

The use of visualization techniques for extracting behavioural patterns in residential customers.

By

Sandra Vegas Antuña



Submitted to the Department of Electrical Engineering, Electronics,
Computers and Systems
in partial fulfilment of the requirements for the degree of
Master Degree in Electrical Energy Conversion and Power Systems
at the

UNIVERSIDAD DE OVIEDO
July 2015

©Universidad de Oviedo 2015. All rights reserved.

Author

Certified by

Pablo Arboleya Arboleya
Associate Professor
Thesis Supervisor

Certified by

Ignacio Díaz Blanco
Associate Professor
Thesis Supervisor

1. Abstract

The use of visualization techniques for extracting behavioural patterns in residential customers.

by

Sandra Vegas Antuña

Submitted to the Department of Electrical Engineering, Electronics,
Computers and Systems
in partial fulfilment of the requirements for the degree of
Master Degree in Electrical Energy Conversion and Power Systems

In this thesis complex visualization techniques are used to represent analyse and cluster hourly load curves obtained from residential customers provided by the company Electra Norte, as well as simulated curves obtained from a simulation software previously developed called GenMix.

The validation of the previously developed software (GenMix) is done comparing the real and simulated data by means of the visualization tool.

The pattern extraction technique is validated using visualization techniques, considering that the parameters for generating the simulated data are known.

Thesis Supervisor: Pablo Arboleya Arboleya
Title: Associate Professor

Thesis Supervisor: Ignacio Díaz Blanco
Title: Associate Professor

2. Acknowledgments

I would like to express my sincere gratitude to my supervisors Prof. Pablo Arboleya and Prof. Ignacio Díaz for their consistent guidance and contributions which they have made throughout the development of this master thesis. I would also like to thank all the professors of the Electrical Energy Conversion Master. In addition, I would also like to express my thanks to my colleagues for their support during this master, specially to Adrián and Emilio.

I am sincerely thankful to family and friends for their continued love and support.

Contents

1. Abstract	iii
2. Acknowledgments	v
3. List of tables	ix
4. List of figures	x
5. Introduction	1
5.1. <i>Introduction</i>	<i>1</i>
5.2. <i>Objectives</i>	<i>1</i>
6. State of art	2
6.1. <i>Domestic load demand curve</i>	<i>2</i>
6.2. <i>Importance of the characterization of domestic electricity demand</i>	<i>5</i>
6.3. <i>Clustering methods</i>	<i>7</i>
6.3.1. Hierarchical Clustering	<i>8</i>
6.3.2. Fuzzy K-means Clustering.....	<i>10</i>
6.3.3. K-means Clustering	<i>11</i>
6.3.4. Follow the leader Clustering	<i>13</i>
6.3.5. Fuzzy Relation Clustering	<i>15</i>
7. Load demand curves	16
7.1. <i>Real data</i>	<i>16</i>
7.2. <i>Simulated data: GenMix software</i>	<i>17</i>
7.2.1. Power demand modeling.....	<i>17</i>
7.2.2. Daily energy consumption.	<i>18</i>
7.2.3. Unoccupied period.....	<i>20</i>
7.2.4. Usage hours for the different loads	<i>20</i>
7.2.5. Consumption profile for the different loads	<i>22</i>
7.2.6. Simulated load demand curves.....	<i>22</i>
8. Visualization tool	28
8.1. <i>Introduction</i>	<i>28</i>
8.2. <i>The interactive Dimensionality Reduction approach</i>	<i>28</i>
9. Load demand profile clustering	32
9.1. <i>Define the best input attributes</i>	<i>32</i>

9.1.1.	Average consumption at each hour of the day.....	33
9.1.2.	Average consumption at periods of the day and histogram.....	39
9.2.	<i>Simulated data load demand profile clustering.....</i>	<i>43</i>
9.3.	<i>Real data load demand profile clustering.....</i>	<i>45</i>
9.4.	<i>Simulated and real data load demand profile clustering.....</i>	<i>50</i>
10.	Conclusions.....	56

3. List of tables

Table 1. Typical energy consumption for a single dwelling in Spain.	18
Table 2. Usage hours for every load for one occupant.....	21
Table 3. Simulated clients type depending on the number of occupants and the unoccupied period.	23
Table 4. Input attributes (average consumption at each hour of the day) for the simulated clients.....	34
Table 5. Input attributes (average consumption at periods of hours of the day and histogram) for the simulated clients.	40
Table 6. Input attributes for the real clients.	46

4. List of figures

Figure 1. Daily load data curve of a client.	2
Figure 2. Daily load data curve of a client.	3
Figure 3. Monthly load data curve of different clients.	5
Figure 4. Hierarchical tree	9
Figure 5. Fuzzy K-Means Clustering	11
Figure 6. K-Means Clustering	12
Figure 7. Follow the Leader Clustering.....	14
Figure 8. Hourly energy consumption of one domestic client in the month of February.	16
Figure 9. Artificial and Natural Light Hours throughout the year for $L=43.53^{\circ}$	21
Figure 10. GenMix simulated demand for client type 1 during one day.	24
Figure 11. GenMix simulated demand for client type 6 during one day.	25
Figure 12. GenMix simulated demand for client type 11 during one day.	26
Figure 13. GenMix simulated demand for client type 16 during one day.	27
Figure 14. Batch mode interaction scheme (left) vs. the iDR approach (right).....	29
Figure 15. Simulated clients visualization depending on their average consumption at each hour of the week.	38
Figure 16. Histogram of one client	39
Figure 17. Simulated clients visualization depending on their average consumption at periods of the day and histogram.....	42
Figure 18. Simulated clients visualization depending on their average consumption at periods of the day.	43
Figure 19. Simulated data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption.	44
Figure 20. Clustering of the simulated clients depending on their unoccupied period.....	45

Figure 21. Clustering of the simulated clients depending on their unoccupied period and number of clients.	45
Figure 22. Real data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption	49
Figure 23. Real and simulated data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption.....	50
Figure 24. Real (green) and simulated (red) data visualization.....	51
Figure 25. Real and simulated data zooms.	51
Figure 26. Real (green) and simulated (red) data visualization Zoom 1.	52
Figure 27. Real (green) and simulated (red) data visualization Zoom 2.	53
Figure 28. Real (green) and simulated (red) data visualization Zoom 3.	54

5. Introduction

5.1. Introduction

Electricity is a unique form of energy such that its supply and use need to occur at the same time. As a result, a large amount of research has been focused on characterizing and forecasting electrical demand at a system demand level, in order to balance supply and demand [2]. Various mathematical techniques have been used to do this, each with their own strengths and weaknesses [3] [4]. Recently, the availability of detailed electricity consumption for the domestic sector also means that it is now possible to apply these techniques to characterise individual dwelling electricity demand [1].

5.2. Objectives

The main goals of this thesis are:

- To use complex visualization techniques to represent analyse and cluster hourly load curves obtained from residential customers provided by the company Electra Norte, as well as simulated curves obtained from a simulation software previously developed called GenMix.
- Validation of the previously developed software (GenMix), which is done comparing the real and simulated data.
- Validation the pattern extraction technique by means of visualization techniques, considering that the parameters for generating the simulated data are known.

6. State of art

6.1. Domestic load demand curve

Load curve data refers to the electric energy consumption recorded by meters at certain time intervals at delivery points or end user points. Load curve data is the “heartbeat” of electricity systems and is one of several most important data sets collected and retained in utilities. However, patterns of electricity use at a system demand level and at an individual dwelling level are very different.

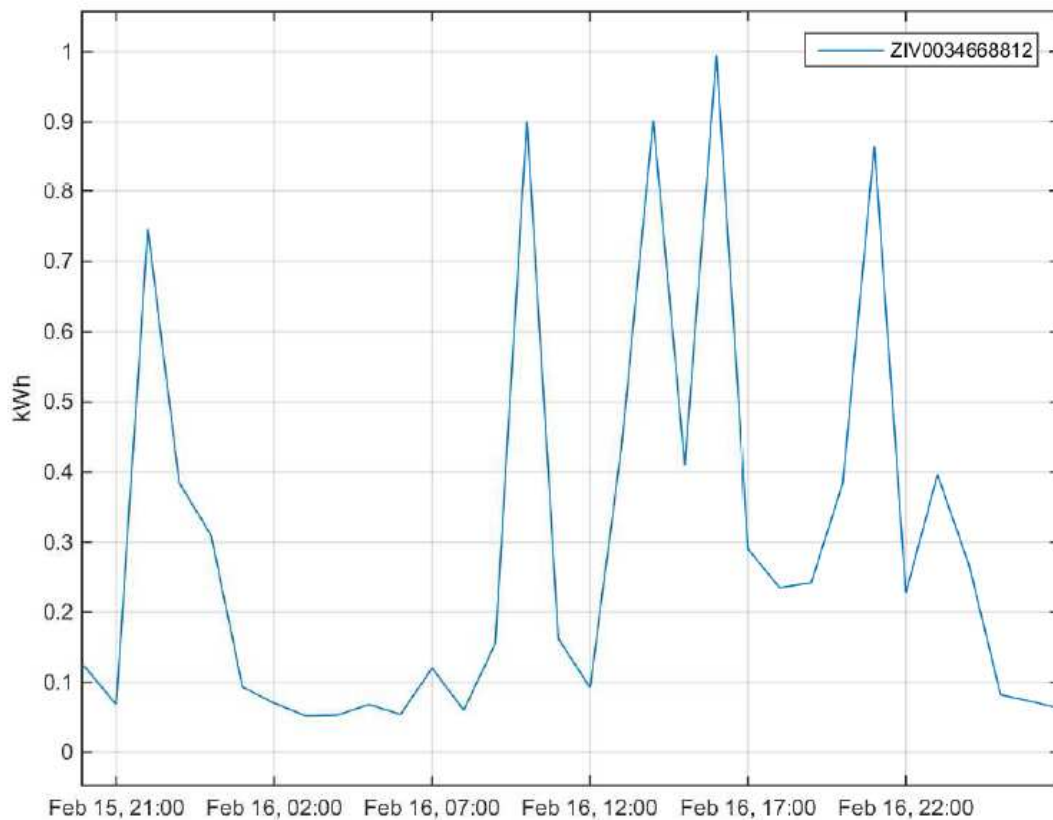


Figure 1. Daily load data curve of a client.

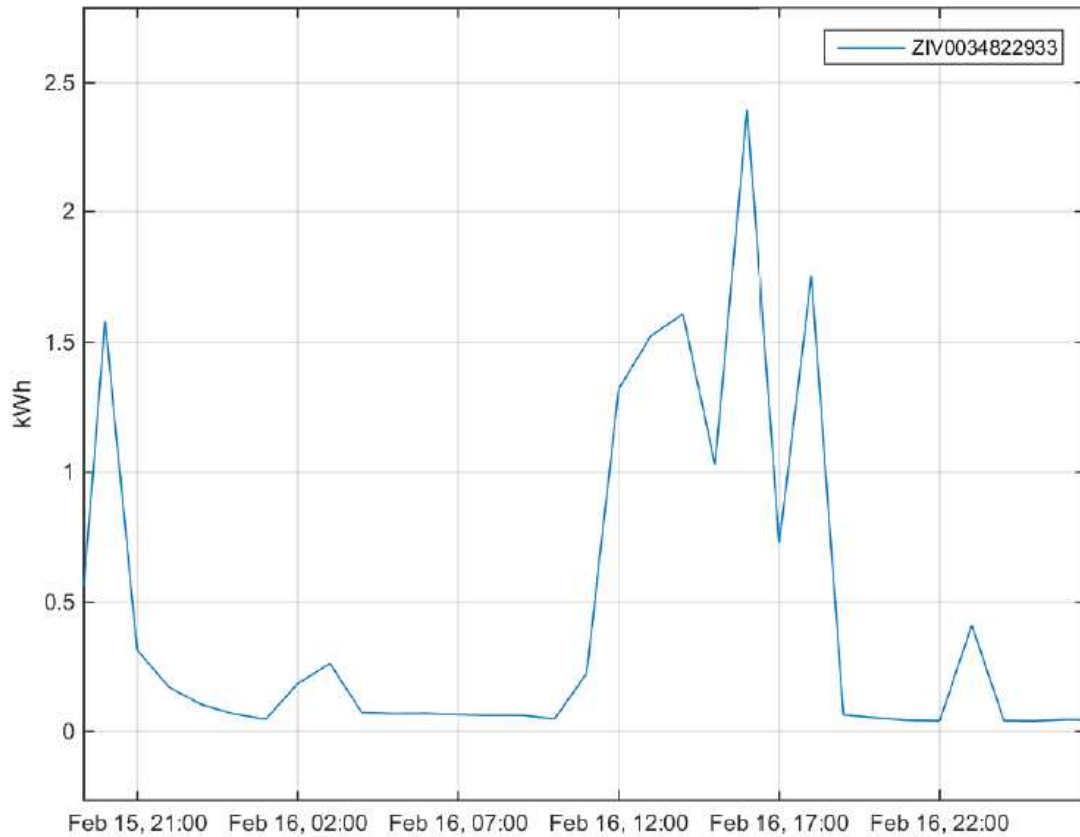


Figure 2. Daily load data curve of a client.

Figures 1 and 2 show the daily load profile of two different dwellings. The profile of figure 1 shows a very low consumption at night, and then a peak in the morning around 9am which lasts approximately 1 hour. Later, there is another two peaks from 1 p.m. to 3 p.m. Finally, there is another peak around 9 p.m. and another small peak around 11 p.m. However, the profile of figure 2 shows a very low consumption at night and three peaks from 12 to 7 p.m. At 11 a.m. there is another small peak.

As figures 1 and 2 show domestic load profiles are in part, based on three events. First, a very small amount of energy is consumed over the night, during this period the consumption is almost zero. Second, when the household occupants wake up there is a peak of consumption in the morning. Finally, as people come from work, there is another increase of energy consumption until the go to bed again and the consumption drops to nearly zero.

However, in practice electricity is consumed far more stochastically across a 24 hour period as was shown in figures 1 and 2. There are similar characteristic deterministic patterns to that of the standard load profile, but these often change on a daily basis and between households. In this manner, a domestic electricity demand load profile can be thought of as a combination of both deterministic and stochastic processes. [17]

Hence, it is known that the standard load profiles used by electricity suppliers only represent an average use pattern for all the customers, but not a detailed representation of each individual household consumption.

The methodology used to characterize the profiles is based on a regression on various parameters across a representative sample of domestic customers [5].

Demand not only varies from each household, but also varies from each day or the week and each month for the same household in terms of the magnitude of electricity consumption and the time at which it is used. Electricity demand also depends on the season, as the temperature and the number of daylight changes for each season. Figure 3, shows how different is the consumption between 6 different households in one month and also how it differs for the same household in different days of the month, in terms of magnitude and profile.

Figure 3 shows the monthly load curve of six different dwellings. It can be noticed that the consumption is different between each dwelling. The maximum energy consumed by each client varies from 1.5kWh to 3.5kWh. It can be appreciated that there are non occupied dwellings, such as dwelling ZIV00034702615 and ZIV00034702617, where the consumption is nearly 0 kWh most of the time. The profile also differs for the same dwelling in different days of the month. It is not repetitive for all the days of the month. The maximums and minimums are not produced at the same time and do not have the same magnitude.

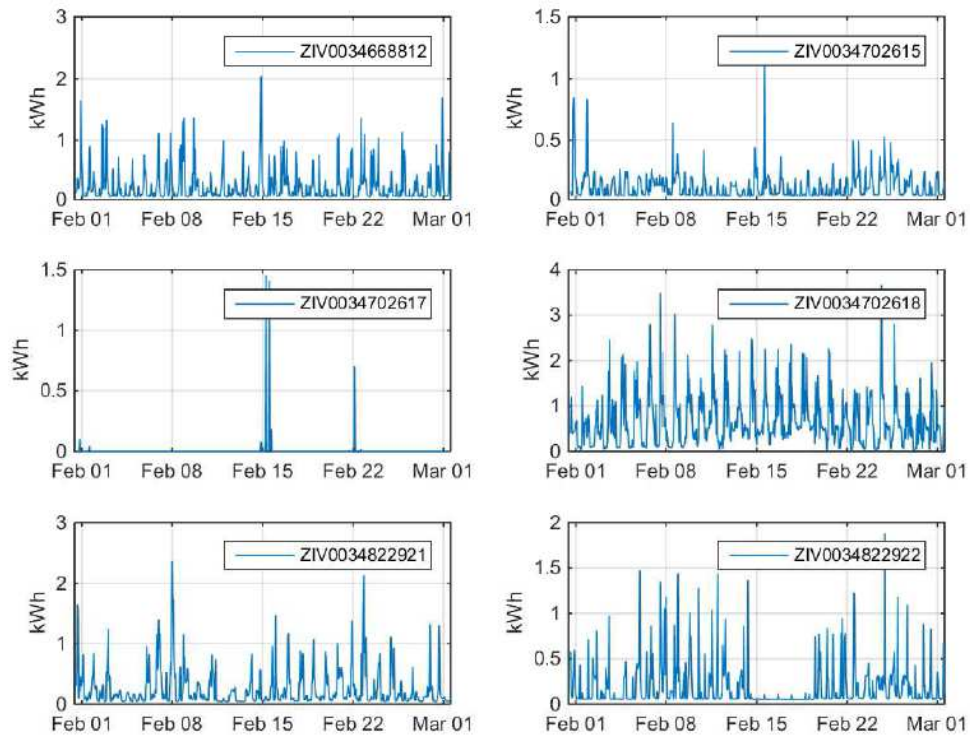


Figure 3. Monthly load data curve of different clients.

6.2. Importance of the characterization of domestic electricity demand

The analysis of load curve data would greatly improve day-to-day operations, system analysis, system visualization, system reliability performance, energy saving, and accuracy in system planning [6]. Two key features in the global vision of smart grid [7] are self-healing from power disturbance events and enabling active participation by consumers in demand response. In deregulated electricity markets information is a crucial factor for the success of all involved players. Indeed, the knowledge about consumption patterns is particularly important for the accomplishment of agreements in the price of the electricity, as well as the definition of marketing policies and innovative contracts and services [9].

