

Application of the k-means clustering method for the detection and analysis of areas of homogeneous residential electricity consumption at the Great La Plata region, Buenos Aires, Argentina



Pedro Chévez*, Dante Barbero, Irene Martini, Carlos Discoli

Research Group 1, IIPAC (Built Environment Policy Research Institute), FAU (School of Architecture and Planning), UNLP (National University of La Plata), Street 47 N° 162, DC 478, La Plata 1900, Argentina

ARTICLE INFO

Article history:

Received 22 December 2016
Received in revised form 13 March 2017
Accepted 27 March 2017
Available online 31 March 2017

Keywords:

Electric consumption
Clustering method
K-means
Homogeneous areas
Households
Socio-demographic information

ABSTRACT

This research aims to detect areas of homogeneous residential electric consumption in the Great La Plata, Buenos Aires, Argentina. This study will identify main socio-demographic factors that impact on the electricity demand and the geographical location of these areas of homogeneous electric consumption. Results were analyzed from groups obtained from the six bimestrial electric average consumption of each of the 1010 census radius which constitutes the Great La Plata, using the K-means clustering method.

The present methodology becomes a plausible mechanism to use in the construction of urban energy scenarios, more precisely to determine areas with homogeneous consumption that can be described based on certain socio-demographic characteristics in what is called the “base year”. This study allowed to identify eight homogeneous areas of electricity consumption and their associated characteristics such as rooms per home, people per home, percentage of homes with unsatisfied basic needs, gas network coverage, housing typologies and quality of construction. In this way, we were able to obtain valuable information that allows to propose energy efficiency strategies and to incorporate renewable energy alternatives using appropriate criteria for each area in different “policy scenarios”.

© 2017 Elsevier Ltd. All rights reserved.

Introduction

Argentina's energy sector is complex in nature due to different transformations that have shaped the current electric and hydro-carbon systems. The norms, regulations, and infrastructure of these systems present various problems. In the case of the electric system the main problems are manifested both in the distribution and generation sector. In particular, the distribution of energy is characterized by the lack of investment in infrastructure that translates into massive supply cuts, mainly in summer (Decree 134, 2015). The generation sector cannot meet consumption peaks observed at specific times of the year caused by the sustained increase in total consumption with a rate close to five percent per year (MINEM, 2016a). In addition, Argentina has a low-diversified electricity generation matrix that is highly dependent on fossil fuels (Garrido, 2016) and presets a context of local declining hydrocarbon pro-

duction in the upstream area (Barrera, Sabbatella & Serrani, 2012). This situation forces the national government to import fuels such as natural gas through pipelines, liquefied natural gas by ships, gas oil and fuel oil, causing a significant outflow of foreign exchange for the trade balance. For example, in 2013 the fuel import produced an output of 7978 million dollars (INDEC, 2014).

Under the aforementioned context the residential sector has a strong impact on the total energy demand in Argentina. In 2014, the total energy demand from the residential sector was 25,98% with a consumption of 15921 kTEP. This value was followed by an industrial demand of 24,98% and a demand from the transportation sector of 27,94%. The lowest demand was observed by the agricultural sector (6,98%), commercial and public buildings (7,76%) and non-energetic (6,76%) sectors. The most used energy sources in the residential sector in 2014 were natural gas by network (64,92%), electricity (23,70%) and liquefied petroleum gas –LPG– (9,65%) followed by charcoal (0,92%), wood (0,70%) and kerosene (0,11%) (MINEM, 2016b). In this sense, the electric consumption in 2014 was driven by 13382765 residential users who consumed 40,386 TWh (MINEM, 2016c) which represents an average of 3017,80 kWh/year. It is notable that 90,28% of the Argentinean dwellings are located in urban areas (INDEC, 2016), so it is necessary to focus

* Corresponding autor.

E-mail addresses: che.pedro@hotmail.com (P. Chévez), dantebbarbero@yahoo.com.ar (D. Barbero), irenemartini@conicet.gov.ar (I. Martini), discoli@rocketmail.com (C. Discoli).

efforts on reducing energy demand in cities. This indicates that working on improving energy efficiency in the residential sector can have significant reductions in the national demand and thus reduce the need to import energy resources.

Since the detection of the national energy problems, the research institute to which the authors of this paper belong to, has been carried out several research projects^(1,2,3). Taking into account the housing density in urban areas our overall objective is to know the usage behavior of various energy resources in the residential and transportation sector in the city and propose different energy savings strategies according to different areas using the city of La Plata (Buenos Aires Province) as our area of study. The purpose of these studies is to analyze possible future scenarios to: (a) design and evaluate various strategies related to energy consumption in residential and transportation sectors and (b) test hypotheses on various combinations of strategies and analyze their implications, effects, and impacts under different urban growth scenarios. The ultimate goal is to improve living conditions and reduce energy demand. The expected impact of these projects is to: (a) raise awareness among the governmental management to internalize methodologies and proposed actions, (b) provide data support to the processes of decision making, planning and management of local sustainability, and (c) improve the interactions between researchers and urban decision makers through appropriate communication platforms.

In this sense, it should be noted that, in general, studies that address the construction of urban scenarios of energy demand do not consider differences within the city's urban structure and approach the studied area as a single element. Although there are studies that obtain potentialities from different city sectors referring to only one type of actions, such as buildings retrofitting (Rodríguez, Martini & Discoli, 2016) or solar energy use (Gadsden, Rylatt, Lomas, & Robinson, 2003; Compagnon, 2004; POLIS, 2012), it is difficult to find studies that include multiple aspects related to energy demand reduction in different sectors of a study area. For this reason, the contribution of Viegas (2010) or Amado, Poggi & Amado (2016) are important, because they provide a methodology for the analysis of improvements of the buildings' envelope and solar potential in different urban sectors.

In addition, previous studies about characterization of residential energy consumption using different methods have been performed. First, we should mention Argentinian studies performed by our research group that have provided energy assessments in Capital Federal and the Great Buenos Aires (Audibaires, IAS, 1983), energy consumption and conservation in the Río Turbio mining village (Rosenfeld et al., 1988), and rational energy use in the Metropolitan Region of Buenos Aires (URE-AM, Rosenfeld, 1999). Studies by other research groups have quantified energy consumption from surveys in the cities of Bariloche (González et al., 2006; González, 2008) and the city of Mar del Plata (Jacob et al., 2012, 2013). Also, there are publications that make diagnoses of the building park and its resulting energy consumption in cities such as Santa Rosa or Comodoro Rivadavia (Filippín, 2005; Vagge et al., 2008; Mercado et al., 2008).

Second, at the international level we can mention different studies where the methods used are very diverse, for example, the work of Swan, Ugursal & Beausoleil-Morrison (2011) where neural net-

work techniques are used to classify the energy consumption of Canadian residential sector. Moreover, in Fan, MacGill & Sproul (2015) statistical methods are used to determine the main factors which characterize the energy demand in the region of Sydney, Australia. Meanwhile, Iwafune & Yagita (2016) characterize the Japanese residential demand from comprehensive data for more than seven hundred homes which attempt to describe the main variables and thereby contribute to develop future programs aimed at efficient use of resources. Moreover, the work of Garrido-Soriano, Rosas-Casals, Ivancic & Alvarez-Del Castillo (2012) provides an energy classification of the different building types for the Catalan region in Spain and then tests energy efficiency measures. Other example is the work of Khatib et al. (2014) where a comparison between the energy uses of dwellings in Italy and Austria is carried out using questionnaires, then they model their demand and test energy management measures. In Zhou & Teng (2013), an annual urban household survey data is used to estimate the income and price elasticities of residential electricity demand along with the effects of lifestyle-related variables at Sichuan Province, China. Also, Tewolde, Longtin, Das & Sharma (2013) propose a high-resolution metering system for residential natural gas meters to determinate the appliance energy usage. McLoughlin, Duffy & Conlon (2012) analyze data obtained from a smart metering and from surveys of a representative cross section of approximately 4200 domestic Irish dwellings. This analysis is followed by a multiple linear regression model to four parameters: total electricity consumption, maximum demand, load factor, and time of use of maximum electricity demand for a number of different dwelling and occupant socio-economic variables. The same authors (McLoughlin et al., 2013) then use two techniques, Fourier transforms and Gaussian processes to characterize individual household electricity demand. In addition, Matsumoto (2016) examines how the socio-economic characteristics of the households affect appliance usage using a conditional demand analysis to Japanese household data. Finally Gouveia & Seixas (2016) make a classification through clusters for the residential sector of the city of Évora in Portugal with high resolution energy data. Groupings of users are generated by different load curves that occur throughout the year (365 records with the daily average) whose data are obtained from smart meters. Complementarily, door to door surveys are used to obtain conclusions about the common elements that determine these groupings. This method is very useful to find users with similar characteristics, but does not allow the identification of homogeneous areas in the city, which would enable the analysis of the behavior of different energy efficiency measures in the territory. In the context of the previously mentioned studies the objectives and scopes of this work are presented below.

2. Objectives and scope

Considering the few studies that deal with energy consumption affected by differences within urban structures of cities this work intends to identify areas within a city with homogeneous electricity intake. These areas are identified from the clustering of the average residential electric consumption obtained from the six bimonthly values recorded in 2014 of each of the 1010 census radius⁴ from the Great La Plata used in the last National Census in 2010. In addition, this study will relate the homogeneous areas with the main socio-demographic variables and will identify the role that these variables play in the electricity demand. The proposed methodology provides fundamental information for the adoption of different

¹ Research project "Construction of urban scenarios from an energy and environmental diagnosis." PICT 2012–2172. Director: Irene Martini.

² Research project "Construction of urban scenarios aimed to improve energy residential and transport sectors". UNLP project 2014–2017. Director: Irene Martini.

³ Research project "Development of technologies and guidelines for the massive recycling of the residential building envelope oriented to rational and efficient energy use in urban areas." PIP CONICET research project 097. Director: Carlos Discoli.

⁴ Census radius are census units defined by a territorial space with geographical limits and a certain number of housing units to be surveyed. The size of the radius in urban areas is approximately three hundred houses.

