

Analysis of the Energy Usage in University Buildings: The Case of Aristotle University Campus

Iliana N. Pappi, Nikolaos G. Paterakis, João P. S. Catalão
Univ. Beira Interior, Covilhã, Portugal, and
INESC-ID, Inst. Super Tecn., Univ. Lisbon, Lisbon, Portugal

Ioannis Panapakidis, Grigoris Papagiannis
Aristotle University of Thessaloniki
Thessaloniki, Greece

Abstract—In this study the case of the energy consumption profile of the Aristotle University of Thessaloniki, in Greece, is presented and statistically analyzed by clustering methods on the basis of seasonal daily load curves and load shape factors, using data from real-time measurements. The results indicate that the categorization of active power demand in university buildings is an extremely useful tool for understanding and predicting the seasonal, hourly and daily energy consumption changes, which is the first step towards adopting energy efficiency policies in such scale premises as well as performing demand-side actions aiming to achieve a more economical and environmentally sustainable energy usage.

Index Terms—Energy Efficiency; Load Profiling; Time Domain Clustering; Statistical Analysis.

I. INTRODUCTION

Research in the field of energy behavior of buildings and energy efficiency has drawn significant attention recently both in theory and in industry. Almost 40% of the global energy production is estimated to be consumed in residential and non-residential buildings [1]. The main loads that need to be electrified in such extensive premises include air-conditioning (A/C) and electrical heating, a large number of computers and other appliances and of course, lighting.

As reported in [2], the behavior of the electricity users has a direct impact on the assessment of the energy usage efficiency of buildings, especially when excessive energy use occurs during times or even whole periods in which it is not needed. As a result, studies in the relevant research field take into account a series of parameters towards reducing the total energy requirements of buildings while ensuring the comfort of its users [2, 3].

In this regard, the way in which a building is heated, cooled and illuminated, the local climate, the average sunshine and its impact on the lighting requirements, even structural and architectural details and elements of the premises are investigated. Naturally, the use of a specific building, as well as the type of work that is conducted by its occupants is directly linked with the electrical and thermal demand of the premises and should be considered in relevant studies [4].

Essential part of the relevant studies is also the macroscopic statistical analysis of electricity consumption data, related to different type of buildings or group of buildings located in different areas or even countries. Several criteria are used to compare buildings, such as their usage, the local climate, cultural elements of the electricity users and the general economic situation of the area they are located in.

Finally, increased environmental sensitivity has also led to studies that examine the carbon footprint of buildings both in the developed and the developing countries. The main impact index considered is the CO₂ emissions [1, 4, 5].

University buildings are also separately examined in many studies. Gul and Patidar [6] examined the relationship between user activities and related energy demand profiles in a sample university building and concluded that Building Energy Management System controlled buildings are negligibly sensitive to user profiles.

Burgas et al. [7] provided a Multiway Principal Component Analysis method based analysis for a university building to model the time dependencies of variables that majorly vary the power profile.

Domingues et al. [8] proposed a power monitoring system based strategy to extract the load profile of a university building including load patterns, fault information etc. to use for predicting future power consumption in order to enable energy savings.

Moran et al. [9] applied a reduction technique to enable the projection of the data that characterize the consumption profile of a university building.

Yarbrough et al. [10] facilitated the relationship between individual buildings and total of overall buildings in a university campus by evaluating the pattern of energy use across time and day affected by different user habits.

Gallachoir et al. [11] explained performance indicators to enhance the examination of energy trends and accordingly the building energy performance in a university building.

Nevertheless, none of the aforementioned studies and others not mentioned here has presented a complete analysis for university buildings regarding the typical consumption patterns.

Hence, the purpose of this analysis is to lead to a vivid depiction and understanding of the behavioral consumption patterns and the load profiling of buildings such as a university campus through clustering techniques, leading to solutions for increasing energy efficiency and lowering electricity bills. More specifically, this study investigates the case of the main campus of Aristotle University of Thessaloniki (AUTH) in Greece. AUTH is the largest university in Greece and in the Balkans with more than 95,000 students, more than 4,000 staff and a campus that covers 334,000 m^2 .

