TÉCNICO LISBOA

Identification of residential electricity consumption profiles for building dynamic simulation through smart-meter data

Catarina Pargana Oliveira Pires

Thesis to obtain the Master of Science Degree in

Mechanical Engineering

Supervisors: Prof. Carlos Augusto Santos Silva

Arqª Cláudia Azevedo de Sousa Monteiro

Examination Committee:

Chairperson: Prof. Edgar Caetano Fernandes

Supervisor: Prof. Carlos Augusto Santos Silva

Member of the Committee: Prof. Susana Margarida da Silva Vieira

November 2017

Abstract

There is a great opportunity to reduce energy consumption in Europe by addressing energy efficient measures applied to residential buildings since they are responsible for 27% of Europe's final energy consumption and are known to be highly ineffective. To evaluate these measures, building energy simulation (BES) models have been widely used. However, these models are very complex and require detailed input data on building characteristics and operation, which is usually hard to collect. The recent investment in smart meters encourages innovative studies addressing residential and non-residential electrical consumption by making high-resolution data available.

A mathematical model is proposed to disaggregate total electrical consumption data into possible end-use profiles for building energy simulation. Data from several smart meters is collected and used to perform cluster analysis via *k-means* algorithm, to determine residential users' daily consumption profile in three reference months (free-float, winter and summer). Building energy simulations are performed to test, if the resulting input profiles for occupancy, lighting, equipment, heating and cooling use can describe the real consumption of a given set of users. Different comfort temperature, heating and cooling systems types are considered in a model calibration step.

The model proved to be effective, as the difference between the total daily consumption from measured and simulated data is small: for free-float (0.22%), for winter (0.90%) and for summer (4.00%). Concluding that even if the measured and the simulated profile consumption do not present a perfect hourly match, the total consumption does present a close approximation.

Keywords

Clustering analysis, Residential user's characterization, End-uses profiles, Building energy simulation

Resumo

Há uma grande oportunidade para reduzir o consumo de energia na Europa, analisando medidas de eficiência energética aplicados a edifícios residenciais, já que são responsáveis por 27% do consumo final de energia da Europa e são altamente ineficazes. É necessário realizar simulações de energia em edifícios para avaliar tais medidas, mas esses modelos são muito complexos e requerem dados detalhados sobre características de construção e operação, geralmente difícies de obter. O investimento em "*smart meters"* disponibiliza dados de alta resolução promovendo estudos inovadores sobre o consumo elétrico residencial e não residencial.

Um modelo matemático é proposto para desagregar dados de consumo elétrico em possíveis perfis de uso-final para simulação de energia em edificios. Dados de "*smart meters"* são utilizados para realizar uma análise de "*clusters"* para determinar perfis de consumo diário de utilizadores residenciais, nos 3 meses de referência *("free-float",* inverno e verão). São realizadas simulações de energia para testar se é possível descrever o perfil de consumo real de um conjunto de utilizadores através desses perfis de *"input"* (ocupação, iluminação, equipamentos, aquecimento e arrefecimento).

O modelo provou ser eficaz, ja que a diferença entre o consumo total diário medido e simulado no pior caso é igual a: para o *"free-float"* (0.22%), para o inverno (0.9%) e para o verão (4.0%). É possível concluir que, mesmo que o consumo de perfil medido e simulado não tenha uma combinação horária perfeita, é possivel concluir que se obtem uma boa aproximação usando o modelo matematico proposto.

Acknowledgments

An expressed gratitude to Claudia Monteiro for 24-hour availability and dedication and to Professor Carlos Silva for the knowledge and insight provided.

Thanks a lot, my family, for the willingness to help and for inspiring me to follow this field of studies. Thank you, my friends and team members, for all the support and understanding.

Sören, thank you for your patience and company.

“Putting energy efficiency first reflects the fact that the cheapest and cleanest source of energy is the energy that does not need to be produced or used.”

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, the Committee of the Regions and the European Investment Bank, 2016, Clean Energy For All Europeans [1]

Contents

Acronyms

BES – Building Energy Simulation

CDD – Cooling Degree Days

CH – Calinski Harabasz Index or Minimum Variance Criterion

CV(RMSE) – Coefficient of Variation of the Root Mean Square Error

DBI – Davies-Bouldin index

DF – Daylight factor

EEI – Equipment Energy Efficiency Index

HDD – Heating Degree Days

IAI – Intra Cluster Index

IPO – Input Process Output

J – Mean Square Error or Error Function

NMBE – Normalized Mean Bias Error

SCOP – Seasonal Coefficient of Performance

SEER – Seasonal Energy Efficiency Ratio

List of Figures

List of Tables

1 Introduction

1.1 Motivation

Several efforts have been made over the past decades to fight climate change and its effects. More recently, at the Paris climate conference (also known as SCOP21), 195 countries signed an historical agreement to address global climate and prevent the increase of events associated with climate change by restricting global warming below 2°C [2].

To accomplish the international climate actions before 2020 and the objectives of the Paris agreement up to 2030, the European Union (EU) efforts should consider both domestic action and international cooperation. This objective will not be achieved without strong initiatives towards a more efficient use of energy from less polluting fossil fuels at all sectors, as well as more well-adjusted transport options, more environmentally responsible land-use and agriculture, more sustainable cities, more climate robust communities and fewer emissions from all sectors of our economy [3].

EU has been recognized as being in the forefront of this challenge. In 2007, the European Council set energy and climate change targets for 2020, namely the reduction of greenhouse gas emissions by 20%, the increase of the renewable energy share to 20% and the improvement of energy efficiency of 20% [4]. In 2014, the EU-28 met their 2020 target for final energy consumption (Figure 1).

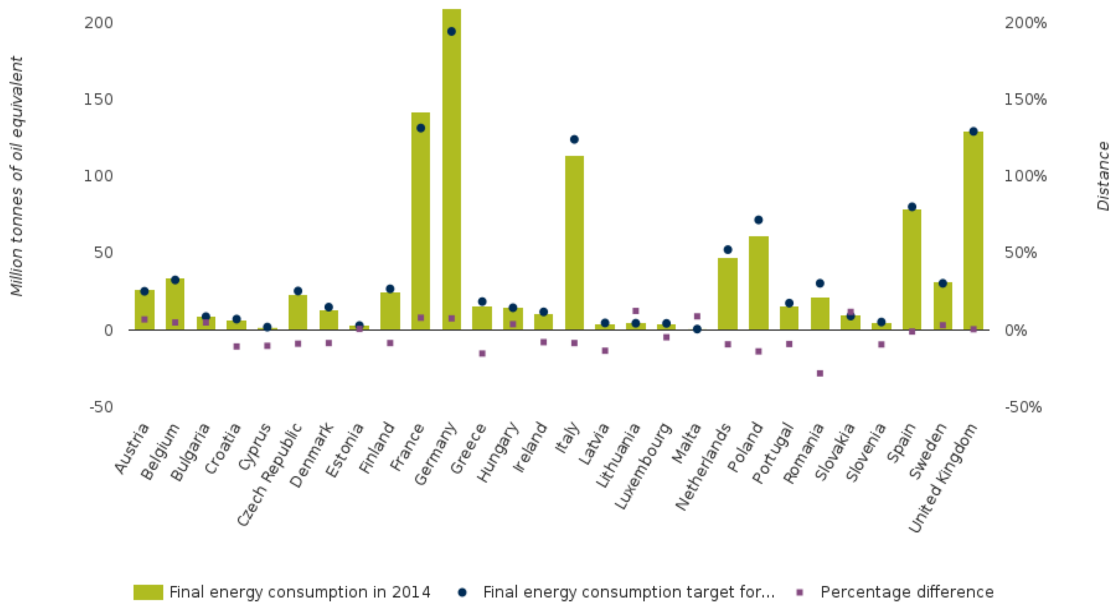


Figure 1 - Member states' progress towards energy efficiency targets for 2020, 2014 [3].

Between 2005 and 2014, the final energy consumption decreased in all sectors (11%), particularly in the industry and household's sectors (16.5% and 14.8%, respectively), influenced by the economic performance, the structural changes in various end-use sectors, the improvements in the end-use efficiency and heat consumption reduction due to the recent years favorable climatic conditions.

In October 2014, new energy efficiency targets were set up to 2030, now including at least 40% cuts in the greenhouse gas emissions (from 1990 levels), a minimum of 27% share for renewable energy and at least 27% of improvement in energy efficiency [5].

The building sector is the largest single consumer of energy in Europe. It accounted for 40% of the final energy consumption in 2012 and 38% of the EU's CO_2 emissions, with 27% of the final energy consumption being represented by residential buildings [6]. The final energy consumption of the building stock, residential and nonresidential, increased by 14% between 1990 and 2012 and the electricity use increased by 60% between 1990 and 2012, largely as a result of the high penetration of consumer appliances and electronic devices (Figure 2) [6].

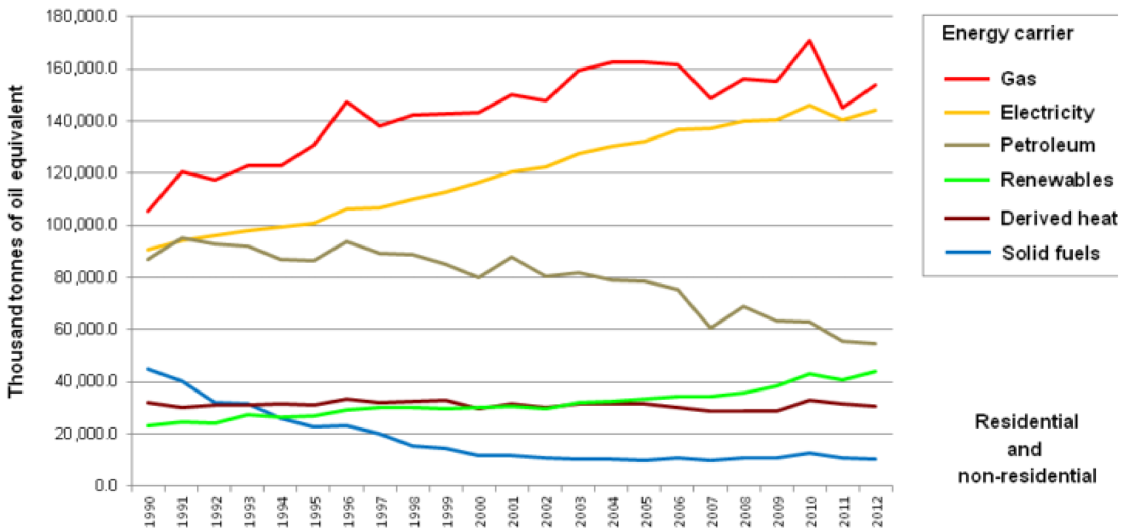


Figure 2 – Buildings' final energy consumption by energy carrier [6]

In general, the energy efficiency in the residential sector can be improved by using more efficient energy equipment, upgrading the energy characteristics of buildings or by inducing changes in the consumers behavior. However, at today's rate of renovation of 1% per year, it would take a century to upgrade the building stock to modern, near-zero energy levels [1]. A great opportunity lies on addressing the building owners awareness of their energy use and potential savings related to different energy efficiency measures paired with the financial and other incentives related to building renovation [7].

To be able to measure the impact of such actions one has to be able to simulate energy savings and therefore accurate dynamic models of buildings energy consumption need to be developed. Buildings are complex systems in which energy consumption is influenced by a combination of factors, including the age and location of the building, the household size and the occupants lifestyle, and the penetration of appliances and electronic devices [6]. Due to the diversity and disaggregation of the buildings sector, it remains challenging to acquire good data on building's characteristics and its energy use, with negative consequences on the perception of the cost-effective energy savings potential[7].

Besides building's related information, data about the occupant's behavior is necessary to be able to perform accurate models of energy consumption, as well as information about the existent equipment in each household, including the type, function, dimension, quantity and efficiency.

Occupants' behavior in residential buildings can be described in three main categories: the occupants' presence, the occupants' operation of systems- lighting, equipment, heating and cooling systems - and their interaction with windows and blinds [8]. However, to obtain this type of information it is necessary to perform occupancy surveys, occupant's monitoring (through sensors or observation) or field measurements and questionnaires, which can be time consuming and intrusive.

With this goal in mind, EU has adopted a number of initiatives aiming to improve energy consumption awareness, including the replacement of at least 80% of electricity meters with smart meters by 2020 [9]. These smart meters will allow for a temporal assessment of the electricity use, which holds the potential to reveal insights about the electricity consumption and the behavioral and technological drivers of that consumption [10]. In this way, the accuracy of the buildings energy models can be increased, enabling the simulation of impactful measures for the improvement of the energy efficiency in the residential sector.

1.2 Objective

Smart meters and smart grid deployments are making the information about energy use more available, which allows the temporal assessment of the electricity use, with the potential to reveal information about the end-use and the factors that affect that consumption. Moreover, it can also improve the operation, planning and control of distribution networks [11] and give more information to the consumer, allowing him to have a knowledgeable behavior towards its electrical consumption.

There has been an increasing use of smart meters data in recent studies, namely to identify various types of consumers for short-term and midterm load forecasting, time of Use (ToU) tariff design and Demand Side Management (DSM) strategies [12]–[18]. Other studies focus only on the residential load characterization [10], [11], [19]–[26], or on inferring about the drivers behind the residential consumption, in terms of socio-economic status, appliances stocks and dwellings characteristics [27]–[29]. Finally, electric consumption disaggregation, appliances, lighting and plug load profiles distinction [30]–[35], as well as occupancy inference and inhabitants routines are other uses of smart meter data [36]–[41].

The objective of this research is to analyze the behavior of different residential electricity consumers to identify input operation parameters to improve the dynamic energy simulation of the building. The clustering k-means method is applied to identify different occupant's profiles. By comparing different seasons, it is then possible to infer about the inhabitant consumptions habits on heating and cooling energy. From the paths describing the inhabitant behavior throughout the different seasons, a parametric model of the household energy use is created, where occupancy profiles are inferred, with disaggregated lighting, equipment, heating and cooling consumption. From that information, a dynamic energy simulation is taken and the model accuracy is evaluated against real energy data.

The main goal of this thesis is to develop a method to give back information about the different end-uses of the hourly electrical consumption (lighting, equipment, cooling and heating) and occupancy profiles, using aggregated smart meter data. The research questions addressed in this work are as follow:

- How can we develop a model to simulate the energy in buildings, as accurate as possible?
- How can we use smart meter data to create this model?
- What information can we retrieve from it?

1.3 Methodology

The methodology proposed in this work consists of six stages: 1) Data cleaning and preprocessing; 2) Typification of energy consumption; 3) Residential consumption characterization; 4) Input profiles determination; 5) Parametric building energy model creation and 5) Model evaluation (Figure 3). Each stage is described in detail hereafter.

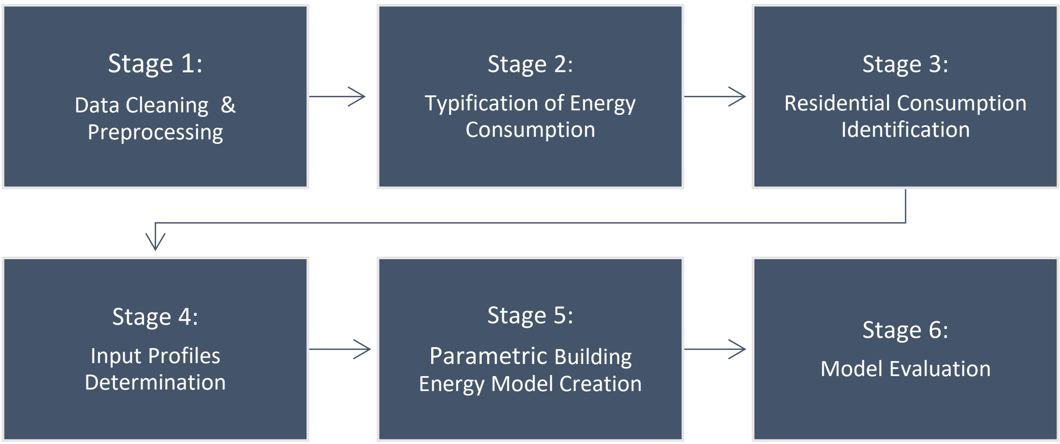


Figure 3 – Methodology framework

Stage 1 refers to the identification of *missing-value users*, *zero-use residential users* and *outliers*, where a method to fill in the empty data is proposed. This stage is performed to ensure a reliable performance of the next stages.

Stage 2 proposes a method to distinguish residential from nonresidential consumption profiles, through clustering analysis. Representative profiles per user are determined, where different data normalization techniques are discussed. Adequacy measures are applied to evaluate the segmentation of the dataset.

In Stage 3, residential users are selected and characterized in 3 reference months: a free-float month, a winter month and a summer month, to identify the changes in the consumption profiles throughout the seasons. Another clustering analysis is performed and the clusters centroids obtained per month are used to create the consumption paths, throughout the year that represents the behavior of a certain number of users throughout the different seasons.

In Stage 4 a mathematical model is developed, using as input the paths consumption profiles, as well as the time of use survey and illuminance values to determine possible *occupancy*, *lighting*, *baseline*, *activity*, *heating* and *cooling profiles*.

In Stage 5, a parametric model of a reference dwelling in the building is created. The disaggregated consumption profiles by end-use obtained from Stage 4 are used as inputs to perform the building energy simulation.

Finally, in Stage 6, a model evaluation is taken to evaluate the reliability of the mathematical model proposed in Stage 4. Model calibration through different heating and cooling systems efficiencies as well as temperature setpoints are tested.

1.4 Structure

This dissertation is organized into six chapters.

In the first chapter, the context, motivation and objectives of this thesis are presented. It starts with the collective need to assure the sustainable use of energy and the commitments that were generated for energy efficiency. The Europe's current situation towards those goals is presented and the link to the need to perform more accurate energy simulations in buildings is introduced. Then the methodology proposed in this thesis is briefly developed. Finally, the thesis main goal, research questions and structure are disclosed.

In the second chapter, a brief review on the relevant literature is introduced. The topics considered on the method developed in this thesis are investigated, namely:

- the smart meter data applicability, where the possible uses of smart meter data referenced in literature are explored;
- the typification of energy consumption, where methods for electrical consumption characterization are presented;
- the clustering analysis, where different cluster techniques and algorithms and its application will be reviewed.

The third chapter presents the Case Study, where the building characteristics and location as well as the available data are described.

The forth chapter describes the methodology developed under this thesis and the applicability to the case study. The challenges and benefits of the developed method are established.

In the fifth chapter, the validity of the results is discussed and analyzed.

In the sixth chapter, the summary of the findings of this thesis and its implications are presented, reviewing the merits of the methodology and highlighting its fragilities. Finally, some thoughts on the potential use of this work and its relevance towards the global sustainability goal are shared.

2 Literature review

2.1 Smart meter data applicability

Electricity consumption data can provide useful information about the consumers and their habits. Also, it can be used to characterize the activities and equipment behind electrical consumption, which can be crucial in building simulations as well as providing energy-related feedback to households. In addition, by linking energy intensive activities to their time profile, it is possible to identify potential for load shifting, demand side management and identification of suitable tariffs.

