-term load forecasting in Smart Grids using three methods: i) a bias correction, ii) top-down approach and iii) bottom-up approach and regression.

Cluster analysis has been employed in the literature for residential electrical load modeling [5]. The clustering was developed based on two models: i) a mixed model and ii) a Markov model for achieving group similarity. The time - use data model has been developed in paper [6] for the construction of load profile in residential electricity and hot-water demand use by assigning appliances to different categories of related activities.

McLoughlin *et al*. [7] described a methodological approach to electricity load profile characterization through a clustering model of a residential electricity load profile.

In all the articles presented, information solutions are distributed, using data mining for the concept of smart metering as well as different algorithms for the determination of load profiles, but they have not managed to propose fast solutions for their implementation involving all actors, distribution operators, equipment manufacturers and ultimately the consumer. Through the data mining procedure, to achieve the daily load curve proposed in this paper, we want to include the consumer as an active participant in the Smart Metering concept and ultimately make the quickest decisions on the free energy market.

2. DATA MINING FOR LOAD PROFILING

The data mining procedure is summarized in FIG. 1. This shows the main steps to follow in view of building up the load profiling.

Starting from a data set at the pre-processing stage presented in section 3, a Statistical Analysis and a Cluster Analysis are performed in section 4. These are then merged into a Server Application, where a decision is made on the suitability of the curve. In case of a positive result, the typical load profile is constructed. Alternatively, the Load Forecasting approach/step is being followed.

A. Statistical analysis

Statistical analysis allows the representation of data as diagrams, the calculation of specific quantities and reducing the amount of data by dividing it into classes and studying statistical characteristics of each of these.

Descriptive statistics is a tool that can characterize a sample X from population (where X is random variable) in the form of a data set of measurements $\{ x_i \}$, where $i = 1, 2, \ldots, n$ [8], where n is number of items from sample X.

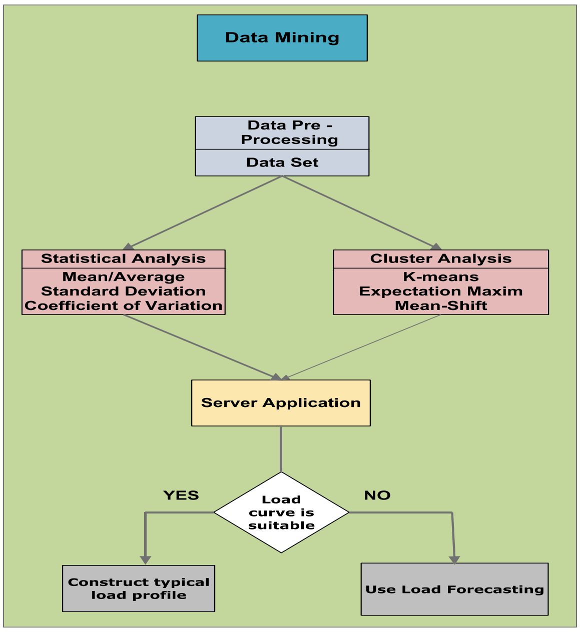

FIG. 1. Schematic representation of the data mining procedure

The mean (or expected value) of a data set of measurements $\{ x_i \}$, where $i = 1, 2, \ldots, n$, is defined by:

$$\mu_x = E(x) = \begin{cases} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i; & x : discrete \\ \int_{-\infty}^{\infty} x \cdot f_x(x) dx; & x : continuous \end{cases} \tag{1}$$

where: f_x is the probability density function of the random variable.

The variance of a data sample from measurements $\{ x_i \}$, where $i = 1, 2, \ldots, n$ is defined by:

$$\sigma_X{}^2 = Var(x) = \begin{cases} \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \mu_X)^2; & x : discrete \\ \int_{-\infty}^{\infty} (x_i - \mu_X)^2 \cdot f_x(x) dx; & x : continuous \end{cases} \tag{2}$$

The standard deviation of a data sample from measurements $\{ x_i \}$, where $i = 1, 2, \ldots, n$, denoted by a σ_X is a positive square root of the variance shown in equation (2).