Load profiles provide information concerning electrical consumers' consumption behavior, useful for both consumers and suppliers. Consumers can participate in the retail market and manage their actual power consumption allowing some change of their habits, by, for instance, the establishment of electric energy interruption contracts. Suppliers could use such information for load management, specify new electricity tariffs and identify load diagrams' peaks [10, 11]. In fact, understanding the demand behavior is vital for decision in power system operation and planning.

The collection of valid load curve data is critical for supporting decision making in a smart grid system. For example, smart meters are an important initiative in smart grid and the quality of data is essential for the success of smart meters.

Another field in which load data curve is important is the field of effectively power management based on Smart Grid, that is recently an emerging issue in the electric power industry. Especially, due to the increased recognition of environment change from increasing levels of carbon dioxide and the lack of fossil fuel, energy management based on Smart grids becomes interesting. Power utilities are trying to implement the Smart Grid System for effective energy management, and to do research on renewable energy such as solar photovoltaic power generation, wind power generation and waterpower generation [13, 14].

To operate effectively, it is important that data about national or regional based real- time power usage is accessible and applicable. Current power grid has facilities for monitoring power usage on power grid, transmission lines and some parts of distribution lines. However, periodic acquisition of power usage of customer as well as distribution lines is necessitated in order to operate power efficiently. [16]

Collecting all load data accurately in fine granularity is a challenging and costly task. There is often missing and corrupted data in the process of information collection and transfer. This is caused by various reasons including meter problems, communication failures, equipment outages, lost data, and other factors. [8]

In order to characterize the electrical consumers, clustering methods should be used expecting that in the near future all, or almost all, of the consumers have installed automatic meter reading (AMR) which typically operates at quarter-hour intervals. This results in a colossal amount of load data to handle and to be analyzed. Therefore, since thousands of customers may come to be analyzed, some kind of automatic data analysis method, as well as clustering analysis, should be used. [12]

Load profile, every 15 minutes based power usage, is currently collected by installation of an electric meter to customers, but it is still hard to obtain load profile from all users because of non- AMR customers. Therefore, research on creating virtual load profile (VLP) of non-AMR users is performed in order to compute the power usage of all the customers. [10]

The accuracy of the VLP is affected by the accuracy of the typical load profile (TLP) which is affected by the clustering method. So, in order to obtain the load profile of all the customers it is important to choose the correct clustering method to classify the AMR customers, as this classification is used to create the VLP of non-AMR users. [16]

6.3. Clustering methods

Clustering can be defined as the process of partitioning a large database into groups (or clusters) based on a concept of similarity or proximity among data. There is a wide variety of clustering algorithms, although there is no single algorithm that can, by itself, discover all sorts of cluster shapes and structures. Good clusters present high similarity within a certain group and low or a very different similarity among objects of others classes [18].

There are several clustering techniques that can be used to classify electricity consumers. All of them start by deriving a matrix of feature vectors for each load profile data set, followed by the procedures as described below:

6.3.1. Hierarchical Clustering

Hierarchical clustering groups data, simultaneously over a variety of scales, by creating a cluster tree. The tree is a multilevel hierarchy, where clusters at one level are joined to clusters at the next level.

To perform hierarchical clustering, it is necessary to find the similarity or dissimilarity between every pair of load profiles in the data and then group them into binary clusters based on the previously computed similarity matrix. The process is iteratively repeated by merging the clusters of each level into bigger ones at the upper level until all samples are grouped into expected clusters.

The advantage of this method is that the original data is kept unchanged in the root of the cluster tree [19].

The hierarchical clustering algorithm can be described in the three following steps [19, 20]:

1. Determine the similarity between every pair of load profile data sets.
2. Group into a pair and create a hierarchical cluster tree based on the determined similarity criterion.
3. Determine the cutting position of the hierarchical cluster tree.

Figure 4 illustrates the hierarchical cluster tree; where the horizontal axis represents the load profile data sets and the vertical axis represent the distance between clusters. A possible cutting position is indicated by a dashed line.

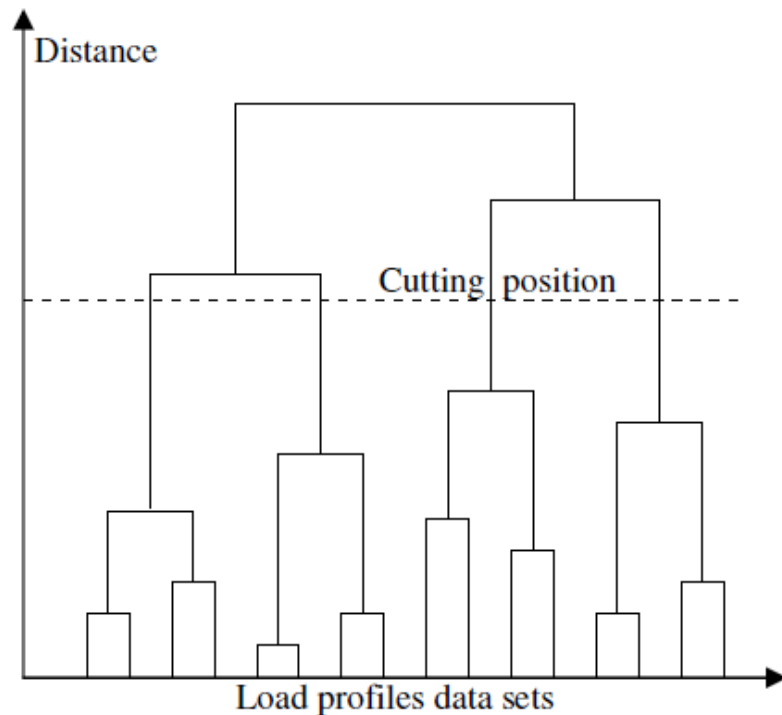


Figure 4. Hierarchical tree [24]

Determining the similarity or the distance between load profiles can be done in various ways such as Euclidean distance, Mahalanobis distance, City Block metric, Minkowski metric and Hamming distance.

Grouping can be processed by linking pairs of load profiles that are in close proximity using linkage criterion such as shortest distance, Average distance, Centroid distance and Ward distance, which use the previously calculated distance. As objects are paired into binary clusters, the newly formed clusters are grouped into larger clusters until a hierarchical tree is formed.

The number of groups is determined by the cutting position in the binary tree which is chosen either by the maximum distance admissible or by selecting directly the distance corresponding to the desired number of clusters.

The hierarchical clustering method is suitable if the specific number of groups is not predetermined. In fact, the cutting position will determine the number of clusters.

6.3.2. Fuzzy K-means Clustering

This method is similar to standard K-means described above, the difference is that each data set has a degree of membership to each initial cluster [19], i.e. each data set belongs to all clusters to some degree. The degrees of membership for each data set to all clusters should sum to one.

The procedure starts with determining the number of clusters and guessing the cluster centre point (most likely incorrect), which is intended to mark the mean location of each cluster, then assigning every data set a membership grade for each cluster.

The next step is updating each cluster centre point and membership grade iteratively until the position of the centre point is stable. In this step the cluster centre point moves iteratively to the correct position within the data sets.

The Fuzzy K-means clustering technique does not create boundaries between data sets for the first iteration, because the clustering process involves all data. The boundaries will automatically evolve when the clustering process is completed.

Compared with the K-means method, the Fuzzy Kmeans process is longer, because the iteration process is not only updating the centre point but also the degree of membership of each data set.

The flowchart for this procedure is shown in Figure 5.

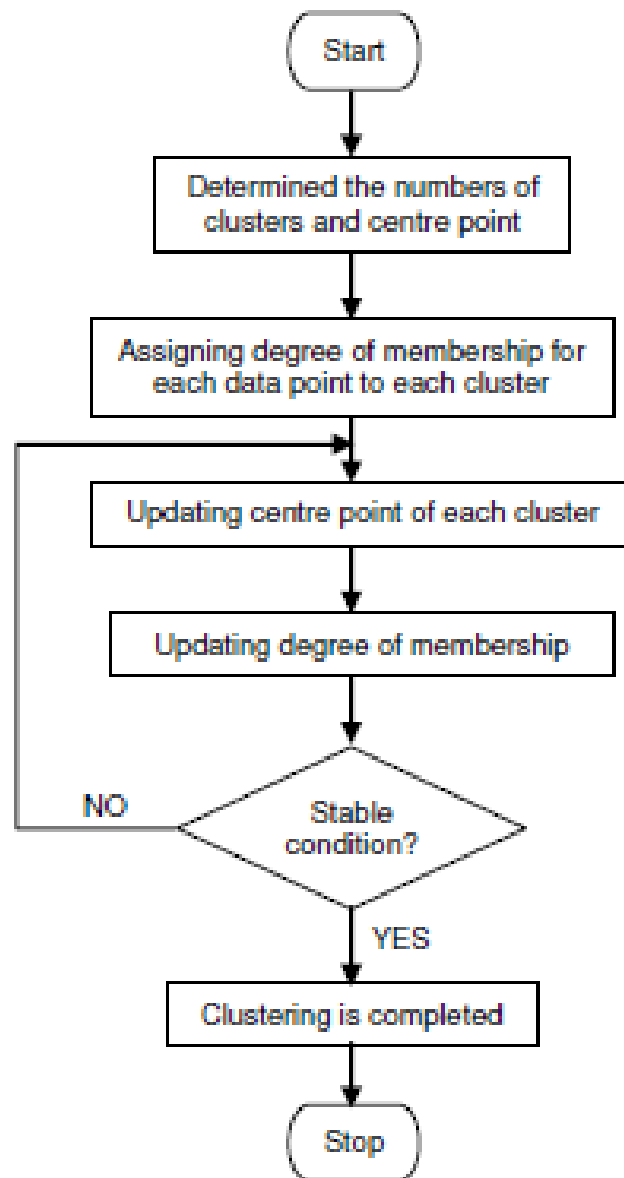


Figure 5. Fuzzy K-Means Clustering [24]

6.3.3. K-means Clustering

The K-means clustering method groups load profile data by determining a certain number of clusters and a centre point for each cluster. After determining the centre point of each cluster, each data set should be assigned to the nearest centre point then a recalculation of the new centre point will be done iteratively until the position of the centre point is stable [19].

Assigning data to a centre point which is evaluated using Euclidean, City Block, Cosine, Correlation, or Hamming distance, automatically creates boundaries between each data set. Each particular data set will become a member of the nearest cluster after the first iteration. The next iteration only has the function of updating the centre point position.

The K-means clustering sequence is described by the flowchart in Figure 6.

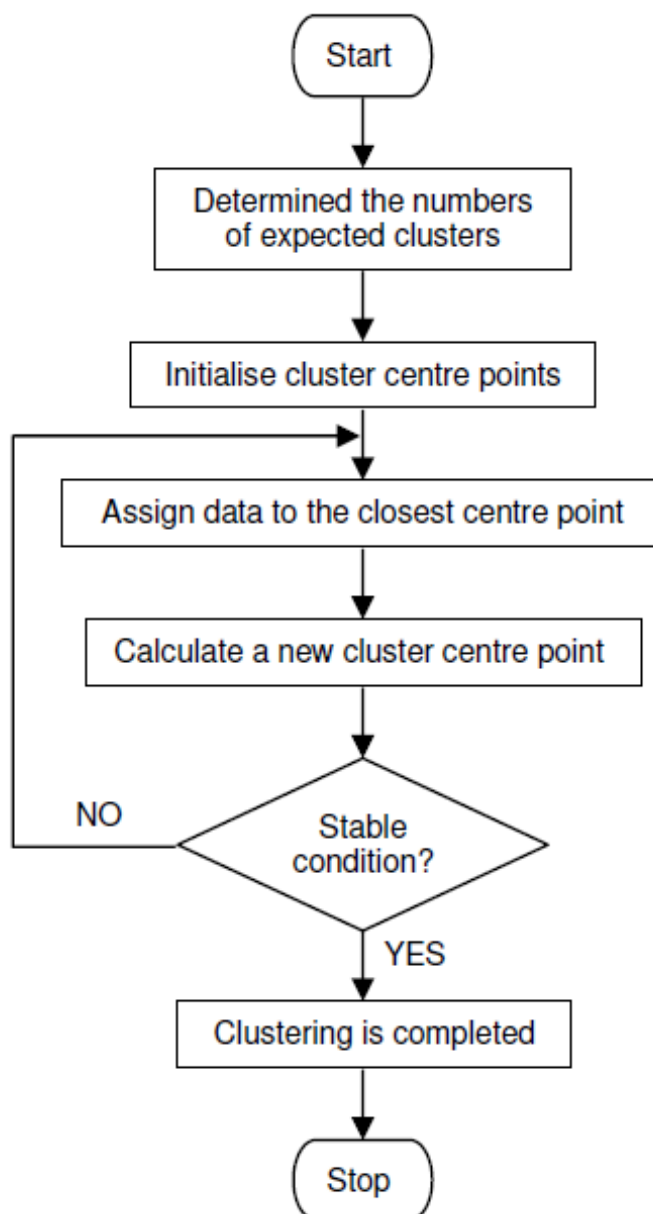


Figure 6. K-Means Clustering [24]

This method does not create a tree structure to describe the groupings of data, but rather creates a single level of clusters.

K-means clustering uses the actual observations of objects or individual data and therefore is more suitable for clustering large amounts of data.

6.3.4. Follow the leader Clustering

The Follow the Leader clustering method has been described in [21, 22]. This technique uses an iterative process to compute the cluster centers and it is not necessary to predetermine the number of clusters. The number of clusters is automatically derived from the determination of the distance threshold. The process stops when the cluster centre point is stable.

The function of the first iteration of the algorithm is to determine the number of clusters and their load profile membership. It means that the first iteration also creates boundaries between load profile data sets.

The subsequent iterations adjust the load profile pattern to the nearest cluster and then update the cluster centre point.

The selection of the distance threshold value should be done by trial and error until the expected number of clusters is produced.

Figure 7 shows the flowchart for the above process as described in [21].

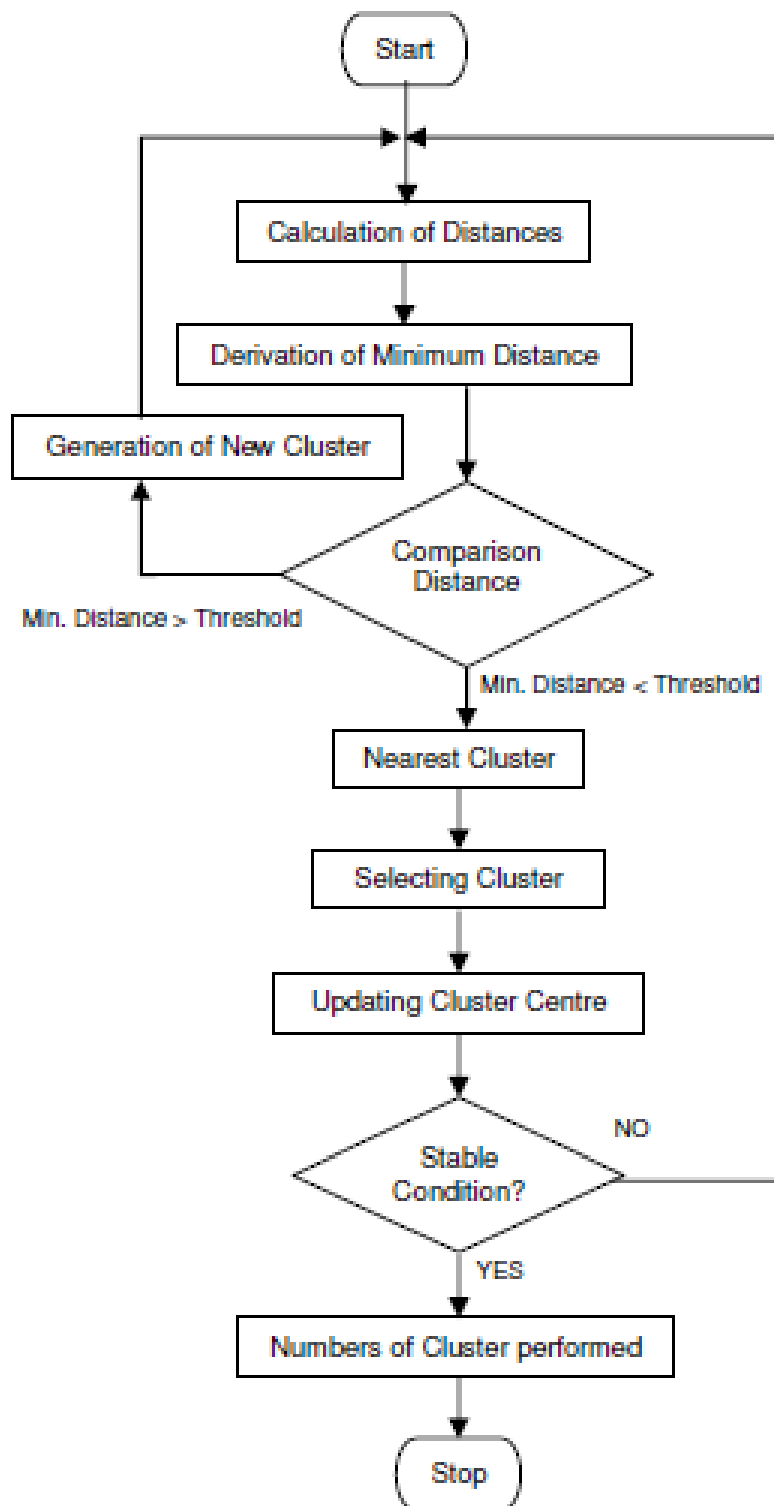


Figure 7. Follow the Leader Clustering [24]

6.3.5. Fuzzy Relation Clustering

Another technique to classify electricity load demand profiles uses Fuzzy Relation clustering as describe in [23].

Actually, this is a complex iterative process that can be simplified into the following steps:

1. Determine the similarity of load profile data sets using Cosine Amplitude Method.
2. Group the load profile data sets by using Max-Min Composition Method.
3. Determine the number of clusters data by using the Lambda-Cuts for fuzzy relation Method.
4. Obtain number of cluster.

Within the above process, there is an assignment of the degree of membership for each data set and an adjustment to the cluster centre position after determining the value of the threshold in the Lambda- Cuts process.

The number of clusters is dependent on the determination of the threshold value in the Lambda-Cuts process. This value can be determined by trial and error to produce the required number of clusters.