According to [28] and [29], smart meter data combined with ground-truth information about the household can be used to estimate which characteristics could be well inferred from the consumption data, like socio-economic status, dwelling properties and appliance stock, with the end objective of using this information in personalized and scalable energy efficiency programs for private households. For example, in [38] the focus of the research was the occupants behavior and electricity consumption patterns studied for 27 dwellings in the Northern Ireland. The conclusion was that the type of dwelling, location, ownership and size, household appliances, number of occupants, income age and occupancy patterns have different, but significant, impacts on electricity consumption. Furthermore, the authors were able to link certain activities with different time periods and different values in consumption: the minimum load typically occurring between 2h and 4h corresponding to the sleeping activity, a minor peak occurring between 6h and 9h when householders were preparing breakfast and performing other morning rituals and a major peak occurring between 5h and 22h, which corresponded to dinner preparation and the maximum usage of lighting.

In another study, a Non-Intrusive Occupancy Monitoring (NIOM) is explored using data from smart meter and ground-truth information about the actual occupancy of two households [39]. It was observed that home´s patterns of electricity usage change when occupants are present due to interaction with electrical loads. The use of electricity meters as occupancy sensors was also evaluated in [36], where five households supplied fine-grained electricity consumption data and ground-truth occupancy information throughout eight months. The authors concluded that power consumption is likely to have a higher mean and standard deviation when the household is occupied, having lower consumption values once occupants might be at home but not using any electrical devices. Using the same dataset, the same authors built on their previous work presenting a detailed analysis of adding an on/off feature representing the activation state of appliances. Through this new feature an accuracy of 94% in occupancy detection from smart meter data was achieved [41].

In [40], the impact of human activity in energy consumption was studied using data from 241 smart meters. Several methods were introduced to identify which days had human activity, to extract daily routines from it, and to identify routine's startup and shutdown. One of the conclusions was that the human activity only accounted for two thirds of the energy consumed daily and the consumption that was not related to the human activity could get as high as 50%.

A methodology to characterize the energy consumption of domestic life was presented in [33] by linking appliance end-use and activities through an ontology built from qualitative data about the households and Non-Intrusive Appliance Load Monitoring (NILM) data. In addition, social science approaches to determine time use, such as interviews and ethnography data, were also used. The total electricity load was disaggregated up to the appliance level and it was combined with related activities. Finally, a sample of ten households was used to show which activities can be inferred from the available data. In [31] one sensor was installed on the main electric supply of the house and one sub-meter targeting the appliances during a week, to find the electric characteristics of the used appliances, allowing to disaggregate the total electricity consumption into major end-uses.

In [32], a method was proposed to distinguish non-electrically heated and cooled homes from broad datasets of whole-house profiles to obtain appliance, lighting and plug load profiles (ALP). This method was based on the comparison of smart meters with a 15-min time-step for 160 homes throughout three years and sub-metered 1-min time-step data, for 23 residential homes. From the sub-metered dataset, Domestic Hot Water (DHW) heating, furnace auxiliary and air conditioner for space cooling were subtracted from the whole house consumption, creating ALP equivalent profiles. By comparing these equivalent profiles with the smart-meter dataset it was possible to establish if the households were using electricity for space cooling and heating or not. From that, a method to identify homes with electric space heating and cooling was proposed by comparing electrical average monthly consumption in months that typically do not require neither cooling or heating, with winter and summer months.

Hourly electricity meter data from 470 household in Norway was merged with weather data and responses from a household survey to examine the impacts of outdoor temperature and different household characteristics on hourly electricity consumption [30]. Two models were compared and it was concluded that daily mean outdoor temperature, represented by heating degree days, is the best suited input variable for modeling aggregate hourly electrical consumption. Further, regression models were used to break the consumption down into space heating and other electrical appliances.

In [34], [35], [37] the authors proposed three models that firstly intended to create occupancy profiles, then lighting profiles and afterwards a combination into a model to infer domestic electricity use. Firstly in [37], a model based upon surveyed time-use data from UK households was presented, describing what people do and in which period and when people are at home and awake. This approach generated statistical occupancy profiles accounting for differences between weekdays and weekends. Then in [35], a model that combines the level of natural light coming in from outdoors, measured by irradiance, and the activity of the households residents, represented by the active occupancy data from the previous work, allowed to construct the lighting electricity demand profiles for individual dwellings. Finally in [34], a high-resolution model of domestic electricity use was presented, based upon a combination of patterns of active occupancy and all the major appliances commonly found in the domestic environment. This model was validated by comparing the synthetic dataset with real electricity consumption data from 22 dwellings, recorded over a year.

In summary, smart meter data can be used for several purposes such as infer which factors have significant impacts on electricity consumption like socio-economic status, dwelling properties and appliance stock, identify the impact of the occupant's behaviour, identify end-use consumption, distinguish non-electrically heated and cooled homes from broad home datasets and study the impact of the weather on the final consumption.

2.2 Typification of energy consumption

Distinct types of users (domestic, commercial, industrial, agricultural, etc.) have different electricity consumption profiles. Even for the same type of users, their electricity consumption patterns may be very different. It is not possible to study individual consumption patterns due to the considerable number of consumers of the electrical system. Therefore, to characterize the behavior of all consumers, it is necessary to represent the entire population by a smaller sample of consumption patterns [12], usually through the Load Classification method.

Load Classification, also known as Load Profiling, is the method used to divide various load consumption curves, or load consumption patterns, into well-defined and separated groups. The load patterns belonging to the same group have more similarities to each other than with the load patterns of the other groups [13]. Characteristic load curves, or Typical Load Curves, are then created, representing the typical load behavior within the same group.

Several studies can be found in literature on this topic. In [14] the authors present an electricity medium voltage (MV) customer characterization framework, focusing on identifying typical load profiles of consumers and developing a rule set for the automatic classification of new consumers. Also, in [17], a similar framework is proposed and applied to a real database of low voltage (LV) consumers of the Portuguese distribution company. In [12], to implement Demand Site Management (DSM), the authors used an analysis of the smart grid data to create load curves patterns for policies decision in Brazil generated by four types of consumers. In [16], a pattern recognition methodology for the classification of the daily chronological load curves of medium voltage electricity customer of the Greek power system is executed, in order to estimate typical days and the respective representative daily load profiles. Also in Greece, in [18] consumption patterns of high voltage (HV) consumers are explored.

As it was previously mentioned, within the same type of users the consumption profiles can differ, and even the same user can behave differently from day to day or month to month, reflecting different behaviors and needs, generated by the variability in weather, work schedule, etc. A process for data-mining the daily electricity consumption profiles of low voltage households in China is proposed in [19]. In the data cleaning and preprocessing phase, the nonresidential profiles were identified and removed from the dataset. In [25], the same author analyzed and clustered 938 household daily electricity consumption profiles, also in China. The final group characteristics were extracted and different electricity consumption patterns were identified. However, while most of the research focus on clustering daily load profiles, in [11] the authors proposed the use of segmented load profiles to estimate missing or future smart meter measurements for the improvement of the Time-of-Use tariff design. Further, the segmentation of 759 Spanish clients throughout 2008 and 2009 was done in [26], with 2 objectives: firstly, using residential dynamic load profiles to evaluate the impact of the changes in normative and laws applied in those years and secondly, to classify groups of users, identify anomalies, evaluate trends and select target groups for DSM action.

The load classification method can be divided into three stages presented in [10], [21], [22], [27]: the determination of the consumer segmentation based on the metering data; the correlation of survey data to the segments identified previously; and the development of statistical and machine learning classification models to predict the users' demand profile. For example, in [27] these 3 stages are applied and 4232 Irish households were segmented, in [22] also by the same authors, using the same data set and the same process but this time extended to create a model to predict the group of a new customer is presented, being characterized by a representative consumption profile and survey answers. Moreover, in [21] the same Irish households were segmented and the process was repeated for each day over a six-month period in order to characterize the diurnal, intra-daily and seasonal variations of domestic electricity demand. In Austin, a dataset with 103 homes, with a temporal resolution of 1 minute, was used in [10] to determine the shape of the seasonal demand residential profiles while combining the information of the survey´s response. The goal was to infer about the possibility of establishing a correlation between the two. A detailed review of the process behind the load classification can be found in [13].

In summary, the stages that describe the typification of energy consumption procedure can be defined as: 1) Data Cleaning and Preprocessing, 2) Creation of representative load curves, 3) Clustering analysis and 4) Post clustering analysis [12], [13], [15], [16], [18], [19], [22], [25], [27].

2.3 Clustering analysis

Load classification is usually done using clustering techniques. Clustering analysis aims at discovering the natural grouping of a set of patterns [42], by distinguishing groups based on their similarities. This means that within a certain group´s data, objects are as similar as possible and dissimilar compared to the other groups [19]. Data clustering analysis has been used to accomplish three main purposes: to understand the underlying structure and gain insight into data; to identify its natural classification by determining the degree of similarity among forms or organisms; and to organize the data and summarize it through cluster prototypes [42].

Two clustering analysis methods are often used in load classification studies: the Hierarchical and the Partitional Clustering Method. The Hierarchical method groups a given dataset of load profiles into the required number of clusters through a series of nested partitions, being the final clusters results hierarchical. On the other hand, the Partitional method aims to group load profiles into several clusters by optimizing a certain objective function but requires the number of clusters to be known in advance. This method is called the center-based method because each cluster is represented by its centroid and this centroid is representative of all the loads contained within that cluster [11]. Other methods for clustering analysis include the density-based methods, grid based methods and model-based methods. Further, the clustering methods can also be characterized as either “crisp” or “fuzzy”. While in the “crisp” clustering each data object can only belong to one group, on “fuzzy” clustering each data object can belong to one or more groups with a certain degree of membership [19]. The typical algorithms used are the k-means algorithm [10]–[12], [14], [17], [21], [22], [27] and the Fuzzy C-means [13], [19], [25].

An overview of the clustering algorithms for load clasification and its performance assement can be found in [15], [16], [18]. In [15], the author identified six main algorithms: the hierarchical agglomerative clustering, k-means, fuzzy k-means, self-organized maps (SOM), probabilistic neural networks (PNN) and follow the leader (FDL) algorithm. He concluded that if the objective is to get the representative classes, the k-means algorithm is appropriate, while if the objective is to find outliers a hierarchical algorithm is best suited. In [14], several algorithms were tested for electricity users characterization and the k-means algorithm was chosen since it performed the best. In [21], three algorithms were also tested for load classification (SOM, k-means and k-method) and the authors found that the SOM algorithm had the best performance. However, k-means also performed good, with slightly lower values on the index used to evaluate the algorithm performance.

Besides the choice of the algorithm, when performing any typical clustering scenario, two questions need also to be addressed: firstly, what is the number of clusters that best divides the dataset and how good is the performance of the clustering itself [43] and secondly, if comparing different methods, which one is the suitable method. To answer those questions, validity indexes are used. Typically, these indexes measure the compactness and the mutual separation between each cluster. Several validity indexes exists and can be found in literature such as: the Mean Index Adequacy (MIA) [17], the Mean square error (J), the Clustering dispersion indicator (CDI), the Similarity matrix indicator (SMI), the Davies-Bouldin indicator (DBI) and Ratio of within cluster sum of squares to between cluster variation (WCBCR) [16]. The use of the Intra-cluster index (IAI), the Modified Dunn Index (MDI), the Scatter Index (SI), the Variance ratio criterion (VRC) was reported by [15]. Moreover, in [27] and [22] using k-means algorithm and the appropriate number of representative groups was inferred using the Dunn Index, the Davies Bouldin Index and the silhouette index. For fuzzy c-means the fuzzy cluster index (PBMF) was used by [19],[25].

In [18], several algorithms and validity indexes were tested, in order to evaluate the best suited validity indexes per algorithm to assess their performance. The authors stated that for k-means clustering method the most efficient adequacy measures are: the *J*-Mean Square Error or Error Function, the *IAI*- Intra Cluster Index and the *CH* – Calinski Harabasz Index or Minimum Variance Criterion.

3 Case study

3.1 Neighborhood characteristics

The building under analysis in this thesis is located in Lisbon, more precisely in Parque das Nações neighborhood (Figure 4).

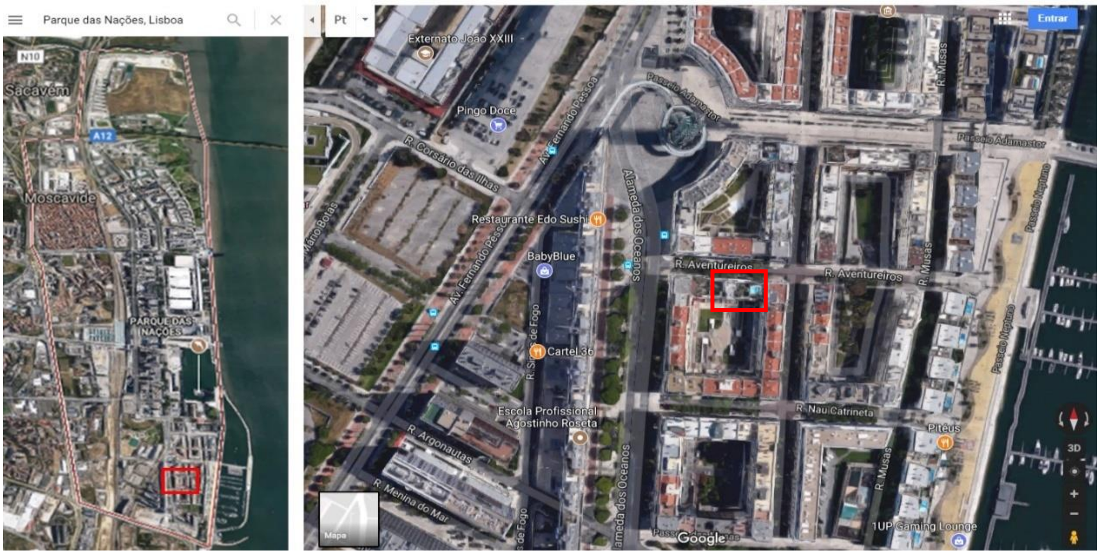


Figure 4 – Parque das Nações case study

The most common age range in the casa-study area is between 35 to 44 years, representing 25%, followed by 45-54 years with 18.5%, 24-34 years with 17.8% followed by >65 with 16.5% and the remaining is distributed by 55-64 with 14.3% and by 18-24 with 7.9%.[44]

Considering the inhabitants employment situation, most of the sample is working full time (57%), followed by 22.8% of retired people. Only 6.6% are unemployed and 6.1% are studying [44]. When it comes to the number of persons per household, 2 members is the most common value with 30.6%, followed by 3 with 24.7%, 4 with 20.6% and 1 with 15.2%, having the remaining distributed into 5 or more members . It is likely to represent a couple without kids as the most common value followed by couple with one to two kids [44].

3.2 Building characteristics

The building under analysis is located in a recent area of intervention, where around 69% of the buildings were constructed after 1991. The typical buildings in this area have 5 and more floors (68%), with 3 or more dwellings per building (74%) and typologies with 5 or more rooms (50%), and 3 to 4 (47%), with average dimensions between 101 and 200m^2 (51%) and 50 to 100m^2 (35%) [45].

The building used as reference for this model has 4 floors (Figure 5) with a floor-to-floor height of 3 meters (more drawings detailing the building sections and elevations can be found in Annex A). Each floor has two dwellings with 3 bedrooms and a living-room (commonly called a 'T3' typology) with an area of 150m^2, and one dwelling with 1 bedroom and a living-room (commonly called a 'T1') with an area of 55m^2. The buildings in this area all share similar characteristics regarding their geometry and construction materials, being the dwelling chosen for simulation representative of the existing dwellings in this neighborhood. The dwelling chosen as reference for this case study is highlighted in red in Figure 5.

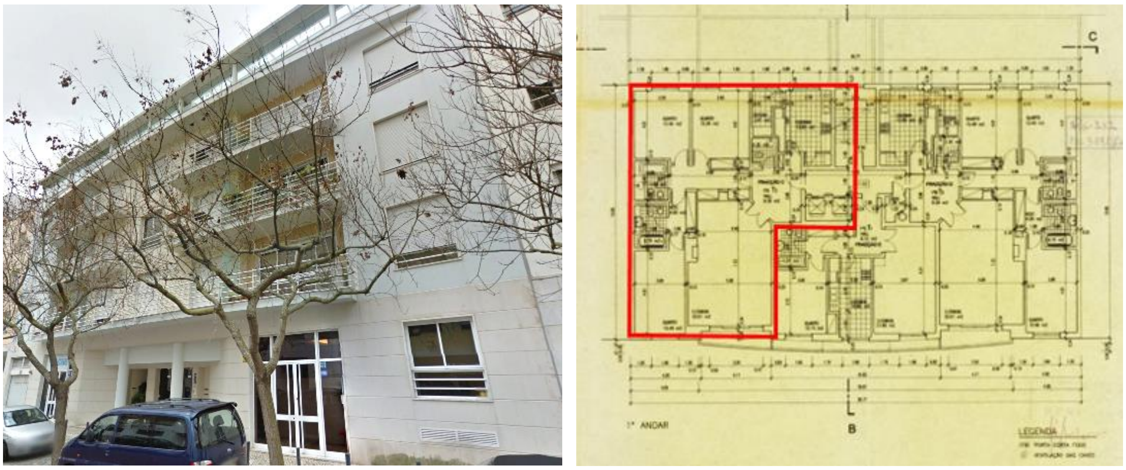


Figure 5- Building and floorplan used as case study

The constructive information used as input parameters in the energy model was retrieved from the energy performance certificates [46] and are presented in Table 1 and Table 2:

Table 1– Walls constructive parameters

Walls	Location	U_{total} (W/(m^2.ºC)
Double wall with thermal insulation in air space	Exterior	0,54
Single or double wall plastered (post 1960)	Interior	1,16

Table 2– Windows constructive parameters

Windows	U(W/(m^2.ºC)	Solar heat gain coefficient
Double glazing with metallic frame with no thermal break	2.9	0.7

For the exterior walls, a total thermal transmittance value of 0.54 W/(m^2.ºC) was considered, corresponding to a double wall composed by outer coating, masonry layer, thermal insulation material completely filling the air space, masonry layer and inter coating, from [47].The materials used in the simulation engine are presented in Table 3.

Table 3 - Exterior Wall Individual Materials

Material	Thickness(m)	Thermal Resistance ((m^2.ºC)/W)	Thermal Conductivity (w/(m.ºC))
Gypsum Plaster	0.019	0.083	0.23
Concrete Blocks	0.11+0.15	0.78	0.33
Extruded Polystyrene (XPS)	0.03	0.81	0.037

For the interior walls, a total thermal transmittance of 1.16 W/(m^2.ºC) was considered, corresponding to a single wall composed by outer coating, masonry layer and inter coating, from [47], . The materials used in the simulation engine are presented in Table 4.

Table 4 - Interior Wall Individual Materials

Material	Thickness(m)	Thermal Resistance ((m^2. ºC)/W)	Thermal Conductivity (w/(m.ºC)
Gypsum Plaster	0.019	0.083	0.23
Concrete Blocks	0.15	0.45	0.33

The renovation of the air through natural ventilation in residential buildings is often carried out by the opening of windows and by uncontrolled air infiltrations through the latches in the enclosure, in particular the shutters, windows and doors [48]. In this case study, it was considered the average of the 18 energy performance certificates from the apartment block where this building is inserted, where no mechanical or hybrid ventilation was reported, only natural ventilation which includes an infiltration value of 0.62 air changes per hour.

From the same energy performance certificates, it is also possible to redraw information about the most common energy systems available in the households (Figure 6) When it comes to gas systems, 64,7% reported the existence of a gas heater for water heating and space heating, with an average efficiency of 0.9, and 11.8% reported a gas heater only for water heating with an efficiency of 0.86. When it comes to electric systems, 11.8% are for water heating and 11.8% for space heating and cooling. The equipment used for water heating was an electric heater with an efficiency of 0.81 and for space heating and cooling was a multi-split with an average SCOP of 2.6 and SEER of 2.5.