The remainder of this paper is organized as follows: Section II describes the data acquisition methodology and highlights the need for a further analysis of the energy consumption data. Then, in Section III the employed clustering methodology is described. Section IV presents and discusses the obtained results. Finally, conclusions are drawn in Section V.

II. DATA ACQUISITION

The data processed in this study were acquired by the measuring system of the Greek Public Power Corporation (PPC) installed in AUTH campus that is connected to the medium-voltage network. Measurements regarding power consumption are recorded with a sufficiently high granularity and span 12 months (June 2010 – May 2011). The PPC meters chosen for the measurements are placed in the 8 main buildings of the campus, namely the departments of Veterinarian School, Dentist, Physics, Engineering, Chemical Engineering, Hydraulics, Law School and Central Library.

The initial approach to the data processing is to conduct and analyze relevant comparisons in order to correlate the consumption patterns of each building with its respective usage. Preliminary macroscopic conclusions were drawn for each department's premises based on the logs of the utility meters spanning all the 12 months of the study horizon. More specifically, concerning active power load curves which have the most evident contribution in such premises electricity bills, the findings indicate an unreasonably large amount of electricity consumption in proportion to expected needs during several hours (nights, non- typical working hours), days or months.

Two comparative figures of peak and base active power consumption (in kW) for the installed 8 meters in the whole examined period are presented in Figs. 1 and 2. Some specific departments evidently have a large amount of base load consumption directly linked to the lack of energy efficiency policies for such premises, even failure of equipment and lack of consideration from the users of the buildings point of view. This is also evident in the electrical energy (in kWh) profile for the whole year portrayed in Fig. 3, for the case of the Engineering Department. In Fig. 3 the division between working and non-working hours was chosen, considering that the typical working hours in Greece are between 8:00am and 16:00pm. The rest of the hours per day were considered as non-working hours. As it can be observed, 56% of the yearly energy is consumed during the non-working hours, a fact that implies both inefficient energy usage and atypical user behavior in comparison to other sectors, i.e. commercial.

These findings highlight that a thorough statistic-based analysis with clustering evaluation computational methods is essential to identify characteristic load patterns. In this study, 2 out of the 8 installed meters are selected to be analyzed, namely the meters corresponding to the Dentists School and the Central Library, as two examples of electrical consumption on the “edge” of the overall behavior of all the 8 departments of the main campus.

III. EMPLOYED CLUSTERING METHODOLOGY

A. Pre-processing of the datasets

There are many ways of pre-processing data introduced in the literature. In this study, the data pre-processing comprises three stages.

In the first stage, the bad measurements that are caused by malfunctioning of the measuring system are removed from the data base [12]. Such measurements are easily identified since they appear as zeros during limited time intervals.

In the second stage, anomalous daily curves are omitted from the sample [13, 14]. For the case of the Dentists School, the weekends were excluded since the functionality of the building is the one of a typical working premise. Furthermore, for the case of the Dentists School, all the calendar holidays for the Greek universities were excluded from the sample. The second dataset referring to the Central Library was handled in a different way. The main reason for this is that the functionality of the Central Library during weekends and also in some cases of public holidays is different than the other university buildings. As a result, only a few elements were excluded from the sample.

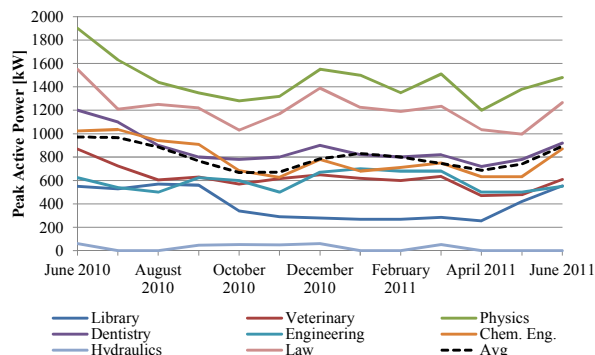


Figure 1. Peak active power demand during the study horizon for each department.