This technique uses fuzzy relation to evaluate the similarities and to group the data sets. Therefore, the boundaries between data sets are created in the final iteration after the threshold value is determined.

This technique is suitable for handling large sets of fuzzy data with complex interactions. The numbers of customer classes being decided by the Lambda-Cuts.

7. Load demand curves

Two types of electric consumption data are going to be analysed, as the amount of real clients is very small, 300 clients are going to be simulated using Gen Mix software and they are going to be compared with the real clients.

7.1. Real data

The real data is provided by the company Electra Norte. The data includes the energy consumption of 46 domestic customers in the month of February. The energy consumption is measured hourly. Figure 8 shows an example of one of these clients.

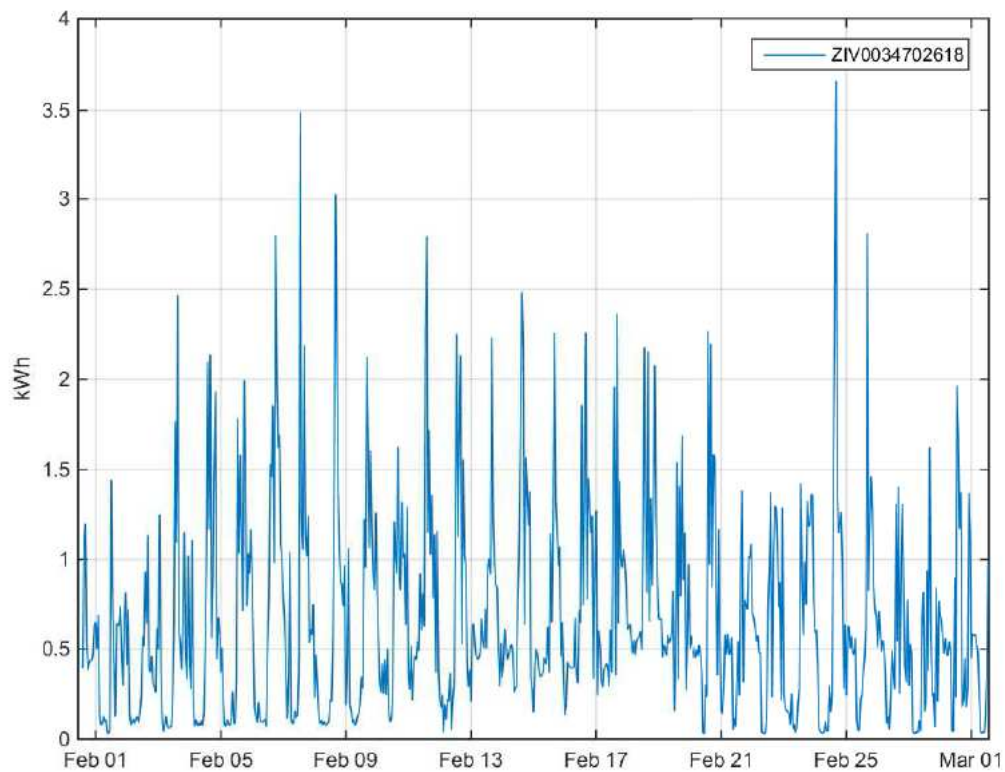


Figure 8. Hourly energy consumption of one domestic client in the month of February.

7.2. Simulated data: GenMix software

The simulated data is obtained from the GenMix software. The main objective of this software tool is to successfully assess professionals, promoters or companies in the selection of the best renewable-distributed generation-mix suitable to be installed in residential buildings considering technical, economic and reliability aspects.

To fulfill the mentioned goal, the following activities will be developed:

1. Discussion and implementation of renewable/distributed power generation models.
2. Discussion and implementation of a power demand model for residential buildings.
3. Discussion and implementation of an energy storage model for distributed generation.
4. Analyze the grid energy exchange for renewable/distributed generation systems in residential buildings.
5. Evaluate and validate the developed computer tool to assess the power generation-mix in residential buildings performing power reliability and economic analyses for study cases. [25]

For this thesis, among all the applications of this software, the discussion and implementation of a power demand model for residential buildings is going to be used to simulate the demand of different kinds of dwellings depending it periods of occupancy and the number of occupants.

7.2.1. Power demand modeling

The daily load profile shape will vary from day to day and apartment to apartment depending in the following factors:

1. Daily energy consumption.
2. Period of the house unoccupied during the day.
3. Individual consumption profile for every appliance. [25]

7.2.2. Daily energy consumption.

The daily energy consumption per dwelling primarily depends in the number of occupants. For every country, there are representative statistical studies that detail the yearly average energy consumption per home for different number of occupants. For the Spanish case, the main organizations providing key information regarding the energy consumption levels and patterns for end-users are the IDAE (Instituto para la Diversificación y Ahorro de la Energía) [26] and REE (Red Eléctrica de España) [27].

Table 1. Typical energy consumption for a single dwelling in Spain. [25, 48]

Number of occupants	Typical yearly energy consumption in a single dwelling in kWh excluding space heating and hot water services	Typical seasonal energy consumption in air conditioning for a single dwelling in kWh
1	1200	300
2	2400	450
3	3250	550
4	3850	600
5	4500	650

Table 1 shows the typical yearly energy consumption (excluding space heating, hot water services and air conditioning) for Spanish dwellings. Similarly, it also details the air conditioning energy employed on summer.

Once this yearly energy consumption E_{year} is selected (from the corresponding column in Table 1 and depending in the number of occupants), we can split this energy into the different days of the year. However, as [29] details, the energy consumption significantly varies between the different days throughout the week, being Saturday the one with the biggest consumption, followed by Sunday (0.98 times Saturday) and the different weekdays (0.96 times Saturday). Bearing this in mind, the amount of daily energy consumption for any of the 365 days of the year can be found by using equations (1) to (3):

$$E_{year} = 261 * E_{weekday} + 52 * E_{Saturday} + E_{Sunday} \quad (1)$$

$$E_{Sunday} = 0.96 * E_{Saturday} \quad (2)$$

$$E_{weekday} = 0.96 * E_{Saturday} \quad (3)$$

Where:

E_{year} = Energy consumed in a year.

$E_{weekday}$ = Energy consumed in a week day from Monday to Friday.

$E_{Saturday}$ = Energy consumed in Saturday.

E_{Sunday} = Energy consumed in Sunday. [25]

On the other hand, the amount of seasonal air conditioning energy used is only split throughout summer days (from the June 1st until the August 31st). This means that the daily energy use for air conditioning (E_{daily_ac}) is obtained by equally dividing this seasonal consumption by 91 which is the number of summer days. This is shown in equation (4).

$$E_{daily_ac} = \frac{E_{Seasonal_ac}}{91} \quad (4)$$

7.2.3. Unoccupied period

The usage pattern is related to the number of occupants and the occupied period. For example, when people are not at home, most appliances will not be used. For the daily appliance electricity profile, at night while the occupants sleep, very little power is used (stand by and fridge-freezer), then may wake up and have breakfast, leave the house during the morning and then return around mid-day for lunch. In the evening, the meal is cooked, television is watched, washer machines and dryers are used, etc.

The four most common scenarios of household occupancy pattern in Spain have been considered:

- Scenario 1: Unoccupied period is from 09:00 to 14:00. The occupants in this type of household may have a part-time job in the morning session.
- Scenario 2: Unoccupied period is from 14:00 to 19:00. The occupants in this type of household may have a part-time job in the afternoon session.
- Scenario 3: Unoccupied period is from 09.00 to 18.00. The occupants in the house have full-time job.
- Scenario 4: The house is occupied all the time. The family of this type of household may have minor children to take care of or retired members.

For weekend days, all dwellings will be considered to be under Scenario 4, this is not having an unoccupied period as it happens in most homes worldwide. [25]

7.2.4. Usage hours for the different loads

Depending on the number of occupants and the occupancy pattern, the number of usage hours for every appliance has to be selected as Table 2 shows for one occupant for example. The daily usage hours for minor loads or appliances (video games consoles, radios, watches, chargers, etc.) named as "Others" in the Table, are randomly selected between 3, 4 or 5 hours. Additionally, the air conditioning load only takes place during summer days. Note that there are similar tables with different usage hours for 2 up to 5 occupants per dwelling. The numerical values in these tables for the different number of occupants have been conjectured from [26].

Table 2. Usage hours for every load for one occupant

Load	Unoccupied period			
	9:00 to 14:00	14:00 to 19:00	9:00 to 18:00	No
Cooker	3	3	3	3
Fridge	12	12	12	12
TV	6	6	4	8
Washer	1	1	1	1
Stand-by	24	24	24	24
Oven	1	1	1	1
PC	4	4	3	6
Dish Washer	1	1	1	1
Dryer	1	1	1	1
Others	3, 4, 5*	3, 4, 5*	2, 3, 4*	4, 5, 6*
Air conditioning	3**	3**	3**	5**

*Are randomly selected for every day

**Only during summer days

A special consideration has been placed to get the number of usage hours for the lighting system as the natural sunlight hours D (known as day length) significantly varies for different latitudes. Figure 9 exposes the number of natural light hours and the selection of artificial light hours during a year for northern latitude equal to 40.40 (Madrid/Spain Latitude). [25]

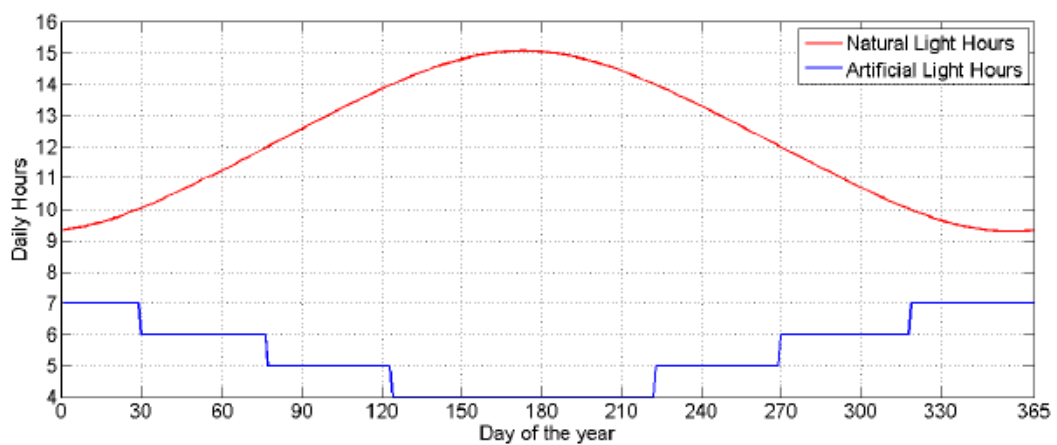


Figure 9. Artificial and Natural Light Hours throughout the year for $L=43.53^\circ$. [1]

7.2.5. Consumption profile for the different loads

Now that the usage number of hours for every load has been acquired, the specific hours of the day where those loads are being used have to be settled. This is done by performing a weighted random selection between the 24 hours of the day, where every hour presents a specific heaviness. Every appliance/load presents its own hourly weighting percentages according to its pattern of use for the different unoccupied periods. The only exceptions to this rule are the Freezer and the Stand-by loads.

Later, the daily individual energy usage for the different loads, E_{daily_loadX} , can be attained as a percentage of the daily energy consumption E_{daily} , as it is shown in equation (5)

$$E_{daily_loadX} = E_{daily} * \%_{loadX} \quad (5)$$

7.2.6. Simulated load demand curves

600 clients are going to be simulated using this tool. Having each 30 of them the same customer profile, their profile depends on the number of occupants (from 1 to 5) and unoccupied period (from 9:00 to 14:00, from 14:00 to 19:00, from 14:00 to 19:00 or non unoccupied period). Hence, there are 20 groups of clients with the same simulated profile.

The data which is going to be input of the visualization tool is simulated during winter months. As the real data that is used in this thesis is from the month of February, we use simulated data of consumption in winter months in order to compare then the real and simulated data.

Table 3 shows the 20 different kinds of clients depending on the number of occupants and the unoccupied period.

Table 3. Simulated clients type depending on the number of occupants and the unoccupied period.

Client type	Number of occupants	Unoccupied period
1	1	9:00 to 14:00
2	1	14:00 to 19:00
3	1	9:00 to 18:00
4	1	No
5	2	9:00 to 14:00
6	2	14:00 to 19:00
7	2	9:00 to 18:00
8	2	No
9	3	9:00 to 14:00
10	3	14:00 to 19:00
11	3	9:00 to 18:00
12	3	No
13	4	9:00 to 14:00
14	4	14:00 to 19:00
15	4	9:00 to 18:00
16	4	No
17	5	9:00 to 14:00
18	5	14:00 to 19:00
19	5	9:00 to 18:00
20	5	No

Following, different types of clients, with different unoccupied period and different number of occupants are going to be simulated with GenMix in order to see the differences between their demand.

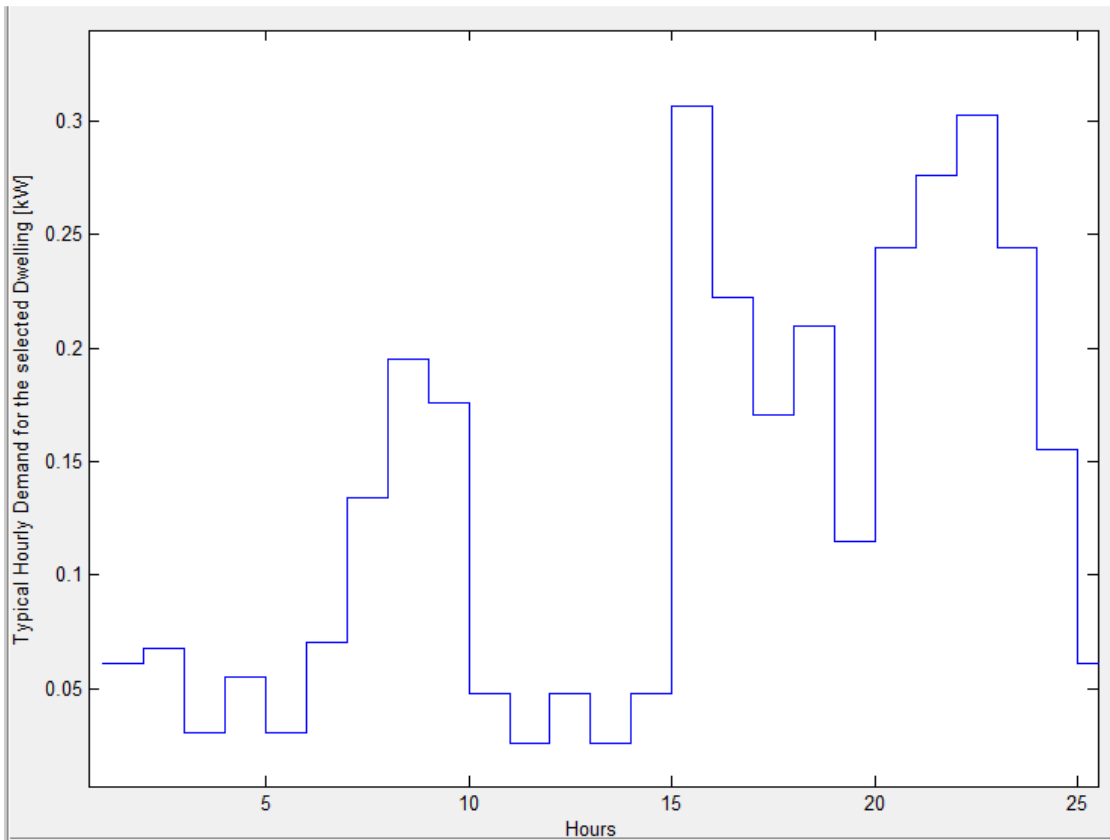


Figure 10. GenMix simulated demand for client type 1 during one day.

Figure 10 shows the demand in one day for client type 1. This client has one occupant and an unoccupied period from 9:00 to 14 :00. It can be seen that the consumption from 00:00 to 7:00 is very low . Then, there is a peak in the morning when the occupant gets up which decreases at 9:00 hours. From 9:00 to 14 :00 the demand ins nearly 0 kWh. Then, the energy demand increases again during the evening, having two peaks around 15:00 and 23:00.

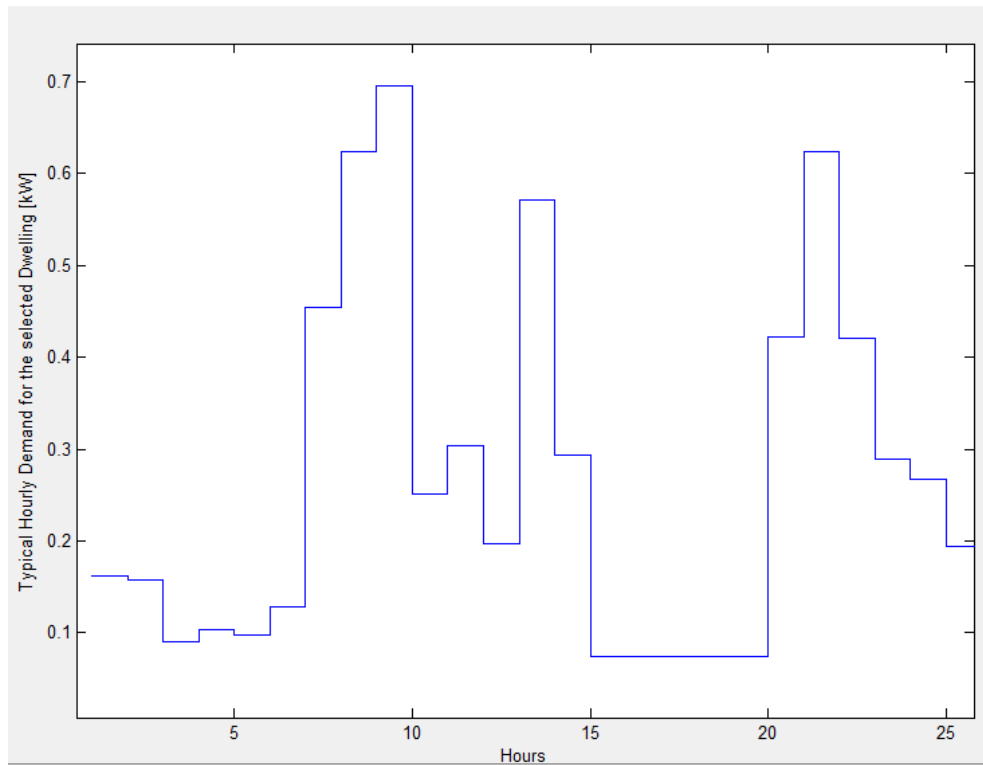


Figure 11. GenMix simulated demand for client type 6 during one day.