The coefficient of variation (*CV*), calculated as a percentage ratio between the standard deviation and the mean, is defined by:

$$CV_{\%} = \frac{\sigma_x}{\mu_x} \cdot 100 \tag{3}$$

The coefficient of variation enables the comparison of the scattering of data with different distributions for variables shown in different units.

In paper [9] the coefficient of variation is presented, considering the following thresholds and can be used as a significance test for the representativeness of the mean:

- $0 < CV \leq 17\,\%$ - the mean is strictly representative;
- $17\,\% < CV \leq 35\,\%$ - the mean is representative;
- $35\,\% < CV \leq 50\,\%$ - the mean is broadly representative;
- $50\,\% < CV$ - the mean is not representative.

In this paper, a procedure analysis is carried out with respect to the power consumption characteristic parameters using descriptive statistics, and experimental data are processed by applying statistical methods to a set of data obtained by smart metering for 6 months (May until October).

B. Cluster analysis

A useful tool for identifying and predicting groups within data sets is cluster analysis. The work presented in paper [10] shows that the main purpose of clustering methods is to define the types of clusters; there are several established cluster analysis algorithms such as: two step clusters, K-Means clustering, Expectation Maximization (EM), and Mean-Shift.

The aim of the K-means algorithm is to divide M points in N dimensions into in a sample into k clusters ($k < N$) so that each object belongs to the cluster with the nearest mean. There are multiple solutions for applying the K-means cluster algorithm, the simplest of them using special software tools. A number of software tools suitable for statistical processing are: Matlab, IBM SPSS, Weka, Statistica, OriginGraph, MedCalc. Paper [11] presents the algorithm of k-means clustering for the IBM SPSS software and was used for load profiling for a gas station with measurements data acquired over a period of one year. Paper [1] presents cluster analysis using K-means++, and the calculation was performed using the Matlab software.

In the current paper, the data mining procedure is performed using Statistica software, with tool cluster analysis. For this tool, the number of clusters introduced must be greater than 1 and less than the number of samples. Objects are placed in the cluster with the nearest mean at each iteration; following re-assignation of objects to clusters, the cluster mean is recalculated.

For quantitative data, the Euclidean distance used, denoted by d_{ij} between object i and cluster j is:

$$d_{ij} = \sqrt{\sum \left(x_i^{(j)} - x_j\right)^2} \tag{4}$$

where: $x_i^{(j)}$ is data point from de cluster j and x_j is the center of the cluster j.

For obtaining some high tech and accurate models, the program users must have knowledge about the underlying modeling algorithms such as: forecasting, classification, segmentation and association detection. Current Automated Meter Management (AMM) systems lack a data mining procedure for load profiling and prediction in their smart metering component.

As it will be seen in the measurement data analysis Section 3, at present, because only components of Automated Meter Reading (AMR) exist, the measured data analysis takes a lot of time and the utilization of other software solutions and the utilization of data mining are mandatory.

C. Load forecasting

As power use increases, so grows the importance of information regarding power consumption and generation.

Having information in advance regarding the power consumption makes it possible to schedule renewable energy generators such that they operate at the lowest possible cost. In order to have information in advance, there is a need for load forecasting. In order to achieve load forecasting, it is necessary to establish characteristics of data set through statistical analysis, to find similarities in all load curves and make clustering analysis for typical load profiles using information about the type of consumer and historical data.

Table 1 presents: the level of accuracy, data acquisition source and the method for calculating the load curve, as reported in literature table 2.

Table 1. Load curve method profiling

TYPE LOAD CURVE	ACCURACY	COST FOR OBTAINING DATA
Dynamic	*High*	Very High
Adjusted Static	*Medium-high*	High
Static	*Medium*	Medium
Similar	*Medium-low*	Medium
Net	*Low*	Low

Table 2. Load curve method profiling

TYPE LOAD CURVE	DATA SOURCE	METHOD
Dynamic	Electronic meter with load curve recording	ANNs , K-means, Wavelet - Fuzzy transform, VSTF
Adjusted Static	Historical database with meteorological factors	Long, medium, short-term load forecasting
Static	Historical measured data	Statistical
Similar	Less Historical measured data	Statistical
Net	System load curve mixed with electronic meter data *Low*	Statistical graphics

From Table 1, it can be noticed that the highest accuracy is achieved when profiling using the dynamic load curves based on data obtained at high cost, because a machine learning algorithm is used for this case. One of the challenges is to address similarity measures that are used to make clusters. Similar load curves are grouped together based on calculating similarity among load curves using similarity measures [12].