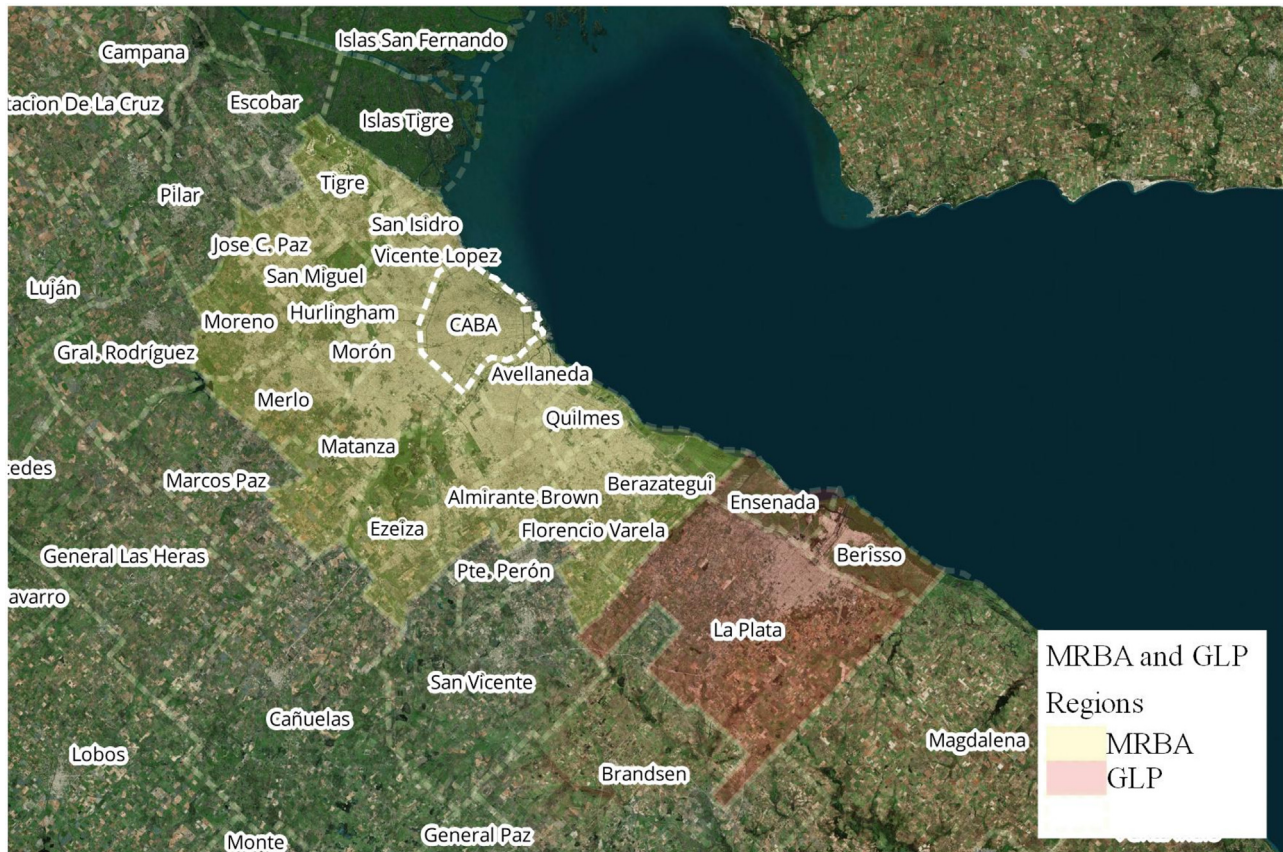


Fig. 1. Metropolitan Region of Buenos Aires (MRBA-Yellow) and Great La Plata (GLP-Red). Scale: 1:1.500.000.

Source: own elaboration

energy saving strategies with appropriate criteria for each area. This will help to identify the hypothetical impact of the application of different energy efficiency strategies and the incorporation of renewable energy systems in various city sectors through the construction scenarios. In this way, the socio-demographic variables in the present work are: rooms per home, people per home, percentage of homes with unsatisfied basic needs, gas network coverage, housing typologies, and construction quality. It is known that other information such as urban morphology, building density, economic and social characteristics, and home equipment are important in the determination of energy consumption. This information will be completed in future studies with door-to-door surveys and other sources of information and it will allow to characterize the potential use of solar energy, domestic equipment, consumption patterns, and others variables. In this way, it will be possible to obtain different potential savings within the city and provide important input for the formulation of policies.

It is necessary to emphasize the importance of identifying homogeneous areas for our case study. In the Great La Plata, and as it happens in great part of the Argentine cities, urban areas share mixed uses including retail and wholesale shops, educational, health, safety or administrative facilities, small industries and housing with well-defined typologies and do not have clear limits on their morphological characteristics. In the residential sector, the city does not provide clear limits regarding the scope of housing typologies except for a few neighborhoods with repetitive typologies. Therefore, finding and delimiting areas with homogeneous electric consumption becomes a fundamental finding of this study. In addition, it is necessary to emphasize that the methodology developed in this work is unprecedented nationwide.

The methodology used in this study, the results and discussion and the conclusions are presented below.

3. Methodology for the detection and analysis of areas of homogeneous residential electric consumption

The following methodology was defined to detect and analyze homogeneous areas of electric consumption. First the case study area is briefly described, second the data used is presented, and third the clustering method and the analysis of results are explained.

3.1. The case study area

The Great La Plata region (GLP) which includes the cities of La Plata, Berisso, and Ensenada was chosen as the case study area. GLP has annual energy consumption values similar to the national average with an annual value of 2710,16 kWh/year (Table 1).

In 2014 the Great La Plata (GLP) region had an estimated population of 833139 inhabitants and it is located adjacent to the Metropolitan Region of Buenos Aires (MRBA) that concentrates more than 13 million people, a 32,06% of the Argentinian population (INDEC, 2016). The geographical location of GLP region is shown at Fig. 1.

3.2. Data used for the clustering

Information provided by the electric distribution company of the Great La Plata was used to obtain the homogeneous areas from the clustering analysis. The information provided contains data from the year 2014 for La Plata, Berisso, and Ensenada (Great La

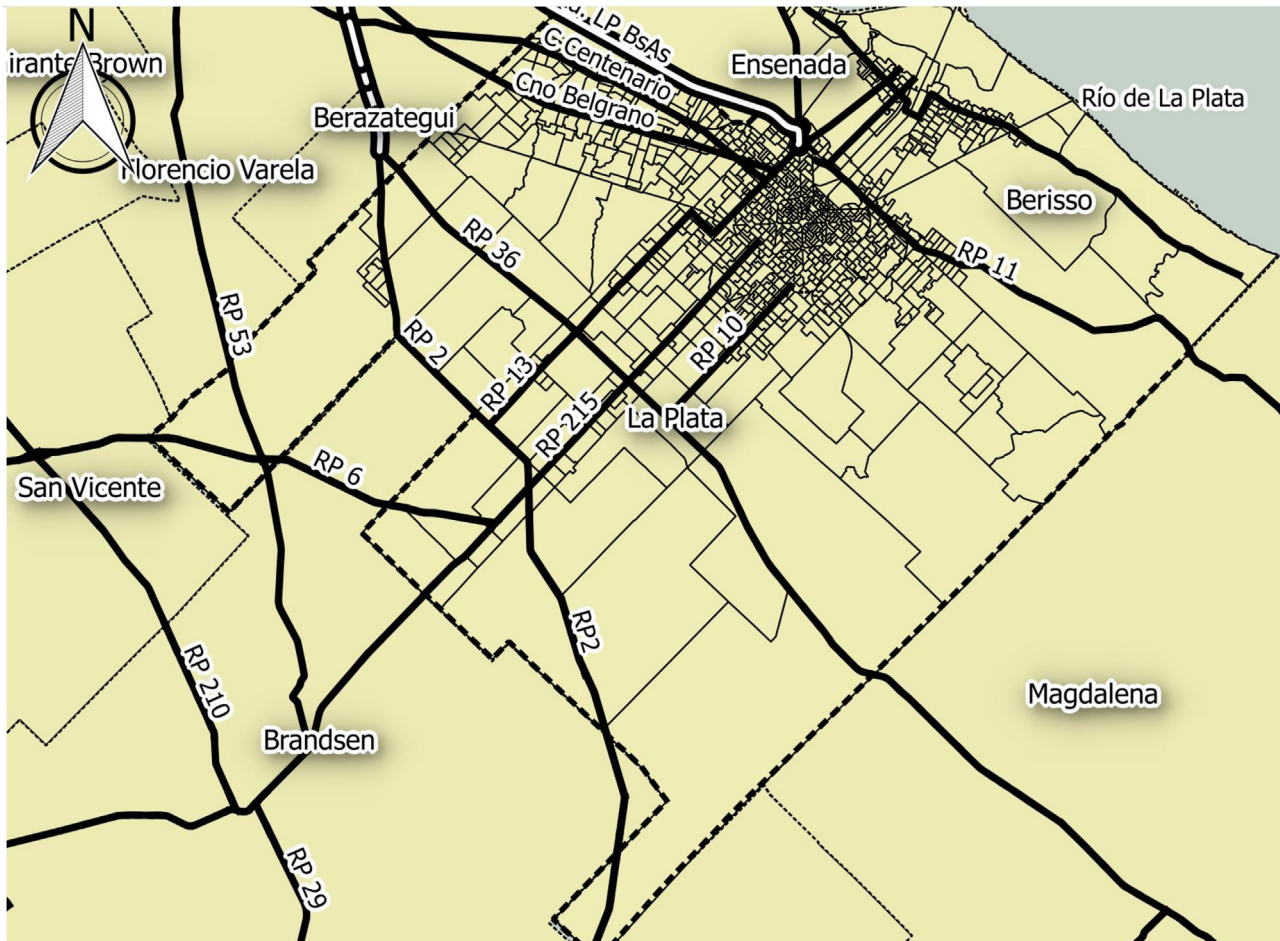


Fig. 2. Census radius used in La Plata at National Census 2010. Scale 1:400,000.

Table 1

Number of users, billed energy and annual average per user for Great La Plata the municipalities. (MINEM, 2016c).

	Residential users in 2014	Residential billed energy [GWh]	Average per user [kWh/year]
La Plata	244003	644,31	2640,59
Berisso	25424	79,06	3109,76
Ensenada	14560	46,27	3178,22
Total Great La Plata	283987	769,65	2710,16

Plata). This information is divided into 1010 census radii defined as previously described for the area of study. For each radius the company calculates the average residential electric consumption from all houses in the radius for the six bimesters of the year. Table 1 shows that a total of 283987 of meters or residential users was reported in 2014 for the three municipalities studied (MINEM, 2016c). The use of this bimonthly information helps to differentiate areas that may have similar annual consumption but that may have different consumption patterns during the year (winter-summer) demonstrating differences in their socio-economic conformation. It is necessary to clarify that distribution companies do not have smart meters and that bimonthly data was the highest degree of temporal disaggregation available in our country until 2016 when the billing began to be monthly.

