

## Pattern Recognition and Classification for Electrical Energy Use in Residential Buildings

Martina Ferrando<sup>1</sup>, Alla Marchenko<sup>2</sup>, Silvia Erba<sup>1</sup>, Francesco Causone<sup>1</sup>, Salvatore Carlucci<sup>2</sup>

<sup>1</sup>Department of Energy, Politecnico di Milano, Milano, Italy

<sup>2</sup>Department of Civil and Environmental Engineering, Norwegian University of Science and Technology, Trondheim, Norway

### Abstract

In the last years, researchers and energy utilities are showing a rising interest in the study and definition of actual buildings' energy uses. A key aspect of this investigation is the description of daily energy use patterns and their variability over the time. This paper discusses the application of machine learning techniques for pattern recognition with the implementation of a Self-Organizing Map (SOM) algorithm coupled with a *k*-means clustering algorithm on a dataset of registered electrical energy use in a residential building located in Milan. In the study, five clusters emerged with different daily patterns, that can be ascribed to different uses of electric appliances by people inside the flats.

### Introduction

The European Union, with the directives and regulations of the last decades, is highlighting the importance of increasing the energy performance of new and existing buildings (Panapakidis et al., 2014). In Italy, in 2016, the most financed energy conservation measure, with 1,17 billion euros spent, was the installation of heat pumps, together with new lighting systems and the increase of thermal insulation of building envelopes. (Joule Assets Europe, 2017). Thus, new and renovated efficient buildings are implementing electricity not only for lighting and appliances, but also for HVAC purposes. Indeed, during the last years, also due to the increase of competitiveness in electricity markets, utility companies are facing new challenges to decrease service costs, while maintaining a high-efficiency distribution (Chicco et al., 2004). Accordingly, energy management in buildings is a fundamental issue to improve energy efficiency, comfort and equipment life, as well as in reducing energy consumption and operational costs (Capozzoli et al., 2018). The temporal assessment of electricity use is of major importance also because it is still expensive to store it, and thus, it is usually produced at the consumption rate (Rhodes et al., 2014). It is fundamental for utility companies to maintain and improve their services continuously, to guarantee a sufficient energy supply to communities (Tso and Yau, 2007). In this electricity scenario, it is highly desirable for decision-makers, to identify and study the electrical behaviour of customers (McLoughlin et al., 2015). The installation overall Europe of advanced metering infrastructures, such as building

automation systems, Internet of Things solutions, smart meters and smart grids, is increasing the information on energy usage that can be registered and analysed (Rhodes et al., 2014). This quantity of collected raw data requires suitable processing and insight to extract fit for use information (Panapakidis et al., 2014). The datasets, which typically have 15-minute or 1-hour granularity (Rhodes et al., 2014), could be exploited to study the dynamic behaviour of electricity use (Capozzoli et al., 2018). To help the energy management of buildings, data mining techniques are available and beneficial tools. For example, clustering algorithms (a typical machine learning technique), are an effective approach to analyse time series data to extract pattern and recurrent behaviours (Rhodes et al., 2014).

This paper will focus on the application of machine learning techniques for pattern recognition with the implementation of a Self-Organizing Map (SOM) algorithm coupled with a *k*-means clustering algorithm on a registered dataset of electrical energy use. The case study is a residential estate composed of two buildings located in the South-East area of Milan, comprehending approximately 70 flats. While industrial parks and offices show regular operation behaviour due to working hours and closing days, thus are well documented (Capozzoli et al., 2017; Chicco et al., 2004; Dudek, 2016), the residential sector remains a more complex and unexplored area of study. The electricity patterns of residential loads depend from numerous variables, such as the number of people composing the family, as well as their activity, age, lifestyle, installed equipment, and appliances (Chicco, 2012). These characteristics are intrinsically stochastic and bring to a variety of daily patterns that cannot be easily analysed. Occupants' behaviour can bring to great differences in energy use among the same typology of buildings (Carlucci et al., 2017; O'Brien et al., 2017). For this reason, in the last years, researchers started to analyse data registered by energy companies and to exploit the results to improve the description of the stochastic nature of energy use in buildings (Carlucci et al., 2016). In particular, energy modellers could take advantage of these results to create internal heat gain schedules due to electrical energy uses to be used in dynamic simulation software.

In literature, numerous researches deal with the clustering and classification of electrical energy use datasets,

however, they are mainly focused on large-scale (city-level, regional, state (Deshani et al., 2014; Dudek, 2016; Tsekouras et al., 2007)) and/or non-residential buildings (Chicco et al., 2004; Hernández et al., 2012). Unsupervised machine learning algorithms are the typical method to identify patterns in datasets with any prior knowledge (Capozzoli et al., 2017).

Even in the less numerous researches about residential buildings, machine learning algorithms emerged as the leading method to perform pattern recognition of energy and electrical uses (Ali et al., 2016; McLoughlin et al., 2015; Rhodes et al., 2014; Viegas et al., 2015). Data mining and unsupervised machine learning methods fitted in the specific case study reported in this paper, because of their capacity of reducing noise in datasets and identifying patterns in multi-dimensional data. The case study here presented concerns a data sample of a residential building with 24 flats and no surveys.