Figure 11 shows the demand in one day for client type 6. This client has two occupants and an unoccupied period from 14:00 to 19:00. It can be seen that the consumption from 00:00 to 7:00 is very low. Then, there is a peak in the morning when the occupants get up, and another peak around 13:00. From 15:00 to 20:00 the demand is nearly 0 kWh. Then, the energy demand increases again, having one peak around 22:00.

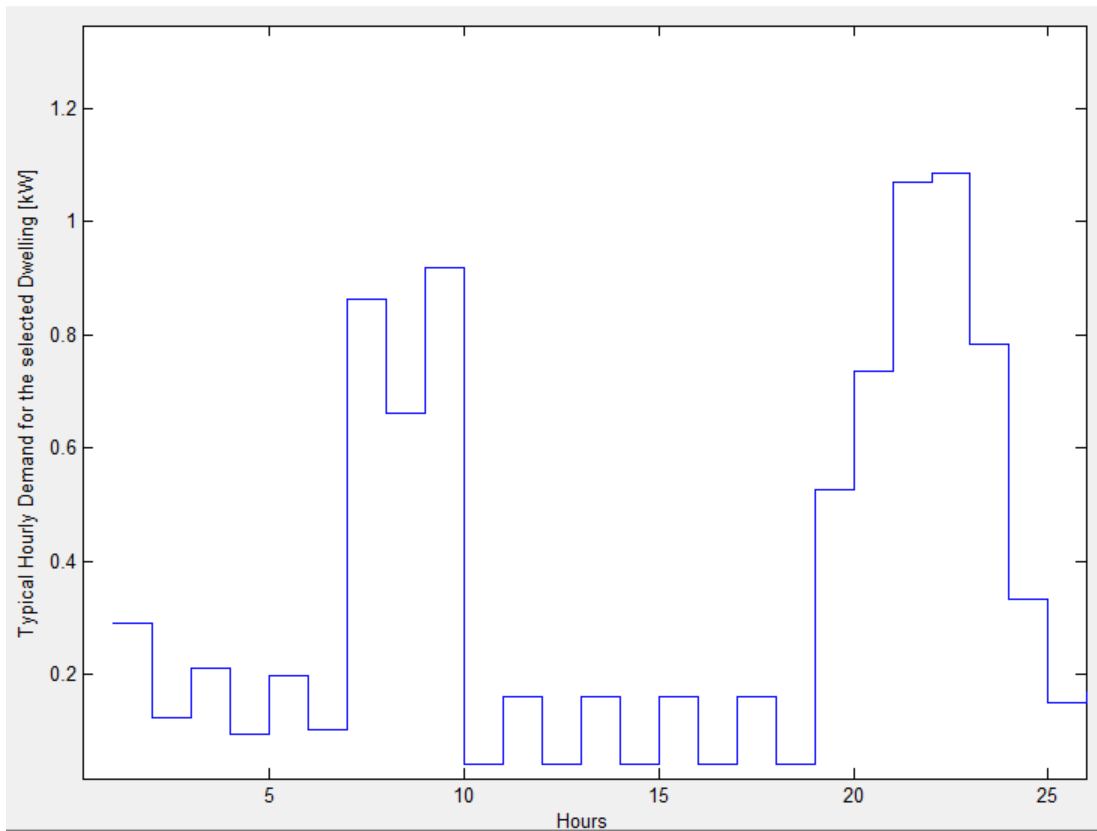


Figure 12. GenMix simulated demand for client type 11 during one day.

Figure 12 shows the demand in one day for client type 11. This client has three occupants and an unoccupied period from 9:00 to 18:00. It can be seen that the consumption from 00:00 to 7:00 is very low. Then, there is a peak in the morning when the occupants get up, until 10:00 when the consumption decreases to nearly zero until 19:00. Then, the energy demand increases, having its highest peak around 22:00.

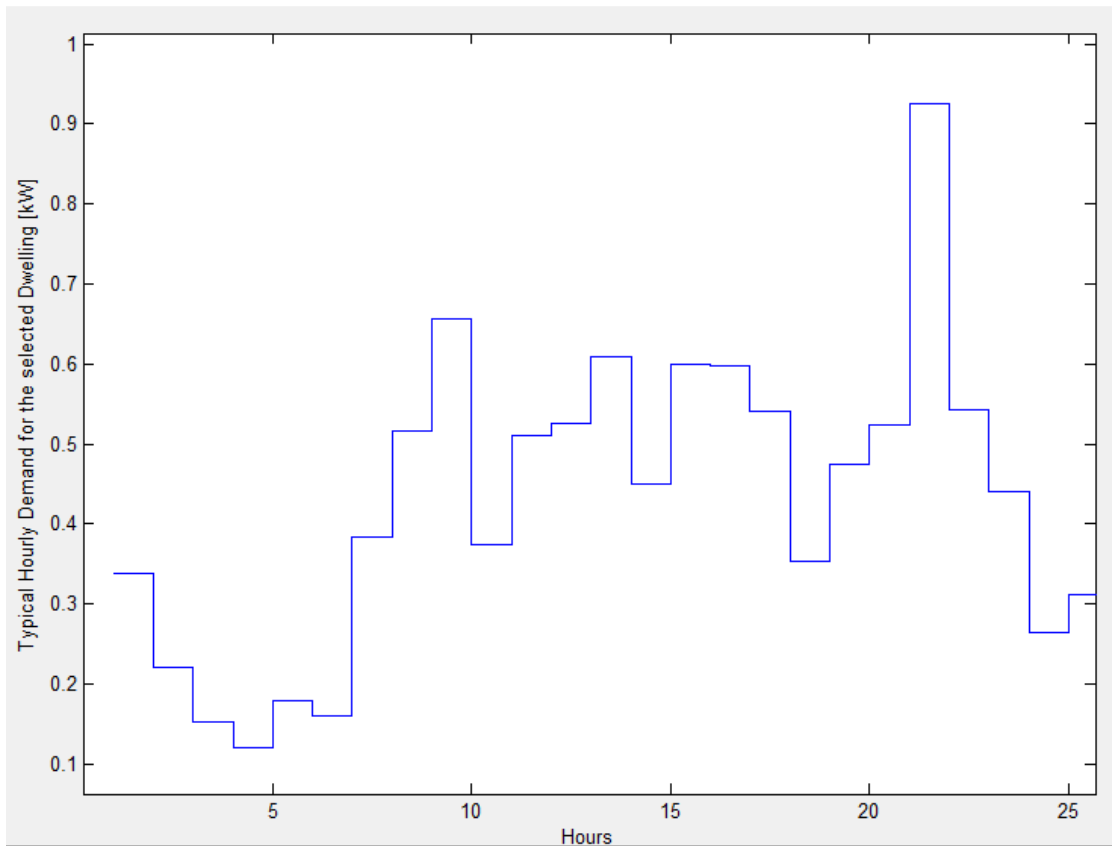


Figure 13. GenMix simulated demand for client type 16 during one day.

Figure 13 shows the demand in one day for client type 16. This client has four occupants and non unoccupied period. It can be seen that the consumption from 00:00 to 7:00 is very low. Then, there is a peak in the morning when the occupants get up. Then, the energy demand increases, being more or less constant, but with small peaks, during the morning and evening, and having its higher peak around 22:00.

8. Visualization tool

8.1. Introduction

Many problems today involve the analysis of large datasets, which also contain a very large number of variables from which the user should be able to find meaningful relationships to acquire knowledge. The mere fact of obeying laws, rules or restrictions arising from the problem domain, leads to dependencies that make the intrinsic dimensionality of the data to be much smaller. Dimensionality reduction (DR) algorithms are able to find low dimensional latent structures hidden in high dimensional data and produce a mapping on a low dimensional space that preserves the underlying structure of data. They are extremely useful tools in the field of visual analytics, since they provide an advanced way for spatialization of data, which allow creating visual representations where spatial proximity between two items y_i and y_j in the visualization represents similarity between x_i and x_j in a high dimensional space.

Another key ingredient in visual analytics is interaction. Interaction techniques {zoom, pan, brushing, etc.} allow the user to reconfigure the visualization to focus on the interesting aspects of data or to discard information that is irrelevant to the available knowledge of the user. In this paper we present a novel approach for data visualization that suggests a low level integration of user interaction into the DR computation and visualization process, by means of the so-called interactive dimensionality reduction (iDR). [30]

8.2. The interactive Dimensionality Reduction approach

In the typical procedure to use DR algorithms for visual analytics, interaction is often done after DR computation on the input dataset. The user sets up an initial configuration for the DR algorithm, runs it until convergence and, after N iterations, the output results are used to produce visualization. The user may later use interaction techniques to reconfigure this visualization or even decide to run the DR algorithm again using another parameterization, starting the cycle again. This approach can be thought of as a batch mode interaction scheme for DR visual analytics.

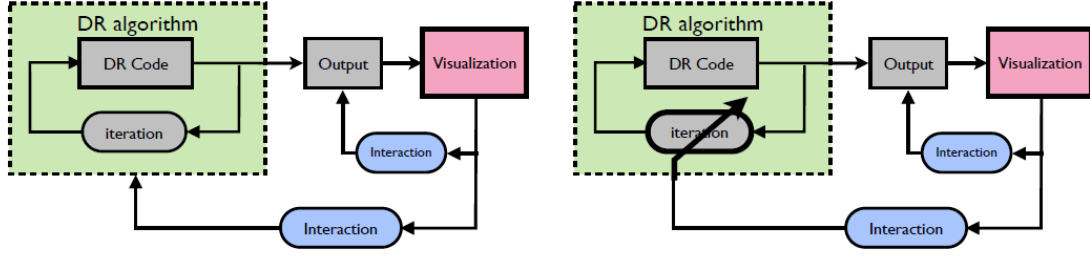


Figure 14. Batch mode interaction scheme (left) vs. the iDR approach (right) [30]

However, interaction can go far beyond this approach, allowing the user to take full control of the DR behaviour by means of iterative reconfiguration of computational algorithms. The right picture in Figure 14, shows the main idea of this approach for DR analysis, where the intermediate results are used to produce a visualization at each iteration. The result is a dynamically changing visualization that allows the user to track changes in the resulting projection under changes in the problem formulation, such as, for instance, user-driven changes of the metric in the input space (e.g. by modifying the weights of the input variables), or under time-varying input data (e.g. in dynamic processes where the elements of the input dataset change with time).

To allow interaction, we shall consider iterative algorithms, such as Stochastic Neighbour Embedding (SNE)

A web based visual interface including an interactive version of the SNE algorithm was implemented in JavaScript. Once the application is launched on a browser, it loads the dataset from a JSON file that includes a data matrix normalized to a $[-1,+1]$ range, the attribute labels, the item labels, the color map used as scale for attribute representation and the min and max values of all attributes for data denormalization. Along the whole session, the SNE algorithm

$$y_i(t+1) = y_i(t) - \eta(t) \frac{\partial C(t)}{\partial y_i} \quad (6)$$

is continuously evolving to minimize the cost function

$$C = \sum_i \sum_j p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (7)$$

where p_{ij} and q_{ij} describe the weighted probabilities of i and j being neighbors, respectively in the input data space and the low dimensional visualization space [30]

This is the code used to convert an excel into a the JSON file that includes a data matrix normalized to a [-1,+1] range:

```
from pylab import *

from pandas import *

options.display.width = 1000

options.display.max_columns = 1000

import locale

# Input of the excel

df = read_excel('REAL_DATA.xlsx')

# normalize to a [-1,+1] range

dfn = (df - df.min())/(df.max() - df.min())*2 - 1

fmin = df.min().tolist()

fmax = df.max().tolist()

J = dict()

J['data'] = [dfn.values.tolist(), (dfn.values*0.9).tolist(), (dfn.values*0.2).tolist()]

J['varnames'] = map(str,dfn.columns.tolist())

J['flimits'] = [fmin, fmax]

J['cmap'] = (cm.RdYlGn(linspace(1,0,200))*255).tolist()

J['labels'] = df.index.tolist()
```

```
f = open('data.json','w')
```

```
json.dump(J,f)
```

```
f.close()
```

9. Load demand profile clustering

In the web based visual interface the weights of the attributes can be changed “on the fly” by the user by moving sliders in the interface. The application works in “free-running mode”, executing one step of the DR algorithm at every frame using the current context (the input data, the weights for each attribute and the σ value) and renders the projection, achieving rates close to 20 frames per second. The system also allows the user to modify an inertia factor λ , which controls the smoothness of the animated transitions.

Using a, in a matter of one or two seconds the user can quickly sweep all the attributes encoded by the point size and a color scale. The different images are displayed as the user moves the slider.

Each point of the screen represents one client, as closer are the clients as similar they are. One of the attributes will be selected to represent the size and the colour of the client.

9.1. Define the best input attributes

The input attributes have to be selected carefully, considering that the clustering of the clients depends on them. For the real data, the smart meters collect a huge amount of data; the demand is measured for each client at each hour. As this amount of data cannot be entered into the visualization tool, some relevant parameters have to be found to represent each client.

First, not only for the real data, but also for the simulated data, a "day type" is computed calculating the average energy consumption at each hour of the day during one month. Reducing so the consumption of one month to one day, which represents the average day of each client.

To define the best input attributes the simulated data is going to be visualised. As table 3 shows, 600 clients are simulated using GenMix software. Among those clients there are groups of 30 clients that have the same number occupants and the same unoccupied period, so their load demand profile is very similar. Having 20 groups of different clients.

Two different kind of input parameters are going to be used, to find the best attributes for the visualization software. First, the input parameters will be the average consumption at each hour of the day and the total consumption in one average day for each client. Second, the average consumption for each hour will be added into periods; the histogram and the total consumption in one average day for each client will be also entered into the visualization tool.

As the group of each simulated client (from type 1 to type 20 shown in table 3) is known beforehand, the best input attributes will be the ones with which the clients with the same profile are closer in the space.

9.1.1. Average consumption at each hour of the day.

First, the attributes used for each client are:

- The average consumption of each client at each hour of the day.
- The total consumption in kWh (labelled as "TOTAL").
- The type of client depending on the type of profile that was used for the simulation, types from 1 to 20 shown in table 3, (labelled as "TYPE"). This last attribute is only used for validation purposes.

Table 4 shows the mentioned input attributes the simulated clients. As the clients are simulated in 20 groups of 30 clients with similar profile, only 20 of the 600 simulated clients, one of each group, are shown in the table. In order to see the approximate magnitude of the input parameters for each kind of client.

Table 4. Input attributes (average consumption at each hour of the day) for the simulated clients.

Attributes	Client 0001	Client 0031	Client 0061	Client 0091	Client 0121
hour 01	0,07184987	0,08672036	0,10647397	0,09084924	0,20461458
hour 02	0,06843927	0,07605491	0,05069458	0,0608951	0,10763047
hour 03	0,03990599	0,05363503	0,06610314	0,03769712	0,14464704
hour 04	0,06348807	0,067338	0,04187601	0,05863275	0,09576153
hour 05	0,03643666	0,03821491	0,0658242	0,04319804	0,12558729
hour 06	0,06794708	0,07169646	0,08205784	0,06463576	0,10852419
hour 07	0,17295799	0,14351536	0,22478143	0,10502215	0,33736451
hour 08	0,27557336	0,33897771	0,22823919	0,0967971	0,4650237
hour 09	0,27785915	0,1927049	0,32142273	0,24742844	0,58938758
hour 10	0,04232929	0,17843725	0,03141572	0,14018099	0,05191788
hour 11	0,03141572	0,13372888	0,04232929	0,13252907	0,09557214
hour 12	0,04232929	0,16253384	0,03141572	0,12665598	0,05191788
hour 13	0,03141572	0,35748798	0,04232929	0,1411805	0,09557214
hour 14	0,04232929	0,14147702	0,03141572	0,17102522	0,05191788
hour 15	0,18258829	0,02595894	0,04232929	0,22970163	0,48456053
hour 16	0,21975216	0,04778607	0,03141572	0,28932423	0,49829803
hour 17	0,08830684	0,02595894	0,04232929	0,25264679	0,42452015
hour 18	0,16110218	0,04778607	0,03141572	0,08129092	0,275648
hour 19	0,18949598	0,02595894	0,30139438	0,15211497	0,32277086
hour 20	0,22085826	0,18654311	0,15747161	0,19000032	0,33592381
hour 21	0,33250769	0,25982673	0,3549556	0,19269688	0,48948599
hour 22	0,22030625	0,27318565	0,40161727	0,12685639	0,37370851
hour 23	0,13785649	0,19380841	0,33247413	0,12037309	0,42679043
hour 24	0,243697	0,1314124	0,19896603	0,10901519	0,36435061
TOTAL	3,26074786	3,26074786	3,26074786	3,26074786	6,52149572
TYPE	1	2	3	4	5

Attributes	Client 0151	Client 0181	Client 0211	Client 0241	Client 0271
hour 01	0,19059193	0,19000505	0,16473079	0,18770923	0,37372661
hour 02	0,10281763	0,10613703	0,08955161	0,24369864	0,17868293
hour 03	0,1016391	0,10487593	0,12616261	0,07605	0,1248406
hour 04	0,06283145	0,10815348	0,09505483	0,18197264	0,09688839
hour 05	0,08465857	0,12195519	0,10210982	0,05255204	0,12644596
hour 06	0,06283145	0,21999707	0,10362529	0,23662409	0,21357447
hour 07	0,56522819	0,45479299	0,17089079	0,30883649	0,60566012
hour 08	0,35776704	0,70937678	0,23755485	0,60355637	0,67581069
hour 09	0,26439163	0,62167527	0,53114969	0,56243737	0,48511252
hour 10	0,267877	0,08465857	0,15712378	0,15897817	0,50522475
hour 11	0,28676269	0,06283145	0,49507965	0,0407479	0,3705316
hour 12	0,44162172	0,08465857	0,33184119	0,15897817	0,22234863
hour 13	0,66407262	0,06283145	0,24489337	0,0407479	0,72012366
hour 14	0,28919113	0,08465857	0,29779452	0,15897817	0,71915841
hour 15	0,08465857	0,06283145	0,40963865	0,61124235	0,11464182
hour 16	0,06283145	0,08465857	0,33486311	0,84864565	0,08508425
hour 17	0,08465857	0,06283145	0,29695504	0,48548646	0,11464182
hour 18	0,06283145	0,08465857	0,16666007	0,43276091	0,08508425
hour 19	0,08465857	0,34062143	0,29036562	0,42188539	0,11464182
hour 20	0,42853657	0,63477572	0,41850909	0,62202088	0,70204501
hour 21	0,73290795	0,56892807	0,60836796	0,68570808	0,66619555
hour 22	0,49334767	0,72980725	0,4502358	0,76605196	0,49118819
hour 23	0,39921601	0,64258411	0,2310558	0,38204917	0,67207291
hour 24	0,34556676	0,29319169	0,16728182	0,5634741	0,3674672
TOTAL	6,52149572	6,52149572	6,52149572	8,83119212	8,83119212
TYPE	6	7	8	9	10