The coverage of the census radii used at Census 2010 is shown in Fig. 2. These census radii also contain information associated with housing, homes, and population (INDEC, 2016) that is used to socio-demographically characterize the clusters obtained.

A representative example of the bimonthly and annual electric power consumption per user by census radius is shown in Table 2.

A histogram of frequencies is built from the data supplied by the distributor to obtain information about the characteristics of residential consumption in the city. The histogram is shown in Fig. 3. These frequencies were obtained by calculating the average of annual consumption per user for each census radius. Fig. 3 shows that the highest frequency was observed for census radii that consume between 2400 and 3100 kWh/y which corresponds to 46.73% of the census radii studied. The map in Fig. 4 shows this information in the territory.

3.3. Clustering method for the detection of homogeneous areas and analysis of the related information

According to the available information the K-means algorithm was used to perform the clustering because it is an unsupervised method and it allows to use real numbers in the values of the vectors. Therefore, classification methods that require integers or

Table 2
Bimonthly and annual average electric consumption per user of a randomly chosen census radius.

Census radius	μ Bimester 1 [kWh]	μ Bimester 2 [kWh]	μ Bimester 3 [kWh]	μ Bimester 4 [kWh]	μ Bimester 5 [kWh]	μ Bimester 6 [kWh]	μ Yearly [kWh]
64412503	407,17	350,34	419,17	476,86	407,27	358,24	2419,08

Source: Local electricity distributor EDELAP S.A.

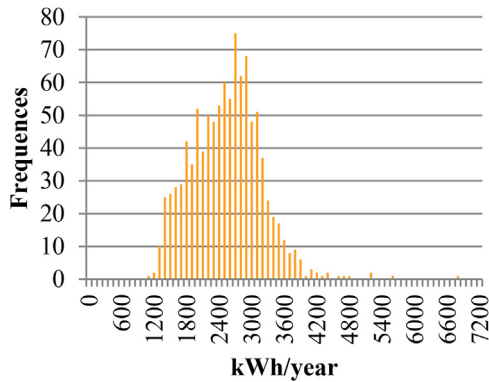


Fig. 3. Histogram of annual electricity consumption per census radius.

binary variables were not taken into account. An unsupervised method is also required because it is useful in an exploratory stage in which it is necessary to obtain initial knowledge of the area of study. In addition, this is a well-known clustering algorithm frequently used in clustering studies of energy patterns showing its efficiency in solving this type of problems.

In terms of its operation mode this algorithm uses as inputs “*n*” vectors with “*d*” dimensions and must know a priori the number of classes “*k*” to detect such as to minimize the sum of the squares within each group using as a metric the Euclidean distance.

First, “*k*” vectors are chosen randomly from the total set of elements. These elements will initially be the centroid of each class (there are “*k*” classes) because each class has initially a unique element. The algorithm iterates until it finds a stable situation like the one detailed below. At each iteration two steps are required:

- Each one of the “*n*” vectors is assigned to the class with the smallest centroid distance.
- The centroid of each class is recomputed.

If the centroids of the current iteration match those of the previous iteration the algorithm finishes (steady state). If the centroids do not match the iteration cycle continues by performing the two previous steps. Finally the clustering membership of each vector is obtained as an output. For this study, the clustering was performed using the IBM SPSS Statistics software (SPSS Inc., 2005).

After the iteration step each of the 1010 census radius (“*n*”) is grouped according to its six bimonthly electric consumption average (“*d*”) and is assigned a membership to one of the eight clusters (“*k*”). In summary: $n = 1010$, $d = 6$, $k = 8$. It should be clarified that the choice of using eight clustering classes ($k = 8$) was made after per-

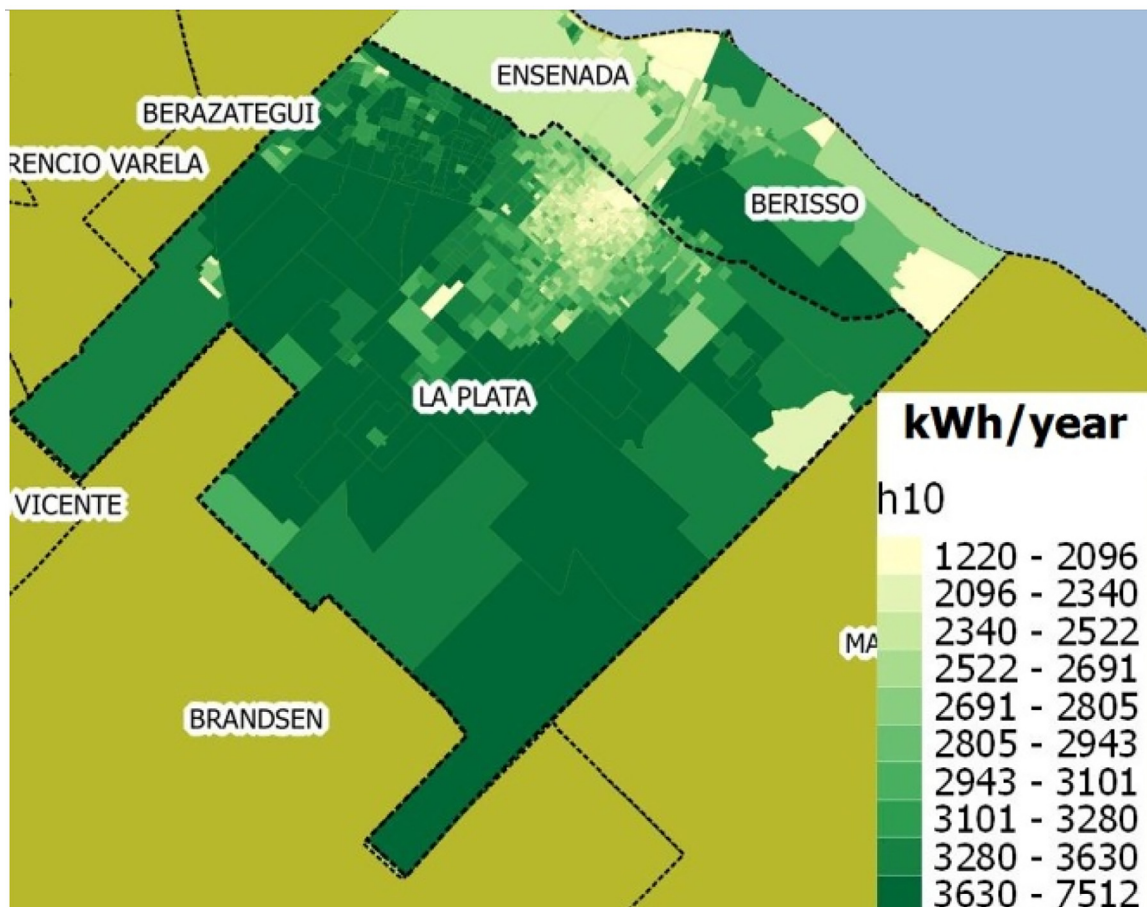


Fig. 4. Average of annual electricity consumption per user and census radius. Classification by deciles. Scale 1: 800.000.

forming several tests. We performed various test covering four to nine classes. Initially, the clustering into four groups generated very large areas that included empirically recognizable consumption heterogeneities. When clustering into five, six, and seven groups we were not able to discriminate areas of high consumption, which were the areas of most interest in our study. As it will be seen later, this area of interest was obtained with clusters F, G, and H. Finally, clustering into nine groups did not provide additional information compared to a cluster of eight groups.

This grouping allowed to obtain a consumption profile for each homogeneous area (cluster) and also to characterize each cluster from the census data. It should be noted here that the areas are homogeneous in terms of their electric energy consumption and that not all of the socio-demographic variables selected are necessarily homogeneous. The census variables associated with this study (INDEC, 2013; INDEC, 2016) are described below:

-Rooms/home

Total number of rooms in the home including those intended for sleeping, dining room, living room, and mezzanines. Bathrooms, kitchens, laundries, garages, corridors, halls, reception rooms, sheds (provided they are not used to sleep), and semi covered areas are not included.

-People/home

Number of people in the household. The census looks at how many people spend the night at home.

-Unsatisfied basics needs (UBN)

Households with UBN are those in which at least one of the following indicators of deprivation is present: more than three persons per room (critical overcrowding); inhabit in an inconvenient dwelling (rented room, precarious house, etc.); live in dwellings that do not have a toilet or have a toilet without water discharge; have a school-age child that does not attend school, and the household head has low level of education (has two or less years of elementary school education).

-Gas network

Gas network coverage is obtained from the query of the main fuel used for cooking where the gas network option can be found.

-Housing typologies (HTyp)

Only private housing was considered. Collective housing such as hotels, prisons, hospitals, and orphanages were disregarded.

House: Dwelling with direct exterior access (its inhabitants do not go through corridors of common use) originally built for people to live.

Flats: Dwelling originally built for people which are part of a building or structure which, with a common entrance, contains at least two dwellings (or a dwelling and one or more premises) accessed through corridors, stairs, halls, or elevators of common use.

Precarious house: Dwelling with direct exterior access originally built for people. They are usually constructed with low quality or scrap materials (urban types) or adobe walls, ground floor and roof of sheet or thatch (rural types).

-Housing constructive quality (HCQ)

This indicator is obtained from the quality of materials used to build the dwelling and internal facilities to provide basic services (water network and drainage).

Satisfactory quality: Refers to houses that have resistant and solid material with adequate insulation. It also has plumbing inside the house and toilet with water discharge.

Basic quality: Houses do not have adequate elements of insulation or they have sheet or fiber cement roof. Similar to the previous one they have plumbing inside the house and toilet with water discharge.

Insufficient quality: This classification includes housings that do not meet either of the two conditions mentioned above.

Two levels of analysis were performed to treat the results obtained. First, clusters were studied by the individual characterization from the mentioned socio-demographic variables and then a cross-sectional analysis was performed to describe the behavior of the main variables studied through the different homogeneous groups detected.