## Dataset

The dataset presented in this paper concerns the anonymous registration of electrical uses, with a 15-minutes time step, from the 1<sup>st</sup> of February to the 31<sup>st</sup> of August 2016, for 24 flats in a residential estate. The multi-storey building built in the '80s, is part of the public housing stock owned by the Municipality of Milan that accounts for about 27 945 flats and 50 500 inhabitants (i.e. 24 684 families) (MM 2014). The population of the public housing stock includes people with different ages: 0-14 (8,36 %), 15-18 (4,49 %), 19-25 (23,50 %), 46-65 (30,54 %), over 66 (32,90 %). A few different nationalities are also represented, with the following share: Italy (83,17 %), Egypt (2,91 %), Morocco (2,30 %), Philippines (1,56 %), Sri Lanka (1,12 %), Peru (1,08 %), Ecuador (0,85 %), other (7,01 %). The multi-storey building follows the characteristics of the building stock, although a specific ethnographic analysis of its population is not available. Since it is made of raw data collected onsite, the database accounts some missing values and possible recording errors. Initially, the dataset has been investigated with Python 3.6 via Jupyter notebook. For this step, the following packages were used: pandas, matplotlib.pyplot and numpy. The loaded dataset has been converted to YYYY-MM-DD hh:mm:ss series format, according to ISO 8601, and all missing values have been changed to NaN (Not a Number). Afterword, line plots have been built for three, randomly chosen, flats. This step makes easier to track missing values in the collected data and to get insides on data spread. As result, in the example Figure 1, it is possible to spot abnormal 'jumps' among available data. After further checks, it was determined that, during dataset creation, some data were organised in the wrong order. To overcome it, data have been sorted using the sort\_index command. This step helped to organize data flow with respect to the day and the time. The new line plots are presented in the example Figure 2. Scatter plots have been, then, built to describe the overall distribution of energy use during the period of recording. As it can be seen in Figure 3, data distribution for each of the apartments has a high variation, and it is not possible

to determine any evident pattern in the data to create clusters. As last step in getting to know the dataset, a kernel density estimation was built for each of the 24 flats. Graphs may be observed at the example Figure 4.

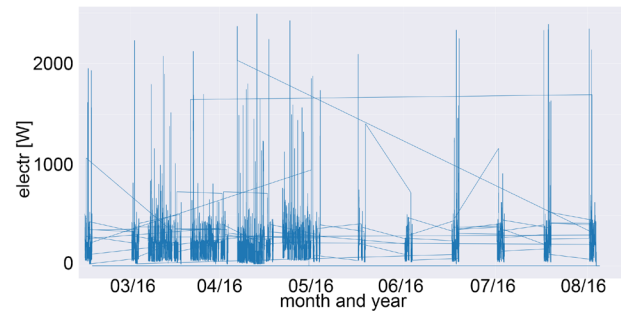


Figure 1: The Line Plot #4 Flat (original).

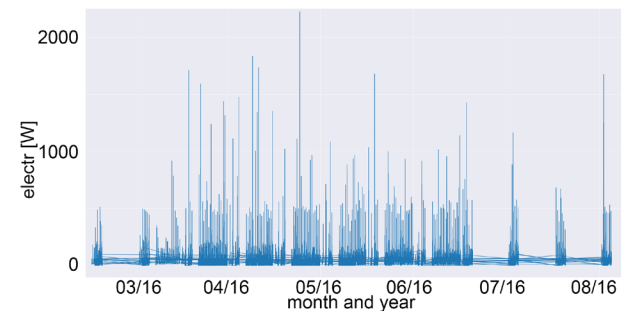


Figure 2: The Line Plot #16 Flat (sorted).

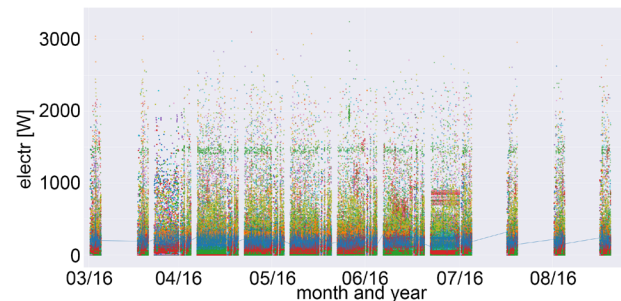


Figure 3: Scatter Plot for 24 Flats.

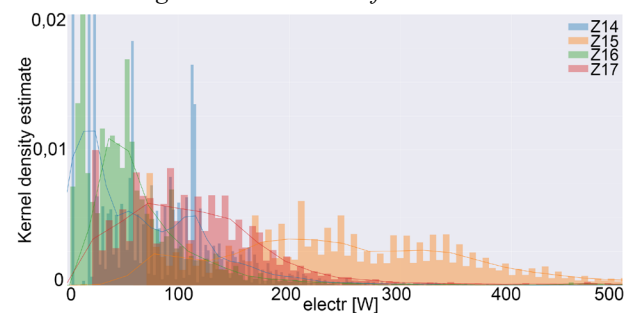


Figure 4: Kernel Density Distribution Flat 14-17.

## Methodology

The aim of the study is to derive from the raw registered data of electric use, different daily clusters resulting from different households' features and occupants' habits. The daily load curves present in the dataset will be clustered into few meaningful groups, each described by the first, the second and the third quartiles. These three curves will define three different scenarios of occupants ascribed to austere, normal and wasteful users. To implement the steps of the methodology, IBM SPSS Statistics 24 and

MATLAB R2017a are used. Figure 5 shows a schematic of the followed methodology.