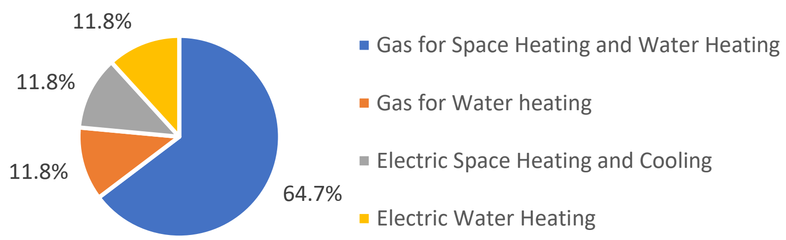


Figure 6 – Energy Systems present in the case study household

3.3 Available data

For this case study, electricity data from smart meters installed in this neighborhood was provided by the local energy supplier. Before any preprocessing and treatment, the dataset was formed by 267 users with a 15-min time-step of electrical consumption values in kWh, from March 2016 to February 2017, inclusive. The floorplans, sections and elevations of the existing buildings in this area were provided by the Municipal Archives [49]. The energy performance certificates were provided by ADENE [46] and included information about the building's envelope, ventilation and energy systems.

4 Model development

Following the methodology presented in Section 1.3, this chapter describes in detail the application of each stage to the case study.

4.1 Initial dataset

For each user of the dataset, the electrical consumption values are combined into total consumption per year. A descriptive statistics analysis is taken (supported by *Minitab* [50]) and the summary results are presented in Figure 7. The graphical summary includes three graphs: the histogram of data, a boxplot, and the 95% confidence intervals for the mean and for the median. The graphical summary also displays a table with the *Anderson-Darling Normality Test statistics*, a test used to verify if the data has a normal distribution and the descriptive statistics and confidence intervals for the mean, standard deviation and the median. From the *Anderson-Darling Normality Test* (for a significance level of 0.05) it is possible to verify that the dataset does not follow a normal distribution since the *p-value* is lower than 0.005. The mean of the yearly consumption corresponds to 2280.5 kWh/year (with a 95% confidence interval of 1999.7 kWh/year and 2561.3 kWh/year) and the median to 1689.1 kWh/year (with a 95% confidence interval of 1523.9 kWh/year and 1963.4 kWh/year). The dataset has a positive skewness due to the presence of high frequency of zero-use residential users and missing-value users, which is visible in the histogram. From the boxplot, it is also possible to verify that there is a severe presence of outliers in the initial dataset.

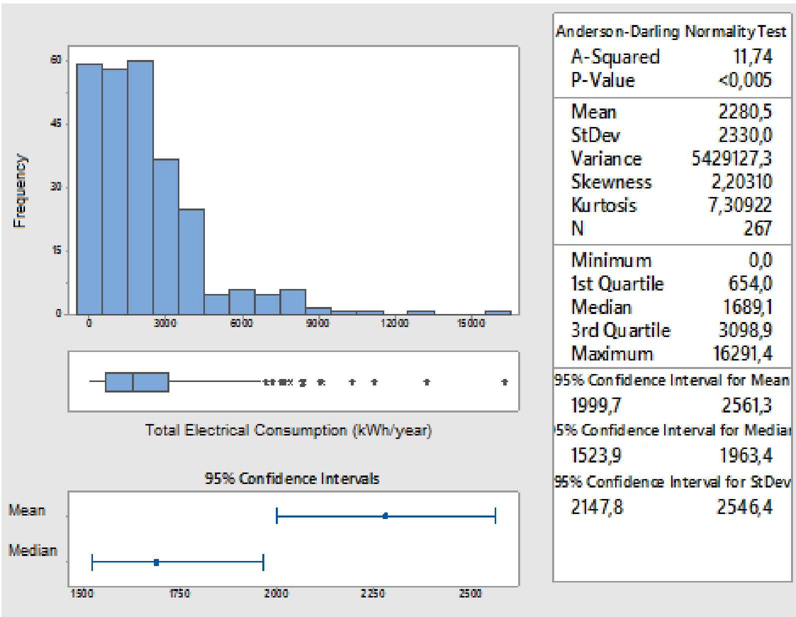


Figure 7 – Graphs and descriptive statistics of the case study initial dataset

4.2 Data cleaning and preprocessing

As presented previously, databases are likely to have discrepancies in data, lack of values or errors in the data and outliers. Therefore, a data cleaning and preprocessing procedure is crucial to ensure that a reliable data analysis is taken [15]-[17]. This procedure is accomplished in two steps: 1) Bad data detection and elimination; 2) Fill in missing data.

4.2.1 Bad data detection and elimination

Missing data on residential hourly electricity consumption is possible due to several reasons, such as equipment failure or external influences and empty households [19]. To create a clustering model as accurate as possible, it is necessary to identify the inconsistencies in the initial dataset, being necessary to perform the following evaluations:

- Missing-value users. To identify users with missing values, the number of days per month with data are counted and users with over 10% of missing data are removed from the dataset. This step is taken to avoid compromising the quality of the data filling procedure.
- Zero-use residential users. To identify zero consumption users, all the users with less than 40 kWh/month of electricity consumption are considered zero-users and therefore removed from the dataset. The threshold of 40 kWh/month was set considering that an occupied household has to have at least the fridge consumption throughout the month [51]. Other criteria can be set as for example in [19] where *zero-use residential users* were defined as a residential user with more than 20 days a month with daily electricity lower than 1 kWh.
- Outliers. To identify outliers, the yearly total consumption per household is plotted into a boxplot graph and outliers are identified. To verify if all outliers were identified the Grubb's test is used with a significance level of 0.05 [52]. The null hypothesis considered that all data values come from the same normal population and the alternative hypothesis considered that the largest data value is an outlier. In this thesis this analysis was also supported by the statistical software *Minitab* [50].

These evaluations were applied to the case study initial dataset. After removing missing-value users and zero-use residential users, the size of the dataset was 101 users. Outliers were identified and removed, achieving a final dataset with 90 users. The final dataset electrical consumption distribution is presented in Figure 8. From the analysis is possible to observe that the mean of the yearly consumption corresponds to 2398.0 kWh/year (with a 95% confidence interval of 216.1 kWh/year and 2628.9 kWh/year) with the median of 2103.8 kWh/year (with a 95% confidence interval of 1856.3 kWh/year and 2472.1 kWh/year), having both increased when compared with the initial dataset.

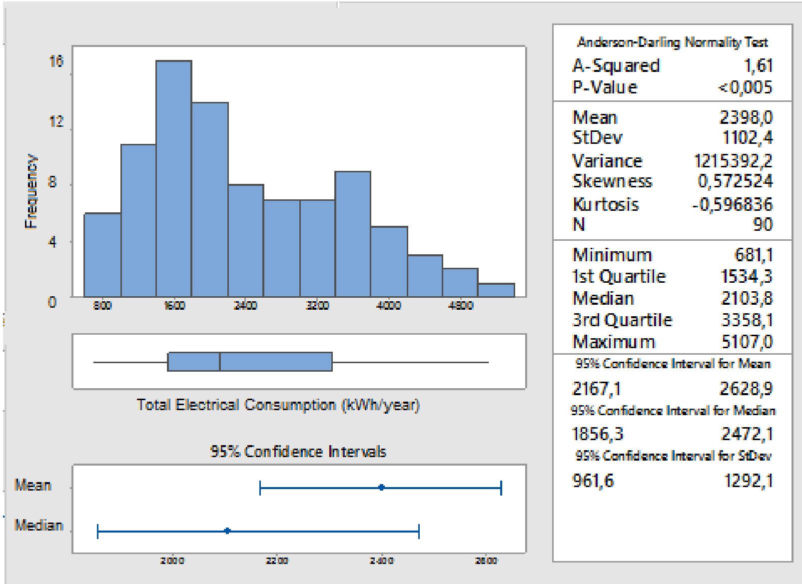


Figure 8 - Graphical summary of the dataset, after data cleaning

From the boxplot in Figure 8 there is no longer any outliers in the dataset. An outlier test was also taken (Figure 9) and as the *p-value* is higher than the significance level of 0,05 which confirms that no outlier is present in the dataset.

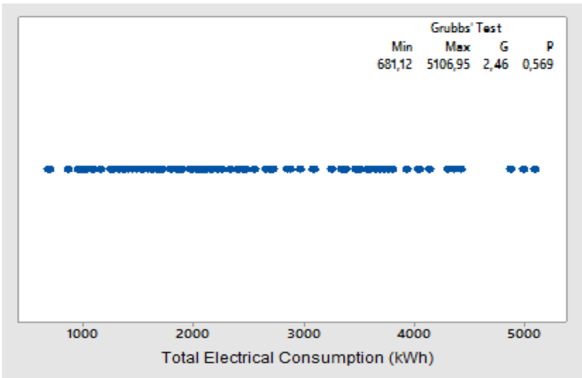


Figure 9 - Outlier test of the dataset, after data cleaning

4.2.2 Filling in missing data

In this data treatment phase, the objective is to fill in the hourly electric consumption for the users with less than 10% of missing data, to make sure that the information from the clustering analysis is reliable. For a reduced dataset size, the method proposed by [15] to input a missing value or replace a bad data at a certain hour, would be the mean of the values available for that hour on the remaining days. The use of the mode instead of the mean is also suggested, if a larger amount of daily data is available. In [53] it is proposed that fill in empty data should start by identifying patterns, by understanding the distribution of the remaining data, and decide on the best procedure to take. As the objective of this thesis is to analyze average daily data, it was verified that the hourly consumption does not vary sharply from hour to hour and that the consumption patterns vary throughout the days of the week.

As the smart meters data is usually available in a 15 minutes time-step, the consumption per hour can be verified knowing that in each hour there should be 4 data points. Every time that the count is less than 4, the electric consumption per hour, $\mathrm{EC_{ik}}$ is estimated as follows:

- If the time-step count, $\mathrm{TSC_{ik}}$, from a certain hour $\mathrm{h_i}$ at a certain day $\mathrm{D_k}$ is 0 and:
  - If $\mathrm{h_{i+1}}$ or $\mathrm{h_{i-1}}$ also have data missing:
    Make the average of the value of $\mathrm{h_i}$ from $\mathrm{D_{k+7}}$ and from $\mathrm{D_{k-7}}$;
  - If $\mathrm{h_{i+1}}$ and $\mathrm{h_{i-1}}$ don´t have data missing:
    Make the average with $\mathrm{h_{i+1}}$ and $\mathrm{h_{i-1}}$;
- If the $\mathrm{TSC_{ik}}$ from $\mathrm{h_i}$ at $\mathrm{D_k}$ is less than 4:

$$\mathrm{EC_{ik}} = \frac{\text{Original electric consumption}_{\mathrm{ik}}}{\mathrm{TSC_{ik}}} \times 4$$

- If the $\mathrm{TSC_{ik}}$ from $\mathrm{h_i}$ at $\mathrm{D_k}$ is equal to 4:

$$\mathrm{EC_{ik}} = \text{Original electric consumption}_{\mathrm{ik}}$$

As an example, a week of a certain user was used to demonstrate how the fill in empty data step was applied in the case study, with the blue being the raw data from the smart meter and the orange being the shape of the profile after filling in the empty gaps (Figure 10).

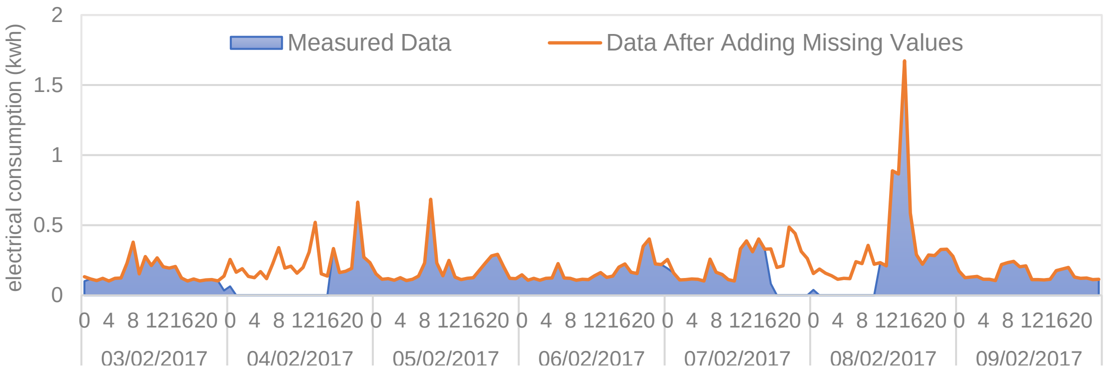


Figure 10 - Comparison of 7 days of a user's measured hourly electrical consumption with the same users filled in electrical consumption

This procedure was applied to the entire dataset. As a result, 70 users had its yearly electrical consumption increased between 1% to 3% when the empty values of consumption were filled in (Figure 11), 15 users had an increase of 3% to 6% and the remaining 5 had an increase higher than 6%. Although this increase in the annual electricity consumption apparently is not significant, for the clustering analysis this step is extremely important.

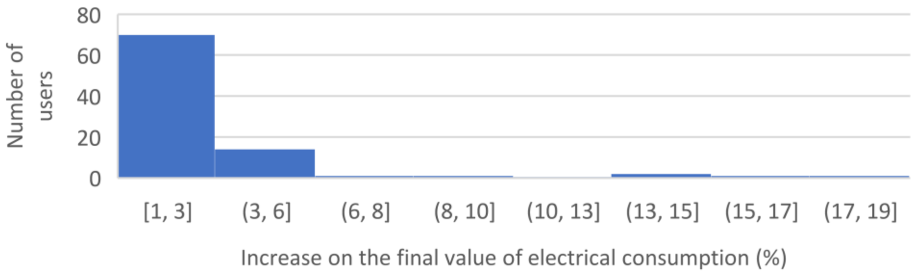


Figure 11 - Histogram of the increase in annual electrical consumption after filling in empty data

4.3 Final dataset

After adding the missing values, the mean and median values increased when compared with *Figure 8.* The mean of the yearly consumption is now 2474.2 kWh/year (95% confidence intervals of 2234.9 kWh/year and 2713.6 kWh/year) and the median is 2152.4 kWh/year (95% confidence intervals of 1902.7 kWh/year and 2661.1 kWh/year). The dataset has a smaller standard deviation when compared with the initial dataset. The minimum consumption value is 691.7 kWh/year and the maximum value is 523.4 kWh/year. The final sample has a significant number of users with consumption between 1200 kWh/year to 2000 kwh/year (Figure 12).

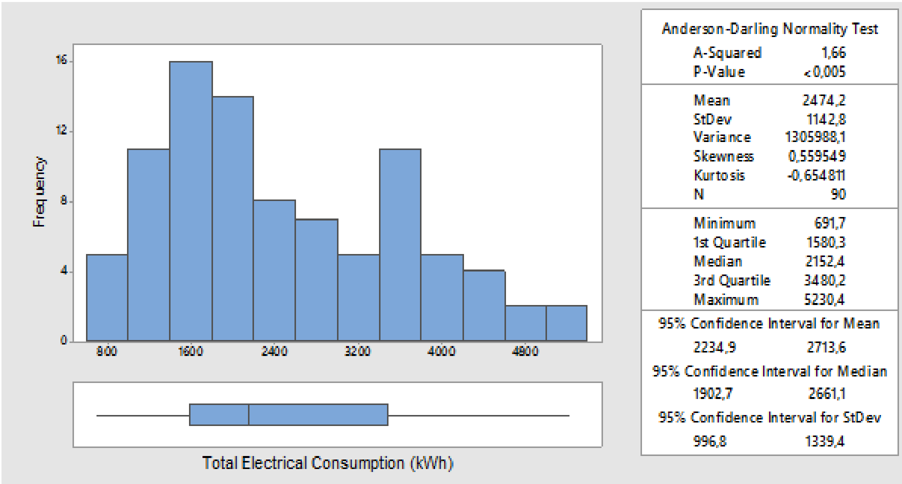


Figure 12 - Graphical summary of the final dataset, after data cleaning and filling in empty data

By having a valid distribution of the dataset, it is possible to go beyond the sample data and make statistical inferences about the larger population that follow the same distribution. So, a probability distribution identification is taken to evaluate the optimal distribution for the final dataset. Based on the probability plots and goodness-of-fit tests, 14 distributions are tested. Several criteria are used to select the appropriate distribution, from [54]:

- 1st criteria: compare the *p-value* to the *α-level*. In this case as α = 0.05, the *p-value* must be equal or higher than 0.05;
- 2nd criteria: choose a distribution which has a calculated *p-value* as similar as possible to Anderson-Darling (AD) statistic value;
- 3rd criteria: for the distribution with an extra parameter, use the for-Likelihood ratio test (LRT P) to determine whether adding the extra parameter improves the fit, if the LRT is smaller than 0,05 then choose the one with extra parameter;
- 4th criteria: use the probability plots to compare the fit of all distributions;
- 5th criteria: compare the values of the percentiles of the sampled data and compare it with the distributions, for 1%, 20%, 50% and 99%;

From the goodness-of-fit test, 9 distributions were eliminated since they failed to meet the first 3 criteria. The remaining distributions were: the Lognormal, the 3-Parameter Weibull, the Largest Extreme Value, the Gamma and the Log-logistic. From the visual inspection, the Largest Extreme Value distribution was eliminated. When comparing the percentiles values of the final dataset with the distribution, the Log-normal values were in the same range as the sampled data having the smaller associated error. The distribution that best describes the dataset is the Log normal, and the probability density function is described by the following function [55]:

$$P(x;\mu;\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\left[-\frac{(\ln(x)-\mu)^2}{2\sigma^2}\right]} \qquad (4.1)$$

The final distribution characterized by two parameters, the mean μ and the standard deviation σ of the variable's natural logarithm, and each has the value of 7.703 and 0.4849, respectively (Figure 13).

When comparing the final value of the mean electricity consumption per dwelling with the Portuguese reference of 3672 kWh/year per dwelling available in [56], the mean value of 2474 kWh/year of this case study is quite small. However, when using the same reference to calculate the mean consumption by summing the electrical consumption by end-use (9.176.569.767 kWh in [56]) and dividing it by the number of dwellings studied to achieved this end-use consumption (3932010 dwellings in [56]), the calculated value corresponds to 2430 kWh/year which is quite close to the obtained value in Figure 13.

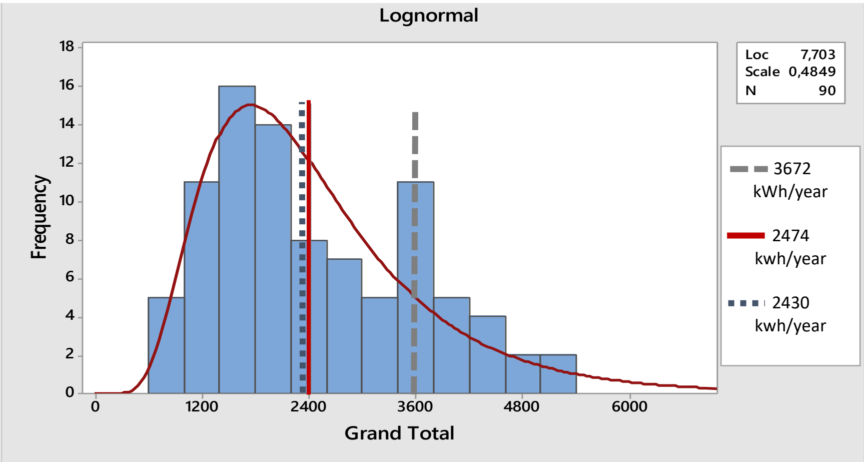


Figure 13- Final dataset distribution

4.4 Typification of energy consumption

In a residential district, usually there are various kinds of low-voltage electric power users, including residential users and nonresidential users. In a low-voltage distribution network, nonresidential electricity generally includes non-industrial use by property management companies (such as for elevators), electricity used for nonresidential lighting (such as for podiums, public corridors, and underground garages), electricity used in commercial buildings (such as shops) and electricity used in office buildings [19].