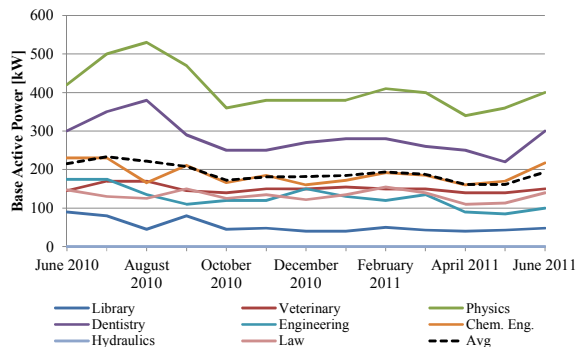


Figure 2. Base active power demand during the study horizon for each department.

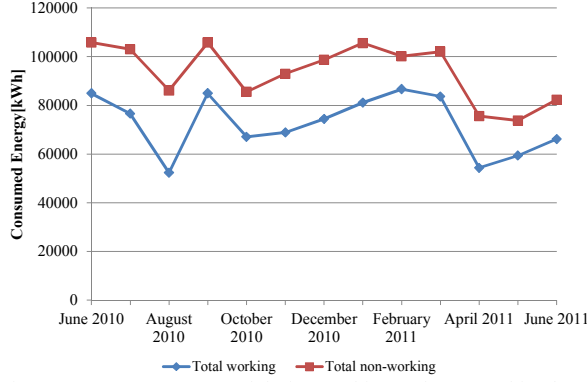


Figure 3. Energy Consumed during working and non-working hours (Engineering School).

The third stage of data pre-processing is the detection of outliers, which are defined as observations that appear to deviate markedly from other observations in the sample. In this case, having excluded false measurements due to temporary failure in the metering system and also outliers due to calendar position, still remains a significant amount of daily load curves which could be considered off-pattern due to random variations etc. In this case the decision of including or not these daily load curves in the sample under analysis should be facilitated by a statistical criterion.

In this study the employed criterion is the *modified z-score criterion* proposed by Iglewicz and Hoaglin [15], which is described by (1).

$$M_i = \frac{0.6745(x_i - \tilde{x})}{MAD} \quad (1)$$

In (1) x_i stands for each element of a vector containing all the observations of a given sample of daily load curves for a given time interval and \tilde{x} denotes the corresponding median. MAD stands for median absolute deviation and is defined by (2).

$$MAD = \text{median}(|Y_i - \tilde{Y}|) \quad (2)$$

This index is a variation of the average absolute deviation that is less affected by extremes in the tail, because the data in the tail have less influence on the calculation of the median than they do on the mean. Iglewicz and Hoaglin recommend that modified z-scores with an absolute value greater than 3.5 should be labeled as potential outliers and further examined. The results of the implementation of the method on the seasonal data sets of the Dentists School and the Central Library are presented in Table I, with modified z-scores suggesting each time a percentage of daily load curves as possible outliers, also indicating in which dates these observations occur.

The interpretation of the results of Table I is twofold. On the one hand, when the percentage of daily load curves with $M_i > 3.5$ is relatively small, it can clearly indicate daily load curves with an out-of-pattern behavior. On the other hand, as this percentage grows greater than 30%-40% it is observed that it underlines the existence of more than one separable groups in the initial sample, especially in the case of the Central Library building the functionality of which and as a result its daily load profiles, are more irregular.