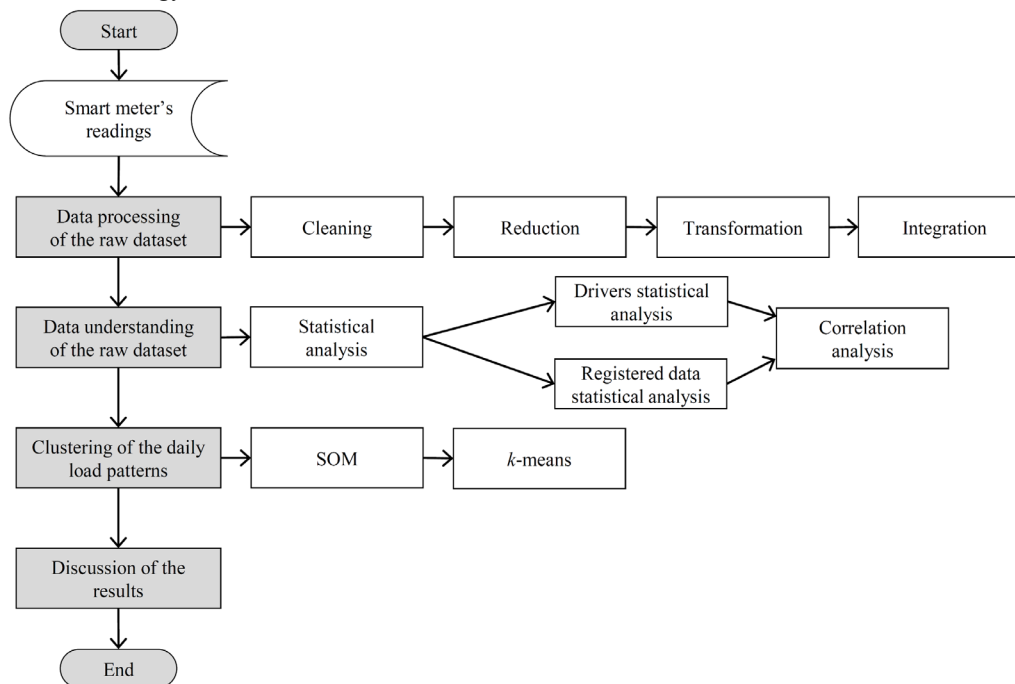


Figure 5: Methodology Schematic.

### Data processing

Data processing is the first and fundamental phase to obtain meaningful results. Usually, the collected raw data is incomplete and contains errors; therefore, several steps are required to create a working dataset that can be used for statistical analyses. The steps are: (i) data cleaning, consisting of identification of outliers and inconsistency removal; (ii) data reduction/discretization, that helps in simplifying data, while maintaining the meaningful characteristics; (iii) data transformation, consisting in the normalization or aggregation of data; and, (iv) data integration, where multiple datasets and attributes are unified into a single useable format.

### Data understanding

Data understanding gives an insight into the dataset through statistical techniques and basic summaries. The goal is to comprehend the dataset features and the relations between the different variables. The hypothesis is that some drivers (characteristics of the family or of the apartment) can trigger and hence explain the use of electricity in the residential sector. Thus, some statistical analyses are performed to validate this hypothesis. The data understanding is organized into four steps: (i) statistical analyses of the possible drivers (using univariate analysis, descriptive statistics, graphs to describe and explore the data); (ii) correlation analyses among drivers (using Pearson and/or Spearman correlations, T-test analysis and ANOVA analysis); (iii) statistical analyses of the registered data (using univariate analysis, descriptive statistics, graphs to describe and explore the data); (iv) correlation analyses between the registered data and the possible drivers (using Pearson and/or Spearman correlations, T-test analysis and ANOVA analysis).

### Clustering

Clustering means grouping a dataset into a  $N$  number of clusters  $C_i$ ,  $I = 1, 2, \dots, N$ . To solve the clustering problem, a two-level approach is implemented, that is formed by a combination of the SOM (Mitchell, 1997) and  $k$ -means (Piech, 2013) algorithm. The coupling of the two methods is effective to minimize the errors, the computational cost and to reduce the noise in the dataset (Hernández et al., 2012; Vesanto and Alhoniemi, 2000). The SOM creates protoclusters that are then grouped with the  $k$ -means in the final clusters. The protoclusters consist of local averages of the original samples and, thus, they are less sensitive to outliers. The set-up of the SOM algorithm, particularly the number of final protoclusters in which the dataset should be subdivided, is a fundamental step that can ease or jeopardize the results. In this paper, a 2-dimensional map with hexagonal lattice is used, sized with the heuristic formula suggested by the SOM Toolbox for MATLAB Report (Vesanto et al., 2000):

$$m = 5\sqrt{n} \quad (1)$$

in which  $m$  is the final number of protoclusters,  $n$  is the number of data sample given as input. The ratio of the side-lengths of the lattice is set as the ratio between the two biggest eigenvalues of the covariance matrix of the given data, and the actual side-lengths are then set so that their product is as close as possible to the desired  $m$ . Moreover, a normalization on the daily maximum is performed to improve the results. To help in defining the number of clusters resulting from the  $k$ -means algorithm the Davies-Bouldin Index is used. After several analyses, the number  $k$  of the final clusters in the  $k$ -means was fixed at 5.