Attributes	Client 0301	Client 0331	Client 0361	Client 0391	Client 0421
hour 01	0,2107037	0,15526201	0,33758441	0,37592832	0,26989728
hour 02	0,16377838	0,19515378	0,17560286	0,14966855	0,18647278
hour 03	0,16472271	0,09088394	0,19853559	0,19171934	0,13457593
hour 04	0,0986354	0,18970027	0,11404834	0,18310736	0,18167191
hour 05	0,18512026	0,08613718	0,19160925	0,17271137	0,15056349
hour 06	0,1686889	0,19172908	0,198545	0,23195202	0,29417241
hour 07	0,56739381	0,35603871	0,55678423	0,71352026	0,74905684
hour 08	0,80182508	0,41171988	1,01544443	0,76922122	1,18825166
hour 09	0,57599517	0,55217059	0,73096712	0,72502092	0,75758695
hour 10	0,07030547	0,59783214	0,10079211	0,28256648	0,13580646
hour 11	0,1294206	0,26924452	0,13580646	0,52637796	0,10079211
hour 12	0,07030547	0,51754275	0,10079211	0,56373963	0,13580646
hour 13	0,1294206	0,34055962	0,13580646	0,89706422	0,10079211
hour 14	0,07030547	0,47610761	0,10079211	0,53790517	0,13580646
hour 15	0,1294206	0,45634951	0,63170737	0,13580646	0,10079211
hour 16	0,07030547	0,64217705	0,91875428	0,10079211	0,13580646
hour 17	0,1294206	0,20759345	0,66247005	0,13580646	0,10079211
hour 18	0,07030547	0,35135412	0,48348838	0,10079211	0,13580646
hour 19	0,88148252	0,34128821	0,42128859	0,13580646	0,7034796
hour 20	1,02415539	0,49382093	0,72652453	0,85953171	1,15602423
hour 21	1,03466953	0,59732622	0,96264271	1,0713687	1,30350772
hour 22	1,03758078	0,55822737	0,53036063	0,55472096	1,04796564
hour 23	0,70926297	0,44922921	0,6934091	0,61624684	0,70369392
hour 24	0,33796779	0,30374394	0,33780993	0,43019142	0,55244494
TOTAL	8,83119212	8,83119212	10,4615661	10,4615661	10,4615661
TYPE	11	12	13	14	15

Attributes	Client 0451	Client 0481	Client 0511	Client 0541	Client 0571
hour 01	0,32952236	0,3178366	0,34410168	0,38442587	0,26514623
hour 02	0,19010464	0,22162397	0,24566561	0,19971507	0,16066126
hour 03	0,17532346	0,14224888	0,12684964	0,13827189	0,18052709
hour 04	0,12975624	0,19866904	0,19320707	0,18154033	0,18791446
hour 05	0,15269837	0,13181827	0,16381388	0,13827189	0,16906436
hour 06	0,13219875	0,15873483	0,31898537	0,3012556	0,21276635
hour 07	0,3331766	0,63226823	0,8365683	0,7633888	0,4574066
hour 08	0,40082008	0,93952242	0,81237724	1,16286782	0,51426982
hour 09	0,90707997	0,8995568	0,54293183	1,07117536	0,91728758
hour 10	0,34195057	0,15873483	0,49233331	0,13827189	0,63624236
hour 11	0,47649022	0,11780896	0,65293817	0,13827189	0,53656937
hour 12	0,33963717	0,15873483	0,54083235	0,13827189	0,63317914
hour 13	0,73067496	0,11780896	0,85278061	0,13827189	0,56427813
hour 14	0,42359571	0,15873483	1,03092466	0,13827189	0,64686884
hour 15	0,51587188	0,81392961	0,09734603	0,13827189	0,50301442
hour 16	0,68419215	0,91385522	0,17919776	0,13827189	0,79946086
hour 17	0,56072576	0,67576048	0,09734603	0,13827189	0,59862118
hour 18	0,46312055	0,43743341	0,17919776	0,13827189	0,39466558
hour 19	0,56438627	0,69038241	0,09734603	0,80694807	0,55218835
hour 20	0,60643099	1,32020676	0,95998939	1,44021935	0,7264576
hour 21	0,65507789	1,05321188	1,04397806	1,62375336	0,74217303
hour 22	0,64351263	0,70613185	1,03335657	1,14570122	0,69985173
hour 23	0,29928091	0,76263327	0,75410791	0,96707597	0,69464415
hour 24	0,40593794	0,50015813	0,63162923	0,65874684	0,43454599
TOTAL	10,4615661	12,2278045	12,2278045	12,2278045	12,2278045
TYPE	16	17	18	19	20

In the visualization software each point represents one client; the colour and the size of each point represent one attribute that can be selected from all the attributes that are an input of the software. In this case, the attribute "TYPE" is selected, which identifies each kind of client with one colour and one size. This attribute is selected to check the validity of the clustering, if the input parameters are valid, the points that are close in the visualization must have the same colour and size.

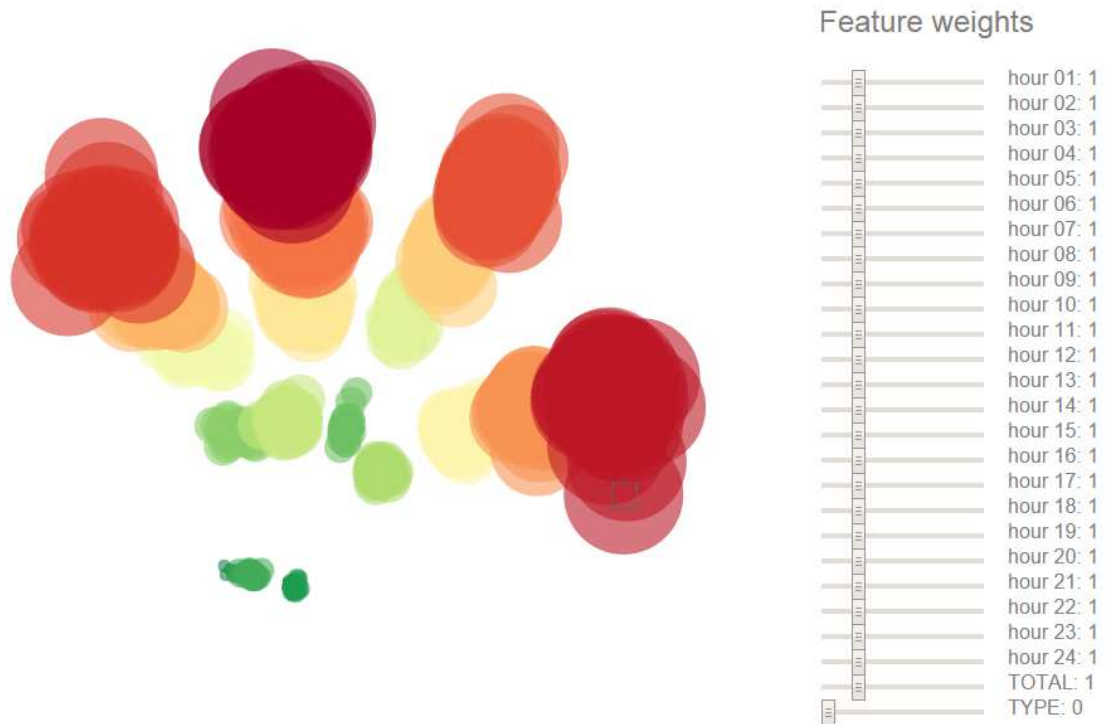


Figure 15. Simulated clients visualization depending on their average consumption at each hour of the week.

As figure 15 shows, the points with the same colour and size, and so the points with the same simulated profile, are gathered in one cluster, differentiating clearly each type of the simulated clients defined in table 3. Twenty clusters are clearly defined, one for each type of simulated client.

From this visualization it is known that this input parameters can define the energy demand of one client and also that the visualization tool is able to identify similarities between clients.

9.1.2. Average consumption at periods of the day and histogram

Second, the attributes that are input of the visualization tool are:

- The average consumption of each client at periods of the day.
- The histogram of each client. The Histogram block computes the frequency distribution of the elements in the input, in this case the frequency for a specific consumption. The histogram value for a given bin represents the frequency of occurrence of the input values bracketed by that bin. Five bins are used.
- The total consumption in kWh (labelled as "TOTAL")
- The type of client depending on the type of profile that was used for the simulation, types from 1 to 20 shown in table 3, (labelled as "TYPE"). This last attribute is only used for validation purposes.

Following the histogram of one of the clients is shown in figure 16, in order to explain what it represents and the values that are input of the visualization tool:

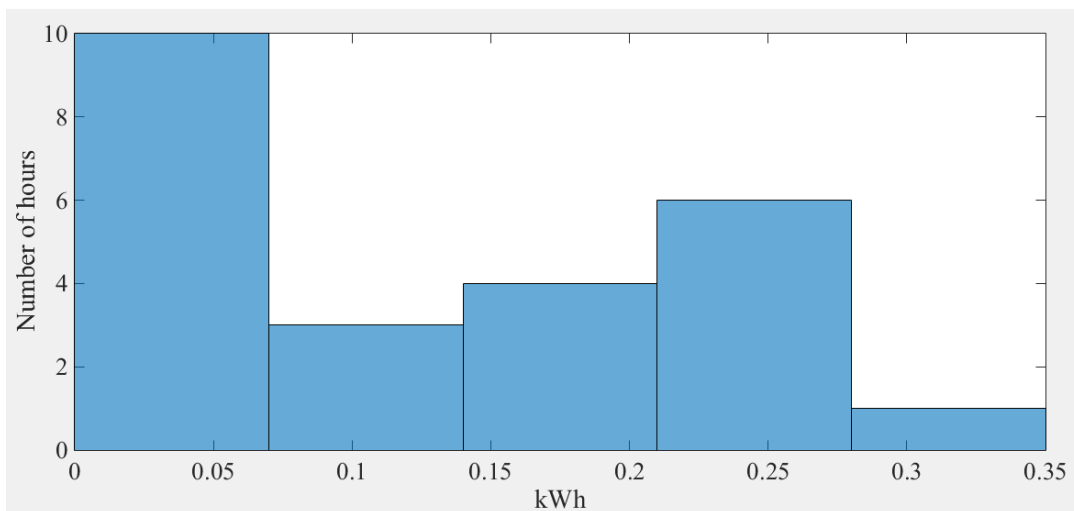


Figure 16. Histogram of one client

The number of bars selected is 5, the first bar means that the client is consuming from 0 to 0.07 kWh during 10 hours, the second bar means that the client is consuming from 0.07 to 0.14 kWh during 3 hours, the third one means that he is demanding from 0.14 to 0.21 kWh during 4 hours, the fourth one means that he is consuming from 0.21 to 0.28 kWh during 6 hours, and the last one means that he is consuming from 0.28 to 0.35 kWh during 6 hours. Hence, the numbers that are going to be input of the visualization tool are the hours that the client is demanding from less to more energy, in this case the numbers would be: 10, 3, 4, 6 and 1.

Table 6 shows the mentioned input attributes of one client of each one of the 20 types of the 600 simulated clients.

Table 5. Input attributes (average consumption at periods of hours of the day and histogram) for the simulated clients.

Attributes	Client 0001	Client 0031	Client 0061	Client 0091	Client 0121
E_01_06	0,34806693	0,39365967	0,41302974	0,35590802	0,7867651
E_07_09	0,7263905	0,67519797	0,77444334	0,44924768	1,39177578
E_10_14	0,18981931	0,97366497	0,17890574	0,71157176	0,34689792
E_15_19	0,84124544	0,17344896	0,4488844	1,00507853	2,00579756
E_20_24	1,15522569	1,0447763	1,44548464	0,73894186	1,99025936
Histo_01	12	11	14	6	10
Histo_02	1	4	2	8	1
Histo_03	4	5	3	4	6
Histo_04	4	2	2	3	3
Histo_05	3	2	3	3	4
TYPE	1	2	3	4	5
TOTAL	3,26074786	3,26074786	3,26074786	3,26074786	6,52149572

Attributes	Client 0151	Client 0181	Client 0211	Client 0241	Client 0271
E_01_06	0,60537013	0,85112375	0,68123493	0,97860664	1,11415895
E_07_09	1,18738685	1,78584505	0,93959532	1,47483023	1,76658333
E_10_14	1,94952516	0,37963862	1,52673251	0,55843031	2,53738704
E_15_19	0,37963862	0,63560147	1,49848249	2,80002075	0,51409396
E_20_24	2,39957496	2,86928684	1,87545047	3,01930419	2,89896885
Histo_01	11	14	10	9	9
Histo_02	4	2	5	3	2
Histo_03	5	2	3	4	3
Histo_04	2	1	4	6	3
Histo_05	2	5	2	2	7
TYPE	6	7	8	9	10
TOTAL	6,52149572	6,52149572	6,52149572	8,83119212	8,83119212

Attributes	Client 0301	Client 0331	Client 0361	Client 0391	Client 0421
E_01_06	0,99164936	0,90886627	1,21592545	1,30508696	1,21735379
E_07_09	1,94521406	1,31992919	2,30319577	2,2077624	2,69489545
E_10_14	0,4697576	2,20128665	0,57398926	2,80765346	0,60900361
E_15_19	1,28093466	1,99876236	3,11770867	0,60900361	1,17667675
E_20_24	4,14363644	2,40234766	3,2507469	3,53205963	4,76363645
Histo_01	15	6	10	11	15
Histo_02	1	3	3	2	1
Histo_03	2	5	4	5	4
Histo_04	2	5	4	4	1
Histo_05	4	5	3	2	3
TYPE	11	12	13	14	15
TOTAL	8,83119212	8,83119212	10,4615661	10,4615661	10,4615661

Attributes	Client 0451	Client 0481	Client 0511	Client 0541	Client 0571
E_01_06	1,10960382	1,1709316	1,39262324	1,34348065	1,17607976
E_07_09	1,64107664	2,47134745	2,19187737	2,99743198	1,88896399
E_10_14	2,31234863	0,7118224	3,56980911	0,69135947	3,01713784
E_15_19	2,78829661	3,53136113	0,6504336	1,36003564	2,84795039
E_20_24	2,61024036	4,34234189	4,42306116	5,83549674	3,29767251
Histo_01	5	11	9	15	6
Histo_02	8	2	2	1	3
Histo_03	5	6	5	3	6
Histo_04	5	4	4	3	7
Histo_05	1	1	4	2	2
TYPE	16	17	18	19	20
TOTAL	10,4615661	12,2278045	12,2278045	12,2278045	12,2278045

In this case, the colour attribute "TYPE" is also selected to check the validity of the clustering, if the input parameters are valid, the points that are close in the visualization must have the same colour and size.

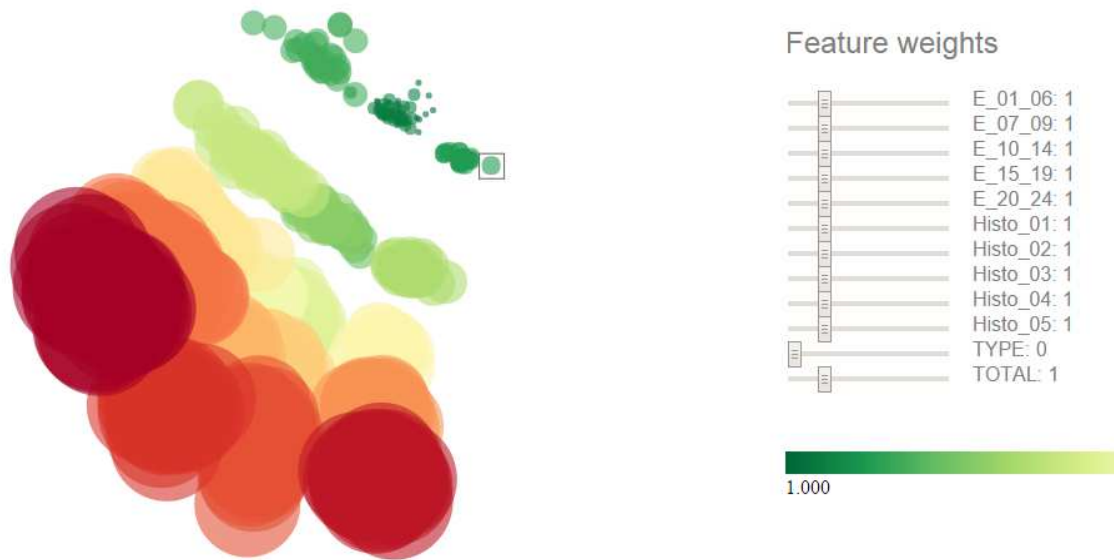


Figure 17. Simulated clients visualization depending on their average consumption at periods of the day and histogram.

As can be seen in figure 17, the clusters are not as well defined for this input attributes as for the input attributes of the average consumption for each hour. The clients with the same simulated profile are more dispersed, and the 20 groups of simulated clients are close together. Only 2 clusters can be defined, instead of the 20 clusters that were defined in the previous visualization.

Figure 18 shows the same visualisation than the previous one, but selecting the weight of the histograms as 0. It is appreciated that the clusters are better defined, and the clients with the same simulation profile are closer in the space. Hence, the histogram is not as good input attribute as the average energy demand at each hour.