TABLE I. PERCENTAGE OF POTENTIAL OUTLIERS IN EACH GROUP OF DATA

Building	Season	Number of daily curves	$M_i > 3.5$
Dentists School	Summer	66	7.58%
	Autumn	64	29.69%
	Winter	52	15.38%
	Spring	56	28.57%
Central Library	Summer	91	15.38%
	Autumn	87	35.63%
	Winter	81	70.37%
	Spring	84	46.43%

B. Clustering Process of seasonal daily load curves

B.1. Direct Clustering of Load Curves

As a first step, the clustering method employed in this study in purpose to create the class representative load curves [12] for each department and each season was the *k-means clustering method*. The curves considered as representative of each cluster are the centroids of k-means method. The k-means algorithm is a centroid model commonly used in load profiling, representing each cluster by a single mean vector, using partitioning of the data space into Voronoi cells.

In this study the set of p observations for each seasonal dataset, after the modified z-scores pre-processing, is described by (3) in which each observation is a 96-dimensional vector.

$$\mathbf{X} = (\vec{x}_1, \dots, \vec{x}_p) \quad (3)$$

K-means clustering aims to partition the observations into k ($k \leq p$) sets by minimizing the within cluster sum of squares (WCSS) as demonstrated by (4) and (5) in which μ_i is the mean of points in S_i .

$$\mathbf{S} = \{S_1, \dots, S_k\} \quad (4)$$

$$\text{arg} \min \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (5)$$

The number of clusters in each k-means clustering procedure was evaluated by the *Calinski-Harabasz criterion*, or as more commonly called the *variance ratio criterion (VRC)*. The Calinski-Harabasz criterion was chosen as it is considered the most suitable for k-means clustering solutions with squared Euclidean distances. In (6), SS_B is the overall-between-cluster variance and is defined by (7) in which k is the number of clusters, m_i is the centroid of the cluster i and m is the overall mean of the sample data. Additionally, SS_W is the overall-within-cluster variance defined by (8) in where k is the number of clusters, x is a data point, S_i is the i -th cluster and m_i is the centroid of cluster i .

$$VRC_k = \frac{SS_B}{SS_W} \times \frac{(N - k)}{(k - 1)} \quad (6)$$

$$SS_B = \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (7)$$

$$SS_W = \sum_{i=1}^k \sum_{x \in S_i} \|x - m_i\|^2 \quad (8)$$

B.2. Load Shape Factor Based Clustering

As a second step, a new clustering procedure was performed for Dentistry and Central Library through the implication of *load shape factors* [12-14]. Load shape factors are normalized features which highlight specific aspects of a load diagram. The shape factors employed in this study for each daily load curve are presented by (9)-(11). The selected shape factors are the ones corresponding better to the load pattern of a university building and the behavior of its users.

$$f_{D1} = \frac{p_{AV,day}}{p_{MAX,day}} \quad (9)$$

$$f_{D3} = \frac{p_{MIN,day}}{p_{AV,day}} \quad (10)$$

$$f_{D4} = \frac{p_{AV,night}}{p_{AV,day}} \quad (11)$$

Indicating the behavior of each building according to minimum, maximum and average daily values, load shape factor vectors corresponding to each daily load curve were created [14], described by (12).

$$\vec{x} = [f_{D1}, f_{D3}, f_{D4}] \quad (12)$$

The clustering process refers to these 3-dimensional vectors representing the amount of average daily power consumption comparatively to the maximum daily consumption, secondly the amount of minimum power consumption comparatively to the average and last, the amount of night hours (22:00-6:00) consumption comparatively to daily average.

IV. RESULTS AND DISCUSSION

A. Results on k-means clustering of daily load curves

The results of clustering on each daily load curve for each season for the chosen buildings are presented in Table II, after using the k-means method evaluated by the Calinski-Harabasz Criterion. Furthermore, in order to visualize the results in a more effective way, the representative load curves which are the centroids of each cluster are provided in Figs. 4 and 5. For the case of Dentists School, a comparative observation of the class representative load curves in Fig. 4 indicates a significant larger consumption during the summer period among all seasons. One of the reasons implicated in this pattern observed in representative load curves is the use of A/C systems due to very high temperatures in this time. Representative load curves of winter present the second larger in line consumption due to similar reasons of A/C systems used for heating, while autumn and spring present the lower representative consumption. This case study indicates how load profiling through statistical analysis facilitates demand-side actions through an expected behavior for university buildings, even electricity companies of improving the tariff offers due to these seasonal changes [13].