## Results and discussion

### Data processing

During this phase, the records affected by inconsistency are deleted. Overall, 12 % of the dataset is missing or affected by errors and the registrations for flat 3 and 13 are completely missing, because, respectively, with closed contract and empty. Regarding the reduction of data, the time-step is increased from 15 minutes to 1 hour; thus energy records, being extensive variables are accumulated for each hour. This action also allows reducing the impact of possible eluded outliers. Afterwards, numerous drivers that are proved to influence the electric use in residential buildings are identified from the scientific literature and contrasted with the data available in the dataset. Table 1 summarizes the results of this analysis and shows the final exploitability in this case study. Table 2 reports the drivers that are integrated into the dataset with intervals of variation or options in order to run statistical analysis.

The used weather dataset is given by A.R.P.A. Lombardia. Precisely, the weather station of Via Juvara in Milan (Agenzia Regionale per la Protezione dell'Ambiente della Lombardia, 2019) is used (the closest registration point to the building site). The registrations include the *External-radiation*, the *External-temperature* and the hourly cumulative *Precipitation*. The *Month*, *Day-of-the-month*, *Day-of-the-week*, *Hour-of-the-day* are just temporal annotative variables that can help to understand the general trend of the electric use. The distinction between *Workdays* or *Not-working days* is set in accordance with the national holidays of the year 2016 and counting Saturday and Sunday as not working days, since usually, people in Italy go to work from Monday to Friday. The driver *Season-heating/cooling* is set according to Art. 9 of D.P.R. 26/08/93 (Presidenza della Repubblica Italiana, 1993). Milan, belonging to the climatic zone E, is characterized by a heating season that starts on the 15<sup>th</sup> of October and ends on the 15<sup>th</sup> of April. The driver *Day/Night*, *Day* is related to the time in which there is solar radiation during the shortest day of the year (from 8 a.m. to 4 p.m. on the 21<sup>st</sup> of December), whilst *Night* is set as the hours without solar radiation in the shortest night of the year (from 10 p.m. to 4 a.m. on the 21<sup>st</sup> of June). A third group is composed by the hours in between, that can change to be day or night during the year. In the building under study, bi-hourly tariffs contracts are used, thus electricity costs less during the nights, weekends and national holidays; the *Availability-of-cost-of-electricity* driver explicates this. The *Orientation*, *Flat-number*, *Floor-number* and *Number-of-bedrooms*, *floor-area* are all drivers used to describe the position of the flat in the building and its characteristics.

### Data understanding

In this step, the dataset is investigated through visualization graphs, descriptive statistics and correlation tests, to understand its characteristics and the relations among data samples. Along the registration period, the

electric use of the building decreases. However, calculating the hourly mean electric load per *Month*, July shows an hourly average comparable to February; January shows the highest value, whilst, August the lowest ones. Calculating the hourly mean electric load per *Day-of-the-week*, it is clear that the weekends correspond to a higher hourly average value compared to the weekdays, probably due to the increase of time spent at home by people. The daily curve of the hourly electric use of the building gives a first glimpse in the general usage pattern. The minimum use is registered around 4 a.m. in the early morning. The electric use increases till lunchtime around noon and stays constant during the afternoon, increasing again, till the maximum, that is registered in the evening, between 7 and 10 p.m. This period shows higher values than the rest of the day, but it is in line with the Italian lifestyle, i.e. people coming home from work in the evening, cooking dinner, using artificial lighting and leisure electric equipment (e.g., television, computers and radio).

The Spearman's rho correlation test is exploited in this case study because, working in ranks, it is able to evaluate the monotonic relationship between two continuous or ordinal variables. In the results, just weak or very weak correlation coefficients are registered (Rumsey, 2016). A moderate and an almost moderate correlation (0,39 and 0,29) is registered respectively with the number of rooms and with the floor area, justifiable with the probability of a higher number of installed electric appliances in larger flats.

### Clustering

In this section, the results of the two-levels approach composed by a SOM with the *k*-means algorithm are described. The final size of the SOM lattice is 8 x 42, for a total of 336 protoclusters. Before running the *k*-means algorithm the average of the normalized daily use in each protocluster is calculated. The final five clusters are shown in Figure 6. The light-grey lines represent the protoclusters in each cluster, and for each, the first, second and third quartiles are underlined.

The SOM plus *k*-mean method is able to subdivide into groups, days with similar daily patterns. 28 % of the days of the original dataset are grouped in cluster 5, 25 % are in cluster 4, 20 % are in cluster 1, 14 % are in cluster 2, and 13 % are in cluster 3. The green line is used for the first quartile and it could be the representation of the austere users (Q1 in the legend), the blue line represents the second quartile and the normal users (Q2 in the legend) and the orange line represents the third quartile and the wasteful users (Q3 in the legend).

Figure 7 shows the distribution of the five clusters in each dwelling. Flat number 9 and 19 are represented for more than 80 % of the days from just one cluster, respectively number 5 and 1. Others (i.e., numbers 6, 7, 12, 13, 14, 17) are mainly represented by two clusters. On the other hand, for some flats (e.g., number 1, 4) none of the five clusters is predominant respect to the others.

Table 1: List of drivers that can affect the electric energy use with related references.