4. Results and analysis of the clustering information

The two levels of analysis previously mentioned in the methodology section are presented below. First, clusters are studied individually. A characterization is made from a series of socio-demographic variables for each cluster and represented spatially. Subsequently, an analysis is performed to describe the behavior of the main variables studied through the different homogeneous groups identified. Finally the discussion of the results is presented.

4.1. Individual analysis of clusters

From the application of the proposed methodology eight clusters are obtained which are sorted according to the average annual electric energy consumption to easily identify patterns that impact the demand. Each cluster is assigned with a letter from A to H according to their consumption and the original number obtained in the software remains in brackets. Fig. 5 shows a box and whisker diagram with the obtained clusters sorted in an ascending order with the assigned name.

Once the clusters are obtained the annual average consumption profiles are calculated and shown in Fig. 6.

Next, the satellite image of the study area is shown in Fig. 7 where the clustering result is represented by assigning the membership of each census radius to a cluster in Fig. 8. Considering the urban agglomerate in Fig. 7 and analyzing the geographical location of the clusters in Fig. 8, it is seen that the cluster QA(1) is located in the original layout where the center of the city of La Plata City is located. This cluster represents the lowest annual consumption of electricity. Then the cluster QB(6) has a higher consumption and surrounds the cluster QA(1). A third ring surrounding the QB(6) can be identified which consists of census radius that belong to QC(3). In addition, it is important to note that the northwest and southeast axis have different clusters but with greater preponderance of radius belonging to QD(2) cluster. Finally, QG(4), QE(5), QH(7) and QF(8) have the highest consumption, less territorial presence, and are located in outlying areas of the largest urban agglomeration. Those identified clusters have in some cases residential uses shared with productive uses, informal settlements that takes energy in an irregularly way, and sectors of high-income, or gated communities that have higher energy consumption.

Not only the clusters have been sorted in ascending order, but also the socio-demographic census data has been incorporated to allow for a description of each individual cluster.

Table 3 shows the ordered clusters from lowest to highest values as a function of the average of annual energy consumption per user and the associated variables and values obtained from the census.

The next step in this study consists on characterizing the clusters with the census information as shown below:

4.1.1. QA(1)

This cluster presents a prevalence of the “Flats” typology (69,9%) over the type “Houses” (28,72%), while the “Precarious house” typology completes the total (1,37%). The large number of flats causes the average of people per home to be low (2,12) as does the number of rooms per household (3,05). Unsatisfied basics needs (UBN) levels are the lowest (2,85%) and the presence of natural

Box and whisker plot of the clusters obtained.

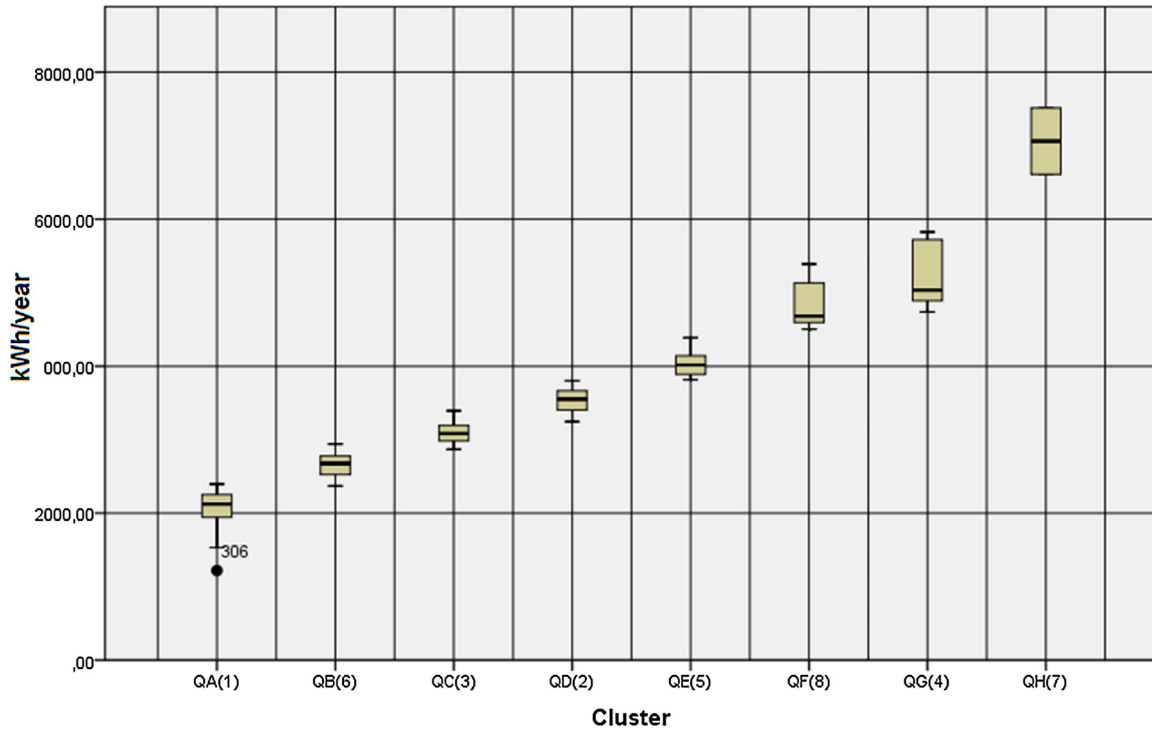


Fig. 5. Box and whisker plot of the clusters obtained ordered in an increasing manner according to the annual electric energy consumption.

Bimonthly average electricity consumption per cluster

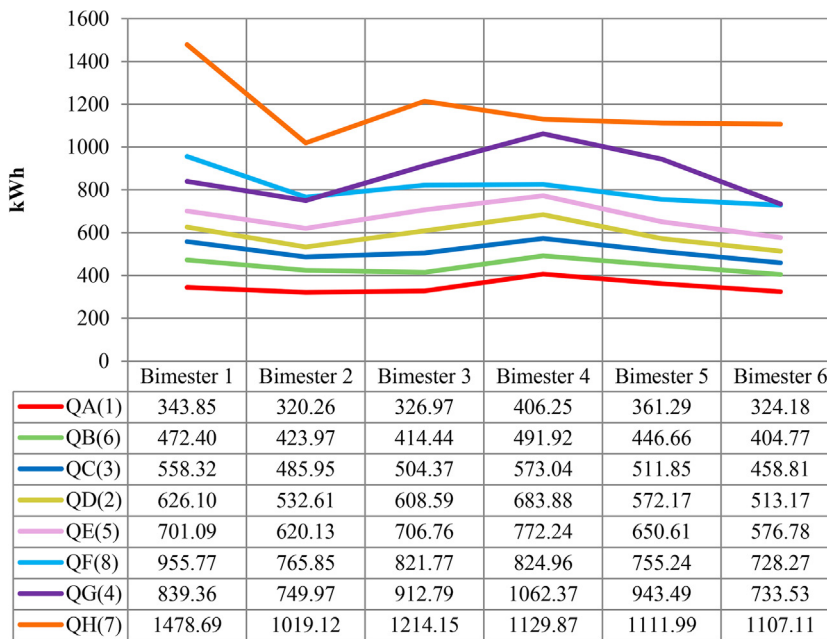


Fig. 6. Bimonthly average electricity consumption for each cluster. Data from 2014.

gas network is the highest which practically covers all households (96,18%). Building housing conditions are the best quality throughout the city, since a 91,49% has a satisfactory quality, 6,35% has basic qualities and 2,15% are unsatisfactory. As mentioned above, census radii identified with this cluster are located in the center of the city. Since these have the highest densities of housing and consequently they tend to be smaller and hold fewer people the high presence of gas network infers less use of electricity to heat the house in winter

and resulting in the lowest annual electricity consumption per user (2082,81 kWh/year).

4.1.2. QB (6)

This cluster is represented by census radii where types of flats and houses coexist, but in this case the “Houses” have a presence of 67,34%, the “Flats” 29,64%, while the rest of the typologies cover all cases (3,02%). The households average 2,69 persons per home and



Fig. 7. Delimitation of the municipalities in a satellite image.

Source: Own elaboration on Bing maps. Scale 1:600.000

3,44 rooms per home. UBN levels in these radius are low (3,96%) and the presence of gas network reaches 88,94% coverage. The housing quality construction is 85,77% satisfactory, 9,26% basic, and 4,97% unsatisfactory. This group presents the second lowest power consumption because homes are small both in area and in number of people; it also has good natural gas network coverage and a high proportion of good quality construction. Again the high coverage of homes with natural gas replaces electricity consumption in winter which results in lower overall consumption (2654,16 kWh/year).

4.1.3. QC (3)

This cluster presents a large majority of “Houses” (85,29%), while the rest is covered by “Flats” (6,91%) and “Precarious houses” (7,48%). The size of households starts growing with 3,26 people per home and 3,27 rooms per home. UBN levels increase and reach a 10,65% while the gas distribution network is reduced and reaches 68,70%. Meanwhile, the construction quality is generally good with 67,03% of satisfactory houses, 19,54% basic houses, and 13,43% unsatisfactory. The average energy consumption is the third lowest (3092,35 kWh/year).

4.1.4. QD (2)

This cluster has an intermediate consumption and is mainly represented by “Houses” (87,23%) and “Precarious houses” (11,28%), while the “Flats” presence is marginal (1,49%). It is characterized by a high number of people per household (3,42) and it has UBN levels which increase with respect to the QC(3) cluster and reach 14,87%. The gas distribution network reaches 51,55% households. This situation increases the annual electricity consumption. The construction quality is not regular with a 58,58% being satisfac-

tory, 21,46% are basics, and 19,96% are unsatisfactory housing. The annual energy consumption is 3536,51 kWh/year.

4.1.5. QE (5)

This cluster reaches an electricity consumption that is considered high for the average of the city. This high electricity consumption is caused by various issues. First, there is a high average number of people per household (3,36), UBN levels become similar to the QC(3) and reach 11,93% of households, while the gas network serves only to 49,81% of households. The types of housing detected are a 89,46% “Houses” and 9,70% “Precarious houses”, while 0,84% corresponds to “Flats”. The quality of construction is 60,41% satisfactory, 19,75% basic, and 19,84% unsatisfactory. The territorial location of this group is mainly in the northwest axis and peripheral points away from the foundational city. The average electric power of this group is about 4027,62 kWh/year.