As the objective of this research is to specifically analyze the residential electrical consumption, once removed the *zero-use* and *missing-value residential users* from the initial dataset, it is also necessary to identify and eliminate the non-residential users. Several steps are taken to accomplish this goal, namely: the identification of the representative consumption profile; the profile consumption normalization; the clustering analysis; the adequacy measure analysis and the identification of different types of profiles. Each of these steps are described in detail hereafter

4.4.1 Representative consumption profile

As each user has 365 daily profiles, it is necessary to reduce the dataset without losing the information required for the clustering analyses [17]. Therefore, the 365 daily profiles were combined into an average daily curve consumption, where each hourly value represents the value of the mean consumption of the whole year at that hour, creating for each user a single daily load representative curve [14], [17], [22]. Even though the weekdays and weekends have different profiles, especially for the nonresidential, the criterion was that the 2/7 of the impact from the weekend would still allow to make a good identification.

The 365 daily consumption profiles per user were averaged into one daily consumption profile. The 90 daily average profiles are presented in Figure 14, and it is possible to verify visually that within the 90 users there are some nonresidential users. However, the majority are residential users.

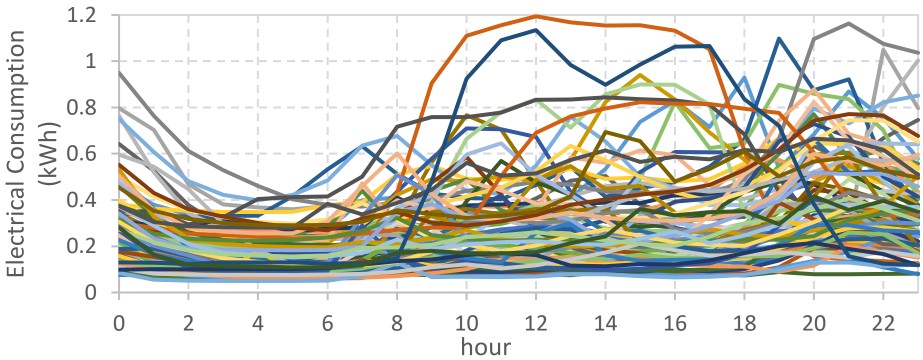


Figure 14- 90 user's average daily consumption profile

4.4.2 Profile consumption normalization

The main reason to normalize the consumption profiles is to have the consumption dynamics independent of the amount of energy consumed. For example, when applying a cluster algorithm to a normalized dataset it will group the consumers by its consumption behavior. On the other hand, with non-normalized data, the cluster algorithm will tend to group consumers by the amount of consumed energy [23].

Several data normalization strategies can be considered. In [23] the author described two different types: one based on the Maximum-Minimum values (4.2) and another based on the peak consumption (Maximum) value (4.3), where the first was preferred as it is aligned with focusing on the consumption dynamics, especially in the variance rather than the peak weight. In [24], besides the normalization strategies already described, the authors also introduced the Zero Mean Normalization and the Mean Normalization where the profiles have unitary mean, ending up choosing the zero mean normalization method as the easiest in identifying consumption peaks and consumers with sudden alteration in the consumptions.

$$\frac{C - \min C}{\max C - \min C} \tag{4.2}$$

$$\frac{C}{\max C} \tag{4.3}$$

The Maximum-Minimum normalization allows variables to have different means and standard deviations but equal ranges. In this case, there is at least one observed value at 0 and 1 end points. One should focus on time, rather than magnitude, if the goal is to aggregate load curves with the same time of peak consumption. On the other hand, as all curves have the same range, it is not possible to distinguish the steady profiles from the most irregular ones [24].

The Maximum Normalization works better than the maximum-minimum method in keeping the variance between the peak and the remaining values of the load profile. It prevents that a profile with a small peak to become scaled and grouped with another normalized profile with a much higher peak. On the other hand, it can lead to wrong interpretations, as there is always a maximum peak of 1 and a lower consumption user's curve will be moved upwards. Compared with Maximum-Minimum normalization method, the relationship between maximum value and minimum is kept [24]. In [10], [14], [15], [17], [18], [22] the normalization with maximum is used since the goal is to analyze the dynamics of the consumption regardless of the magnitude, maintaining the original shape of the profiles. In [16] the normalization used was Maximum-Minimum Normalization.

As an example, 3 consumption profiles were chosen and normalized with the methods previously described and compared with raw data. In Figure 15 it is possible to verify that in the raw data graph, profile 'A' has a steadier consumption when compared with profile 'D'. However, that information is lost when applying a Maximum-Minimum normalization.

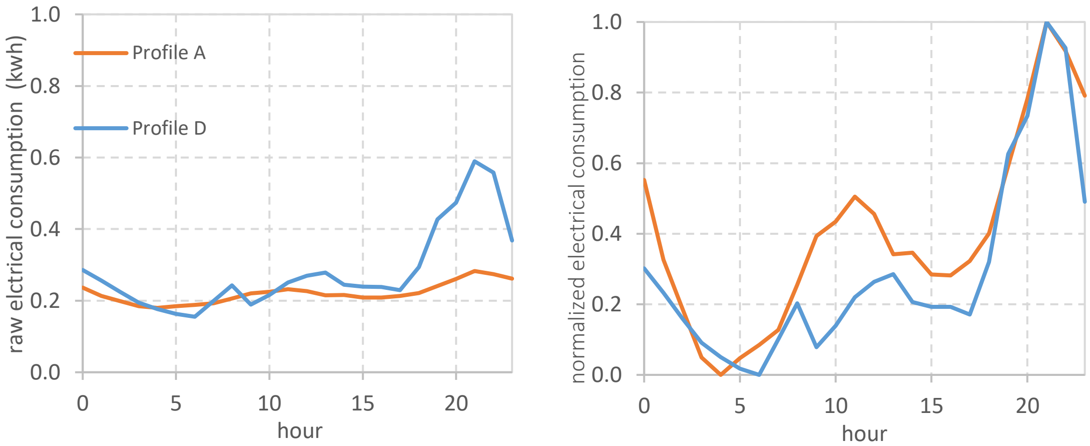


Figure 15 - Comparison Raw data (left) and Max-Min normalization (right) consumption profiles

From Figure 16 raw data graph (left), Profile 'B' has a higher consumption when compared with Profile 'A'. However, when comparing the normalized profiles (righ), the consumption curves are almost coincident, not being possible to verify that the profile 'B' has a higher consumption when compared to profile 'A'.

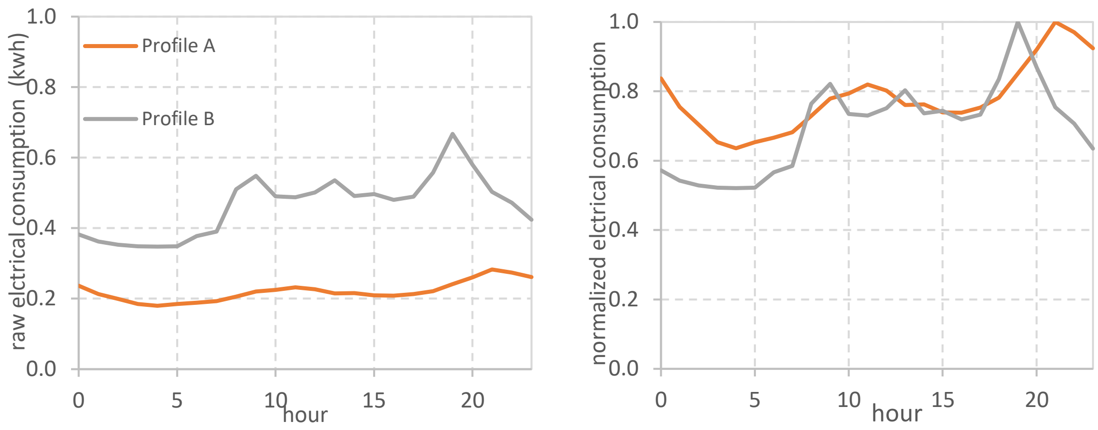


Figure 16 - Comparison Raw data (left) and Max normalization (right) consumption profiles

Both methods have pros and contras, so it is required a deeper study by testing the normalization methods in the clustering analysis.

4.4.3 Clustering analysis

After selecting the best method to normalize the data, a clustering analysis is performed in order to identify the different types of user's profiles. For this case study it was selected the normalization with Maximum value. Nonetheless, the study of the impact of choosing different normalization methods in the clustering analysis is discussed again atthe end of this section.

Partitional methods are highly efficient for clustering large and high dimensional datasets, and they are preferred for clustering daily load profiles of residential users [11]. The most popular and the simplest partitional algorithm is the k-means, because it is easy to implement, simple, efficient and with empirical success [42]. Also, this method emphasizes the minimization of the internal discrepancy for grouping, maximizing the distance in relation to other groups [12].

The k-means algorithm is based on the minimization of the internal distance between the patterns of a group or cluster. Due to this characteristic, the k-means clustering is considered as an algorithm of unsupervised data mining [12]. The k-means algorithm groups data around centers called centroids. All of this occurs in a cycle that ends only when the partitions cannot be improved, or until they reach a predetermined level of precision. The centroids are at the core of each class [12]. The k-means method is one of the most common cluster analysis methods. This method is an iterative process that groups n load profiles into k clusters, by minimizing the intra-cluster sum of squares J (4.4).

$$J = \sum_{k=1}^{K} \sum_{i=1, i \in k}^{n} \|X_i - W_k\|^2 \qquad (4.4)$$

Let $\mathrm{X_i} = \{\mathrm{x_i}, \mathrm{i} = 1, .., \mathrm{n}\}$ be the set of n load profiles to be clustered into a set of k clusters, $\mathrm{W_k} = \{\mathrm{w_k}, \mathrm{k} = 1, .., \mathrm{K}\}$, where $\mathrm{x_i} = \{\mathrm{x_i(t)}, \mathrm{t} = 1, .., \mathrm{T}\}$ is the $\mathrm{i^{th}}$ load profile and where $\mathrm{w_k} = \{\mathrm{w_k(t)}, \mathrm{t} = 1, .., \mathrm{T}\}$ represents the $\mathrm{k^{th}}$ cluster centroid. The inputs of the proposed k-means clustering module include load profiles of residential users and the number of clusters. The maximum number of clusters is always equal to the number of input load profiles and in extreme each load profile will have its cluster centroid. At each iteration of the k-means, the average Euclidean distance is calculated between the load profiles and the cluster centers according to (4.5) . This results in the assignment of each load profile to the cluster that has the nearest center [11].

$$d(x_i, w_k) = \sqrt{\frac{\sum_{t=1}^{T}(x_i - w_k)}{T}} \qquad (4.5)$$

The key steps of the k-means clustering algorithm are: 1) selecting an initial partition with k-clusters; 2) producing a new partition by assigning each pattern to its closest cluster center; 3) computing new cluster centers; and 4) repeat steps 2 and 3 until cluster membership stabilizes [42]. The disadvantage of this method is its sensitivity in the selection of the initial clusters that can converge to a local minimum [12].

The clustering analysis presented in this work was taken through the Minitab 17 statistical software [50]. A multivariable analysis using the cluster k-means algorithm is taken. The first input is a matrix with the normalized representative profile curve, with the size of 24 hours by N, the number of users. The number of clusters or initial partition column (where it is specified a column identifying the cluster membership) is a requirement to initialize the process.

The software then evaluates each observation, moving it into the nearest cluster, being the nearest cluster the one with the smallest Euclidean distance between the observation and the centroid of the cluster. The centroid is the vector of variable means for the observations in that cluster and it is used as a cluster midpoint. When a cluster changes, by losing or gaining an observation, Minitab recalculates the cluster centroid. This process is repeated until no more observations can be moved into a different cluster. At this point, all the observations are in their nearest cluster according to the criteria. Unlike hierarchical clustering of observations, it is possible for two observations to be split into separate clusters after they are joined together.

As output, one can obtain the number of observations in each cluster, the within cluster sum of squares, the average distance from observations to its cluster centroid, and the maximum distance of observation to the cluster centroid. The centroids for the individual clusters and the distances between their centroids (as Euclidean distances) are highlighted as well and the cluster membership can also be obtained and stored.

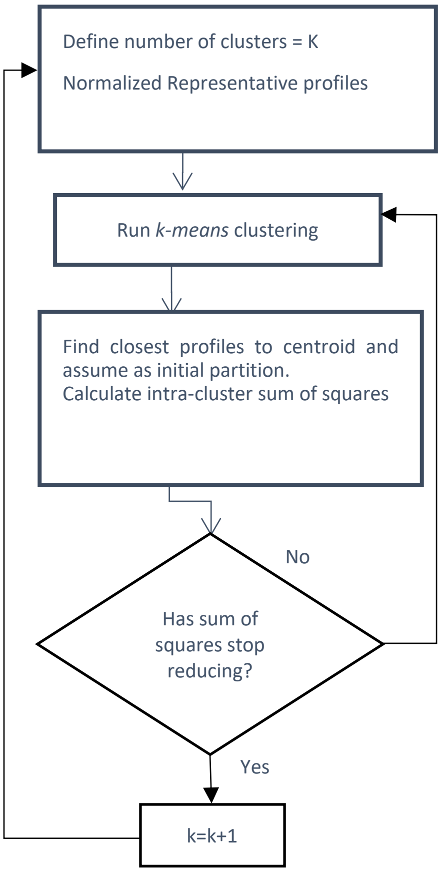


Figure 17 – Clustering Analysis Framework

The final partition will depend to some extent on the initial partition that is used, as k-means procedures may not perform as well when the initializations are done arbitrarily. However, if it is provided a good starting point, k-means clustering may be quite robust.

To overcome this issue, in this work the first step was to run the program and define the initial number of clusters. After having the distances from each profile to the centroids, an initial partition was created using the profile closest to each centroid center. The program ran again, changing the initial partition through the same method until the intra-cluster sum of squares stops reducing (Figure 17). In general, a cluster with a small sum of squares is more compact than one with a large sum of squares. This process is repeated from $k = 2$ to $k = \sqrt{N}$ [17].

As the goal of the clustering analysis at this stage is to identify residential and non-residential users, the dynamic of the profiles is the most important feature to keep, which both normalization methods previously described do. A deeper study is necessary, and it is done by using the clustering method with raw data (Figure 18*)* and the two types of normalization (Figure 19 and Figure 20) to compare results and be sure that the type of normalization chosen was the most adequate.

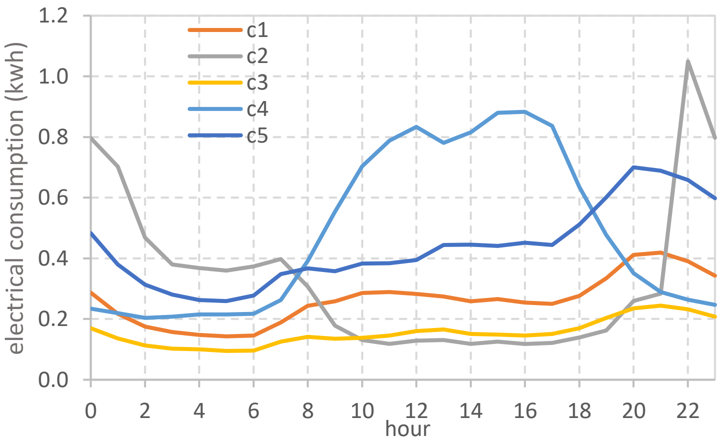


Figure 18 - Cluster centroids with raw data

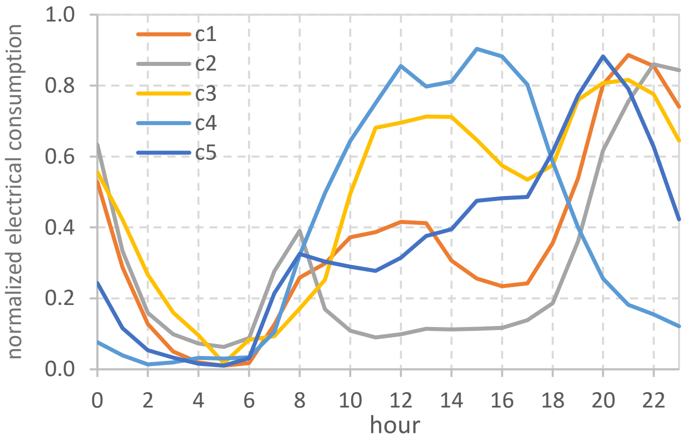


Figure 19 - Cluster centroids using normalized data with Maximum and Minimum

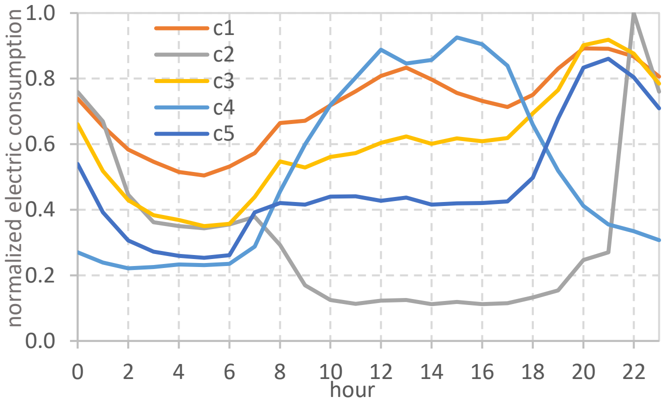


Figure 20 - Cluster centroids using normalized data with Maximum

For a 5 clusters result, it is possible to verify that, regardless of the normalization method chosen, the cluster centroid ‘c4’ corresponds to a nonresidential profile. However, when considering the cluster centroid ‘c2’, although in Figure 18 and Figure 20 the curve suggests a nonresidential profile (due to having an increase of 75 % of consumption in only one hour), in Figure 19 these distinction is not clear. In fact, ‘c2’ can easily be classified as residential when compared with ‘c3’, ‘c4’ and ‘c5’ (which are residential profiles). When using both types of normalization, ‘c4’ has 8 members, while the ‘c4’ obtained with raw data only had 7 members (Figure 21*).* That happened because when using raw data, a nonresidential profile with lower consumption, ’16.007’, was grouped in a residential centroid, ‘c1’ (Figure 18), since it had a similar magnitude of consumption to the remaining profiles.