Class of drivers	Driver	Reference
Location/Weather/Habit	External radiation <sup>♦</sup>	(Mardaljevic et al., 2009)
	External temperature <sup>♦</sup>	(Sandels et al., 2015)
	Workdays / Holidays <sup>♦</sup>	(Paatero and Lund, 2006)
	Day of the week <sup>♦</sup>	(Buttitta et al., 2017)
	Precipitation <sup>♦</sup>	-
	Cost of electricity <sup>♦</sup>	-
	Hour of the day <sup>♦</sup>	(Paatero and Lund, 2006)
	Heating/Cooling season <sup>♦</sup>	(Paatero and Lund, 2006), (Sandels et al., 2015)
	Renewables energy source available on site <sup>♦</sup>	(Galvin, 2016)
Flat and physical characteristics	House demand limit <sup>♦</sup>	(Capasso et al., 1994)
	Main orientation <sup>♦</sup>	(Mardaljevic et al., 2009)
	Floor number <sup>♦</sup>	(Menezes et al., 2012)
	Number of rooms <sup>♦</sup>	(Yohanis et al., 2008)
	Floor area <sup>♦</sup>	(Yohanis et al., 2008)
	Typology <sup>♦</sup>	(Yohanis et al., 2008)
	Insulation level <sup>♦</sup>	(Sandels et al., 2015)
Indoor environmental parameters	g-Value <sup>♦</sup>	(Mardaljevic et al., 2009)
	Shading type <sup>♦</sup>	(Tzempelikos and Athienitis, 2007)
Indoor environmental parameters	Indoor air temperature <sup>♥</sup>	(Sandels et al., 2015)
	Daylight Factor <sup>♥</sup>	(Mardaljevic et al., 2009)
Individual or family feature	Number of people <sup>♥</sup>	(Capasso et al., 1994), (Yohanis et al., 2008)
	Sex <sup>♥</sup>	(Capasso et al., 1994), (Yohanis et al., 2008)
	Age <sup>♥</sup>	(Shimoda et al., 2004)
	Income <sup>♥</sup>	(Capasso et al., 1994), (Yohanis et al., 2008)
	Occupation <sup>♥</sup>	(Capasso et al., 1994), (Yohanis et al., 2008)
	Shading operation <sup>♥</sup>	(Tzempelikos and Athienitis, 2007)
	Appliances Efficiency <sup>♥</sup>	(Menezes et al., 2012)
	Availability of an electric car <sup>♥</sup>	(Clement-Nyns et al., 2010)
Kind and number of Installed equipment <sup>♥</sup>	(Capasso et al., 1994), (Menezes et al., 2012)	
♥ drivers that are not available.		
♦ drivers that are not exploitable in this case study because they are constant all over the dataset.		
♦ drivers that are exploitable in our data sample and mark a difference among hours.		
♦ drivers that are exploitable in our data sample and mark a difference among flats.		

Table 2: List of selected drivers and their features.

	Name	Unit of measure	Type of variable	Range of variation - Continuous [Interval, step] - Categorical {discrete values}
AMONG HOURS/DAYS	External-radiation	W/m <sup>2</sup>	Continuous	[0 ≤ x ≤ 931,3]
	External-temperature	°C	Continuous	[1,6 ≤ x ≤ 33,8]
	Precipitation	mm	Continuous	[0 ≤ x ≤ 29,6]
	Month	-	Categorical	{2-8}
	Day-of-the-month	-	Categorical	{1-31}
	Hour-of-the-day	-	Categorical	{0-23}
	Day-of-the-week	-	Categorical	{1-7}
	Day/Night	-	Categorical	{-1; 0; 1}
	Workdays/Not-working	-	Binary	{-1; 1}
	Season-heating/cooling	-	Binary	{-1; 1}
Availability-of-cost-of-electricity	-	Categorical	{-1; 1}	
AMONG FLATS	Orientation	-	Categorical	{1-4}
	Flat-number	-	Categorical	{2; 4-7; 14-29}
	Floor-number	-	Categorical	{0-3}
	Number-of-bedrooms	-	Categorical	{1; 2; 3}
	Floor-area	m <sup>2</sup>	Continuous	[37,9 ≤ x ≤ 95,3]