4.1.6. QF (8)

This cluster has an average of 3,52 people per household, high UBN levels (18,69%), 42,98% gas network coverage which increases the average power consumption, and a representation of “Precarious houses” of 15,77%, while “Houses” reach 83,82% and “Flats” complete the totally (0,41%). Levels of construction quality are poor and 25,47% of the houses are unsatisfactory, while 19,83% are basic and 54,70% are satisfactory. The resulting annual demand is about 4851,87 kWh/year and its location is mainly rural.

4.1.7. QG (4)

This cluster has the largest number of people per household (3,62), the highest levels of UBN (21,66%), the lowest gas network coverage (9,32%), the most precarious houses (17,17%) and the least

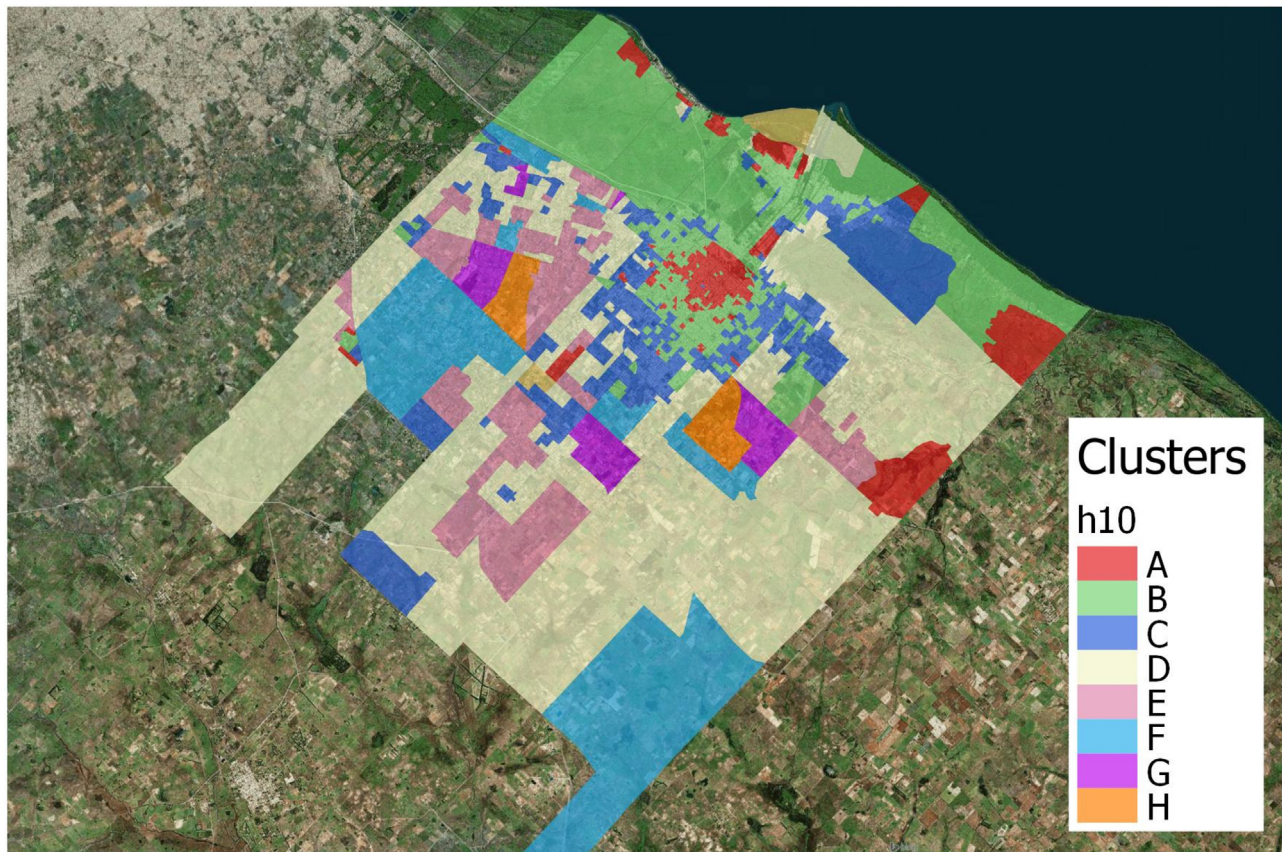


Fig. 8. Identification of different clusters in the territory. Scale 1:600,000.

amount of satisfactory housing (46,58%). These features make these census radii the most vulnerable which have the second largest city consumption that has been detected in this study (5241,51 kWh/year). Fig. 5 shows that the annual power curve of this cluster has a much higher consumption in winter months compared to the rest of the year, which can be directly linked to the low presence of natural gas, the existence of very low efficiency heating equipment, and the low construction quality of dwellings mentioned above.

4.1.8. QH (7)

This is the cluster with highest consumption and with 3,44 people per household. It also has high levels of UBN (21,23%), a gas network coverage of 40,12%, a high share of “Precarious houses” (12,84%) and a slight improvement in terms of building levels over the previous cluster with 52,17% of satisfactory housing. It is noteworthy that inside these census radii there are houses with both high and low income, so this increases annual energy demand to very high values of around 7060,92 kWh/year. This value is a good representation of reality since the census radii belonging to this cluster contain gated communities with high incomes and high energy consumption and in turn low-income areas where recognizable illegal connections are detected by a single meter that supplies multiple homes. Meanwhile, the annual electricity demand curve (Fig. 5) denotes high consumption in the summer months, which can be associated with massive use of air conditioners, pumps pools, among others from the high income radii. These are sectors with low consolidation which are characterized by big houses with higher consumptions.

After the identification and classification of the clusters the following step was performed to execute a cross-sectional analysis of the variables used in this work and their individual behavior in different clusters.

4.2. Cross-sectional analysis of the variables evaluated in this study

This section provides an analysis where the behavior of the main socio-demographic characteristics used in this study are described across the different homogeneous groups identified.

The purpose of clustering is to obtain sectors with similar patterns of annual consumption. This implies that given the heterogeneity of building typologies, housing quality, and socio-economic level that may exist in the same group, the correlation between consumption data and any of the aforementioned variables may be weak. In addition, there is a certain relationship (in the sense of non-independence) between pairs of variables, such as UBN and persons per household, or UBN and quality of housing. Thus, when analyzing the different groups identified through clustering heterogeneity in socio-economic and demographic variables can be observed which is due to characteristics of the city and its inhabitants. This is different from what occurs in sectors of cities in other latitudes where the value of such variables can be largely predicted.

First, the average number of people per household for each cluster is calculated with Census 2010 data (Fig. 9). It is observed that as the average household size is growing the electricity demand grows quickly. A scatter plot between people per home and its energy consumption is shown in Fig. 10 with an exponential trend line ($R^2 = 0,69$).

Fig. 11 shows the types of housing (INDEC, 2016) that characterizes each of the clusters. It can be seen that clusters QA(1), QB(6), and QC(3), which represent the lowest consumptions, have the highest amount of “Flats” types. Their presence decreases as the clusters are analyzed from low to high. Consequently, the type “House” increases its participation as it advances through clusters. Meanwhile housing type “Precarious” have low participation, but

Table 3
Census data for each cluster.

Cluster	Rooms/Home	People/Home	UBN [%]	Gas Network [%]	Housing typologies (HTyp)		Housing constructive quality (HCQ)			Yearly consumption [kWh/year]	
					Houses [%]	Flats [%]	Precarious House [%]	Satisfactory [%]	Basics [%]		Unsatisfactory [%]
QA(1)	3.05	2.12	2.85	96.18	28.72	69.90	1.37	91.49	6.36	2.15	2082.81
QB(6)	3.44	2.69	3.96	88.95	67.35	29.65	3.01	85.77	9.26	4.96	2654.16
QC(3)	3.27	3.26	10.65	68.70	85.29	6.92	7.79	67.03	19.54	13.43	3092.35
QD(2)	3.33	3.42	14.87	51.55	87.23	1.49	11.28	58.58	21.46	19.96	3536.51
QE(5)	3.51	3.36	11.93	49.81	89.46	0.84	9.70	60.41	19.75	19.84	4027.62
QF(8)	3.61	3.52	18.69	42.98	83.82	0.41	15.77	54.70	19.83	25.47	4851.87
QG(4)	3.21	3.62	21.66	9.32	81.68	1.16	17.17	46.58	18.43	35.00	5241.51
QH(7)	3.99	3.44	21.23	40.12	86.92	0.25	12.84	52.17	9.70	38.14	7060.92

Source: own calculations based on INDEC (2016).

Average people per home by cluster

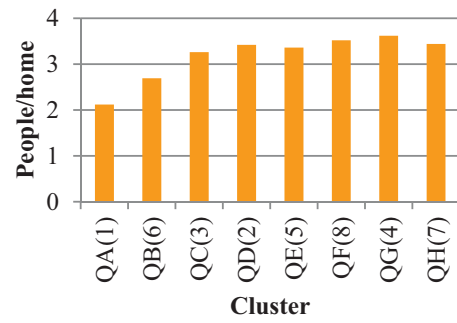


Fig. 9. Average number of people per home for each cluster.

People/home vs. kWh/year

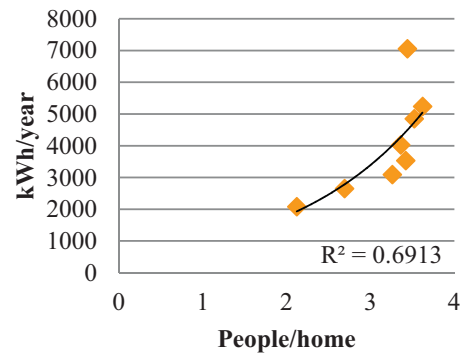


Fig. 10. Scatter plot of people/home vs. kWh/year of clusters.

Type of housing per cluster

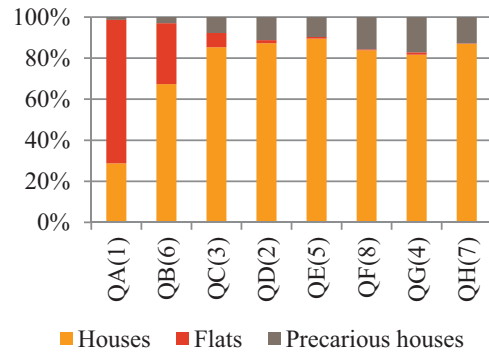


Fig. 11. Proportion of household types in each cluster.

tend to have greater presence in the clusters with high consumption. This may represent the strong orientation of users of these types towards electricity due to the lack of gas network, which is discussed below, but it can also be related to homes with high amount of people and even to several homes connected clandestinely to a single energy supply. In any case, it is noteworthy that although the types “Precarious houses” have a participation in clusters of high consumption it does not mean that their consumption are really high or that they are close to the average cluster. Fig. 5 shows in the box and whisker diagram that the amplitude of clusters with samples of greater demand is higher than those clusters of lower consumption.