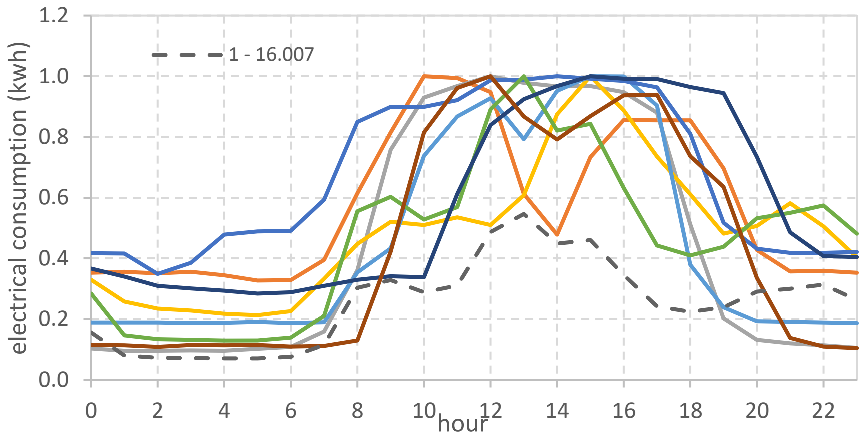


Figure 21 - Raw data profiles from c4, with the first seven profiles from clustering with raw data, and the 8 profiles from clustering with normalized data

Even though the identification of the higher consumption profiles is tougher, that it is not a crucial part of this clustering analysis. What it is crucial, is that the variation is kept as well as the dynamics of the profiles allowing for the nonresidential profiles to be identified. Therefore, using a Maximum normalization to perform the clustering analysis will yield better results as the different consumption profiles are well separated.

4.4.4 Adequacy measure

To verify what is the number of clusters that best divides the dataset and how good is the performance of the clustering algorithm, adequacy measures, also known as validity indexes are used. In [18] the authors stated that for k-means clustering method the most efficient adequacy measures are: the J_{MSE}- Mean Square Error or Error Function, the *IAI*- Intra Cluster Index and the *CH* – Calinski Harabasz Index or Minimum Variance Criterion. While Mean Square Error and Intra Cluster Index measure the compactness of the cluster and are also appropriate for extracting the optimal number of clusters by the "knee" detection method, the Calinski Index measures the compactness and the mutual separation between each cluster. The optimal number of clusters is the solution with the highest Calinski-Harabasz index value, since well-defined clusters have large IAI and a small IEI, meaning the larger the ratio, the better the data partition [57]. In [21] the measure used to evaluate how well the dataset was segmented was the Davies-Bouldin (DBI) validity index since it corresponds to the average similarity between each cluster to its most similar cluster. Therefore, low values of DBI indicate good partitions [14].

As IAI is already represented by the Calinski Harabasz Index it will not be used separately, so by combining the information of the remaining indexes, it is possible to get the right number of cluster that create a set of compact clusters and provide a good separation between cluster groups, when J_{MSE} starts to stabilize, when CH is as big as possible and DBI is as low as possible. These indexes are presented in the following equations:

The J_{MSE}- mean square error- distance of each patterns from its cluster centroid is defined as:

$$J_{MSE} = \frac{1}{N}\sum_{l=1}^{N} d^2(\vec{x}_i, \vec{w}_{k:\vec{x}_i \in S_k}) \quad (4.6)$$

where, *X* is the dataset and *N* is defined as its population, S_k is the subset of *X* that includes the population of the *k^th^* cluster where the $\vec{x}_i$ symbolizes the *n^th^* input vector and the $\vec{w}_{k:\vec{x}_i \in S_k}$ expresses the cluster´s center. The $d^2(\vec{x}_i, \vec{w}_{k:\vec{x}_i \in S_k})$ is the square of the Euclidean distance between the input vector which belong to the *k*th cluster and the centroid.

The IAI- Intra Cluster Index indicates the similarity level of the input patterns that belong to the same cluster and is defined as:

$$IAI = \sum_{k=1}^{K}\sum_{x_i \in X}(x_i - w_k).(x_i - w_k)^t \quad (4.7)$$

where, *X* is the dataset and *N* is defined as its population, *k* is defined as the obtained number of clusters, the $\vec{x}_i$ symbolizes the *n* input vector and the $\vec{w}_k$ expresses the cluster´s center.

The CH - Calinski Harabasz Index, is the ratio of the dispersion among the different clusters and the dispersion within the same cluster:

$$CH = \frac{N-K}{K-1}.\frac{IEI}{IAI} \quad (4.8)$$

where, *N* is defined as its population, K is defined as the number of clusters, *IAI* represents the Intra Cluster Index and *IEI* represents the Inter Cluster Index.

the IEI represents the Inter Cluster Index, which is defined as:

$$IEI = \sum_{k=1}^{K} N_k.(w_k - p).(w_k - p)^t \quad (4.9)$$

Where N_k is the population of the k^{th} cluster and p is the arithmetic mean of the set *X*. The Davies-Bouldin Index, which relates the mean distance of each cluster with the distance to the closest cluster [18]:

$$DBI = \frac{1}{K}\sum_{s,t=1}^{K}\max_{s\neq t}\left\{\frac{\hat{d}(W_s)+\hat{d}(W_t)}{d(w_s, w_t)}\right\} \quad (4.10)$$

Where $\hat{d}(W_s)$ denote as the mean distance of the load profiles belonging to cluster W_s and its centroid w_s, and $d(w_s, w_t)$ denote as the distance between the centroid of the cluster W_s and the centroid of the cluster W_t.

To select the appropriate number of clusters *k* that provide a good segmentation of the dataset, adequacy measures are used, but not always all the indexes agree. Therefore, the best solution is the compromise between the individual best results. From Figure 22, *k* should be equals to 3 or 5 to minimize DBI Index (left), for *k* equal or higher than 4 the J_{MSE}stabilizes (center), and the CH Index is highest when equal to 3 or as close (right). Since from mean square error k has to be equal or higher than 4, from that only the second-best option is available for DBI (k equal to 5), leaving the fourth best option for CH, the final number of clusters considered was 5, which also provides more diverse cluster centroids.

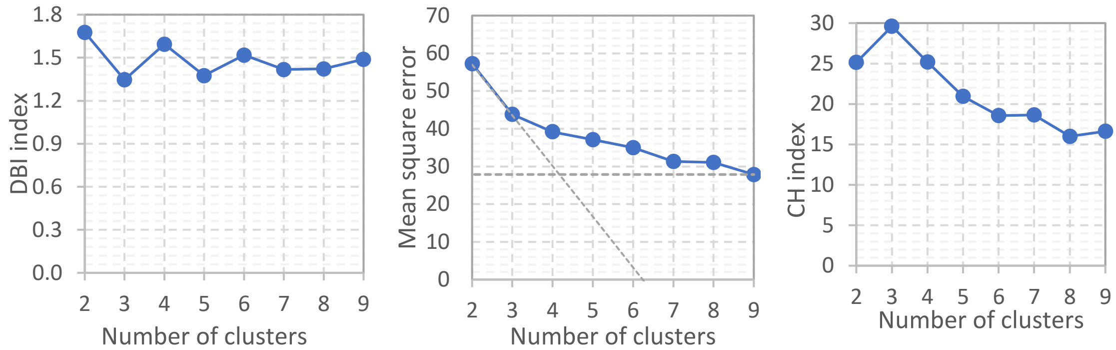


Figure 22 - Number of cluster selection methods: DBI index (left), Mean square error (center) and CH index (right)

4.4.5 Types of profiles

As a result, the final cluster centroids are presented in Figure 23 (considering the normalization with Maximum). From the final 90 profiles, 16 members are in *c1, 1* in *c2, 34* in *c3, 8* in *c4 and 31* in *c5 (*cluster centroids and it members presented in Annex B)

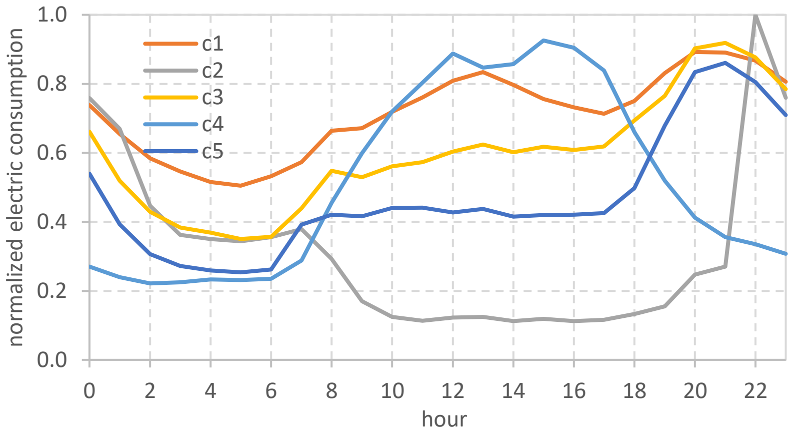


Figure 23 –Final Cluster centroids

From a literature review, it is possible to graphically identify nonresidential [12]–[14] and residential [12]–[14] [21], [22] profiles of load consumptions. The nonresidential profiles reflect commercial spaces, office spaces and restaurants/cafes which tend to have high and constant consumption throughout the day, with a significant reduction in consumption during the nigh time. Further, street lights would have a profile with a steadier consumption during the night, while lighting or common interior areas could have peaks during the day. The residential profiles tend to have a low consumption from 0 to 6 h when the consumption starts to increase, then the consumption stays steady until around 15 h when it starts to increase and the hours of higher consumption are usually from 18 h to 22 h.

Through the visual inspection it was therefore possible to identify the cluster centroid *c2* and *c4* as nonresidential profiles*.* By removing these users from the dataset (which means to remove 9 profiles corresponding to 10% of the total analyzed profiles), the final sample considers 81 residential consumption profiles, which are represented in Figure 23 by *c1, c3* and *c5.*

4.5 Residential consumption characterization

Since the goal of this work is to determine the operation schedules for residential buildings, nonresidential profiles were excluded from the analysis. In this step, the objective is to characterize the residential users and identify the changes throughout the days, weeks and seasons. To do so, three reference months have to be analyzed: a free-float month with no heating or cooling needs [58]; a winter month, with heating needs and a summer month, with cooling needs. For each of the reference months, every household will be represented by two daily average profiles: one for the weekdays and other for the weekends. The steps included in this stage are: month selection; the profile representation; the clustering analysis and the paths analysis.

In this stage, it was chosen not to normalize the representative profiles, for three reasons. First, in this stage the outliers, the *zero-use* residential profiles, the profiles with *missing values* and the nonresidential profiles are no longer present in the dataset, which improves the performance of the analysis. Secondly, in the previous analysis, the risk of using raw data is that it generates groups based on magnitude, losing the distinctive shapes of the different costumers. In residential profiles there are no longer such diverse shapes and the magnitude can give useful information about the inhabitant's behavior. Lastly, in the next stage of this research, the cluster centroids will be used to create input to make an energy simulation and then will be compared with the modeled values. Besides, by not normalizing the dataset, the Clustering analysis proceeds in the same way as described in Section 4.4.3 (illustrated in Figure 24) and the only difference is that the profiles are not normalized. For the validity index, the indexes were normalized to be able to compare the best solution for all the cases. Each index normalization was made with its maximum value. The month selection and path analysis procedures are further developed below.

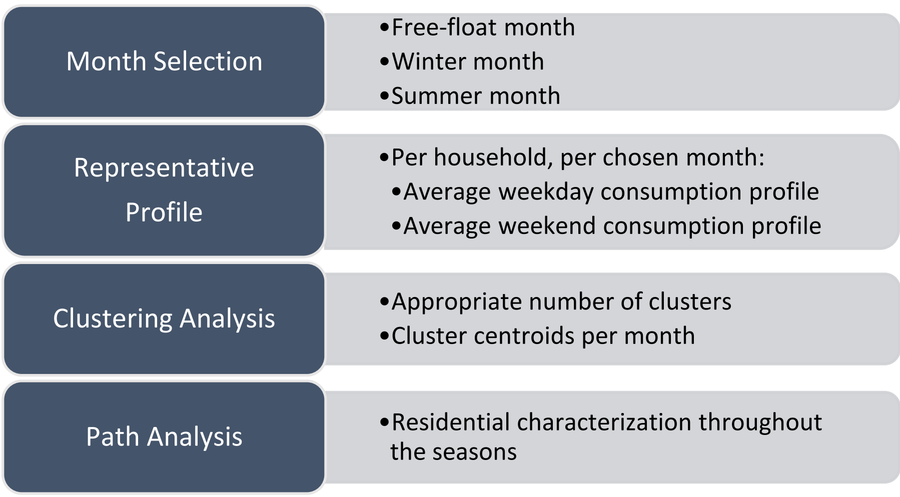


Figure 24- Workflow for residential consumption profile characterization

4.5.1 Month selection

The selection of the appropriate months can be done using the Heating Degree Days (HDD) and Cooling Degree Days (CDD) methodologies [59]. The degree days provide a measure of how much (in degrees), and for how long (in days), the outside temperature was below, in case of the heating degree days, and above, in the case of cooling degree days, a base temperature. HDD and CDD are calculated as follow [60]:

$$\mathrm{HDD(T_{base})} = \frac{1}{24}\sum_{i=1}^{8760}(\mathrm{T_{base}} - \mathrm{T_{exterior}}, 0) \qquad (4.11)$$

$$\mathrm{CDD(T_{base})} = \frac{1}{24}\sum_{i=1}^{8760}(\mathrm{T_{exterior}} - \mathrm{T_{base}}, 0) \qquad (4.12)$$

The free-float month is therefore the month with the lowest amount of cooling and heating needs, being the winter month, the one with the highest heating needs, and the summer month the one with the highest cooling needs. However, besides the degree days, other indicators may be considered in the selection process, such as other climate related factors like precipitation [61], as well as occupants behavioral factors like large vacation periods. Since the Degree Days concept has several limitations because only exterior temperature is considered, it is best suited when solar gains and internal gains do not affect significantly the heat balance, as for example, for Nordic countries where the major driver for space heating is the outside temperature [62]. From [63], in the winter season the effect of precipitation is negative because it can further intensify the use of heating equipment. In summer, the effect of precipitation is positive because temperature will decrease and become pleasant, so less air conditioning and other cooling appliances will be used by domestic consumers. For example, when it comes to selecting the free-float month, if multiple months present similar values of heating and cooling degree days, the selection can be based on the lowest precipitation values since that factor will also affect electrical consumption [64]. When it comes to selecting the summer month, if two or more months presents similar values in terms of cooling degree days, the month that is not typically associated with large periods of vacations should be the choice. After the month selection, representative profiles are created per month, and then clustering analysis is applied to each month.

In this case study, to select the appropriate reference months, using the degree days methodology, the base temperature considered for the heating degree days was 15.5ºC, based on [65], and 22ºC for the cooling degree days, considering the average summer temperatures in Portugal. For the free-float month selection, is possible to verify from the sum of HDD and CDD (Figure 25) that the months with lowest value are May and October. Therefore, other factor will be taken into consideration for this selection. For the winter month, there is a clear peak in the HDD for January, and for the summer August is the month with highest amount of cooling degree days, closely followed by July.

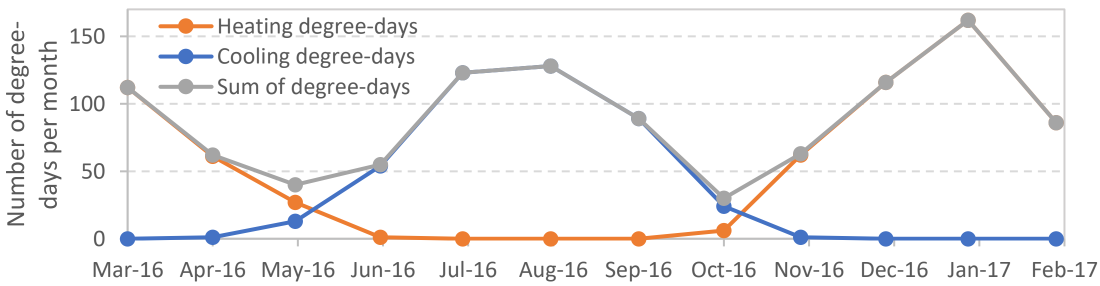


Figure 25 – Sum of degree days per month

From Figure 26, it is possible to observe that the average values of rainfall are typically the triple in October than in May, when comparing with October, May presents to be better suited to represent the free-float month. When it comes to selecting the summer month, vacation periods must be taken into consideration to avoid having empty households. Therefore, July was selected due to Agosto being the typical month of summer vacations. However, it is possible to have also in July a percentage of households in vacations.

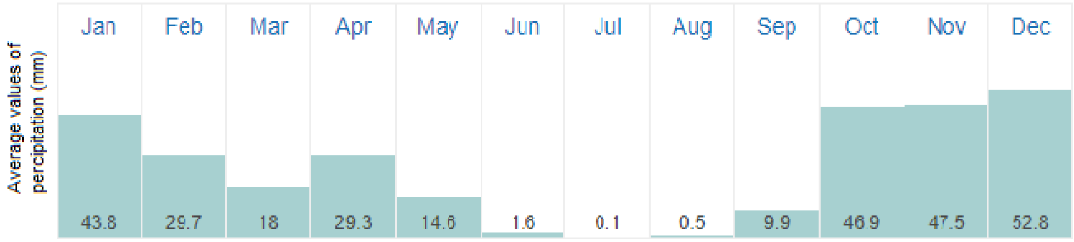


Figure 26– Average rainfall values from Lisbon Portela Airport weather station, from 1985-2015 [66]

In summary, the reference months considered in this work is May for the free-float month, January for the winter month and July for the summer month.

4.5.2 Number of clusters

Considering the months selected in the previous section, the dataset was filtered to perform another clustering analysis, now using the data corresponding to the reference months. Before starting the clustering analysis, the appropriate number of clusters have to be defined. Two indexes were used, DBI index and Mean square error. When it comes to the mean square error (Figure 27), *J* starts to stabilize from *k* equal to 4. When it comes to DBI index (Figure 28), the lower the *k* the better, so when combining the information from the two indexes it was chosen to have 4 clusters.

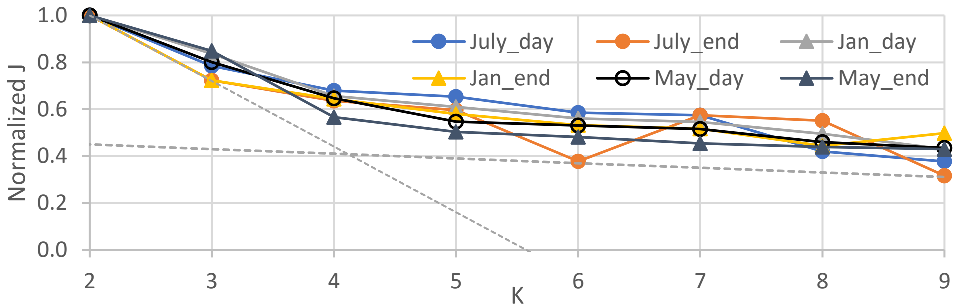


Figure 27- Normalized Mean square error by the number of clusters, for the selected months

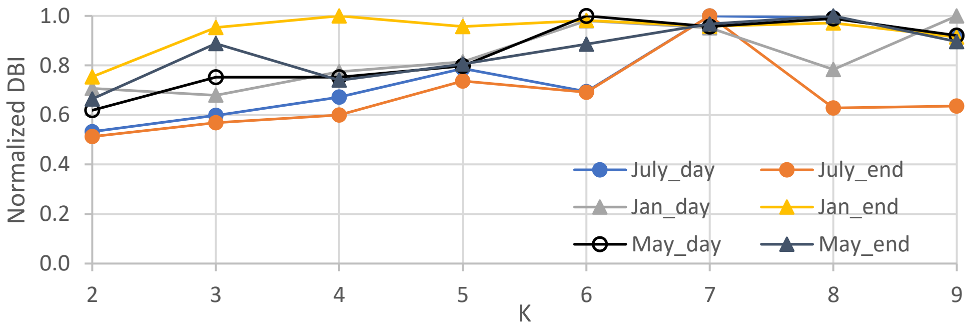


Figure 28 – Normalized DBI index per number of clusters

4.5.3 Residential profiles clusters for reference months

After performing the clustering analysis (following the same methodology described in Section 4.4.3), the results are presented for the free-float month (Figure 29), winter month (Figure 30) and summer month (Figure 31).