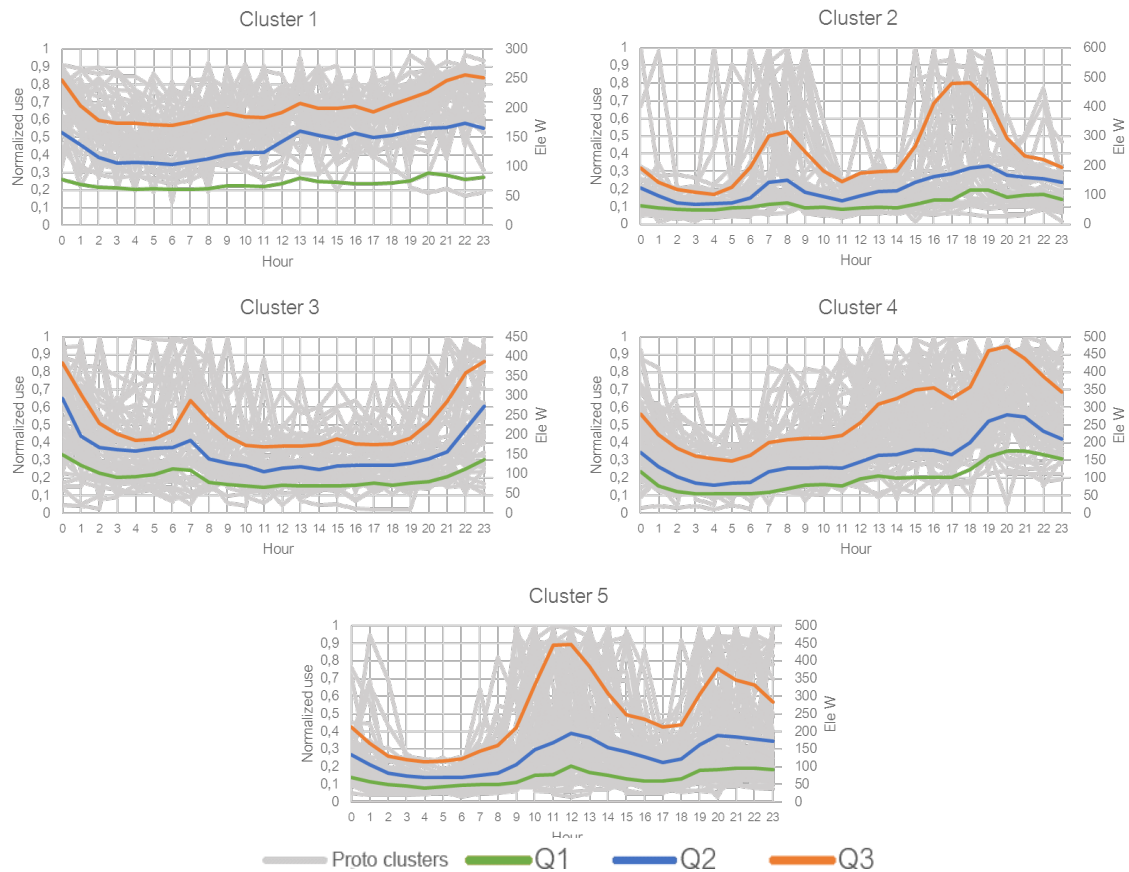


Figure 6: Scenarios and protoclusters in each cluster.

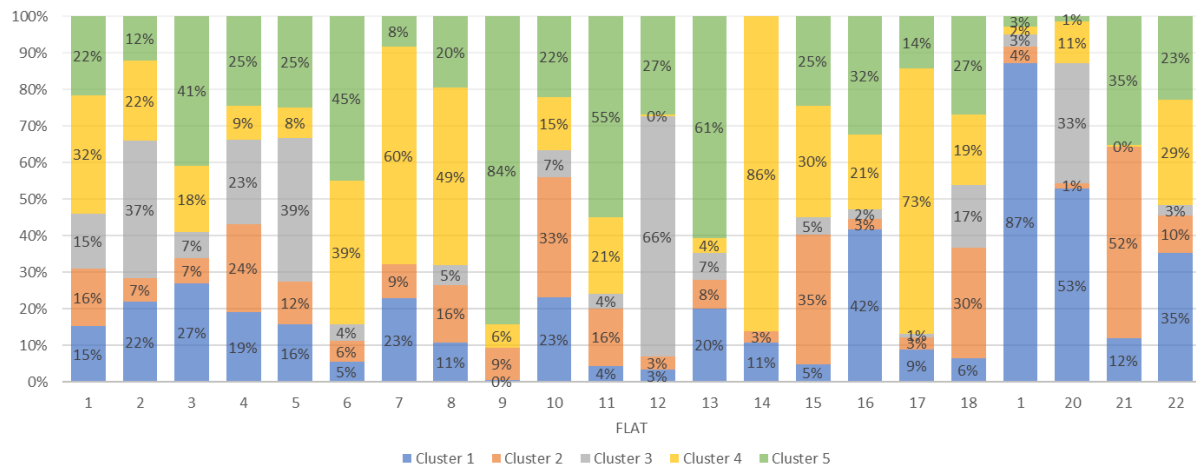


Figure 7: Distribution of clusters for each household.

### Description

Cluster 1, in Figure 6, is characterized by all the days with no evident peaks. These types of daily loads could be representative of the days in which the house is completely empty or the ones in which the dwellers are at home constantly, but using not so many electric appliances. For example, a couple of retired people could stay at home all day long but using a few electric devices. Around lunchtime and in the evening, the first graph shows, in all the three quartiles, a slight increase in the electric demand.

Cluster 2, shown in Figure 6, presents a two-peaks daily load. The average electric use is not far from the value of

cluster 1, but the third quartile shows higher values, that reach almost 500 W. The load rises around 8 a.m. and then increases progressively during the afternoon with maximum uses around 6 p.m. This cluster could be representative of days in which the occupants go out in the morning and gradually go back home during the afternoon. This could be a family with children, that go back to school in the afternoon, and preparing dinner around 7 p.m. Looking at the protoclusters, which visually appear noisier, the same trend is present, and just a few days show peaks during lunchtime and in the late evening. Cluster 3, described in Figure 6, shows a daily load without very high peaks. This cluster is characterized by

a low consumption all day, with a small peak in the morning around 7 a.m., but quite high values during the night. This cluster could be representative of people that are out all day long, going out in the morning and coming home in the late evening, having dinner outside or around 9 p.m. The increase in the electric load during the night could be attributable to the use of a washing machine, dishwasher and/or appliances used for leisure in these hours.