In paper [10] authors presented the similar load curve, these curves are constructed using historical data from similar days in the past. The clear advantages for this method are that is using less historical data and it takes account of meteorological data as show in Table 2. Net load curve this type of load curve uses also a unique curve for modeling consumers and captures one aspect of forecasting variability. The precision can become much higher if more electronic meters would be installed data as show in table 2.

In this work, static and dynamic profiling method will be used and then the k-means algorithm is chosen to construct the typical daily shape.

3. EXPERIMENTAL DATA ANALYSIS

A. Data set

The smart metering system installed in Romania has developed only the concept of AMR and uses different software, distinctive for each metering equipment producer. Typical data collected using the AMR systems for a commercial consumer are presented in Table 3. Only the data related to one working day and one non-working day were selected for representation. From the AMR system can be selected as electrical parameters: active and reactive power, active and reactive energy and power factor. Measurements can be displayed at 15 minutes, 30 minutes and 1 hour. To achieve case study the authors chose as the parameter the active power in kW.

Table 3. Measured data

Hour	kW	kW	Hour	kW	kW
1	24.89	28.89	13	41.36	38.1
2	25.59	28.3	14	47.24	34.24
3	25.5	25.88	15	33.93	27.29
4	24.31	25.92	16	34.51	31.34
5	22.24	25.04	17	38.67	28.28
6	24.54	26.02	18	38.09	27.95
7	25.35	24.27	19	34.06	24.49
8	26.75	28.82	20	40.77	31.23
9	21.37	30.61	21	53.37	25.52
10	24.54	26.03	22	37.78	26.92
11	33.16	29.27	23	34.41	30.29
12	41.27	37.71	24	26.55	32.49

In the process of creating load curves, some criteria must be taken into consideration; one of this is to select data from a working day (WD) and a non-working day (NWD). In the study presented in paper [13], the authors have performed an analysis of the daily diagram of the supply voltage for a working day and for a non-working day. The choice to perform the analysis separately for working days and for non-working days was made in order to obtain higher accuracy of the results.

B. Data set analyses

FIG. 3 shows the load curve for WD for 22 days over 24 hours for the month of May 2016 for a commercial consumer. The maximum value for WD in the month of May was 56.24 kW on the 30th May at 10.00 am and the minimum value was 19.24 kW on May the 24th at 10.00. As it can be observed in FIG. 3, it is difficult to identify curves with high similarity.

The data set analysis is useful both statistically and with respect to clustering as it provides an overall picture of the power consumed and the energy needed from renewable energy sources.

FIG. 4 shows the graphical representation of the mean values from table 2. We can notice that the curve represented in FIG. 4 has two peaks, which are very important when one wants to buy energy on the electricity market. Also, from FIG. 4 it can be observed that the values of the minimum and maximum averages are similar to those in FIG. 3 representing the load curves for working days in the month of May.