On the other hand, the class representative load curves depicted in Fig. 5 for the Central Library of AUTH, indicate a significant variation from a typical load curve of a building used within a usual working timetable.

This is easily shown by the high consumption represented by certain cluster centroids even in the late evening hours and an unpredicted behavior of the users. Also in this case the highest consumption depicted by 2 cluster centroids, for summer and autumn each, is easily explained by the use of A/C systems and also the inclusion of two large examination periods, in June and September respectively, in the Greek Universities.

B. Results on load shape factors k-means clustering

The results of the clustering procedure of load shape factor vectors for each load curve are summarized in Table III. Given the centroids of each seasonal cluster, the building load profiling is grouped by shape pattern. The observations are reinforced by Figs. 6 and 7.

TABLE II. CLUSTERING RESULTS OF DAILY LOAD CURVES

Season	Dentists School		Central Library	
	Cluster Centroids	Daily load curves per centroid – perc(%)	Cluster Centroids	Daily load curves per centroid – perc(%)
Summer	1	30 (46.15%)	1	59 (64.84%)
	2	25 (38.46%)	2	32 (35.16%)
	3	10 (15.38%)		
Autumn	1	18 (28.57%)	1	29 (33.33%)
	2	39 (61.9%)	2	58 (66.67%)
	3	6 (9.52%)		
Winter	1	40 (76.92%)	1	22 (27.16%)
	2	12 (23.08%)	2	59 (72.84%)
Spring	1	18 (33.96%)	1	27 (33.75%)
	2	35 (66.04%)	2	53 (66.25%)

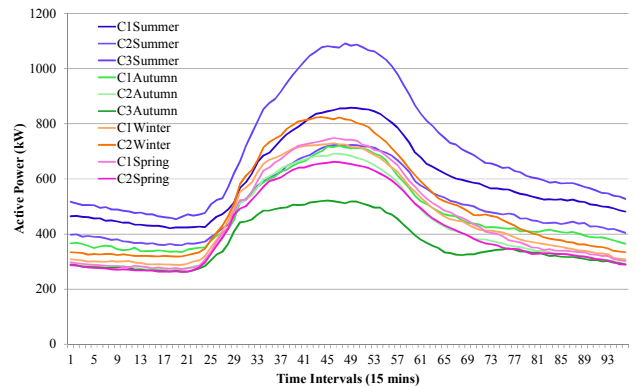


Figure 4. Representative Load Curves – Dentist School.

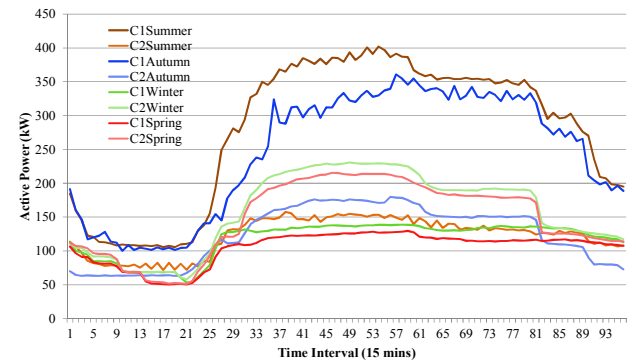


Figure 5. Representative Load Curves – Central Library.