Cluster 4, in Figure 6, shows peaks during the evening, around 8 p.m. This daily load could be typical of dwellers that are out in the morning, and in the afternoon. In terms of pattern and values it is not very different from cluster 1, except from the absence of the morning load and the shifting of the evening peak from 7 p.m. to 8 p.m. This could be an indication of different habits in the same typology of family composition. For example, cluster 1 could be characterized by electric usage during the morning, such as for the television, radio, razors, or kitchen tools, differently from cluster 4. In addition, the dinnertime could be different for the two cases. Even if the general trend is similar to cluster 1, cluster 4 shows a higher difference in terms of hourly values along the day.

Cluster 5, shown in Figure 6, is characterized by a two-peaks load. The maximum values are reached around noon and 8 p.m. These could be respectively the lunch and dinner time for any typology of family. During the afternoon a relatively low electric use is registered, almost similar to the one registered around 4 a.m. in the morning. This could be a sign of the absence of people inside the house, or of the limited use of electric appliances during these hours.

## Conclusions and future outlooks

In this paper, a clustering methodology is presented and is applied to the energy recordings of a multi-family residential building located in Milan, Italy. It aims at determining a few representative electrical energy use profiles generated from the case study, and, potentially, it is extendable to the larger residential sector in Italy, to provide results with a higher statistical value. The implementation of machine learning techniques was found to be appropriate for the nature of the data sample and its complexity. A two-levels approach was implemented; it couples a SOM with the *k*-means algorithm. The analysis results in five daily energy profiles, that can be ascribed to different types of family and/or habits. In each cluster, the first, second and third quartiles are highlighted and considered representative of three types of energy use: austere, normal and wasteful.

These profiles, and further ones derived from larger datasets, might be used by energy modellers to, indirectly, include occupant behaviour, in their simulations.

## Acknowledgements

The study was developed in the framework of the project SHAR-LLM (Sharing Cities), which has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 691895 and the EnergiX R&D project 269650 "Utvikling av

metoder og system for automatisk effektkontroll i bolig". We thank our colleagues from Sikom AS and the IEA EBC Annex 79 "Occupant-centric building design and operation" who provided useful discussion points.

## References

- Agenzia Regionale per la Protezione dell'Ambiente della Lombardia. (2019). ARPA.  
[http://www.arpalombardia.it/Pages/ARPA\\_Home\\_Page.aspx](http://www.arpalombardia.it/Pages/ARPA_Home_Page.aspx). Accessed 30 January 2019
- Ali, U., Buccella, C., & Cecati, C. (2016). Households Electricity Consumption Analysis with Data Mining Techniques. *Industrial Electronics Society, IECON 2016 - 42nd Annual Conference of the IEEE*, 3966–3971.  
doi:10.1109/IECON.2016.7793118
- Buttitta, G., Neu, O., Turner, W., & Finn, D. (2017). Modelling Household Occupancy Profiles using Data Mining Clustering Techniques on Time Use Data. *IBPSA Building Simulation 2017*, 2.
- Capasso, A., Grattieri, W., Lamedica, R., and Prudenzi, A. (1994). Bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, 9(2), 957–964. doi:10.1109/59.317650
- Capozzoli, A., Piscitelli, M. S., and Brandi, S. (2017). Mining typical load profiles in buildings to support energy management in the smart city context. *Energy Procedia*, 134, 865–874.  
doi:10.1016/j.egypro.2017.09.545
- Capozzoli, A., Piscitelli, M. S., Brandi, S., Grassi, D., and Chicco, G. (2018). Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. *Energy*, 157, 336–352. doi:10.1016/j.energy.2018.05.127
- Carlucci, S., Causone, F., Pagliano, L., and Pietrobon, M. (2017). Zero-Energy Living Lab. In J. Littlewood, C. Spataru, R. J. Howlett, & L. C. Jain (Eds.), *Smart Energy Control Systems for Sustainable Buildings* (Smart Inno.). Cham: Springer International Publishing.
- Carlucci, S., Lobaccaro, G., Li, Y., Catto Lucchino, E., and Ramaci, R. (2016). The effect of spatial and temporal randomness of stochastically generated occupancy schedules on the energy performance of a multiresidential building. *Energy and Buildings*, 127, 279–300. doi:10.1016/j.enbuild.2016.05.023
- Chicco, G. (2012). Overview and performance assessment of the clustering methods for electrical load pattern grouping. *Energy*, 42(1), 68–80.  
doi:10.1016/j.energy.2011.12.031
- Chicco, G., Napoli, R., Piglion, F., Postolache, P., Scutariu, M., and Toader, C. (2004). Load pattern-based classification of electricity customers. *IEEE Transactions on Power Systems*, 19(2), 1232–1239. doi:10.1109/TPWRS.2004.826810
- Clement-Nyns, K., Haesen, E., and Driesen, J. (2010). The impact of Charging plug-in hybrid electric vehicles on a residential distribution grid. *IEEE*