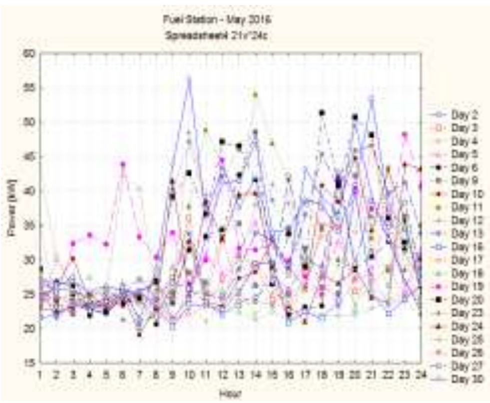

FIG. 3. Load curve for the month of May 2016

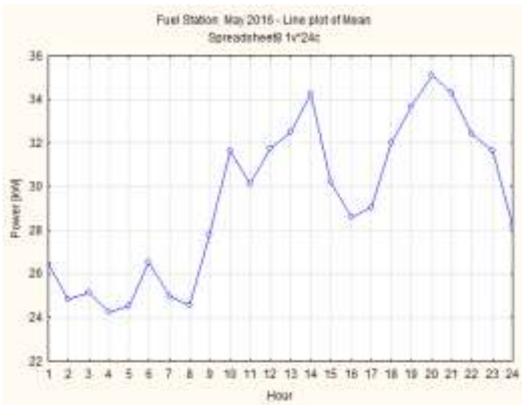

FIG. 4. Load curve with mean values

Using equations 1-3, table 4 presents the statistical analysis (minimum, mean maximum, standard deviation and coefficient of variation) for all WDs in the month of May 2016.

According to Section 2, the maximum value of the coefficient of variation is 30.12 % at 10.00 am; this means that the parameter mean is slightly representative.

Table 4. Statistical analysis for WD in month May

Hour	Min [kW]	Mean [kW]	Max [kW]	Standard deviation	Coefficient of variation [%]	Hour	Min [kW]	Mean [kW]	Max [kW]	Standard deviation	Coefficient of variation [%]
1	21.54	26.33	43.66	4.78	18.15	13	22.31	33.45	53.42	8.65	25.87
2	21.50	24.76	30.53	2.30	9.29	14	21.27	34.71	54.08	9.83	28.32
3	22.12	25.10	32.30	2.47	9.83	15	22.46	30.11	46.93	6.00	19.92
4	21.87	24.21	33.56	2.42	10.00	16	20.89	28.94	42.09	6.59	22.79
5	22.24	24.49	32.16	2.11	8.60	17	21.07	28.98	43.21	5.67	19.57
6	21.21	26.40	43.85	5.50	20.83	18	19.41	31.43	51.31	8.26	26.27
7	19.17	24.78	40.20	4.58	18.50	19	21.87	33.33	42.62	6.58	19.73
8	20.57	24.92	32.49	2.89	11.59	20	21.91	34.81	50.67	9.32	26.78
9	20.27	28.28	43.71	7.51	26.57	21	22.90	34.39	53.37	8.36	24.31
10	22.01	31.53	56.24	9.50	30.12	22	22.14	32.20	43.47	6.55	20.34
11	21.01	30.58	48.91	7.42	24.25	23	24.01	31.46	48.22	6.91	21.97
12	21.94	31.84	47.07	8.42	26.45	24	21.86	28.01	43.16	5.82	20.77

5. SIMULATION AND RESULTS

In order to perform the simulation, the data set was divided into two categories: WD and NWD, in the period from May 2016 until October 2016. Using the Statistica software, the cluster analysis with the k-means algorithm was performed. The data set used as input into the Statistica software package was all WD and NWD in the period from May 2016 until October 2016.

The decision to select this particular period of 6 months is based on the fact that in Romania the peak electricity consumption has moved from the winter months into the summer months. This is due to the preferential use of natural gas for heating (relevant mostly to winter months) versus the use of electricity for cooling (through the use of air-conditioning units) in the summer months.

Table 5 shows the statistical analyses for mean values of each month for all WDs in a 6 months period. The coefficient of variation has many values lower than 17%; therefore, the mean (average) for 6 month is representative and can be used as cluster mean. FIG. 5 shows the details of the statistical graphs using Box plots for mean (average) on 6 months. This diagram, referring to the results of Table 5, aims to achieve a qualitative data representation by summarizing the five statistical parameters: maximum, minimum, average, median and standard deviation.

In FIG. 5 it can be observed that, for the period analyzed, the lowest WD consumption is in the month of July and the highest is in September. Generally, the load variation was achieved in the summer months.

Table 6 shows the statistical analysis of the mean values of each month for all NWDs in a 6 months period.