Meanwhile, the quality of housing construction was analyzed and it was found that in clusters with lower annual consumption

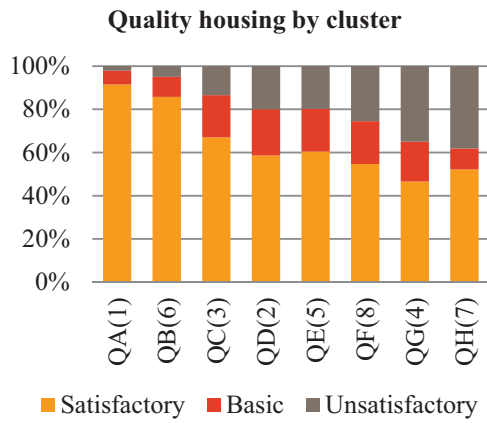


Fig. 12. Proportion of housing quality detected in each cluster.

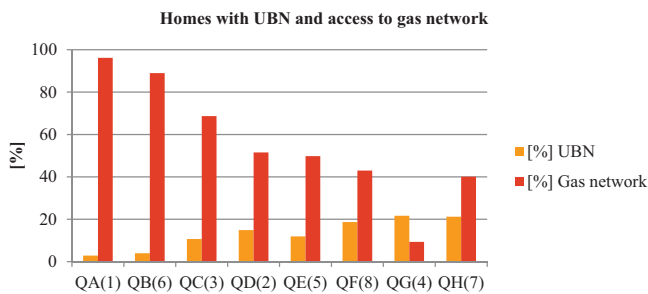


Fig. 13. Percentage of households with Unsatisfied Basic Needs and percentage of homes connected to gas network in each cluster.

housing is mainly satisfactory and that the number of basic and unsatisfactory houses increases in clusters with greater demand. This behavior could be explained with the existence of large energy losses in HVAC that consequently affect the consumption of electric heating and with the high average of people per household in those clusters. Figs. 11 and 12 show the typology housing and the construction quality among different clusters.

It is notorious that as we analyze clusters with greater electricity demand the households with gas network proportion decreases. Therefore it confirms that those homes that do not have gas connection must rely on electric power for applications such as water heating, cooking and heating to replace the more expensive liquefied natural gas. In addition, highest levels of UBN were detected in higher energy consumption. These results suggest that electricity consumption could be inefficient or that they replace the lack of gas network. These aspects are analyzed in Fig. 13.

Several analyses were performed to detect relationships between the aspects mentioned above. A scatter plot was performed for the proportion of households with gas network and the annual demand for electricity and a trend line was drawn using an exponential equation with an R^2 of 0,73 (Fig. 14). The scatter plot of the household proportion with UBN and the annual demand for electricity was fitted using a polynomial equation with an R^2 of 0,84 (Fig. 15).

4.3. Discussion of the results

While some results could be predictable such the direct relationship between the number of persons per household and the increase in energy consumption per user, some unexpected results were obtained. For example, it can be observed that high electric consumption is observed in clusters with high percentage of “Precarious house” typology, with insufficient or basic quality construction, and high levels of UBN. This occurs mainly because of

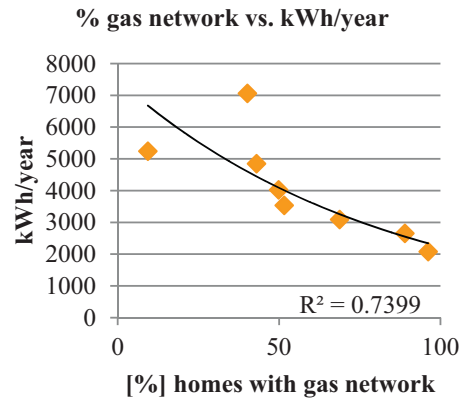


Fig. 14. Scatter plot of percentage of homes with gas network and kWh/year.

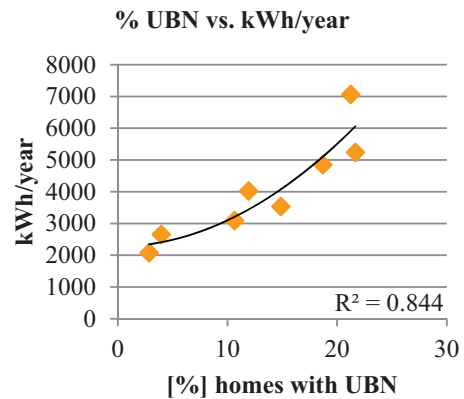


Fig. 15. Scatter plot of percentage of households with UBN and kWh/year.

the strong variation of amount of people per home and the scarce gas network coverage which deviate energy consumptions to electricity. This phenomenon indicates that high energy consumption does not always mean overconsumption. Instead, it is probable that some households are unable to achieve internal comfort due to critical building conditions that favor heat losses and social precariousness translated into overcrowding, low efficiency of their equipment, or difficulties to meet the electric service payments, caused by rates that aim to penalize high consumption.

The main variables studied are represented by three-dimensional territorial graphics that allow us to quickly visualize their behavior in different sectors of the city. Fig. 16 shows the average number of people per household, which is normalized between 0 and 1 to achieve a better discrimination. A depression in the central area and an increase in peripheral sectors is observed which is coincident with the housing density and consequently the household size. In Fig. 17 the presence of gas network is shown where the graph shows an inverted behavior compared to the previous one, i.e. the central area has the largest coverage. Unsatisfied basic needs are characterized in Fig. 18 where the highest levels are observed in the peripheral census radii and lower levels are found in the center of the city. Finally, the annual average power consumption per user is shown in Fig. 19 where QF(8), QG(4) and QH(7) clusters have higher consumption, followed by QD(2) and QE(5) clusters that form the northwest and southeast axis. Finally, there is a depression in the average consumption per user in the center of the city which, as mentioned before, has a high presence of gas network, low levels of UBN, household with lower average people per household and a high contribution of flats and high proportion of satisfactory housing construction.

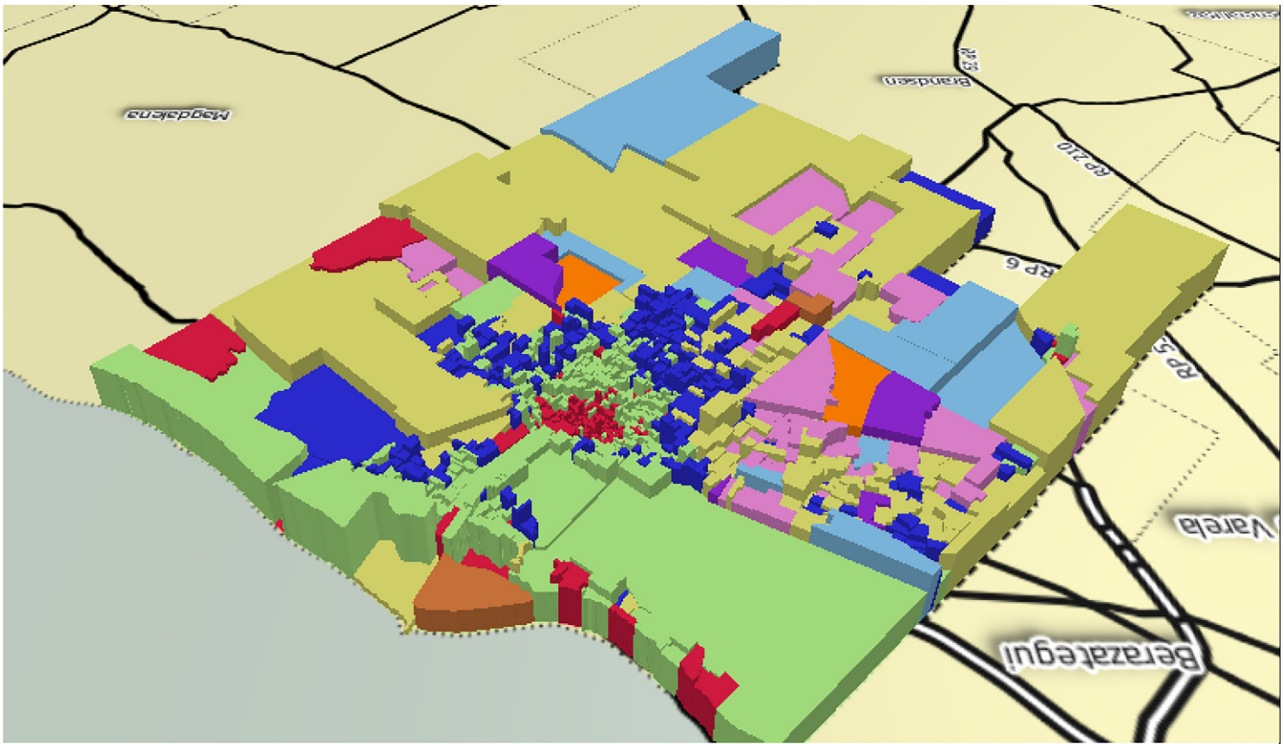


Fig. 16. Normalized average number of people per household (0–1).