- Transactions on Power Systems*, 25(1), 371–380.  
doi:10.1109/TPWRS.2009.2036481
- Deshani, K. A. D., Hansen, L. L., Attygalle, M. D. T., & Karunaratne, A. (2014). Improved Neural Network Prediction Performances of Electricity Demand: Modifying Inputs through Clustering. *Second International Conference on Computational Science & Engineering*, 137–147.  
doi:10.5121/csit.2014.4412
- Dudek, G. (2016). Neural networks for pattern-based short-term load forecasting: A comparative study. *Neurocomputing*, 205, 64–74.  
doi:10.1016/j.neucom.2016.04.021
- Galvin, R. (2016). *The Rebound Effect in Home Heating*. New York: Routledge.
- Hernández, L., Baladrón, C., Aguiar, J. M., Carro, B., and Sánchez-Esguevillas, A. (2012). Classification and clustering of electricity demand patterns in industrial parks. *Energies*, 5(12), 5215–5228.  
doi:10.3390/en5125215
- Joule Assets Europe. (2017). The White Paper on Opportunities and Challenges in the Italian Energy Efficiency Market.
- Mardaljevic, J., Heschong, L., and Lee, E. (2009). Daylight metrics and energy savings. *Lighting Research and Technology*, 41(3), 261–283.  
doi:10.1177/1477153509339703
- McLoughlin, F., Duffy, A., and Conlon, M. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied Energy*, 141, 190–199.  
doi:10.1016/j.apenergy.2014.12.039
- Menezes, A. C., Cripps, A., Bouchlaghem, D., and Buswell, R. (2012). Predicted vs. actual energy performance of non-domestic buildings: Using post-occupancy evaluation data to reduce the performance gap. *Applied Energy*, 97, 355–364.  
doi:10.1016/j.apenergy.2011.11.075
- Mitchell, T. (1997). Chapter 4: Artificial Neural Networks. In *Machine Learning* (p. 432). McGraw-Hill Science/Engineering/Math; (March 1, 1997).
- MM. (2014). Casa - Metropolitana Milanese SPA. <https://www.mmspa.eu/wps/portal/mmspa/it/home/mm-per-milano/casa>. Accessed 13 May 2019
- O'Brien, W., Gaetani, I., Gilani, S., Carlucci, S., Hoes, P. J., and Hensen, J. (2017). International survey on current occupant modelling approaches in building performance simulation. *Journal of Building Performance Simulation*, 10(5–6), 653–671. doi:10.1080/19401493.2016.1243731
- Paatero, J. V., and Lund, P. D. (2006). A model for generating household electricity load profiles. *International Journal of Energy Research*, 30(5), 273–290. doi:10.1002/er.1136
- Panapakidis, I. P., Papadopoulos, T. A., Christoforidis, G. C., and Papagiannis, G. K. (2014). Pattern recognition algorithms for electricity load curve analysis of buildings. *Energy and Buildings*, 73, 137–145. doi:10.1016/j.enbuild.2014.01.002
- Piech Chris. (2013). K Means. <http://stanford.edu/~cpiech/cs221/handouts/kmeans.html>. Accessed 1 July 2018
- Presidenza della Repubblica Italiana. (1993). D.P.R. 412, 513(1), 16–19.
- Rhodes, J. D., Cole, W. J., Upshaw, C. R., Edgar, T. F., and Webber, M. E. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461–471.  
doi:10.1016/j.apenergy.2014.08.111
- Rumsey, D. J. (2016). *Statistics For Dummies, 2nd Edition*. For Dummies.
- Sandels, C., Widén, J., Nordström, L., and Andersson, E. (2015). Day-ahead predictions of electricity consumption in a Swedish office building from weather, occupancy, and temporal data. *Energy and Buildings*, 108, 279–290.  
doi:10.1016/j.enbuild.2015.08.052
- Shimoda, Y., Fujii, T., Morikawa, T., and Mizuno, M. (2004). Residential end-use energy simulation at city scale. *Building and Environment*, 39(8), 959–967. doi:10.1016/J.BUILDENV.2004.01.020
- Tsekouras, G. J., Hatziaargyriou, N. D., and Dialynas, E. N. (2007). Two-stage pattern recognition of load curves for classification of electricity customers. *IEEE Transactions on Power Systems*, 22(3), 1120–1128. doi:10.1109/TPWRS.2007.901287
- Tso, G. K. F., and Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761–1768.  
doi:10.1016/j.energy.2006.11.010
- Tzempelikos, A., and Athienitis, A. K. (2007). The impact of shading design and control on building cooling and lighting demand. *Solar Energy*, 81(3), 369–382. doi:10.1016/j.solener.2006.06.015
- Vesanto, J., and Alhoniemi, E. (2000). Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3), 586–600. doi:10.1109/72.846731
- Vesanto, J., Himberg, J., Alhoniemi, E., and Parhankangas, J. (2000). *SOM Toolbox for Matlab 5*. (SOM Toolbox Team, Ed.). Espoo: Libella Oy.
- Viegas, J. L., Vieira, S. M., Sousa, J. M. C., Melicio, R., & Mendes, V. M. F. (2015). Electricity demand profile prediction based on household characteristics. *International Conference on the European Energy Market, EEM, 2015-Augus*, 0–4. doi:10.1109/EEM.2015.7216746
- Yohanis, Y. G., Mondol, J. D., Wright, A., and Norton, B. (2008). Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use. *Energy and Buildings*, 40(6), 1053–1059.  
doi:10.1016/j.enbuild.2007.09.001