By analyzing the sample of users within the clusters, 'c3' has 4 users in the weekday and 2 users in the weekend for the free-float month, being this the centroid that the represents the energy-intensive consumers. It is possible to observe consumption peaks around the meal time, which could indicate the use of electric hob and oven, and the use of electric water heating. The remaining users were equally distributed in the other clusters of the free-float month. "*c1"* represents a low consumption household where it is hard to identify peaks of consumption, maybe representing a smaller household or/and more efficient equipment. The "*c2"* and "*c3"* have quite similar profiles, where the peaks have different magnitudes (lower in "*c2").* From both profiles, it is possible to identify the beginning of the activity around 7h and the peak of consumption around 21h, which could be related to cooking and/or other activities as watching TV. When comparing the weekday with the weekend profile, the first peak of consumption, which in the weekday occurs in early morning, in the weekend shifts for the lunch time.

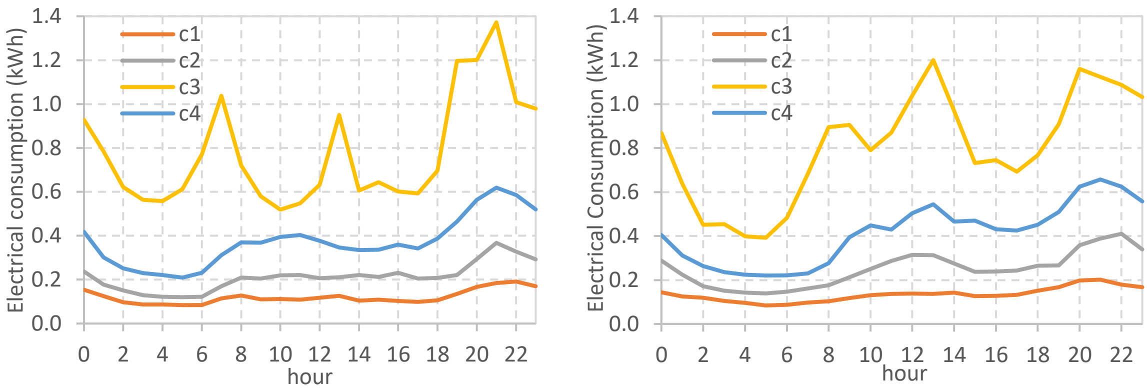


Figure 29- Cluster centroids for free-float month Weekday (left), Weekend (right)

For the winter month, the same scenario occurs, with "*c3"* having 3 users in weekdays and 4 in weekends. The remaining users were distributed among the clusters as follows: for the weekdays, "*c1"* has 26 users, "*c2"* 35 users, and *"c4"* 16 users*;* for the weekends, "*c1"* has 18, "*c2"* has 28 and*, and "c4"* has 32 members. The clusters centroids in the winter are distributed in the same way as the free-float month, with "*c3"* representing energy intensive consumers and "*c1"* low energy consumers. The peaks also occur at the same time as the free-float reference. However, the peak value increased probably due to artificial lighting need in the early morning, in winter, and due to heating needs at night.

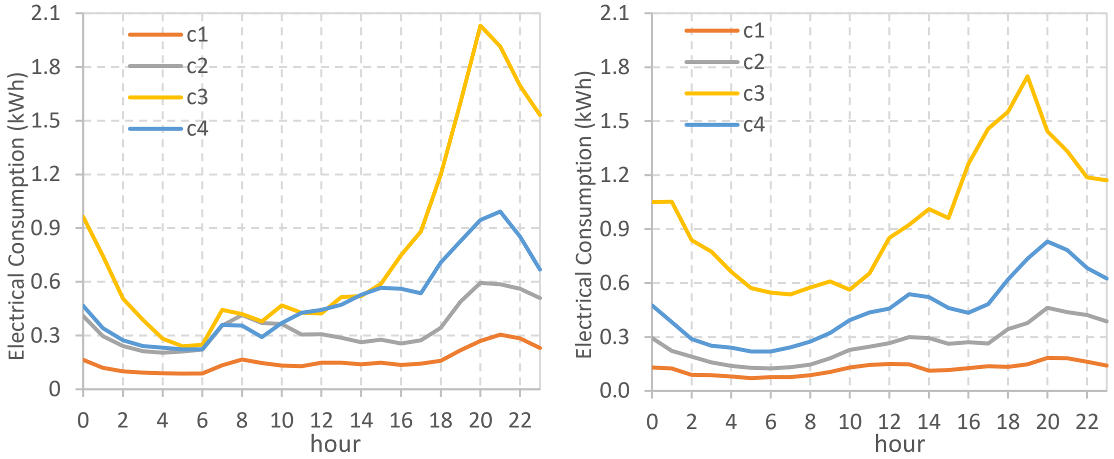


Figure 30 - Cluster centroids for winter month Weekday (left), Weekend (right)

In the summer month, the same scenario occurs, with *"c3"* having 7 members in weekdays and 6 in weekends. The remaining users were distributed among the clusters as follows: for the weekdays, *"c1"* has 36 users, *"c2"* has 23 users and *"c4"* 15 users; for the weekends, *"c4"* has 33 users, *"c2"* has 27 users and *"c4"* 15 users. The major difference in the summer profile is that, for example, in the *"c4"* cluster there is an increase in the consumption at 16 h, which could mean people at home due to vacations or this increase could be related to the use of space cooling.

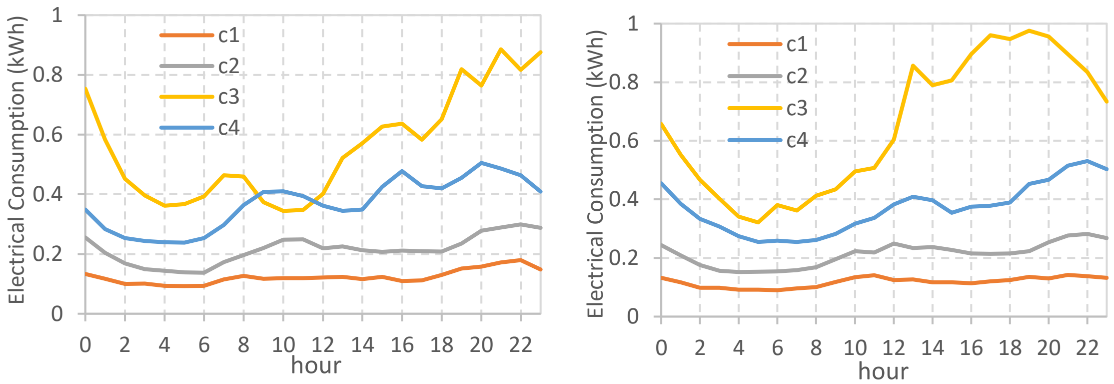


Figure 31- Cluster centroids from summer month Weekday (left), Weekend (right)

4.5.4 Paths analysis

After performing the clustering analysis and having the correspondent cluster centroids for the 3 reference months, paths for weekdays and weekend-days are then created as the combination of the cluster centroids, describing the way different household groups behave at these 3 moths: free-float month, winter and summer.

As an example, a certain number of households is represented in the weekdays by the cluster centroid 1 during the free-float month, cluster centroid 2 on the winter month and on the summer month by cluster centroid number 3. Therefore, the path for the weekdays would be defined as *123*. The same process is applied to create weekend paths. An example of a week path creation is presented in Figure 32 to illustrate that some deviation can also occur: the 22 households represented by cluster centroid 1 during the free-float month are distributed, in the winter month, by the cluster centroid number 1 with 86.3% of the initial members, cluster centroid number 2 with 4.5% and cluster centroid number 3 with 9%. Those first 86.3% are then represented by the cluster centroid number 1 and 2 in the summer month, creating paths 111 and 112, each representing 72.2% and 13.6% of the initial users. The process is repeated for all possible paths. When it comes to combining the week paths with weekend paths, they could also divide itself into different weekend paths, for example week path 111 could have half of its members represented by weekend path 222 and the others by weekend path 123.

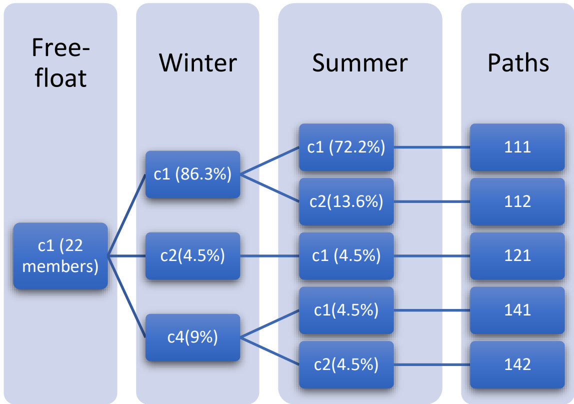


Figure 32 - path construction, starting on cluster centroid number one

In this thesis 27 paths for the weekdays were obtained but only 5 were studied. This selection was made based on the number of households represented by each path, having as minimum threshold 5%, since from this value the scope would be less than 4 households. 6 paths were above this minimum. However, one of the paths was not analyzed since it had very similar profiles when compared with one of the 5 selected. When combining these 5 weekday paths with the correspondent weekend paths, 25 final paths (Annex C) would be created. Due to the extensive number of final paths it was decided that only the most representative weekend path, for each week path, would be analyzed. The final 5 paths in analysis are Path 111 (20%), Path 212 (16%), Path 222 (10%), Path 422 (7%) and Path 444 (6%), which are discussed hereafter.

4.5.4.1 *Path 111*

Path 111 represents 19.75% of the total residential users (Figure 33*)* having also path 111 in the weekends. It is possible to verify that this path represents a low consumption user profile, where the winter month has a higher consumption than the remaining profiles. The summer profile has a similar consumption when compared with the free-float month, which can indicate that these users might not be using electricity for cooling or that the household was empty for a certain period, due to vacations, for example. This would affect the shape of the average profile, making it a steadier profile. During the weekend apparently heating or cooling equipment is not used.

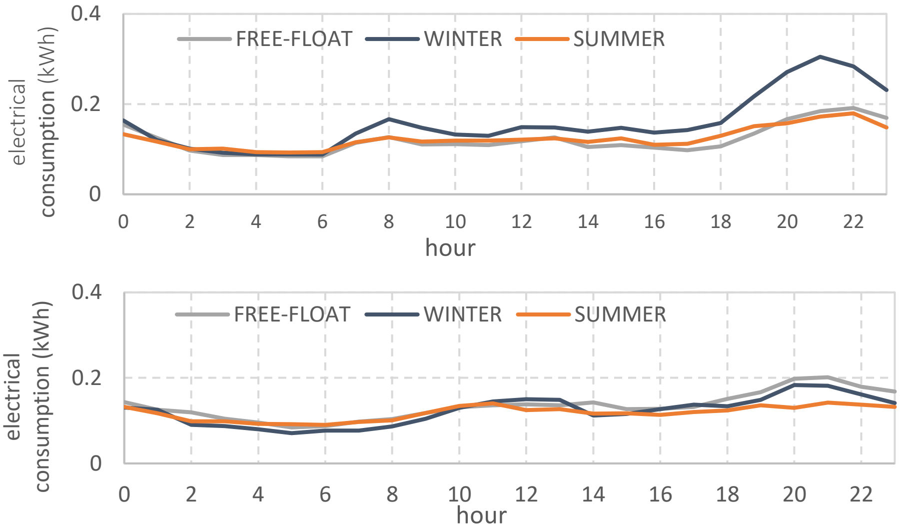


Figure 33- Path 111 from Weekdays (top) and Path 111 from Weekends (bottom)

4.5.4.2 *Path 212*

The path 212 represents 16.5% of the residential users on the weekdays and on weekends it is path 222 (Annex C). From Figure 34, the winter consumption is smaller than the remaining profiles which could be related to winter vacations or/and that the source of energy for heating is gas instead electricity. The summer profile has a morning peak around 10h, while in winter and free-float that occurred at 8h, which could indicate that at least part of the household is not working. This profile has a magnitude of consumption like the free-float month which could indicate no use of cooling system. The weekend path, in the free-float and winter, has an increase of consumption when comparing to the week path, especially in the winter profile. For example, from 15h a higher consumption can be observed which can indicate that the household is occupied for longer periods or that an electric system is used for heating. It is also visible that the first peak shifts from the early morning to around 12 h in the weekend.

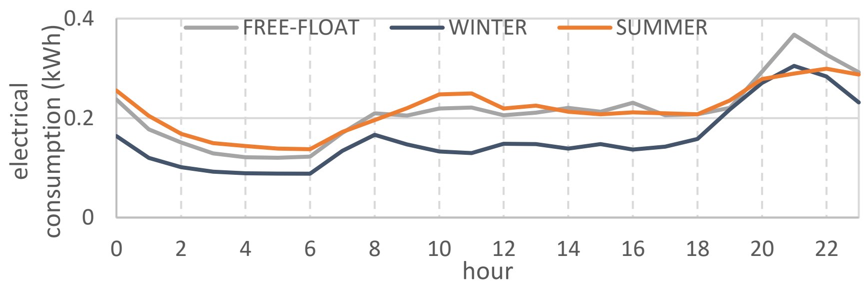


Figure 34 – Path 212 (Weekdays)

4.5.4.3 *Path 222*

The path 222 represents 9.88% of the residential users on the weekdays and has the path 222 as the weekend path (Annex C). As it can be seen in Figure 35*,* in general this path has more severe peaks of consumption when compared to the path 111. The winter profile has a much higher consumption when compared with free-float month, which could indicate that the source for heating consumption is electricity. The summer morning peak occurs at 10 h which could indicate that the household members, or at least some part, is on vacations. In the weekend path, the winter still has high consumption but not as high as in the weekday, and the summer profile is smaller than the free-float which indicates that no cooling equipment is used.

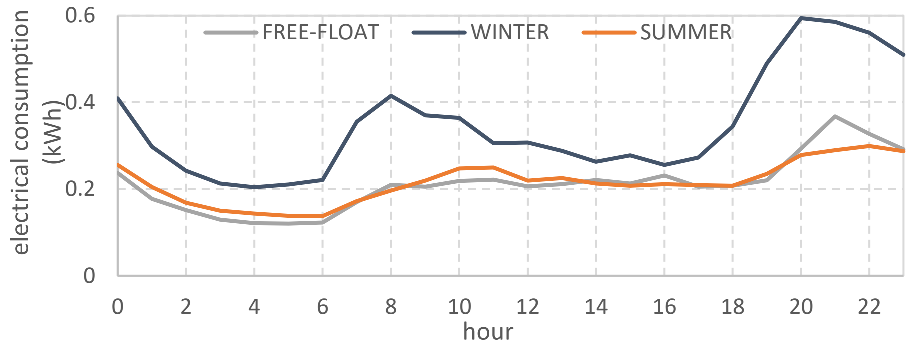
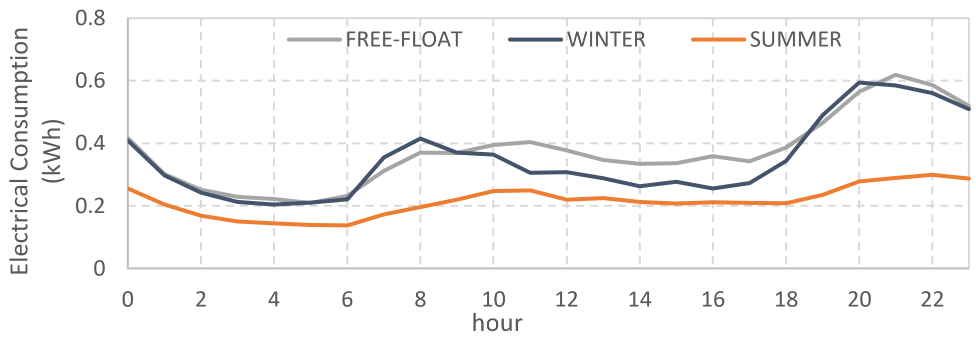


Figure 35 – Path 222 (weekdays)

4.5.4.4 *Path 422*

Path 422 represents 7.4% of the residential users on the weekdays, the correspondent profile for the weekend is 444 (Annex C). In Figure 36, the summer profile has smaller consumption when compared with the remaining months, which could indicate that the household was empty, possibly due to vacations. From the free-float month a peak occurs around 11h which could indicate that these household has one of the occupants not working full time or that it has a housekeeper. As the winter profile is quite similar to the free-float profile it is not possible to identify the use of heating equipment. When comparing with the path 222, the only difference is in the free-float profile which increases in this path, being in this case similar to winter profile. For the weekend, the winter profile has higher consumption than the remaining profiles from 16 h on, which could indicate the use of equipment for heating. The first peak occurs around 13 h, which could indicate the use of electric equipment for cooking.



.

Figure 36 – Path 422 (weekdays)

4.5.4.5 *Path 444*

The Path 444 represents 6.17% of the residential users on the weekdays, the correspondent profile for the weekend is the path 444 (Annex C). In Figure 37, the winter profile has a much higher consumption when compared with free-float month, suggesting the use of an electrical equipment for heating with low efficiency system and/or in a high percentage of rooms. The weekend profiles maintain the same dynamic as the week profile, having a slightly lower peak at 21h (0.8 kWh) in the winter when compared with the correspondent week profile.

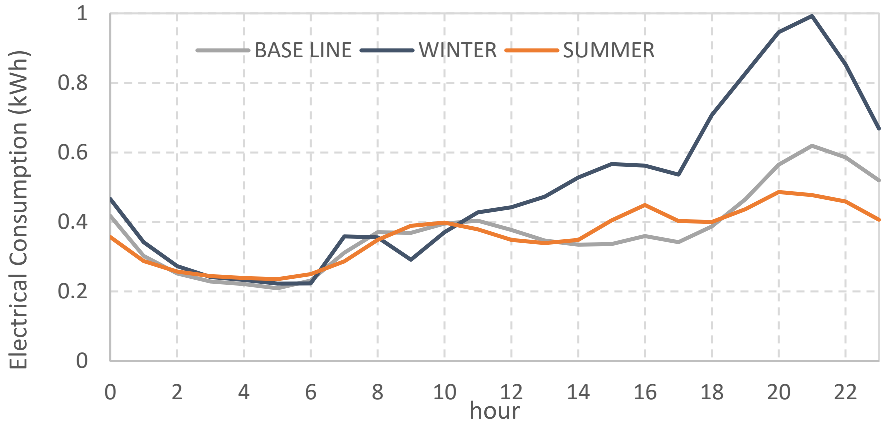


Figure 37 - Path 444 (weekdays)

In summary, from this analysis we can distinguish diverse user types profiles: two lower consumption profiles, one with heating use (Path 111) and another with no heating and no cooling use (Path 212); two medium consumption profiles with heating use (Path 222) and another without heating use (Path 422) and finally a high consumption profile, with heating use (Path 444). Also, two profiles have one or more members in vacations (Path 212 and Path 222), one household is occupied for longer periods than the remaining paths (Path 444), and one empty household in the summer month (Path 422).

4.6 Input profiles determination

To create a parametric model of the energy consumption of a building, certain information related to the occupants' behavior is mandatory such as: the occupancy profile, the lighting profile, the equipment profile (baseline and activity related consumption), and the heating and cooling profile (Figure 38). However, from the smart meter data this information is not directly assessable.