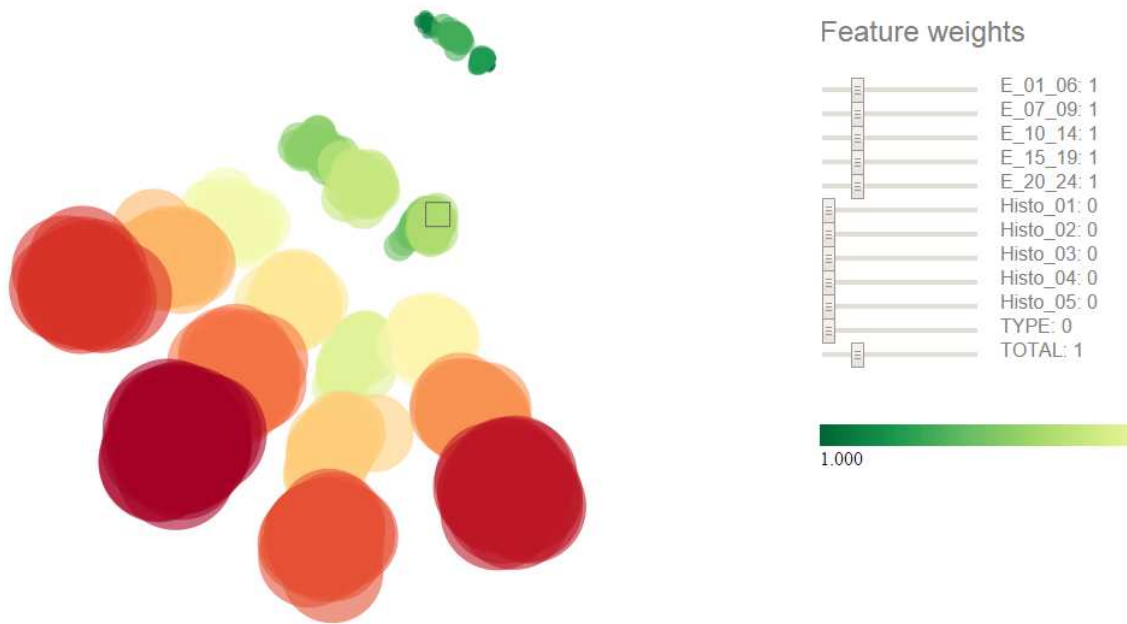


Figure 18. Simulated clients visualization depending on their average consumption at periods of the day.

It is known from both visualizations that the best input parameters for the visualization tool are:

- The average consumption of each client at each hour of the day.
- The total consumption in kWh (labelled as "TOTAL").

9.2. Simulated data load demand profile clustering

The simulated data is entered into the visualization tool as shown in table 4.

The input parameters for the visualization tool are:

- The average consumption of each client at each hour of the day.
- The total consumption in kWh (labelled as "TOTAL").

Figure 19 shows the first visualization, in which the colour attribute "TOTAL" is selected, which represents the total demand of each client in one average day. It shows different clusters of clients, from less (green) to more total energy consumption (red).

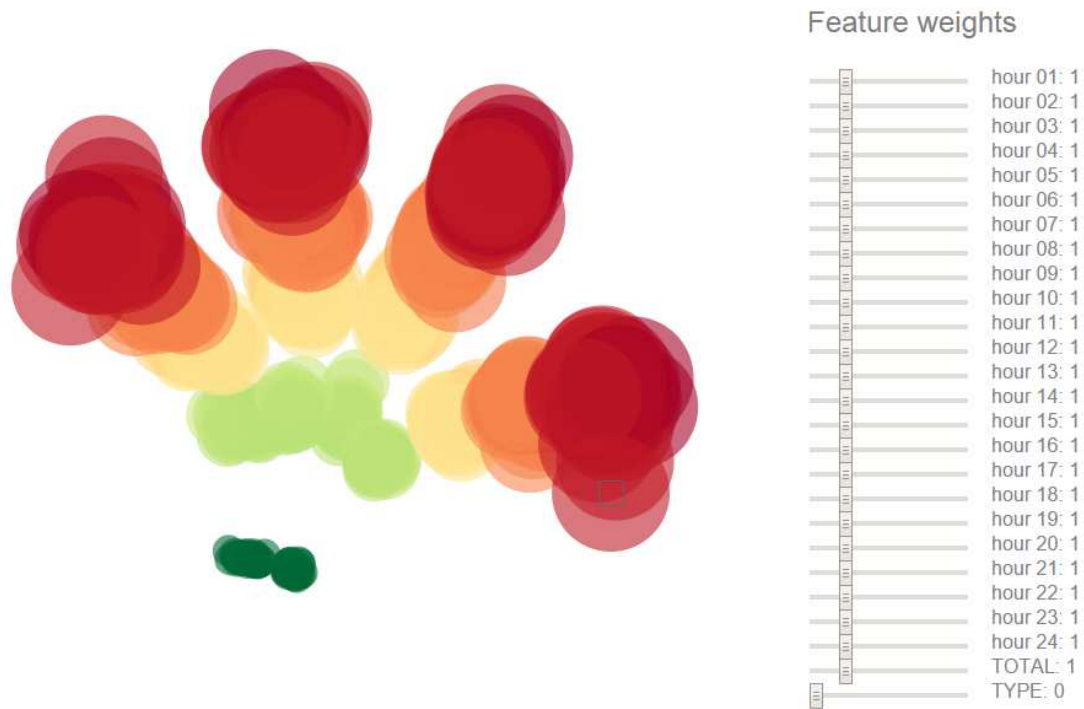


Figure 19. Simulated data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption.

Figure 20 shows the clusters of clients. It can be seen that there are 4 different rows of clients identified by lines of different colours. Each row represent one unoccupied period:

- From 9:00 to 14:00: Represented by the black line. It includes clients type 1, 5, 9, 13 and 17.
- From 14:00 to 19:00: Represented by the yellow line. It includes clients type 2, 6, 10, 14 and 18.
- From 9:00 to 18:00: Represented by the brown line. It includes clients type 3, 7, 11, 15 and 19.
- Non unoccupied period Represented by the blue line. It includes clients type 4, 8, 12, 16 and 20.

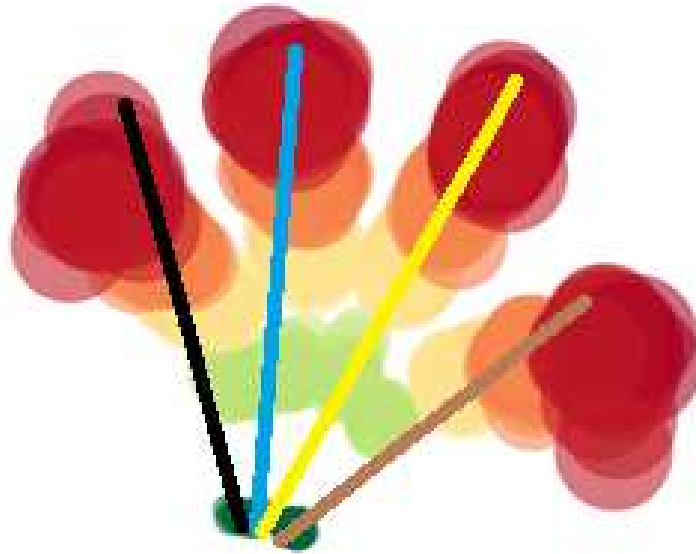


Figure 20. Clustering of the simulated clients depending on their unoccupied period.

Figure 21 shows the number of each type of client described in table 3 for each cluster.

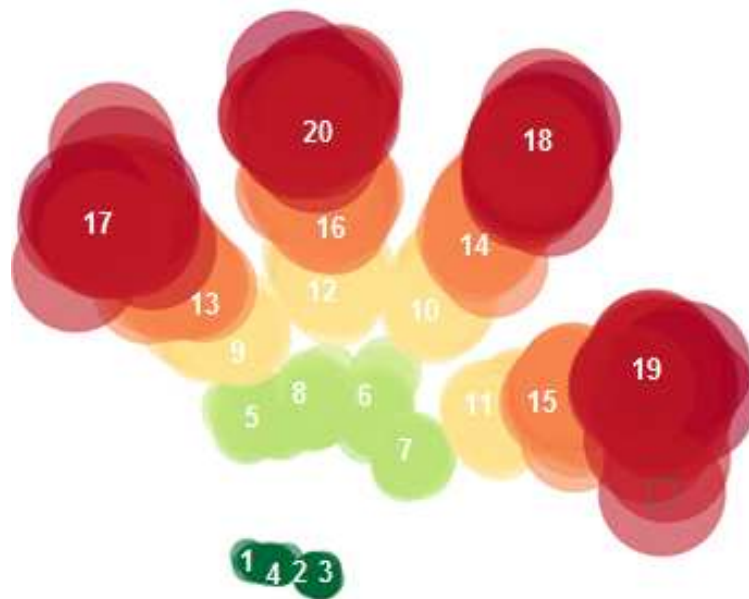


Figure 21. Clustering of the simulated clients depending on their unoccupied period and number of clients.

9.3. Real data load demand profile clustering.

The input parameters for the visualization tool are:

- The average consumption of each client at each hour of the day.
- The total consumption in kWh (labelled as "TOTAL").

The parameters of the real clients entered to the visualization tool are shown in table 6.

Table 6. Input attributes for the real clients.

Attribute	hour 01	hour 02	hour 03	hour 04	hour 05	hour 06	hour 07	hour 08
Client 01	0,27755	0,1116	0,06735	0,06255	0,0596	0,06105	0,05745	0,0633
Client 02	0,30435	0,4384	0,45055	0,49715	0,1414	0,11505	0,1163	0,12375
Client 03	0,037	0,0392	0,0346	0,04015	0,0365	0,0395	0,03455	0,0394
Client 04	0,1855	0,1151	0,04865	0,04665	0,04685	0,0454	0,0441	0,0428
Client 05	0	0,00005	0,0001	0,00015	0,0002	0,00025	0,0003	0,00035
Client 06	0,0936	0,04325	0,03735	0,03255	0,02885	0,02925	0,02765	0,0276
Client 07	0,47995	0,41055	0,3262	0,26835	0,2318	0,18015	0,1248	0,0762
Client 08	0,22485	0,0852	0,06735	0,06385	0,0615	0,06095	0,06675	0,1129
Client 09	0,11155	0,0841	0,04175	0,04115	0,0259	0,0339	0,0285	0,04245
Client 10	0,1085	0,1248	0,0548	0,0497	0,05045	0,0498	0,0836	0,06895
Client 11	0,1529	0,0999	0,046	0,03645	0,03165	0,0391	0,03755	0,1392
Client 12	0,1624	0,0619	0,05045	0,0488	0,04715	0,04325	0,04415	0,1791
Client 13	0,15575	0,09225	0,06675	0,06465	0,06765	0,06625	0,0904	0,0691
Client 14	1,051	0,4862	0,1349	0,07425	0,0721	0,05705	0,05145	0,0535
Client 15	0,3203	0,1953	0,0954	0,04205	0,0474	0,0427	0,0461	0,04285
Client 16	0,2412	0,1596	0,17655	0,07995	0,0488	0,04615	0,0367	0,0425
Client 17	0,1418	0,0515	0,03655	0,0376	0,0479	0,042	0,0415	0,0553
Client 18	0,4296	0,36745	0,2505	0,20525	0,1507	0,1289	0,101	0,09405
Client 19	0,18345	0,13755	0,10565	0,078	0,07485	0,0645	0,0659	0,06195
Client 20	0,46285	0,2755	0,16245	0,15495	0,15265	0,15995	0,1571	0,21635
Client 21	0,1101	0,0594	0,04705	0,0401	0,0411	0,0377	0,03945	0,04895
Client 22	0,0223	0,02235	0,022	0,02135	0,0218	0,02365	0,0217	0,0222
Client 23	0,075	0,0642	0,0621	0,0685	0,0755	0,0687	0,0713	0,0745
Client 24	0,26755	0,19985	0,1347	0,1012	0,0888	0,08915	0,17795	0,1003
Client 25	0,261	0,17685	0,1379	0,10255	0,0815	0,0777	0,0918	0,1489
Client 26	0,73525	0,2699	0,20445	0,15435	0,1278	0,1204	0,12865	0,16885
Client 27	0,10605	0,07265	0,07	0,0737	0,0735	0,07615	0,07045	0,27225
Client 28	0,0216	0,0251	0,0226	0,0192	0,02335	0,0281	0,0274	0,027
Client 29	0,0076	0,00795	0,00805	0,00815	0,00795	0,00805	0,0081	0,00825
Client 30	0,21105	0,20005	0,18185	0,18735	0,1478	0,0888	0,05045	0,0444
Client 31	0,1632	0,07485	0,05325	0,0468	0,04535	0,1125	0,0529	0,13805
Client 32	0,33065	0,2026	0,1909	0,1721	0,1549	0,14165	0,15125	0,1624
Client 33	0,2937	0,11895	0,07025	0,0639	0,07375	0,071	0,0843	0,12185
Client 34	0,0004	0,0005	0,0005	0,0006	0,0006	0,00065	0,0008	0,00075
Client 35	0,2229	0,13155	0,09185	0,07885	0,09065	0,0873	0,0851	0,08555
Client 36	0,40825	0,30745	0,2792	0,2059	0,14455	0,13825	0,1161	0,0939
Client 37	0,0021	0,0023	0,0024	0,00235	0,00245	0,0024	0,00245	0,0027
Client 38	0,1807	0,1637	0,0859	0,06785	0,05885	0,05775	0,0525	0,0612
Client 39	0,0301	0,02855	0,0284	0,02885	0,0276	0,0284	0,0304	0,02865
Client 40	0,021	0,0245	0,0234	0,02015	0,0226	0,02395	0,02245	0,02095
Client 41	0,40385	0,27775	0,2765	0,17355	0,08435	0,0755	0,07255	0,0686
Client 42	0,04935	0,036	0,0432	0,0378	0,04125	0,04065	0,03555	0,0324
Client 43	0,446	0,3199	0,17865	0,1367	0,14105	0,1348	0,1225	0,1024
Client 44	0,0262	0,02705	0,02725	0,0297	0,02745	0,02655	0,03095	0,0315
Client 45	0,0561	0,0478	0,0386	0,03165	0,03085	0,0346	0,03155	0,03425
Client 46	0,2909	0,30675	0,31645	0,3391	0,3042	0,301	0,3326	0,35975

Attribute	hour 09	hour 10	hour 11	hour 12	hour 13	hour 14	hour 15	hour 16
Client 01	0,05395	0,29135	0,15625	0,1004	0,1361	0,10455	0,25235	0,16755
Client 02	0,1179	0,3496	0,40545	0,4533	0,48835	0,35795	0,32615	0,3727
Client 03	0,0305	0,04245	0,03985	0,0332	0,04295	0,0322	0,045	0,03905
Client 04	0,0436	0,088	0,1454	0,1324	0,0699	0,07565	0,0724	0,06255
Client 05	0,0004	0,00045	0,0005	0,00055	0,0006	0,00065	0,0007	0,00075
Client 06	0,03785	0,0409	0,02885	0,03055	0,03985	0,03755	0,08755	0,0857
Client 07	0,0736	0,11105	0,11165	0,1112	0,1143	0,10905	0,1101	0,17785
Client 08	0,16995	0,1701	0,1763	0,20245	0,28825	0,2789	0,2293	0,1826
Client 09	0,04025	0,14775	0,1324	0,1113	0,0483	0,16455	0,34805	0,0789
Client 10	0,0577	0,0536	0,06185	0,14	0,34045	0,4272	0,335	0,3128
Client 11	0,20095	0,09285	0,02895	0,0249	0,04345	0,0456	0,02965	0,03545
Client 12	0,0893	0,0515	0,16315	0,14585	0,0686	0,1323	0,08645	0,1807
Client 13	0,08445	0,1252	0,1426	0,1106	0,0832	0,09395	0,1595	0,1245
Client 14	0,1928	1,4079	0,51315	0,3122	0,16965	0,2467	0,3872	0,40805
Client 15	0,05045	0,08505	0,1191	0,2056	0,3615	0,22365	0,1803	0,14515
Client 16	0,03855	0,5882	0,4341	0,47075	0,2235	0,2355	0,28305	0,2501
Client 17	0,06885	0,0532	0,07075	0,0419	0,0467	0,03735	0,04355	0,04245
Client 18	0,09005	0,0935	0,1137	0,08405	0,0967	0,17145	0,25925	0,64025
Client 19	0,0733	0,09665	0,1197	0,12945	0,2888	0,27855	0,2389	0,48915
Client 20	0,29175	0,15005	0,197	0,3469	0,2663	0,21845	0,2395	0,2411
Client 21	0,063	0,07375	0,12725	0,12365	0,0562	0,09685	0,0946	0,0576
Client 22	0,02345	0,0239	0,02195	0,02125	0,0225	0,03005	0,0272	0,02515
Client 23	0,19815	0,08255	0,1053	0,12255	0,11225	0,0659	0,0651	0,0604
Client 24	0,0929	0,09165	0,09565	0,15805	0,3003	0,37365	0,27555	0,6805
Client 25	0,36115	0,1731	0,15645	0,19755	0,17625	0,28295	0,7406	0,2219
Client 26	0,8355	0,37685	0,2238	0,2027	0,1958	0,2185	0,2227	0,4515
Client 27	0,3765	0,6884	0,23925	0,0754	0,06955	0,0719	0,07035	0,0707
Client 28	0,02515	0,0323	0,0375	0,03035	0,027	0,0247	0,0317	0,02765
Client 29	0,00825	0,00825	0,0084	0,00855	0,00895	0,0091	0,00885	0,009
Client 30	0,04555	0,05695	0,06855	0,0981	0,10975	0,0902	0,06355	0,15215
Client 31	0,2117	0,18965	0,081	0,0671	0,1011	0,08135	0,08635	0,0887
Client 32	0,2107	0,3087	0,3598	0,63155	0,55075	0,49425	0,4138	0,40375
Client 33	0,42885	0,0846	0,0623	0,0839	0,07555	0,07105	0,0924	0,0758
Client 34	0,00085	0,00085	0,00095	0,0009	0,0011	0,0011	0,00105	0,00125
Client 35	0,11565	0,1431	0,17065	0,25815	0,48105	0,45055	0,3758	0,2596
Client 36	0,1014	0,10475	0,1624	0,1412	0,14495	0,2011	0,48295	0,92265
Client 37	0,0025	0,00265	0,00265	0,00265	0,00275	0,0029	0,0028	0,0031
Client 38	0,05555	0,1279	0,07115	0,09045	0,1144	0,13935	0,1457	0,16195
Client 39	0,03195	0,033	0,0347	0,0355	0,0344	0,03695	0,03715	0,0311
Client 40	0,02705	0,02195	0,0279	0,01565	0,02285	0,0199	0,0248	0,0226
Client 41	0,1023	0,10295	0,0798	0,0903	0,07375	0,0694	0,08645	0,0997
Client 42	0,12795	0,1878	0,16315	0,2268	0,1856	0,21885	0,22515	0,1787
Client 43	0,1161	0,33085	0,3296	0,29835	0,33445	0,2129	0,26915	0,4468
Client 44	0,0291	0,0284	0,0286	0,03	0,0283	0,02805	0,0287	0,02895
Client 45	0,0314	0,02945	0,04455	0,0367	0,0291	0,0331	0,0295	0,0316
Client 46	0,31665	0,29635	0,2465	0,2385	1,1406	0,3567	0,39785	0,4118