Table 5. Statistical analyses for values of means for WD

WD	Mean May	Mean June	Mean July	Mean Aug.	Mean Sept.	Mean Oct.	Mean 6 Month	Coefficient of variation
[H]	[kW]	[kW]	[kW]	[kW]	[kW]	[kW]	[kW]	[%]
1	26.33	30.17	26.40	25.78	26.30	29.14	27.35	4.68
2	24.76	27.18	25.20	25.74	24.25	28.31	25.91	2.88
3	25.10	25.32	25.19	25.56	23.81	28.69	25.61	5.49
4	24.21	24.08	24.13	24.64	24.30	29.62	25.16	5.31
5	24.49	23.59	24.52	24.78	24.76	29.26	25.23	8.52
6	26.40	23.89	24.29	24.90	24.32	29.61	25.57	7.88
7	24.78	24.48	25.14	25.56	25.94	31.09	26.16	12.56
8	24.92	24.48	22.94	25.09	26.57	30.43	25.74	7.05
9	28.28	23.68	27.05	26.31	26.62	34.85	27.80	13.27
10	31.53	26.94	30.00	31.06	31.98	40.05	31.93	19.22
11	30.58	34.13	32.40	33.82	33.56	40.15	34.11	8.01
12	31.84	33.30	34.60	36.78	35.07	37.85	34.91	11.16
13	33.45	35.35	36.05	36.11	35.20	36.22	35.40	6.03
14	34.71	36.23	37.34	37.42	37.37	41.24	37.38	7.95
15	30.11	38.13	37.70	37.12	37.27	39.72	36.67	11.36
16	28.94	38.27	39.56	36.91	36.57	40.69	36.82	12.55
17	28.98	36.92	36.34	38.23	36.89	36.29	35.61	9.88
18	31.43	35.48	40.04	35.84	39.29	35.52	36.27	12.19
19	33.33	36.74	37.78	36.93	37.99	39.09	36.98	3.73
20	34.81	34.59	41.56	40.96	38.85	40.72	38.58	12.05
21	34.39	40.72	40.57	40.52	44.08	40.13	40.07	11.02
22	32.20	37.76	37.55	37.18	36.55	35.55	36.13	3.49
23	31.46	38.49	36.65	31.53	32.64	30.64	33.57	10.72
24	28.01	35.19	32.18	29.30	29.65	28.22	30.43	7.28