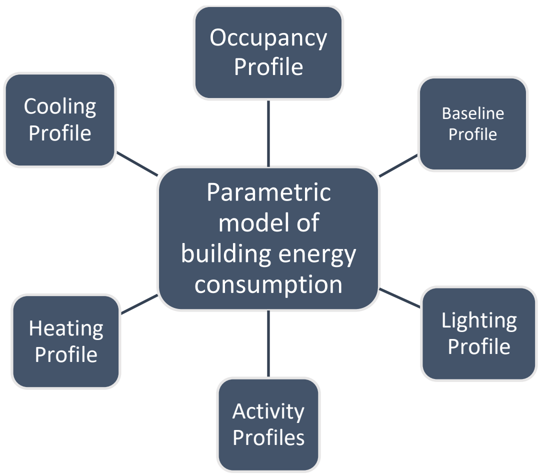


Figure 38 – Inputs required for the energy simulation model

The objective of this step is to use the paths identified in the previous section to create the input for the building energy simulation. To do so, it is necessary to create a mathematical model that can identify the occupancy patterns, distinguish the consumption related to certain activities from the occupant's behavior, create the lighting profiles and identify the use of electric consumption for heating and cooling use. The input-process-output is illustrated in Figure 39 and described in detail in the next sub-sections.

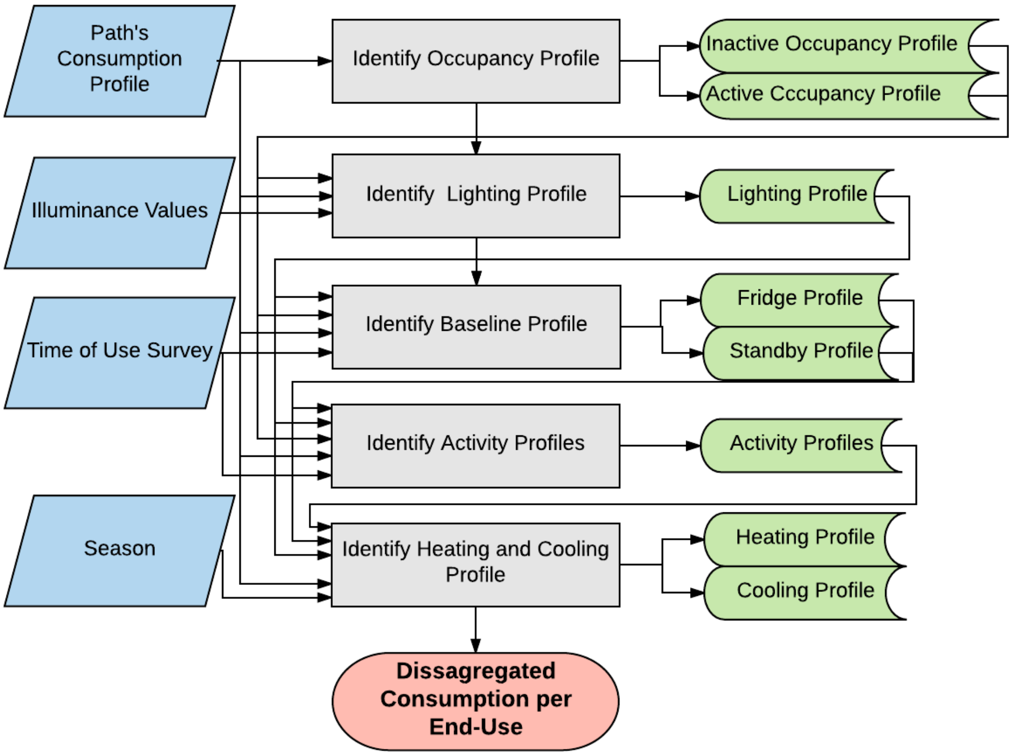


Figure 39- Input-Process-Output for the mathematical model

The starting point was a data-driven (inverse) approach based on parameter estimation, which in the context of building energy modelling, refers to methods which use monitored data from the building to produce models which are capable of accurately predicting the system behavior [67]. Each of the stages of the mathematical model developed in this thesis are detailed in the following subsections.

4.6.1 Occupancy profile

The method starts by identifying the startup and shutdown events by using the method proposed in [40], to determine the interval of time of possible active occupancy (i.e. the house is occupied and the occupants are not sleeping). That possible active occupancy is combined with higher consumption periods and turning on *events* to create an occupancy profile as showed in the input-process-output (IPO) from Figure 40. While doing this analysis, it was possible to verify that some difficulties would emerge when inferring the occupancy profiles: firstly, the time-step of electrical consumption used by the models to infer occupancy is between 1 min to 15 min time-step. However, the profiles obtained in the paths have 1-hour time-step; secondly, the consumption profile is a cluster centroid of a monthly average. Therefore, some information on the dynamics of the profile could be lost if all the inhabitants represented by that cluster centroid do not have the same schedules throughout the entire month.

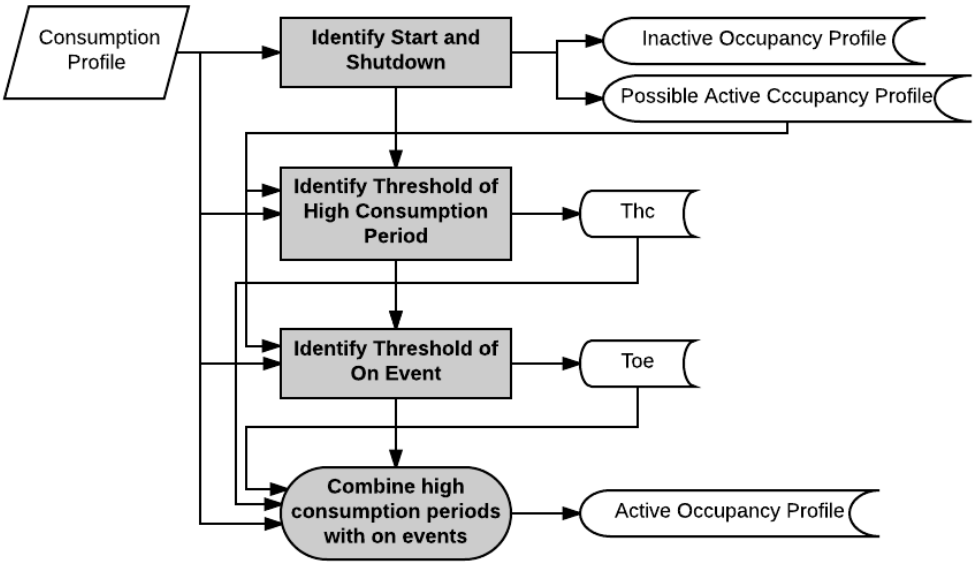


Figure 40 – IPO behind Identifying Occupancy Profiles

4.6.1.1 *Identify start and shutdown*

To infer about the possible active occupancy, the activity startup of and shutdown is determined using the method proposed in [40]. First, the difference between two consecutive intervals is calculated as follows:

$$\Delta i_{(d,t+1)} = i_{(d,t+1)} - i_{(d,t)} \tag{4.13}$$

Where $i_{(d,t)}$ represents the interval, data registered for day *d* and time *t.*

Second, $\bar{\Delta}_t$ is calculated to merge all the $\Delta i_{(d,t)}$ from different days, by calculating the median of $\Delta i_{(d,t)}$ for a certain time, this can be seen as the consumption routine for a median day of a particular meter.

In this thesis, the values available to calculate $\Delta i_{(d,t+1)}$ are the ones retrieved from the cluster centroids of a monthly average. Therefore $\bar{\Delta}_t$ is considered equal to $\Delta i_{(d,t+1)}$.

In [40] it is assumed that the startup takes place between 5h and 11h. As smaller spikes maybe due to some appliance turning on or off only significant spikes are considered to infer startup and shutdowns. The variation index must be higher than 0.5 to have enough variation to figure out a possible daily routine.

So, let $P(X, i)$ be the *i*th percentile of X, $0 < i \ll 100$. To identify a spike that is significant enough to mark the beginning of a daily routine, the follow rule set is proposed:

$$\text{startup} = \left\{ t: \bar{\Delta}_t > P(|\bar{\Delta}_t|, 80), t \in \{05{:}00 + 00{:}15k, k = 0, \dots, 24\} \right\} \tag{4.14}$$

The startup is identified as all the intervals for which $\bar{\Delta}_t$ is higher than 80% of all other spikes taken as absolute value. Since each set can have more than one element, startup time is the earliest of all the times in the set.

In [40], it is assumed that the shutdown takes place between 17h and 4h. Inferring a routine´s shutdown is more challenging than the startup as electricity consumption slowly decreases with night-time and there is not a clear moment where one can say that the occupants go to sleep. To overcome this difficulty, two different methods are proposed, choosing the one that yielded better results.

Interval data is used without calculating the difference between two consecutive intervals. After, the consumption values are grouped using the median of the values of the same interval, with $\bar{\Delta}_t$ equal to the median of the $i_{(d,t)}$. In this thesis, the values available of $i_{(d,t)}$ are the ones retrieved from the cluster centroids of a monthly average and therefore $\bar{\Delta}_t$ is considered equal to $i_{(d,t)}$. Having the variation index calculated as follows:

$$\Delta_t = \frac{\max(\bar{\Delta}_t) - \text{median}(\bar{\Delta}_t)}{\max(\bar{\Delta}_t)} \qquad (4.15)$$

If $\Delta_t < 0.2$ it is considered that there is not enough variation to figure out a possible shutdown. Otherwise the rule goes as follows:

$$\text{shutdown} = \left\{ t: \bar{\Delta}_t > P(\bar{\Delta}_t, 30), t \in \{05{:}00 + 00{:}15k, k = 0, \dots, 44\} \right\} \qquad (4.16)$$

There can be more than one element in the set so the final shutdown time is the latest of all elements.

To Identify possible active occupancy profile and inactive occupancy profile, the time when shutdown and the startup occurs are used. As a result, a profile of the possible active occupancy, $\text{Opa}_i = \left\{ \text{Opa}_i(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\} \right\}$, with the *i* representing the 3 different seasons can be determined:

if $t_{\text{shutdowni}} < t_{\text{start-upi}}$ then:

$$\text{Opa}_i(t) = \begin{cases} 0, & \text{if } t_{\text{shutdowni}} < t < t_{\text{start-upi}} \\ 1, & \text{if } t \gg t_{\text{start-upi}} \end{cases} \qquad (4.17)$$

If $t_{\text{shutdowni}} > t_{\text{start-upi}}$ then:

$$\text{Opa}_i(t) = \begin{cases} 0, & \text{if } 0 < t < t_{\text{start-upi}} \\ 1, & \text{if } t_{\text{shutdowni}} \gg t \gg t_{\text{start-upi}} \end{cases} \qquad (4.18)$$

The inactive occupancy profile, meaning occupants are in the household but are sleeping, $\text{Oina}_i = \left\{ \text{Oina}_i(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\} \right\}$, if $t_{\text{shutdowni}} < t_{\text{start-upi}}$ then:

$$\text{Oina}_i(t) = \begin{cases} 1, & \text{if } t_{\text{shutdowni}} < t < t_{\text{start-upi}} \\ 0, & \text{if } t \gg t_{\text{start-upi}} \end{cases} \qquad (4.19)$$

If $t_{\text{shutdowni}} > t_{\text{start-upi}}$ then:

$$\text{Oina}_i(t) = \begin{cases} 1, & \text{if } 0 < t < t_{\text{start-upi}} \\ 0, & \text{if } t_{\text{shutdowni}} \gg t \gg t_{\text{start-upi}} \end{cases} \qquad (4.20)$$

This step is applied to the three analyzed months, creating 3 inactive occupancy profiles: $\text{Oina}_{\text{free-floath}}$, Oina_{winter} and $\text{Oina}_{\text{summer}}$.

4.6.1.2 *Identify threshold of high consumption periods*

Electricity usage patterns change when occupants are present due to their interaction with the electrical loads. That is visible in higher average or variability of power usage, and a larger absolute range in power when compared with an unoccupied household. In

Figure 41 (where one indicates at least one occupant is present and zero indicates no occupants are present, in terms of the occupancy values) it is possible to verify an increase in consumption in both scenarios from 17h, when the household is occupied [39].

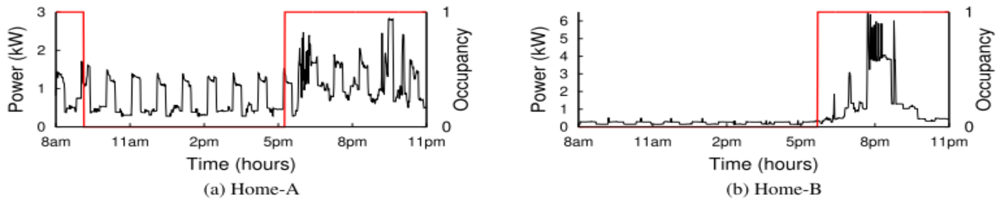


Figure 41 – Overlay of average power usage every minute (black) with binary occupancy (red) [39]

Figure 42 shows the probability to see a measurement during the presence or the absence of the occupants. One can see that the power consumption is likely to have a higher mean and standard deviation when the household is occupied. During the periods of absence, the electricity power is centered around 100 Watt and can be clearly distinguished from the overall daytime curve. When the household is occupied, the probability of higher consumption increases, even if, lower consumption values are yet quite frequent, when no electrical devices are being used [36]

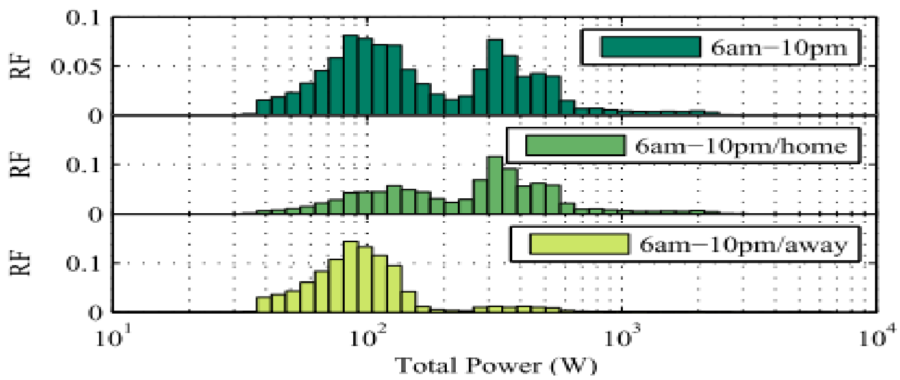


Figure 42 – Relative Frequencies of various total power consumptions (sum of all phases) values over the whole day (dark green) and divided into presence(light green) and absence (yellow) respectively [36]

The method to infer high consumption periods combines the information from the consumption profile and possible active occupancy profile. For each month, the first step is to calculate the median value of consumption, $\mathrm{Cm_i}$, being $\mathrm{C_i} = \{\mathrm{C_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$ the consumption profile:

$$\mathrm{Cm_i} = \mathrm{median}\ (\mathrm{C_i}) \qquad (4.21)$$

A profile that represents the percentage of the difference between the consumption at a certain hour when compared with $\mathrm{C_m}$, the $\mathrm{Cd\%_i} = \{\mathrm{Cd\%_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$ is determined as follows

$$\mathrm{Cd\%_i(t)} = \frac{\mathrm{C_i(t)} - \mathrm{C_m}}{\mathrm{C_m}} \times 100 \qquad (4.22)$$

After, $\mathrm{Cd\%_i}$ is combined with the possible active occupancy $\mathrm{Opa_i}$, to determine $\mathrm{Cd\%\_Opa_i}$:

$$\mathrm{Cd\%\_Opa_i(t)} = \begin{cases} \emptyset, & \text{if } \mathrm{Opa_i(t)} = 0 \\ \mathrm{Cd\%_i(t)}, & \text{if } \mathrm{Opa_i(t)} = 1 \end{cases} \qquad (4.23)$$

To identify the occupancy from a high consumption period, a threshold value must be set, $\mathrm{Thc_i}$, by averaging the values of $\mathrm{Cd\%\_Opa_i}$:

$$\mathrm{Thc_i} = \mathrm{Average}(\mathrm{Cd\%\_Opa_i}) \qquad (4.24)$$

4.6.1.3 *Identify threshold of on/off events*

In [41] an on/off feature representing the activation state of appliances is proposed as one of the factors to infer occupancy. The authors detected an on/off event by calculating the difference between a sample and its predecessor and then compare it with a certain threshold for an established amount of time.

In this thesis the method to infer about turning on event combines information from the consumption profile and the possible active occupancy profile. For each month, the first step is to calculate the difference between a sample and its predecessor, on/off values, $\Delta\mathrm{C_i(t)}$:

$$\Delta\mathrm{C_i(t)} = \mathrm{C_i(t)} - \mathrm{C_i(t-1)} \qquad (4.25)$$

A turning on event profile, $\Delta\mathrm{Cto_i}$, corresponds to all the positive values of on/off, so only positive values of $\Delta\mathrm{C_i(t)}$ are considered:

$$\Delta\mathrm{Cto_i(t)} = \begin{cases} \emptyset, & \text{if } \Delta\mathrm{C_i(t)} \ll 0 \\ \Delta\mathrm{C_i(t)}, & \text{if } \Delta\mathrm{C_i(t)} > 0 \end{cases} \qquad (4.26)$$

The turning on event profile is combined with the possible active occupancy with $\Delta\mathrm{Cto_i} = \{\mathrm{Cto_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$, $\Delta\mathrm{Ctopo_i}$:

$$\Delta\mathrm{Cto\_Opa_i(t)} = \begin{cases} \emptyset, & \text{if } \mathrm{Opa_i(t)} = 0 \\ \Delta\mathrm{Cto_i(t)}, & \text{if } \mathrm{Opa_i(t)} = 1 \end{cases} \qquad (4.27)$$

To identify the occupancy from a turning on event a threshold value must be set, $\mathrm{Toe_i}$, by averaging the values of $\Delta\mathrm{Ctopo_i}$:

$$\mathrm{Toe_i} = \mathrm{Average}(\Delta\mathrm{Cto\_Opa_i}) \qquad (4.28)$$

4.6.1.4 *Combine high consumption periods and turning on events*

The last step combines the values of thresholds from the two steps described earlier (4.6.1.2 and 4.6.1.3).

When it comes to high consumption periods the following method is proposed. For a certain time, *t*:

- If $\mathrm{Cd\%_i(t)} > \mathrm{Thc_i}$, then it is a high consumption period
- If $\mathrm{Cd\%_i(t)} > 0.5 \times \mathrm{Thc_i}$ or $\mathrm{Cd\%_i(t)} > 0.25 \times \mathrm{Thc_i}$ could still mean a high consumption period but that does not occur every day of the week
- If $\mathrm{Cd\%_i(t)} < 0.25 \times \mathrm{Thc_i}$ it is not a high consumption period

When it comes to turning on events the following method is proposed. For a certain time, *t*:

- If $\Delta\mathrm{Cto_i(t)} > \mathrm{Toe_i}$, then it is a turning on event
- If $\Delta\mathrm{Cto_i(t)} > 0.5 \times \mathrm{Toe_i}$ or $\Delta\mathrm{Cto_i(t)} > 0.25 \times \mathrm{Toe_i}$ could still mean a turning on event that doesn´t occur every day of the week
- If $\Delta\mathrm{Cto_i(t)} < 0.25 \times \mathrm{Toe_i}$ it is not a turning on event

4.6.1.5 *Active occupancy profile*

The final objective is to create active occupancy profile, $\mathrm{Oa_i} = \{\mathrm{Oa_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$. As the number of occupants is not a known input the following assumptions are applied:

- $\mathrm{Oa_i(t)} = 0$, if no active occupant is in the household, at a certain time *t*
- $\mathrm{Oa_i(t)} = 1$, if there are at least one active occupants in the household, at a certain time *t*
- $\mathrm{Oa_i(t)} = \mathrm{or}(0.75; 0.5; 0.25)$, meaning that a certain routine is not the same every day of the week. The house can be occupied ¾, or ½, or ¼ of the days of the week, at that time, or an occupant is at home but doing a lower consuming activity.