Attribute	hour 17	hour 18	hour 19	hour 20	hour 21	hour 22	hour 23	hour 24
Client 01	0,27795	0,12145	0,20095	0,2756	0,3379	0,3183	0,4871	0,445
Client 02	0,2917	0,2212	0,23625	0,2732	0,29935	0,5317	0,4482	0,339
Client 03	0,03185	0,0422	0,0385	0,03385	0,0453	0,0368	0,04365	0,0404
Client 04	0,06085	0,06835	0,0675	0,0815	0,0991	0,17805	0,1845	0,1813
Client 05	0,0008	0,00085	0,0009	0,00095	0,001	0,00105	0,0011	0,00115
Client 06	0,04595	0,04595	0,04845	0,05225	0,04705	0,05895	0,1479	0,18455
Client 07	0,1437	0,15455	0,1382	0,17805	0,23865	0,2568	0,39325	0,529
Client 08	0,20785	0,13585	0,16185	0,19575	0,25545	0,1914	0,39295	0,29135
Client 09	0,1775	0,18225	0,08835	0,09605	0,1505	0,20515	0,17	0,1528
Client 10	0,30165	0,261	0,21655	0,24635	0,20165	0,24935	0,45155	0,31175
Client 11	0,0544	0,0643	0,0665	0,0663	0,07375	0,0667	0,1288	0,1816
Client 12	0,11135	0,0974	0,13925	0,1271	0,18705	0,1572	0,2263	0,3987
Client 13	0,198	0,14745	0,21925	0,2714	0,2878	0,34455	0,2334	0,20555
Client 14	0,2408	0,292	0,24155	0,168	0,2416	0,3722	0,5622	0,9037
Client 15	0,11715	0,1997	0,1456	0,1371	0,2447	0,3037	0,7892	0,5762
Client 16	0,3211	0,34395	0,1697	0,11975	0,1074	0,4108	0,40965	0,3034
Client 17	0,0387	0,0714	0,0614	0,1586	0,1607	0,1985	0,30485	0,22445
Client 18	0,59155	0,43685	0,42095	0,42915	0,42415	0,45425	0,5988	0,4804
Client 19	0,32725	0,17095	0,08875	0,08105	0,0995	0,21155	0,4094	0,30645
Client 20	0,3115	0,26835	0,2543	0,275	0,41605	0,64545	0,7121	0,56295
Client 21	0,0491	0,06145	0,0858	0,06085	0,1106	0,1928	0,24255	0,1248
Client 22	0,03795	0,0229	0,0247	0,02805	0,05765	0,02295	0,02405	0,0269
Client 23	0,15745	0,10995	0,09375	0,21175	0,1731	0,1453	0,1829	0,20675
Client 24	0,4547	0,336	0,19435	0,2381	0,253	0,2212	0,3113	0,2565
Client 25	0,2074	0,42235	0,31415	0,42055	0,42015	0,4143	0,47125	0,28155
Client 26	0,73245	0,60505	0,42875	0,42875	0,42755	0,4753	0,6501	0,785
Client 27	0,10785	0,1163	0,20425	0,46045	0,574	0,77845	0,8918	0,26845
Client 28	0,0276	0,02805	0,0275	0,02585	0,02095	0,02305	0,02655	0,02465
Client 29	0,0089	0,0091	0,009	0,02225	0,02305	0,01465	0,01335	0,0136
Client 30	0,08905	0,057	0,04875	0,13175	0,10475	0,1199	0,12805	0,2002
Client 31	0,08885	0,14485	0,19245	0,29565	0,3887	0,37535	0,325	0,26595
Client 32	0,4455	0,6244	0,3573	0,2352	0,2996	0,4024	0,50655	0,4355
Client 33	0,07605	0,11825	0,1946	0,28435	0,42435	0,48845	0,5636	0,56615
Client 34	0,00115	0,0035	0,0149	0,00465	0,0015	0,00135	0,00155	0,00165
Client 35	0,17315	0,2003	0,2145	0,23925	0,35295	0,3498	0,39385	0,28825
Client 36	0,68195	0,52505	0,565	0,45575	0,2468	0,3117	0,81955	0,71185
Client 37	0,0029	0,00295	0,0031	0,0031	0,00325	0,0033	0,0032	0,0033
Client 38	0,17355	0,13525	0,113	0,13945	0,14785	0,1406	0,14305	0,1252
Client 39	0,02985	0,03195	0,02985	0,0332	0,03005	0,032	0,02955	0,026
Client 40	0,0814	0,1283	0,12285	0,0671	0,0235	0,0299	0,0567	0,0319
Client 41	0,0986	0,10845	0,10175	0,0848	0,09445	0,1946	0,34395	0,48815
Client 42	0,1456	0,11465	0,0926	0,1429	0,17015	0,16645	0,16575	0,08475
Client 43	0,64175	0,55765	0,6629	0,41095	0,38755	0,6434	0,83665	0,59265
Client 44	0,0318	0,0301	0,0316	0,0295	0,02865	0,0291	0,0282	0,02795
Client 45	0,0516	0,0451	0,0383	0,0485	0,05165	0,064	0,0514	0,03945
Client 46	0,4046	0,46155	0,4186	0,3394	0,4053	0,4993	0,3463	0,32785

As the real data is made up only of 46 clients it is difficult to make clusters with this small amount of data. Figure 22 shows 5 clusters of clients depending on the total consumption: red, orange, yellow, light green and dark green.

Comparing this data with the simulated data we can classify each real client inside of one of the 20 groups of the simulated client. Knowing hence, the number of occupants and the unoccupied period of each real client.

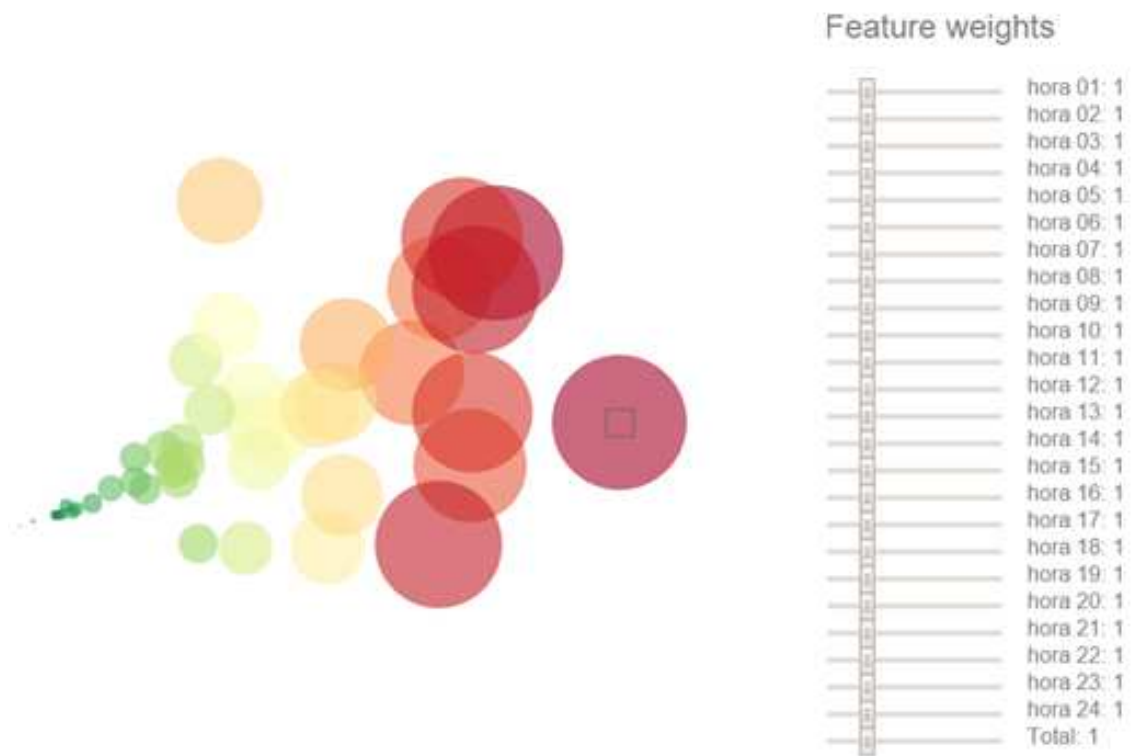


Figure 22. Real data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption

Following, the real clients are going to be visualize with the simulated ones in order to compare them and extract more information about the real data, such as the period of occupancy and the number of occupants.

9.4. Simulated and real data load demand profile clustering.

The real and simulated data are visualize together in order to classify the real data depending on the number of occupants and the unoccupied period.

The data that is entered into the visualization tool is shown in tables 4 (simulated data) and 6 (real data).

The input parameters for the visualization tool are:

- The average consumption of each client at each hour of the day.
- The total consumption in kWh (labelled as "TOTAL").
- The type of client depending on the type of profile that was used for the simulation, types from 1 to 21 shown in table 3, (labelled as "TYPE"). This last attribute is only used for validation purposes.
- The type of client, differencing real and simulated clients.

Figure 23 shows the real and simulated clients, representing the colour and size of each point the total consumption in one day.

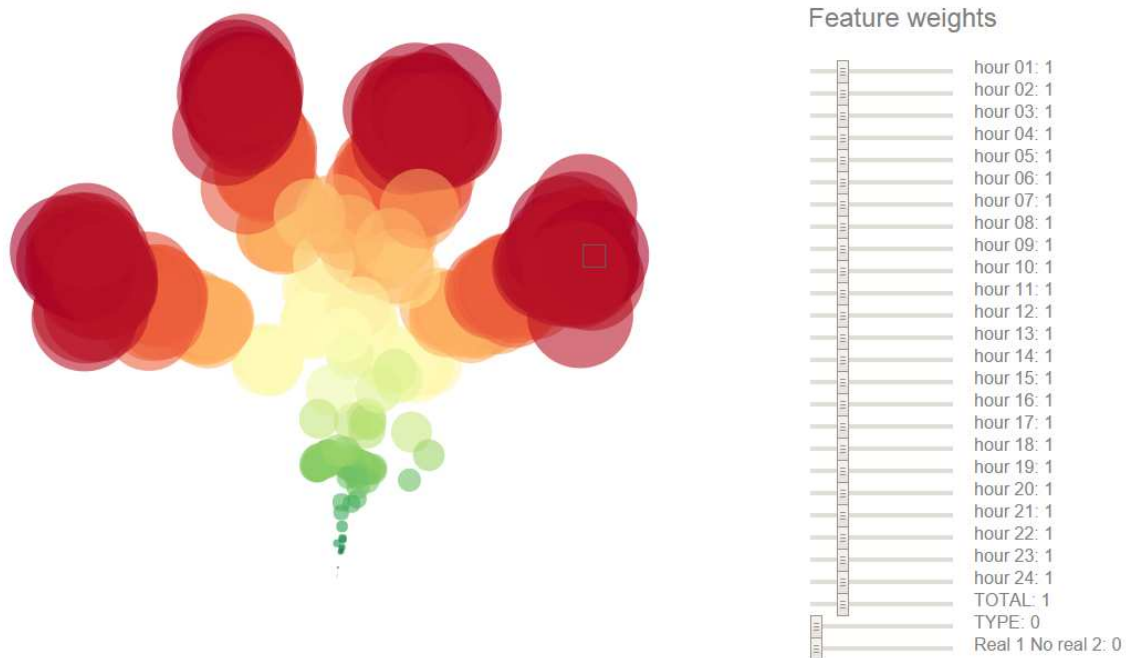


Figure 23. Real and simulated data visualization identifying the total average daily consumption by colours and sizes from more (red) to less (green) consumption

Figure 24 shows that the real clients (green) and the simulated clients (red) have similar profiles, both data are mixed together in the visualization.

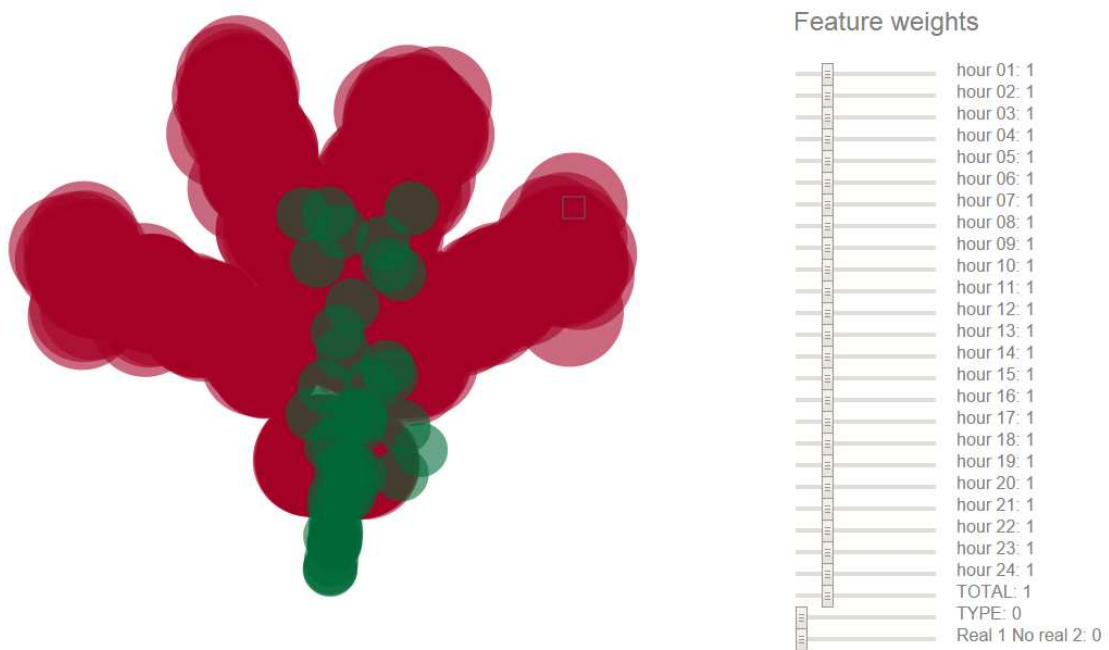


Figure 24. Real (green) and simulated (red) data visualization.

Figure 25 shows the two zooms that are made in order to see which real client belongs to which simulated type of client, described in table 3.

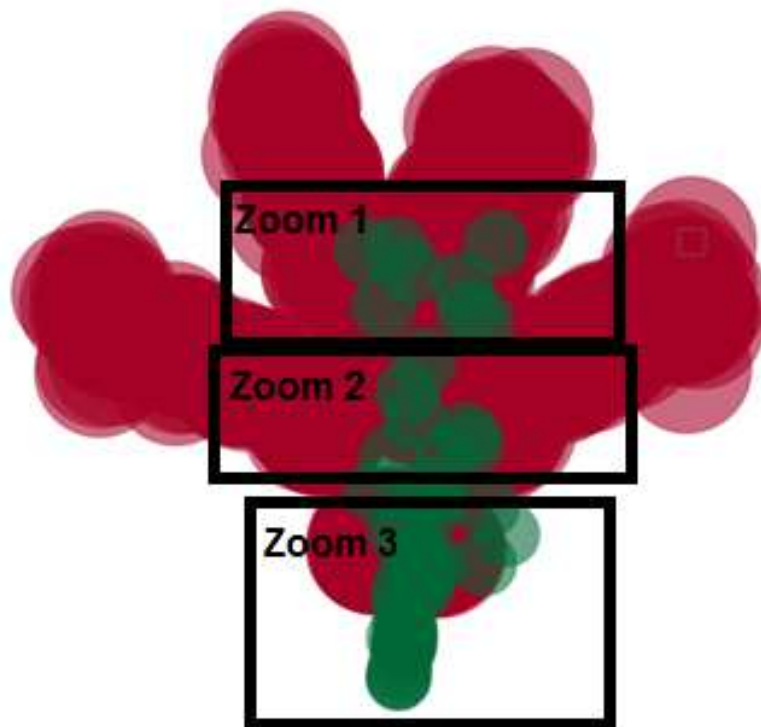


Figure 25. Real and simulated data zooms.

Figure 26 shows that the four simulated clusters in which the number of occupants is 1. A group of real clients belong to the cluster of clients type 4, another group to the type 2 and another client type 1. There is also a group of clients which do not belong to any cluster. This cluster represents the dwellings that are unoccupied, which is know from their load data curves, where the consumption is nearly zero most of the time.