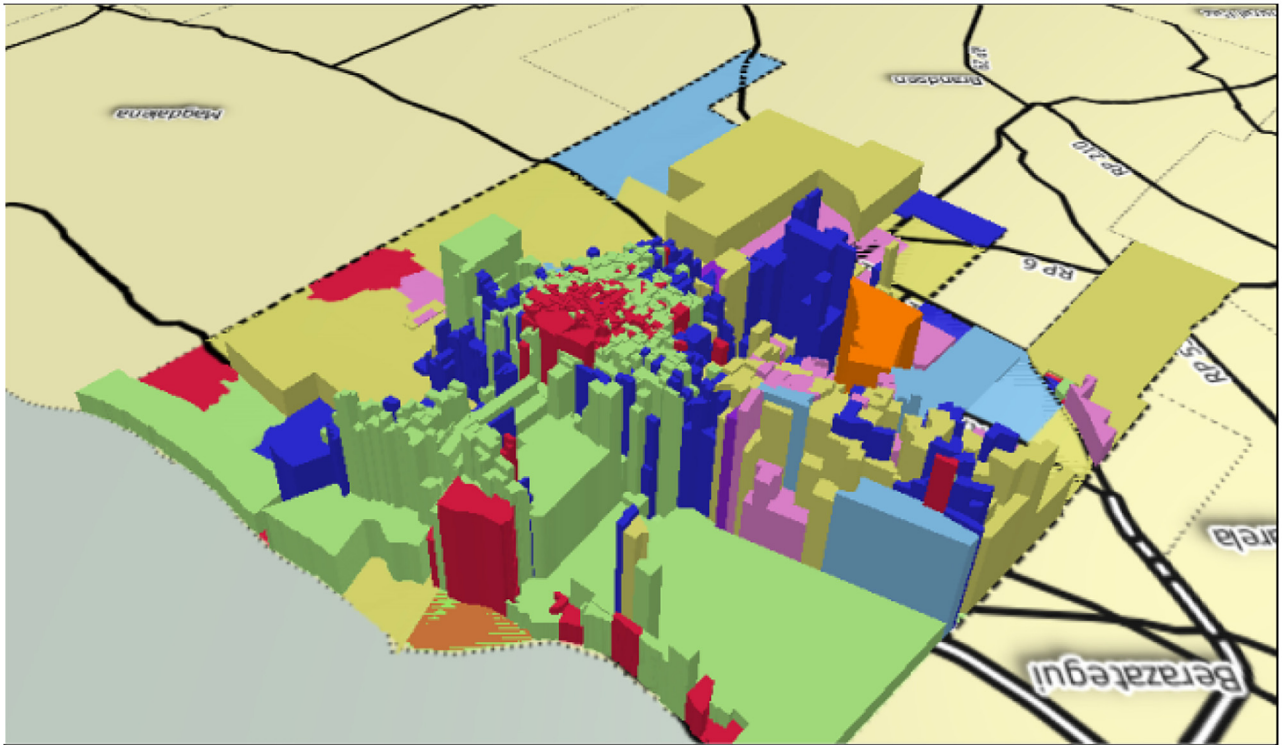


Fig. 17. Percentage of presence of Gas Network.

Finally, as previously mentioned, the proposed methodology will allow to identify the impact of the hypothetical application of different energy efficiency strategies and to incorporate renewable energy in the different sectors of the city through the construction of various scenarios. The potential energy savings in the different areas depends on the characteristics studied in this research and on

other variables that will be studied in future works such as home equipment and urban morphology, among others. Therefore, different saving measures can be considered in each identified cluster, and in turn, the impact of a general measure in each one can be analyzed. This will allow us to deepen our knowledge of the potential energy saving in various cities.

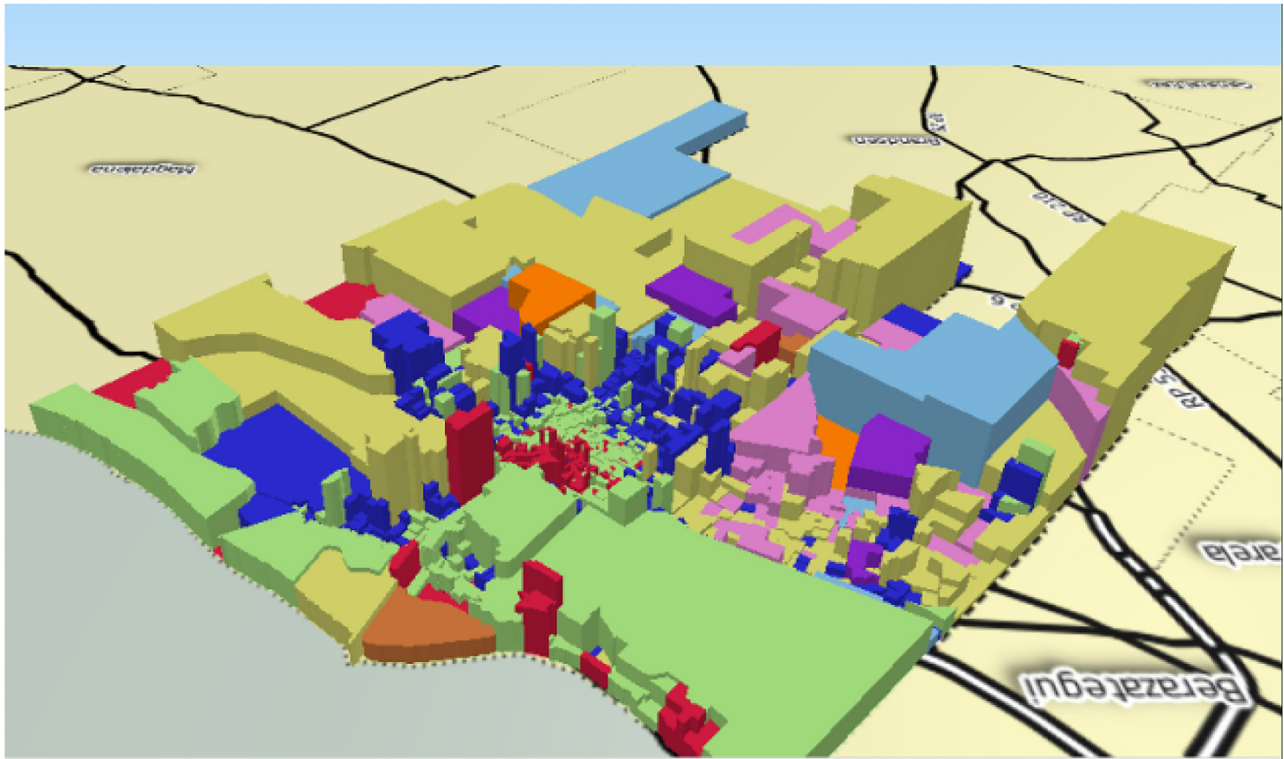


Fig. 18. Percentage of households with unsatisfied basic needs.

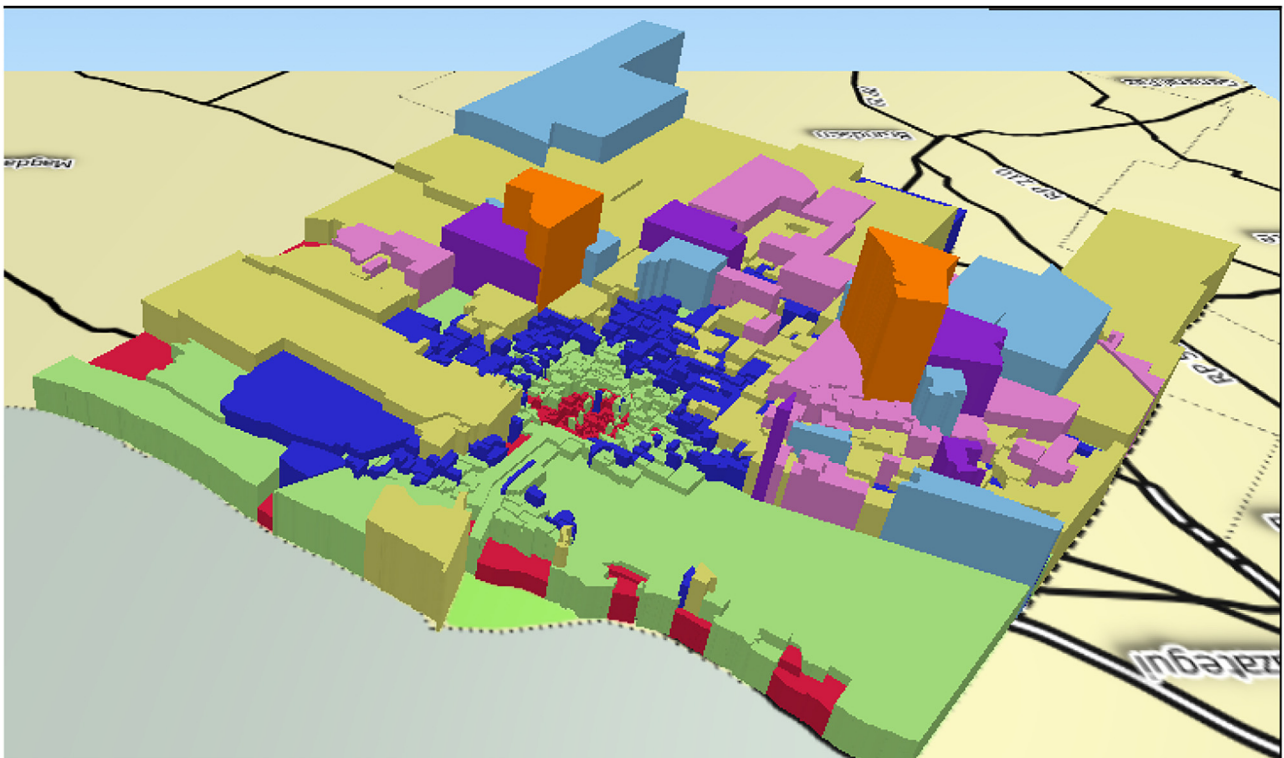


Fig. 19. Annual average electricity consumption per user.

As an example, we can mention that the strategies used in QA (1) clusters should be very different to the ones used in for example QE (5). On one hand, QA (1) cluster will have less potential to massively incorporate renewable energy due to the high presence of flats and consequently the limited availability of sunny surfaces.

However, for the same reasons, it is easier to implement building retrofitting since apartment buildings offer the fastest solutions. On the other hand, QE (5) has practically all its dwellings of “House” typology, therefore it has a greater feasibility for the incorporation of solar systems both active and passive. Almost half of the

households in this cluster have natural gas network, so in these sectors some programs of substitution of electric stoves (kitchens with electric resistance by induction cookers) or substitution of stoves for firewood by equipment of greater efficiency would be appropriate.

In the case of strategies for general application such as light replacement the potential savings will be different in QA than QE clusters since the incidence of illumination in both sectors is likely to be different due to the typologies found (“Flats”- “Houses”). Thus, this work will also allow to calculate the number of surveys to perform in each area based on the internal behavior of some variables from which important information will be incorporated in future works. In this manner we will account with more tools to correctly decide the strategies to implement in each city area.