TABLE III. CLUSTERING RESULTS OF LOAD SHAPE FACTORS

	Dentists School		Central Library	
	Cluster Centroids [f_{D1} f_{D3} f_{D4}]	Days per cluster-perc%	Cluster Centroids [f_{L1} f_{L3} f_{L4}]	Days per cluster-perc%
Summer	[0.66 0.67 0.26]	22 (33.8%)	[0.64 0.31 0.20]	46 (50.55%)
	[0.62 0.59 0.24]	23 (35.38%)	[0.73 0.55 0.28]	24 (26.37%)
	[0.73 0.75 0.28]	20 (30.77%)	[0.50 0.43 0.22]	21 (23.08%)
Autumn	[0.59 0.61 0.23]	38 (60.32%)	[0.59 0.36 0.20]	49 (56.32%)
	[0.66 0.67 0.26]	25 (39.68%)	[0.77 0.73 0.29]	9 (10.34%)
Winter	-	-	[0.64 0.52 0.22]	29 (33.33%)
	[0.60 0.58 0.22]	23 (44.23%)	[0.79 0.45 0.28]	23 (28.4%)
Spring	[0.61 0.61 0.23]	26 (48.06%)	[0.63 0.33 0.22]	57 (71.25%)
	[0.60 0.50 0.22]	2 (3.77%)	[0.79 0.46 0.30]	23 (28.75%)
	[0.60 0.58 0.22]	21 (39.62%)	-	-
	[0.66 0.70 0.26]	4 (7.54%)	-	-

The help of load shape factors to define the seasonal load pattern is noticeable. It is a second view in the analysis of clustering by representative load curves which was performed in the first step. In the case of the Central Library building, the majority of the curves belong to a rather smooth shape, with the relatively lowest distance between the maximum, minimum and average of the day consumption measurement. This is obvious by the number of vectors, representing the respective daily load curves, belonging to the 1st and 3rd cluster in summer, 1st and 3rd cluster in autumn, 2nd cluster in winter and 1st cluster in spring. Another key observation results from the group of clusters in the case of the Dentists School depicted in the 4th cluster where a small group of vectors present a significant distance between minimum and average values. These vectors represent their respective load curves which are characterized by a steeper shape.

V. CONCLUSIONS

In this study a thorough analysis of the energy consumption behavior of AUTH buildings was performed, initially by a yearly overview using 15-minute signals reported by the medium-voltage metering system in 8 buildings of the University Campus. This first comparison of yearly load curves indicated the need of a deeper research on the typical daily load profiling of such scale premises, as the consumption of electricity in these buildings significantly affects the electricity bills, especially by the fact of frequently observed inefficient energy usage. A two – step clustering of the yearly daily load curves in two major buildings of AUTH clearly depicted an expected energy consumption behavior through representative load curves and shape load factor cluster categorization and could contribute to demand-side management actions, along with being an important asset for energy pricing policies for major consumers by the power distribution enterprises.

ACKNOWLEDGMENT

This work was supported by FEDER funds (EU) through COMPETE, by Portuguese funds through FCT, under Projects FCOMP-01-0124-FEDER-020282 (Ref. PTDC/EEA-EEL/118519/2010) and UID/CEC/50021/2013, and by funds from the EU 7th Framework Programme FP7/2007-2013 under grant agreement no. 309048.

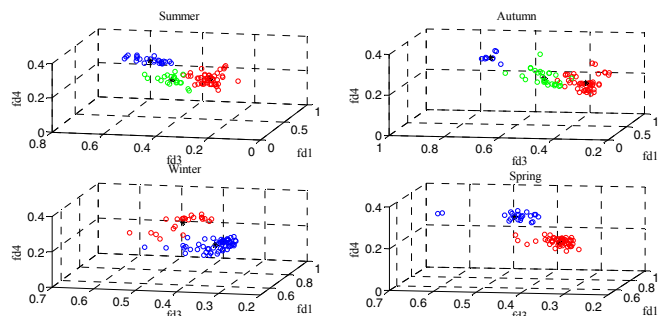


Figure 6. Load shape seasonal clustering – Central Library.