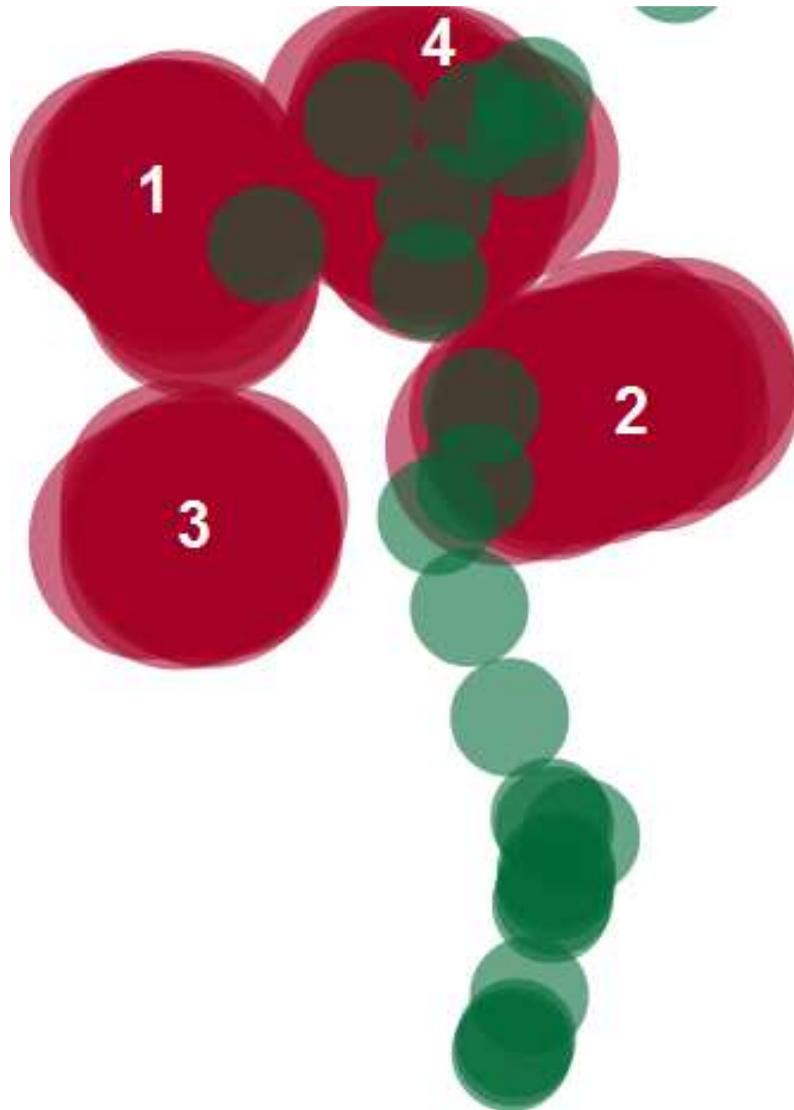


Figure 26. Real (green) and simulated (red) data visualization Zoom 1.

Figure 27 shows that there is a group of real clients that there are between clusters number 1, 2, 4 (1 occupant) and 8 (2 occupants) . Each client can be included into the nearest cluster.

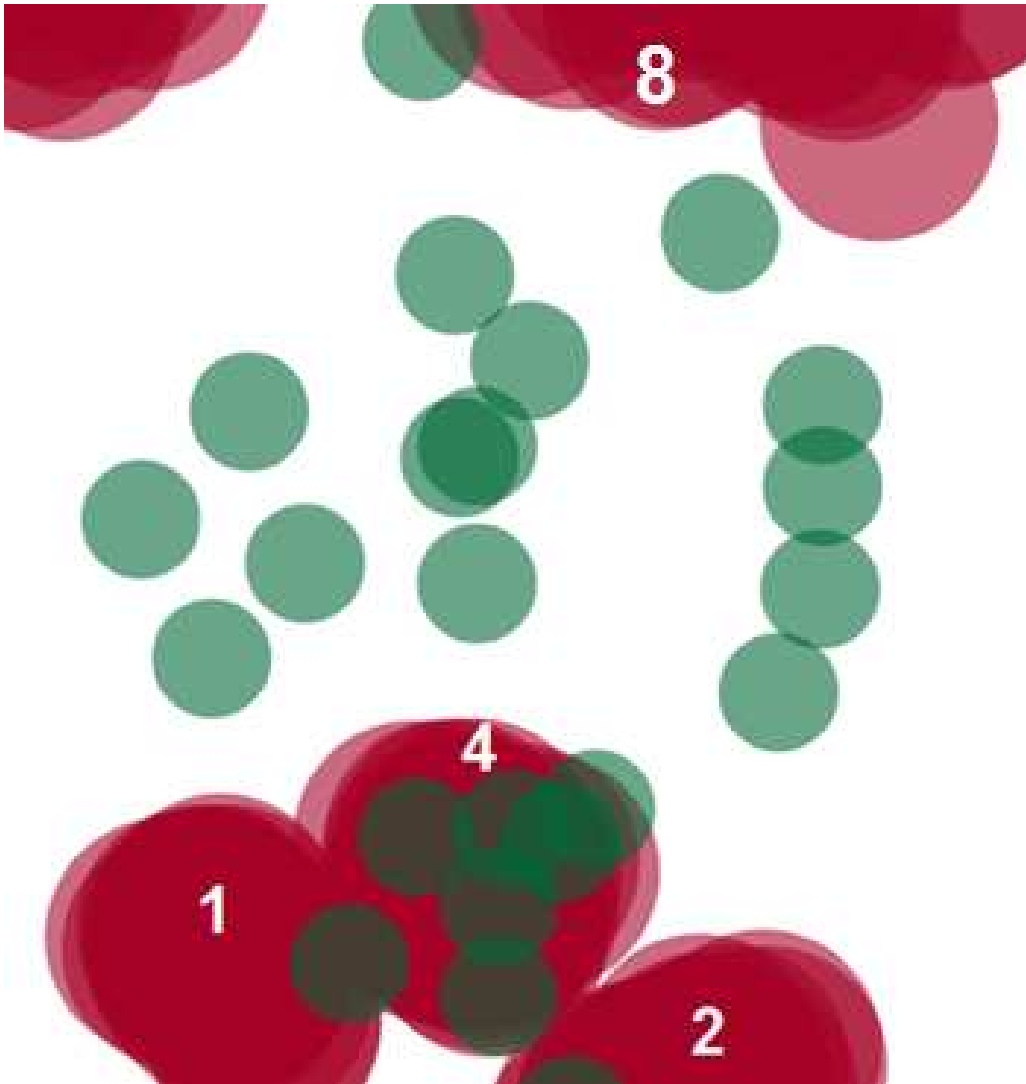


Figure 27. Real (green) and simulated (red) data visualization Zoom 2.

Figure 28 shows the four simulated clusters in which the number of occupants is 2, clusters number 5, 6,7 and 8, and two simulated clusters in which the number of occupants is 3, clusters number 9 and 12.

The visualization shows a group of real clients belong to cluster type 8. Another group of clients belongs to cluster type 12. There is another group of clients between clusters types 9 and 12.

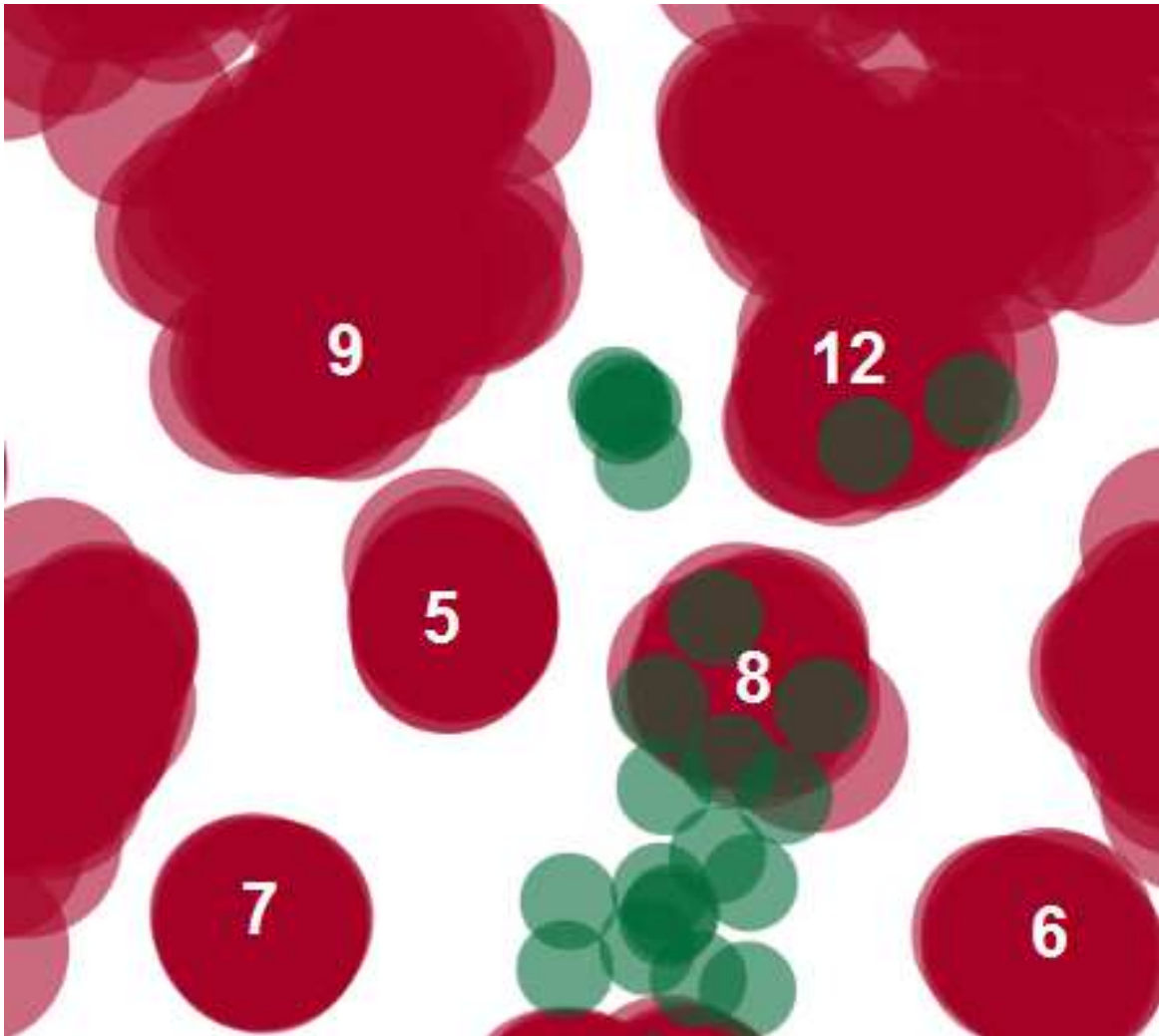


Figure 28. Real (green) and simulated (red) data visualization Zoom 3.

All the real clients are inside the region of the simulated clients, which means that the behaviour of the real clients is similar to the simulated ones. Although some real clients are not inside one simulated cluster, each client can be classified into the nearest type.

It is known from the visualization that the real clients behave as dwellings with 1, 2 or 3 occupants and there are also unoccupied dwellings. Most of the real clients belong to the clusters type 4, 8 and 12, which are the simulated clients with non unoccupied period. Hence, the majority of the real clients have non unoccupied period.

As the visualization tool does not make clusters with the clients, the clustering can be done manually. Using the visualization tool to identify each real client with one type of the simulated profile.

Table 7 shows the real clients identified as each type of the 20 simulated profiles.

Table 7. Profile of the real clients

Number of clients	Identified as type	Profile	
		Unoccupied period	Number of occupants
9	Unoccupied	Unoccupied dwelling	0
7	2	14:00 to 19:00	1
3	1	9:00 to 14:00	1
9	4	No	1
13	8	No	2
5	12	No	3

10. Conclusions

- Complex visualization techniques can be used to represent analyse and cluster hourly load curves obtained from residential customers provided by the company Electra Norte, as well as simulated curves obtained from a simulation software previously developed called GenMix.
- The previously developed software (GenMix) is validated comparing the real and simulated data. The real data appears among the simulated data in the visualization tool. Hence, both data have similar profile. The simulated load data curves are comparable to the real load data curves provided by Electra Norte.
- The pattern extraction technique by means of visualization techniques is validated using the known simulated data. The clients who have the same simulation profile appear together in one cluster. Hence, this tool is useful to identify similar clients and group them into clusters.

Bibliography

- [1] Irish Social Science Data Archive, “Data from the Commission for Energy Regulation (CER) - Smart Metering Project.” [Online]. Available: <http://www.ucd.ie/issda/data/commissionforenergyregulationcer/>. [Accessed: 14-Aug-2013].
- [2] J. V Ringwood, D. Bofelli, and F. T. Murray, “Forecasting Electricity Demand on Short , Medium and Long Time Scales Using Neural Networks,” *Journal of Intelligent and Robotic Systems*, vol. 31, pp. 129–147, 2001.
- [3] J. Lourenco and P. Santos, “Short Term Load Forecasting Using Gaussian Process Models,” *Instituto de Engenharia de Sistemas e Comput adores de Coimbra*, pp. 1–12, 2010.
- [4] L. Amaral, R. Souza, and M. Stevenson, “A smooth transition periodic autoregressive (STPAR) model for short-term load forecasting,” *International Journal of Forecasting*, vol. 24, no. 4, pp. 603–615, Oct. 2008.
- [5] Elexon, “Load Profiles And Their Use In Electricity Settlement.” [Online]. Available: http://www.elexon.co.uk/ELEXON Documents/load_profiles.pdf.
- [6] W. Li, *Risk Assessment of Power Systems: Models, Methods, and Applications*. New York: IEEE Press—Wiley, 2005.
- [7] Smart Grid, U.S. Dept. Energy [Online]. Available: <http://www.oe.energy.gov/smartgrid.htm>
- [8] Automated Load Curve Data Cleansing in Power Systems. Jiyi Chen, Wenyuan Li, Fellow, IEEE, Adriel Lau, Jiguo Cao, and Ke Wang, Senior Member, IEEE
- [9] V. Figueiredo, F. Rodrigues, Z. Vale, B. Gouveia, “An Electric Energy Consumer Characterization Framework based on Data Mining Techniques”. *IEEE Transactions on Power Systems*, Vol. 20, N.2, pp. 596-602, May 2005.
- [10] N. Anuar, Z. Zakaria, "Cluster validity analysis for electricity load profiling". *IEEE International Conference on Power and Energy (PECon)*, pp.35-38, November 29 2010-December 1, 2010.
- [11] S. Ramos, Z. Vale, J. Santana, J. F. Duarte, "Data Mining Contributions to Characterize MV Consumers and to Improve the Suppliers- Consumers Settlements" *IEEE Power Engineering Society General meeting (PES GM)*, Tampa, Florida, USA, 24-28 July, 2007.
- [12] Sérgio Ramos, Isabel Praça, Zita Vale, Tiago M. Sousa, Vera Faria GECAD – Knowledge Engineering and Decision-Support Research Center Institute of Engineering – Polytechnic of Porto (ISEP/IPP) Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, Portugal {scr, icp, zav, tmsbs, vefa}@isep.ipp.pt
- [13] J.W. Hughes, and D. W. Von Dollen, "developing of an integrated energy and communications systems architecture: the initial steps" in *Proc. 2004 IEEE Power Systems conference and Expositions*, pp. 1651-1654, 2004.
- [14] Seon Gu Kim, Geon Il Hur, and Yeoung_Jin Chae, "Smart grid and its implications for electricity market design", *Journal of Electrical Engineering and Technology*. vol. 5, no. 1, pp 1-1, march, 2010.
- [15] Young -Il Kim, Jong_Min Ko, Jae-Ju Song, and IL-Kwon Yang, "Customer classification method using customer common attribute", in *Proc. 2010 International Conference on Convergence and Hybrid Information technology*, pp-174-178, August, 2010.

- [16] Young -Il Kim, Shin-jae Kang, Jong_Min Ko, and Seung-Hwan Choi, " A study for clustering method to generate typical load profiles for smart grid", 8th International Conference on power Electronics, 2011.
- [17] Characterising Domestic Electricity Demand for Customer Load Profile Segmentation, Fintan McLoughlin, Dublin Institute of Technology.
- [18] Fayyad, U., G. Piatetsky-Shapiro, P.J. Smyth, R. Uthurasamy, "From Data Mining to Knowledge Discovery: An Overview". In Advances in Knowledge Discovery and Data Mining, pages 1-34. AAAI/MIT Press, 1996.
- [19] . Chicco, G., Napoli, R. and Piglione, F, Application of Clustering Algorithms and Self Organising Maps to Classify Electricity Customers, Proc.IEEE Bologna Power Tech 2003, Bologna, Italy, June 23- 26, 2003.
- [20] Gerbec, D., Gasperic, S., Smon, I., Gubina, F., An approach to customers daily load profile determination, Power Engineering Society Summer Meeting, 2002 IEEE, Volume 1, 2002, pp. 587 – 591.
- [21] Yu, I. H., Lee, J. K., Ko, J. M., and Kim, S. I., A Method for Classification of Electricity Demands using Load Profile Data, Computer and Information Science, 2005. Fourth Annual ACIS International Conference on 2005, pp. 164 – 168.
- [22] Chicco, G., Napoli, R., Postolache, P., Scutariu, M., and Toader, C., Customer Characterization Options for Improving the Tariff Order, IEEE Trans. Power System, vol. 18, February 2003, pp.381 – 387.
- [23] 9. Ozveren, C. S, Birch, A, P, & Vechakanjana, C, Fuzzy Classification of Electrical Load Demand Profiles – A Case Study, Power Management and Control, 17-19 April 2002, Conference Publication no. 488, IEEE 2002.
- [24] Iswan Prahastono, Dr. D. King and Dr. C.S. Özveren, " A Review of Electricity Load Profile Classification Methods"
- [25] GenMIX USER GUIDE Computer Tool for assessing the power generation-mix in residential Nearly Zero-Energy Buildings, Félix Manuel Lorenzo Bernardo, Edwin Xavier Domínguez Gavilanes, Pablo Arboleya Arboleya
- [26] [Instituto para la Diversificación y Ahorro de la Energía. "www.idae.es". Emissions in Europe (REMODECE)". In: ISR-University of Coimbra (2008)
- [27] Red Eléctrica de España. "www.ree.es".
- [28] Departamento de Planificación y Estudios. "Informe Final Proyecto SECH SPAHOUSEC". In: IDAE (2011)
- [29] J. Ladero. "Sistematización del diseño de un sistema de poligeneración para el abastecimiento energético de un conjunto de viviendas". In: Universidad Pontificia de Comillas, Proyecto de Fin de Carrera (2011)
- [30] Interactive Dimensionality Reduction for Visual Analytics Ignacio Díaz, Abel A. Cuadrado, Daniel Pérez, Francisco J. García and Michel Verleysen, ESANN 2014 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning. Bruges (Belgium), 23-25 April 2014