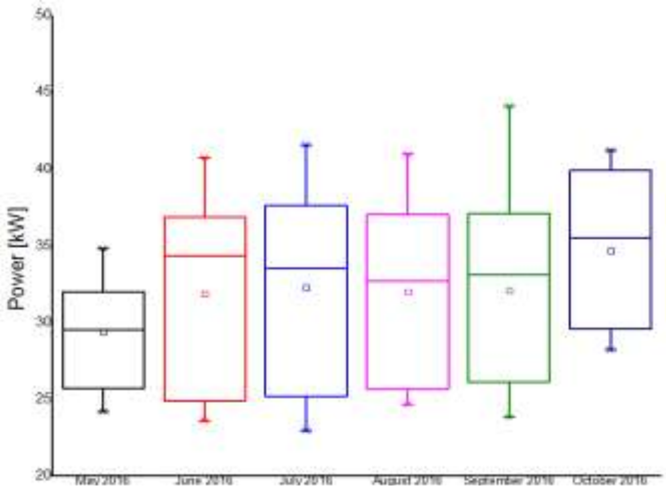

FIG. 5. Statistical graphics with Box plots for WD

Table 6. Statistical analyses for values of means for NWD

NWD	Mean May	Mean June	Mean July	Mean Aug.	Mean Sept.	Mean Oct.	Mean 6 Month	Coefficient of variation
[H]	[kW]	[kW]	[kW]	[kW]	[kW]	[kW]	[kW]	[%]
1	26.33	30.17	26.40	25.78	26.30	29.14	27.35	4.68
2	24.76	27.18	25.20	25.74	24.25	28.31	25.91	2.88
3	25.10	25.32	25.19	25.56	23.81	28.69	25.61	5.49
4	24.21	24.08	24.13	24.64	24.30	29.62	25.16	5.31
5	24.49	23.59	24.52	24.78	24.76	29.26	25.23	8.52
6	26.40	23.89	24.29	24.90	24.32	29.61	25.57	7.88
7	24.78	24.48	25.14	25.56	25.94	31.09	26.16	12.56
8	24.92	24.48	22.94	25.09	26.57	30.43	25.74	7.05
9	28.28	23.68	27.05	26.31	26.62	34.85	27.80	13.27
10	31.53	26.94	30.00	31.06	31.98	40.05	31.93	19.22
11	30.58	34.13	32.40	33.82	33.56	40.15	34.11	8.01
12	31.84	33.30	34.60	36.78	35.07	37.85	34.91	11.16
13	33.45	35.35	36.05	36.11	35.20	36.22	35.40	6.03
14	34.71	36.23	37.34	37.42	37.37	41.24	37.38	7.95
15	30.11	38.13	37.70	37.12	37.27	39.72	36.67	11.36
16	28.94	38.27	39.56	36.91	36.57	40.69	36.82	12.55
17	28.98	36.92	36.34	38.23	36.89	36.29	35.61	9.88
18	31.43	35.48	40.04	35.84	39.29	35.52	36.27	12.19
19	33.33	36.74	37.78	36.93	37.99	39.09	36.98	3.73
20	34.81	34.59	41.56	40.96	38.85	40.72	38.58	12.05
21	34.39	40.72	40.57	40.52	44.08	40.13	40.07	11.02
22	32.20	37.76	37.55	37.18	36.55	35.55	36.13	3.49
23	31.46	38.49	36.65	31.53	32.64	30.64	33.57	10.72
24	28.01	35.19	32.18	29.30	29.65	28.22	30.43	7.28

The statistical graphs using Box plots for mean (average) for a 6 months period for NWD is shown in FIG. 6. This shows that for the period analyzed, the lowest NWD consumption was in the month of September, whereas the highest relates to the month of June.

Due to the high variability of the load, the second stage of the data mining, the cluster analysis, is considered as a prediction of future typical daily shapes. Such information is useful for all the participants in the energy market as it shows the peaks and dips related to the next day, as resulting from the cluster analysis calculations using the Statistica software package, based on the k-means algorithm. The number of clusters is set to 2 and the number of iterations is set to a maximum of 10.

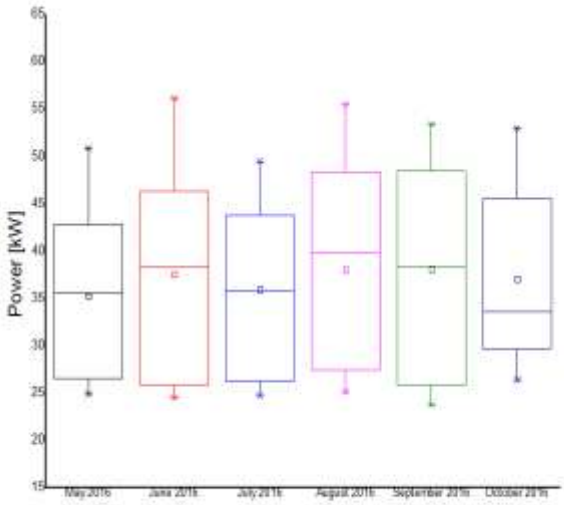

FIG. 6. Statistical graphics with Boxplots for NWD

Table 7 shows the result for cluster analysis with mean (average) for WD in 6 months. In the data mining process, the software Statistica has put in cluster 1 higher values and in cluster 2 lower values.

Using statistical analysis, the mean (average) for WD was verified. The data mining procedure was applied and the mean values can be used for energy load management and price forecasting as typical daily shapes.