5. Conclusions

This study allowed to detect eight homogeneous areas of residential electricity consumption with their corresponding spatial location and their characteristic energy demand. From the cluster analysis performed a greater understanding of the consumption logic was achieved which strongly depends on the geographical location and on related socio-demographic variables.

This initial clustering allowed us to understand characteristics included in the national census of the electric homogeneous areas and recognize their limits and their numbers of dwellings, which will help determine statistical samples. To complement this study an analysis of urban mosaics will be performed that will provide morphological and constructive information, as well as door-to-door surveys that will supply information about level of income, age, gender, equipment, usage habits, among others.

Consequently, it was possible to characterize each homogeneous area with bimonthly electricity consumption, the number of people per household, the number of rooms per household, levels of quality of housing construction, housing typology, UBN levels, and percentage of gas network coverage. A cross-sectional analysis of socio-demographic variables was also possible and a clear trend of their behavior based on energy consumption increase or decrease was found.

Perhaps the most important results are that as the average household size grows the electricity demand also grows quickly, both in people per home and in rooms per home. At the same time, as the network gas coverage descends the electricity consumption increases proportionally. In terms of typologies, it was found that the sectors with lower electric energy demand have a high presence of flats, while those with greater consumption show a considerable presence of precarious housing. This is in line with what has been detected with respect to the constructive quality of the dwellings, where it is observed that the areas of high consumption have a high proportion of houses with poor construction quality. It is also unexpectedly found that in sectors with a high level of UBN the demand for energy is high, which is explained by the values found for the other variables, mainly in relation to the number of people and the gas coverage.

Meanwhile, results obtained in this study can help propose possible scenarios for specific energy saving strategies for each sector considering the main socio-demographic aspects identified. In turn, they can also be analyzed territorially and integrated with other variables which would provide a holistic and comprehensive approach to energy problems.

Acknowledgements

This research received financial support from the following institutions:

Agencia Nacional de Promoción Científica y Tecnológica (ANPCyT). PICT Research project 2012–2172.

Universidad Nacional de la Plata. UNLP Research project 2014–2017.

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET). PIP CONICET Research project 097.

References

- Amado, M., Poggi, F., & Amado, A. (2016). *Energy efficient city: a model for urban planning*. *Sustainable Cities and Society*, 26, 476–485.
- Barrera, M., Sabbatella, I. y Serrani, E., (2012). *Historia de una privatización. Cómo y por qué se perdió YPF*. Buenos Aires : Capital Intelectual.
- Compagnon, C. (2004). *Solar and daylight availability in the urban fabric*. *Energy and Buildings*, 36, 321–328.
- Decree, 134. Emergencia energética. Boletín Oficial de la República Argentina, Buenos Aires, Argentina, december 16th, 2015.
- Fan, H., MacGill, I., & Sproul, A. (2015). *Statistical analysis of driving factors of residential energy demand in the greater Sydney region*. *Energy and Buildings*, 105, 9–25.
- Filippin, C. (2005). *Energy use of buildings in Central Argentina*. *Journal of Building Physics*, 29(1), 69–89.
- Gadsden, S., Rylatt, M., Lomas, K., & Robinson, D. (2003). *Predicting the urban solar fraction: a methodology for energy advisers and planners based on GIS*. *Energy and Buildings*, 35, 37–48.
- Garrido, S. (2016). *Energías renovables y geopolítica*. In: Guzowski, C. (coord.), *Políticas de promoción de las Energías Renovables*. Bahía Blanca: EDIUNS.
- Garrido-Soriano, N., Rosas-Casals, M., Ivancic, A., & Álvarez-del Castillo, M. (2012). *Potential energy savings and economic impact of residential buildings under national and regional efficiency scenarios. A Catalan case study*. *Energy and Buildings*, 49, 119–125.
- González, A. D., Crivelli, E., & Gortari, S. (2006). *Eficiencia en el uso del gas en viviendas unifamiliares de Bariloche*. *Avances en Energías Renovables y Medio Ambiente*, 10, 07. 01–07.08.
- González, A. D. (2008). *Aumento de eficiencia térmica en la ciudad de Bariloche: propuesta de plan de mejoras con dirección de subsidios a la inversión, y no al consumo*. *Avances en Energías Renovables y Medio Ambiente*, 12, 07. 57–07.64.
- Gouveia, J., & Seixas, J. (2016). *Unraveling electricity consumption profiles in households through clusters: combining smart meters and door-to-door surveys*. *Energy and Buildings*, 116, 666–676.
- Hargreaves, A., Cheng, V., Deshmukh, S., Leach, M., & Steemers, K. (2016). *Forecasting how residential urban form affects the regional carbon savings and costs of retrofitting and decentralized energy supply*. *Applied Energy*, 186, 549–561 [Part 3].
- IAS (1983). *Programa de conservación de la energía en la vivienda. Plan piloto de evaluaciones energéticas en la zona de Capital Federal y Gran Buenos Aires*. Audibaires. Contrato Secretaría de Energía N° 1399/83. La Plata: Instituto de Arquitectura Solar.
- INDEC (2013). *Censo Nacional de Población, Hogares y Viviendas 2010. Censo del Bicentenario Serie Base de datos Censo 2010. Abril 2013. Base de datos REDATAM*. Buenos Aires: Instituto Nacional de Estadística y Censos de la República Argentina.
- INDEC (2014). *Comercio exterior argentino 2013*. Buenos Aires : Dirección de Difusión del Instituto Nacional de Estadísticas y Censos.
- INDEC (2016). *Censo Nacional de Población, Hogares y Viviendas 2010. Censo del Bicentenario*. Buenos Aires: Instituto Nacional de Estadística y Censos.
- Iwafune, Y., & Yagita, Y. (2016). *High-resolution determinant analysis of Japanese residential electricity consumption using home energy management system data*. *Energy and Buildings*, 116, 274–284.
- Jacob, S. B., Strack, J. L., Branda, J., & Suárez, J. A. (2012). *Evaluación del consumo eléctrico en el sector residencial de Mar del Plata. Propuestas de ahorro y eficiencia energética*. *Avances en Energías Renovables y Medio Ambiente*, 16, 06. 65–06.72.
- Jacob, S. B., Strack, J. L., Branda, J., Murcia, G. J., & Suárez, J. A. (2013). *Evaluación del consumo eléctrico en el sector residencial de Mar del Plata. Aspectos económicos y ambientales*. *Avances en Energías Renovables y Medio Ambiente*, 17, 01. 35–01.42.
- Khatib, T., Monacchi, A., Elmenreich, W., Egarter, D., D'Alessandro, S., & Tonello, A. (2014). *European end-user's level of energy consumption and attitude toward smart homes: a case study of residential sectors in Austria and Italy*. *Energy Technology & Policy*, 1, 97–105. <http://dx.doi.org/10.1080/23317000.2014.977406>
- MINEM (2016a). *Series Históricas de Energía Eléctrica. Serie Facturación en MWh 1970–2012*. Buenos Aires : Ministerio de Energía y Minería de la Nación.
- MINEM (2016b). *Balance Energético Nacional 2014*. Buenos Aires : Ministerio de Energía y Minería de la Nación.
- MINEM (2016c). *Informe Estadístico del Sector Eléctrico 2014*. Buenos Aires: Ministerio de Energía y Minería de la República Argentina. Available at: <https://www.minem.gov.ar/www/830/25452/informe-estadistico-del-sector-electrico-2014.html>.
- Matsumoto, S. (2016). *How do household characteristics affect appliance usage? Application of conditional demand analysis to Japanese household data*. *Energy Policy*, 94, 214–223.

- McLoughlin, F., Duffy, A., & Conlon, M. (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: an Irish case study. *Energy and Buildings*, 48, 240–248. <http://dx.doi.org/10.1016/j.enbuild.2012.01.037>
- McLoughlin, F., Duffy, A., & Conlon, M. (2013). Evaluation of time series techniques to characterise domestic electricity demand. *Energy*, 50, 120–130. <http://dx.doi.org/10.1016/j.energy.2012.11.048>
- Mercado, V., Esteves, A., & Filippín, C. (2008). Estrategias bioclimáticas en viviendas de índole social en Comodoro Rivadavia, Chubut, Argentina. *Avances en Energías Renovables y Medio Ambiente*, 12, 05. 129–05.136.
- POLIS (2012). *Identificación y Movilización del potencial solar mediante estrategias locales. Directrices basadas en las experiencias piloto de Lisboa, Lyon, Malmö, Múnich, París y Vitoria-Gasteiz. Identification and mobilization of solar Potentials via local strategies*. Lisboa. Sitio Web: http://www.polis-solar.eu/IMG/pdf/polis-guidelines_es_final_web.pdf.
- Rosenfeld, E., Fabris, A., Ravella, O., Discoli, C., Lozano, S., Martínez, S., et al. (1988). *Consumo y conservación de la energía en el sector residencial de la Villa Minera de Río Turbio*. Actas 13^a Reunión de Trabajo de ASADES. Tomo, 2, 273–280.
- Rosenfeld, E., (1999). *Políticas de uso racional de la energía en el área metropolitana y sus efectos en la dimensión ambiental*. URE AM (2000/2002). La Plata, PIP CONICET 4717.
- SPSS Inc. (2005). *Released 2005. SPSS for windows version 14 0*. Chicago: SPSS Inc.
- Swan, L., Ugursal, V., & Beausoleil-Morrison, I. (2011). Occupant related household energy consumption in Canada: estimation using a bottom-up neural-network technique. *Energy and Buildings*, 43, 326–337.
- Tewolde, M., Longtin, J. P., Das, S., & Sharma, S. (2013). Determining appliance energy usage with a high-resolution metering system for residential natural gas meters. *Applied Energy*, 108, 363–372 [108 issue C]
- Vagge, C., Filippín, C., & Czajkowski, J. (2008). Auditorías energéticas en Santa Rosa, La Pampa. Análisis del comportamiento energético y consumo de gas natural en edificio de vivienda multifamiliar. *Avances en Energías Renovables y Medio Ambiente*, 05. 57–05.64.
- Viegas, G., (2010). *Evaluación del potencial energético e intervenciones de mejoramiento del entorno edilicio en áreas urbanas de media y baja consolidación. La ciudad de La Plata como caso de estudio*. PhD thesis of the National University of Salta. Salta: UNSA.
- Zhou, S., & Teng, F. (2013). Estimation of urban residential electricity demand in China using household survey data. *Energy Policy*, 61, 394–402.