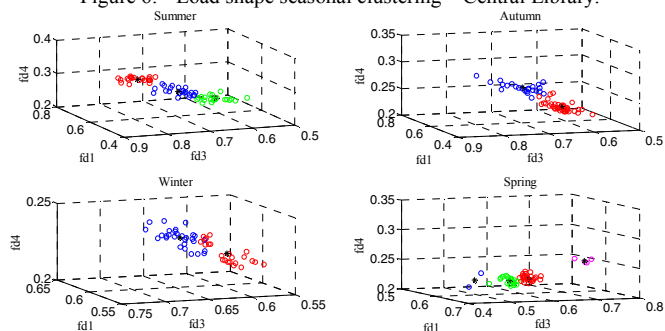


Figure 7. Representative Load Curves – Dentists.

REFERENCES

- [1] R. Saidur, "Energy consumption, energy savings, and emission analysis in Malaysian office buildings," *Energy Policy*, vol. 37, pp. 4104-4113, Oct. 2009.
- [2] O. T. Masoso, and L. J. Grobler, "The dark side of occupants' behaviour on building energy use," *Energy and Buildings*, vol. 42, pp. 173-177, Feb. 2010.
- [3] J. F. Nicol, and M. A. Humphreys, "Adaptive thermal comfort and sustainable thermal standards for buildings," *Energy and Buildings*, vol. 34, pp. 563-572, July 2002.
- [4] L. Perez-Lombard, J. Ortiz, and C. Pout, "A review on buildings energy consumption information," *Energy and buildings*, vol. 40, pp. 394-398, 2008.
- [5] W. L. Lee, and F. W. H. Yik, "Regulatory and voluntary approaches for enhancing building energy efficiency," *Progress in Energy and Combustion Science*, vol. 30, pp. 477-499, 2004.
- [6] M. S. Gul, and S. Patidar, "Understanding the energy consumption and occupancy of a multi-purpose academic building," *Energy and Buildings*, vol. 87, pp. 155-165, Jan. 2015.
- [7] L. Burgas, J. Melendez, and J. Colomer, "Principal component analysis for monitoring electrical consumption of academic buildings," *Energy Procedia*, vol. 62, pp. 555-564, 2014.
- [8] M. Dominguez, J. J. Fuentes, S. Alonso, M. A. Prada, A. Moran, and P. Barrientos, "Power monitoring system for university buildings: Architecture and advanced analysis tools," *Energy and Buildings*, vol. 59, pp. 152-160, April 2013.
- [9] A. Moran, J. J. Fuentes, M. A. Prada, S. Alonso, P. Barrientos, I. Diaz, and M. Dominguez, "Analysis of electricity consumption profiles in public buildings with dimensionality reduction techniques," *Engineering Application of Artificial Intelligence*, vol. 26, pp. 1872-1880, Sept. 2013.
- [10] I. Yarbrough, Q. Sun, D. C. Reeves, K. Hackman, R. Bennett, and D. S. Henshel, "Visualizing building energy demand for building peak energy analysis," *Energy and Buildings*, vol. 91, pp. 10-15, March 2015.
- [11] B. P. O. Gallachoir, M. Keane, E. Morrissey, and J. O. Donnell, "Using indicators to profile energy consumption and to inform energy policy in a university-a case study in Ireland," *Energy and Building*, vol. 39, pp. 913-922, Aug. 2007.
- [12] G. Chicco, R. Napoli, F. Piglion, P. Postolache, M. Scutariu, and C. Toader, "A review of concepts and techniques for emergent customer categorization," in *2002 TELMARK Discussion Forum European Electricity Markets*.
- [13] G. Chicco, R. Napoli, P. Postolache, and M. Scutariu, "Customer characterization options for improving the tariff offer," *IEEE Trans. Power Systems*, vol. 18, pp. 381-387, Feb. 2003.
- [14] I. Panapakidis, M. Alexiadis, and G. Papagiannis, "Evaluation of the performance of clustering algorithms for a high voltage industrial consumer," *Engineering Applications of Artificial Intelligence*, vol. 38, pp. 1-13, Feb. 2015.
- [15] B. Iglewicz and D. Hoaglin (1993), "Volume 16: How to Detect and Handle Outliers", in *The ASQC Basic References in Quality Control: Statistical Techniques*, E. F. Mykytka, Ed.