Table 7. Results for clustering analyses and means for WD

Hour	Cluster 1	Average	Cluster 2	Hour	Cluster 1	Average	Cluster 2
1	28.77	27.36	25.91	13	41.10	35.26	29.22
2	26.94	25.91	24.86	14	44.53	37.30	29.85
3	26.25	25.61	24.95	15	43.80	36.72	29.41
4	25.60	25.14	24.67	16	43.24	36.78	30.11
5	25.87	25.22	24.54	17	42.34	35.68	28.82
6	25.70	25.56	25.41	18	41.81	36.37	30.75
7	26.68	26.17	25.65	19	42.22	37.04	31.69
8	26.23	25.67	25.09	20	43.95	38.64	33.17
9	29.50	27.66	25.76	21	46.74	40.09	33.24
10	35.84	31.89	27.82	22	41.40	36.20	30.82
11	39.34	34.02	28.54	23	36.70	33.60	30.40
12	40.99	34.91	28.64	24	32.93	30.47	27.93

The values for clusters and mean (average) for NWD over 6 months are shown in table 8.

Table 8. Results for clustering analyses and means for NWD

Hour	Cluster 1	Average	Cluster 2	Hour	Cluster 1	Average	Cluster 2
1	29.12	28.47	28.40	13	50.88	47.83	47.60
2	26.36	26.70	26.70	14	52.60	49.11	48.91
3	26.10	26.15	26.14	15	54.59	50.51	50.33
4	25.54	25.25	25.31	16	52.82	47.92	47.77
5	25.65	25.87	25.88	17	50.64	46.69	46.54
6	25.88	26.26	26.37	18	50.21	45.87	45.89
7	25.96	26.08	26.10	19	49.40	45.63	45.52
8	26.65	26.37	26.48	20	47.06	43.90	43.77
9	30.03	28.84	28.95	21	45.34	42.36	42.16
10	38.66	36.09	35.96	22	42.78	40.62	40.36
11	43.55	40.00	39.96	23	33.82	33.41	33.32
12	51.17	46.88	46.70	24	31.00	30.55	30.64

The curves corresponding to the data shown in table 7 are represented using 3D graphics in FIG. 7. The mean (average) for NWDs can be used as typical daily shape. FIG. 8 shows the 3D graphics for the data set values in table 8.

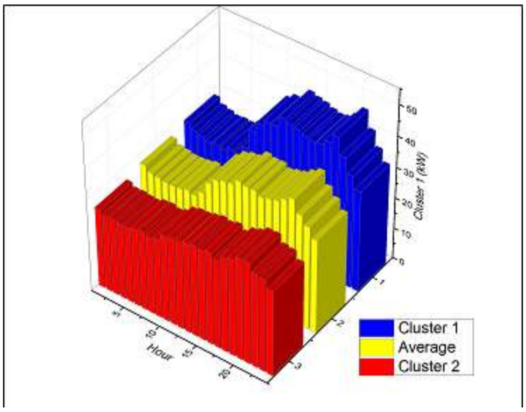

FIG. 7. The 3D graphics for cluster and mean in WD

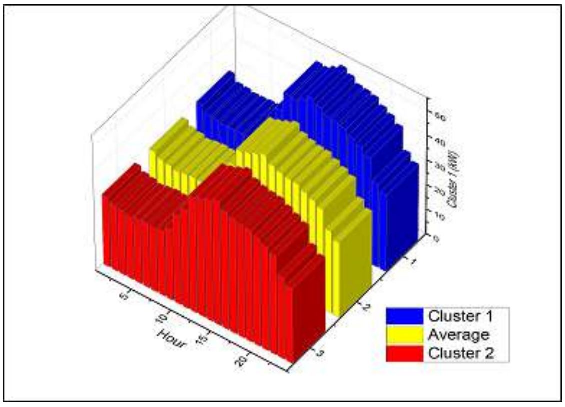

FIG. 8. The 3D graphics for cluster and mean in NWD

It was necessary to use a two weeks time frame in order to perform this analysis because component AMR does not have the possibility to arrange and to format data, so in order to prepare the data set, several attempts were necessary. With our proposed method of data mining for a typical daily shape, the AMR system can easily be implemented into intelligent metering systems as solution for smart metering.

5. CONCLUSIONS

The data set achieved through the Automated Meter Management (AMM) system is the most important part in Smart Metering for electrical energy consumption patterns. The current paper presented a data mining procedure for generating typical daily shapes, designed and tested for a commercial consumer.

From the commercial consumer statistical analysis accomplished, it results that the first stage of the data mining procedure introduces only the average standard deviation parameters and standard deviation to indicate whether the average is representative for the data set analyzed, these parameters being easy to interpret for all actors involved in smart metering.

The second stage of the data mining procedure includes the cluster analysis using an automated learning algorithm, necessary for generating typical daily shapes. This analysis requires the use of specific software which makes it troublesome for some of the users who need to familiarize themselves with these tools.

By using the statistical programs through the k-means clustering algorithm analysis, the production of typical daily shapes was achieved with the commercial consumer, and, as a result, the average values of the data set for the six-month period under analysis may be used as typical daily shapes.

For the case when the average values are different from the accomplished clusters, the analysis is performed again using algorithms for load forecasting. The main advantage of this procedure is that it uses a mathematical apparatus which is easy to understand by all actors involved in energy load management and price forecasting in order to support the concept of Smart Grids. For energy load management and price forecasting, this procedure brings the market actors closer to the final consumer, by offering the possibility of holding information related to the financial management of the energy consumed or produced in Smart Grids.

REFERENCES

[1] R. Al-Otaibi, N. Jin, T. Wilcox and P. Flach, Feature Construction and Calibration for Clustering Daily Load Curves from Smart-Meter Data, *IEEE Transactions on Industrial Informatics*, vol. 12, no. 2, pp. 645-654, April 2016.

[2] L. I. Minchala-Avila, L. Garza-Castañón, Y. Zhang and H. J. A. Ferrer, Optimal Energy Management for Stable Operation of an Islanded Microgrid, *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1361-1370, Aug. 2016.

[3] A. Khosravi and S. Nahavandi, Load Forecasting Using Interval Type-2 Fuzzy Logic Systems: Optimal Type Reduction, *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1055-1063, May 2014.

[4] C. E. Borges, Y. K. Penya and I. Fernández, Evaluating Combined Load Forecasting in Large Power Systems and Smart Grids, *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1570-1577, Aug. 2013.

[5] W. Labeeuw and G. Deconinck, Residential Electrical Load Model Based on Mixture Model Clustering and Markov Models, *IEEE Transactions on Industrial Informatics*, vol. 9, no. 3, pp. 1561-1569, Aug. 2013.

[6] J. Widén, M. Lundh, I. Vassileva, E. Dahlquist, K. Ellegård, E. Wäckelgård, Constructing load profiles for household electricity and hot water from time-use data-Modeling approach and validation, *Energy and Buildings* 41, no. 7, pp. 753-768, 2009.

[7] F. McLoughlin, A. Duffy, M. Conlon, A clustering approach to domestic electricity load profile characterization using smart metering data, *Applied energy*, 141, pp.190-199, 2015.

[8] C. Mihai, E. Helerea, L. Sângerzan, Characterization of voltage variations through the application of experimental statistics, *International Conference on Optimization of Electrical and Electronic Equipment (OPTIM)*, Bran, pp. 198-203, 2014.

[9] C. Mihai, I. Lepadat, E. Helerea, D. Calin, Load curve analysis for an industrial consumer, *12th International Conference on Optimization of Electrical and Electronic Equipment*, Basov, pp. 1275-1280, 2010.

[10] C. Mihai, D. Ilea, P. M. Mircea, Use of load profile curves for the energy market, *International Conference on Development and Application Systems (DAS)*, Suceava, 2016, pp. 63-70, 2016.

[11] C. Mihai, C. Popa, P. M. Mircea, Load profiling for gas stations using cluster techniques, *IEEE International Power Electronics and Motion Control Conference (PEMC)*, Varna, Bulgaria, pp. 1041-1048, 2016.

[12] T. Teeraratkul, D. O'Neill, S. Lall, Shape-Based Approach to Household Load Curve Clustering and Prediction. arXiv preprint arXiv:1702.01414, 2017.

[13] B. Dolník, Contribution to analysis of daily diagram of supply voltage in low voltage network: Working days versus non-working days, *2015 16th International Scientific Conference on Electric Power Engineering (EPE)*, Kouty and Desnou, 2015, pp. 373-376.