So, the active occupancy profile is created as follows:

$$\mathrm{Oa_i(t)} = \begin{cases} 1, & \text{if } (\mathrm{Cd\%_i(t)} > \mathrm{Thc_i}) \vee (\Delta\mathrm{Cto_i(t)} > \mathrm{Toe_i}) \\ 0.75, & \text{if } (\mathrm{Cd\%_i(t)} > 0.5 \times \mathrm{Thc_i}) \wedge (\Delta\mathrm{Cto_i(t)} > 0.5 \times \mathrm{Toe_i}) \\ 0.5, & \text{if } (\mathrm{Cd\%_i(t)} > 0.5 \times \mathrm{Thc_i}) \vee (\Delta\mathrm{Cto_i(t)} > 0.5 \times \mathrm{Toe_i}) \\ 0.25, & \text{if } (\mathrm{Cd\%_i(t)} > 0.25 \times \mathrm{Thc_i}) \vee (\Delta\mathrm{Cto_i(t)} > 0.25 \times \mathrm{Toe_i}) \\ 0, & \text{otherwise} \end{cases} \qquad (4.29)$$

For each path, there are 3 different occupancy profiles, corresponding to the free-float month, $\mathrm{Oa_{free-floath}}$, the winter month, $\mathrm{Oa_{winter}}$, and the summer month, Oa_{summer}, to accomplish changes in routines throughout the seasons.

4.6.1.6 *Application to the case study*

Due to the substantial number of paths, the path 111, presented in the path analysis section, will be used to demonstrate the application to the case study. The free-float will be used to illustrate the intermediate steps presented in the previous subsections and the final occupancy profiles for the three reference months is also showed. After identifying the startup and shutdown, high consumption periods and the turn-on events are identified.

In Figure 43, at 7h it is possible to identify a turning on event, by looking at the orange bar indicating the household members are waking up. At 18h a smaller turn on event occurs, which could indicate the household has occupants again. At 23h there is no turning on event as the power difference between two consecutive hours is no longer positive. However, there is still a high consumption period, when looking at the grey bar, meaning that it is very likely that the household is occupied. From this analysis, it is possible to observe that to create a complete occupancy profile, the information of both turning on events and high consumption periods is necessary.

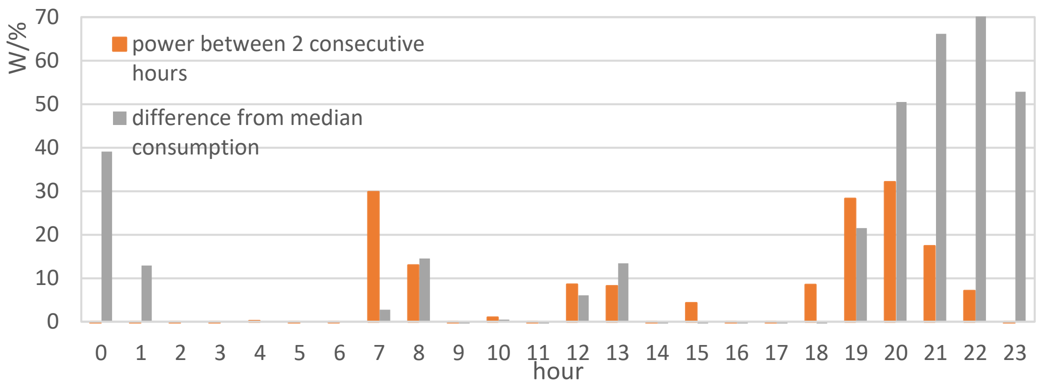


Figure 43 – Combination of the information behind inferring a turning on event and/or high consumption period.

After combining the values from Figure 43 with the thresholds of turning on event and high consumption period, the output is presented in Figure 44. As previously said at 7h the household is assumed to be fully occupied due to turn on event. At 18h half of the household is occupied or half of the week is fully occupied at that time, due to having a value lower but is still higher than half of the threshold of turning on event. At 23h, a high electric consumption period is identified and the household is assumed to be occupied. The values of the difference from median consumption and the power between 2 consecutive hours, for example at 4h and 10h (Figure 43) were not considered actively occupied due to being smaller than 0.25 for both thresholds. The values from

Figure 44 are combined into an active occupancy profile with the ruleset presented in 4.6.1.5.

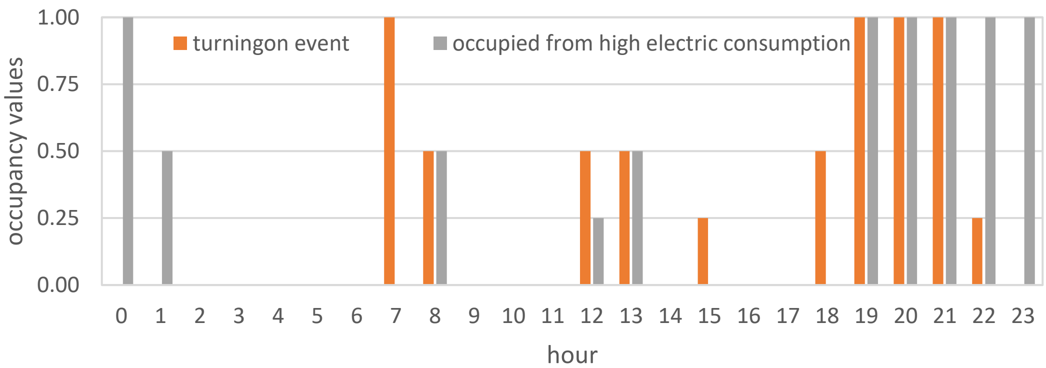


Figure 44 - Output of turning on event and high consumption period

The active occupancy profile is closely related to the consumption profile as showed in Figure 45. This figure shows the relationship between the consumption profile (line) and the active occupancy schedule (column). Between 6h and 7h, it is possible to observe a slight increase in the total consumption, suggesting that the household is occupied and an interaction whit the electric equipment's or lighting may have occurred. From 19h until 0h, it is possible to observe from the consumption profile a high consumption period suggesting also an active occupancy in that period.

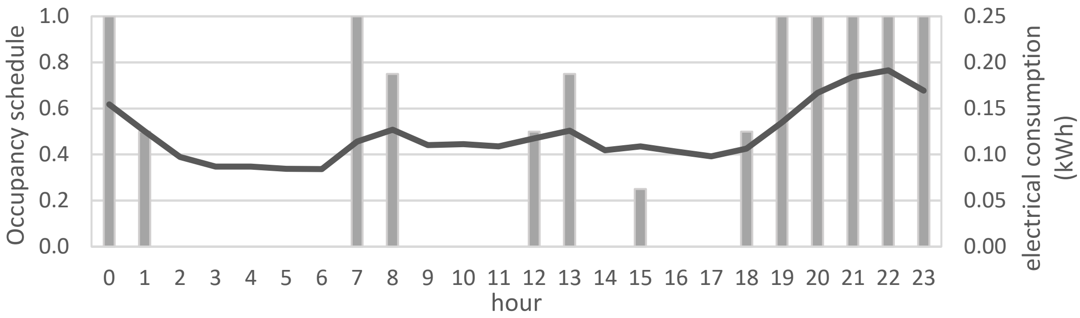


Figure 45 –Free-float consumption profile and active occupancy profile

The same process was applied to the winter and summer months and the results are showed in Figure 46 and Figure 47, respectively. From the three active occupancies profiles, it is also possible to verify that the occupants do not have the same schedules throughout the entire year. However, some resemblance is found as the waking up hour at 7h and a full occupation at 19h of the household. The occupancy values between 12h to 15h, that vary between 0.25 to 0.75, indicating that part of the household is occupied during the lunch time in some days of the week or that the household has a housekeeper some days of the week.

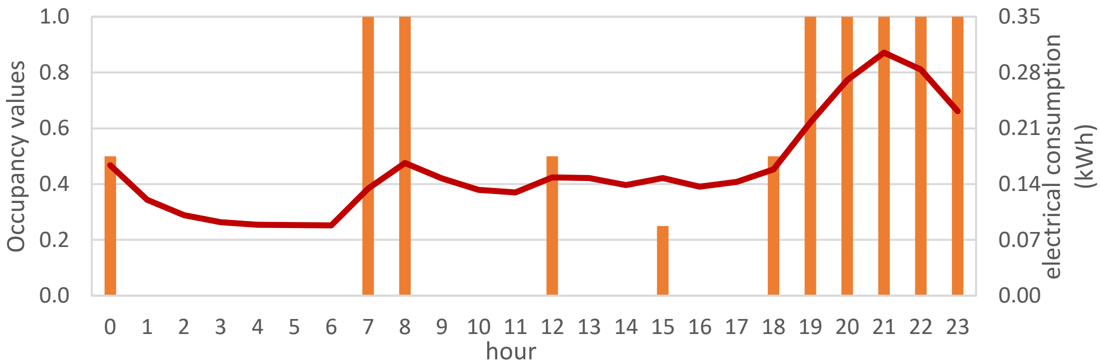


Figure 46 – Winter consumption profile and active occupancy profile

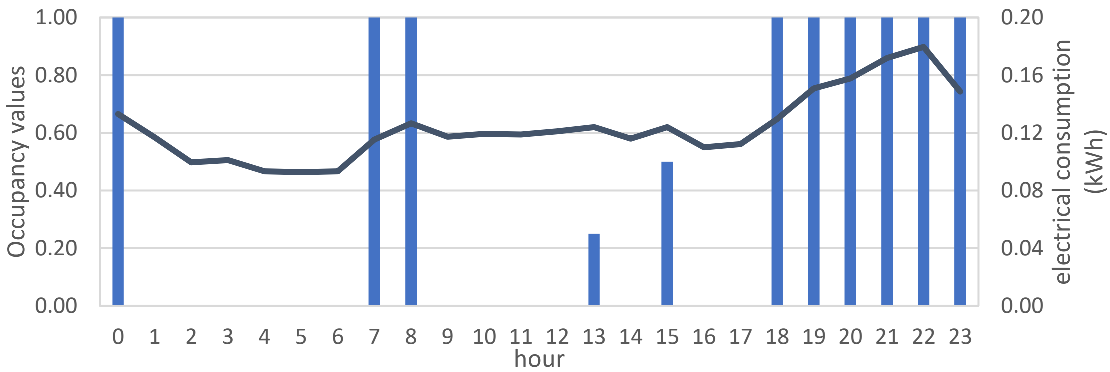


Figure 47 - Summer consumption profile and active occupancy profile

The remaining occupancy profiles are presented in Annex E.

4.6.2 Lighting profile

In [35], a model is proposed to create lighting profiles by combining active occupancy data series with outdoor irradiance data series and by randomly allocating different values of installed lighting units in each household. The main drivers of the lighting consumption are illuminance, occupancy and occupant's behavior and installed lighting units.

To create the lighting profile in this thesis, the active occupancy profile determined previously is combined with the illuminance values and the values of the installed lighting units are calculated to make sure that the total consumption accounted for the lighting agrees with the percentage value from [56], [68], [69] attributed to lighting.

4.6.2.1 *Indoors Illuminance values*

To overcome the unavailability of sensors to obtain accurate indoor illuminance information, a strategy is used based on the available data of the average daily values of global irradiance on a fixed plane $(\mathrm{W/m^2})$ (obtained from PVGIS [70] for the 3 months chosen in Section 4.5.1).

Illuminance is a measure of photometric flux per unit area, or visible flux density (typically measured in lumens per square meter), while irradiance is a measure of radiometric flux per unit area, or flux density (expressed in $\mathrm{W/m^2}$)[71]. There is no simple conversion from lux to $\mathrm{W/m^2}$, since it depends on the wavelength or color of the light. However, a simplification was adapted: for the sun light there is an approximate conversion of 0.0079 $\mathrm{W/m^2}$ per lux [72], valid for direct sunlight, while for daylight the appropriate conversion drops to 10% of that value [73]. This difference is based on the fact that sunlight is a portion of the electromagnetic radiation given off by the Sun, in infrared, visible, and ultraviolet light while daylight, is the combination of all direct and indirect sunlight during the daytime, including direct sunlight and diffuse sky radiation. Therefore, to convert irradiance into illuminance the following equation is used:

$$\mathrm{Illuminace_i\ (lux)} = \mathrm{Eo_i} = \frac{\mathrm{Irradiance_i}\ \left(\mathrm{W}/_{\mathrm{m^2}}\right)}{0.079} \tag{4.30}$$

To convert outdoor values of illuminance into indoors values of illuminance, the daylight factor (DF) is used. The DF is expressed as the percentage of the natural light falling on a work surface, $\mathrm{Ei_i}$, compared to that which would have fallen on a completely unobstructed horizontal surface under the same sky conditions, $\mathrm{Eo_i}$. The DF is analyzed at a point, but these values are often averaged across an entire room or visualized on a grid [74]. DF can be determined as follows:

$$\mathrm{DF}[\%] = \frac{\mathrm{Ei_i}}{\mathrm{Eo_i}} \times 100 \tag{4.31}$$

4.6.2.2 *Needs for artificial lighting*

The need for artificial lighting are related to indoor illuminance levels, since daylight levels are low. From [75], the probability of turning on the lights is proposed. The lowest the illuminance levels the highest the probability of turning on lights.

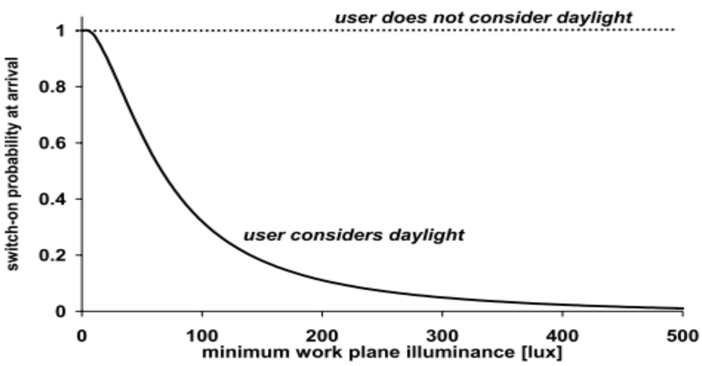


Figure 48 - Probability of switching on lights [75]

The values retrieved from Figure 48 were combined with indoor illuminance to infer the needs for artificial lighting at each time, $\mathrm{Nal_i} = \{\mathrm{Nal_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$. It was assumed that the occupants would turn on the light with the same proportion of values of the probabilities. So $\mathrm{Nal_i}$ is calculated for the three months selected in 4.5.1 as follows:

$$\mathrm{Nal_i(t)} = \begin{cases} 1, & \text{if } \mathrm{Ei_i} = 0 \\ \mathrm{Pso(Ei_i(t))}, & \text{if } 500 > \mathrm{Ei_i} > 0 \\ 0, & \text{if } \mathrm{Ei_i} > 500 \end{cases} \quad (4.32)$$

Where $\mathrm{Pso(Illuminace_i)}$ switch-on probability values for each illuminance values from the Figure 48 and $\mathrm{Ei_i(t)}$ the hourly values of indoor illuminance.

4.6.2.3 *Lighting profile*

To create the lighting profile, $\mathrm{L_i} = \{\mathrm{L_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$: the needs for artificial lighting are combined with the active occupancy profile. Besides that, a residual value of lights left on was assumed regardless artificial lighting needs and occupancy. This is done to represent the use of lights in room without windows or lights that could have been on even if not needed. This value is 5% of the light units, $\mathrm{Ilight\_units}$:

$$\mathrm{L_i(t)} = (\mathrm{Nal_i(t)} \times \mathrm{Oa_i(t)} + 0.05) \times \mathrm{Ilight\_units} \quad (4.33)$$

The total value of light units was obtained by combining different amount and types of bulbs to achieve a close value to the electrical consumption reference for Portugal, [56] proposes the value of 13.8% while in [69] 12% is proposed. A comparison of the total value of lighting consumption with the total electrical consumption is made, represented by the sum of the consumption profile:

$$\%\mathrm{lighthingconsumptioni} = \frac{\sum \mathrm{L_i}}{\sum \mathrm{C_i}} \quad (4.34)$$

The value of light units is calculated for the free-float month and the same value is assumed in the summer and winter months. If a month has more than 20% of lighting consumption, the $\mathrm{Ilight\_units}$ will be reduced to 75% of $\mathrm{Ilight\_units}$ installed in the free-float month. The $\mathrm{L_i}$ will produce 3 different profiles: one corresponding to the free-float month, one to the winter month and one to the summer month, due to different illuminance values and occupancy profiles.

4.6.2.4 *Application to the case study*

To determine the DF, one should use information about the glazing facades characteristics of the buildings in this area, available in the energy performance certificates [46], and cross with DF reference values from literature. A study focusing on obtaining different DF values for different areas in Portugal and different glazing facades is used [76] and the DFs for the case study are determined as follows:

$$\text{Average DF} = \frac{W}{A} \times \frac{T\theta}{(1 - R^2)} \tag{4.35}$$

where: W is the area of the windows (m^2), A is the total area of the internal surfaces (m^2), T is the diffuse visible transmittance of the glazing, including corrections for dirt on glass and any blinds or curtains, θ is the visible sky angle in degrees from the center of the window and R is the average reflectance of area A.

In [76], it was considered a window area of 6.25 m^2 (2.50m x 2.50m), a room floorplan area of 25m^2 (5.0m x 5.0m), and a floor-to-ceiling height of 2.80m. The DF is proportional to the window area and inversely proportional to the total area of the internal surfaces. Even though in the present work the window area is higher than in the reference used, there is also an increase in the total area of the internal surface indicating that the values retrieved from [76] are a good approximation. Therefore, the Indoor illuminance was calculated as follows:

$$Ei_i(\text{lux}) = Eo_i(\text{lux}) \times \frac{6}{100} \tag{4.36}$$

The relationship between the illuminance values and the lighting needs, for the free-float month, are presented in Figure 49. It is possible to observe that the lighting needs are quite small from 6h:30 to 18h. The same occurs in winter, between 7h:30 and 17h (Annex D) while in summer the light needs are small from 6h to 18h (Annex D).

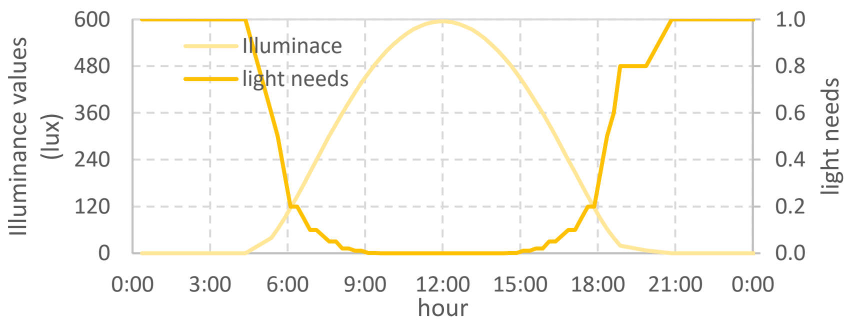


Figure 49 – Illuminance values and lighting need through the day

The lighting needs are crossed with the active occupancy profile and the resulting light profile for the free-float month is presented in Figure 50.

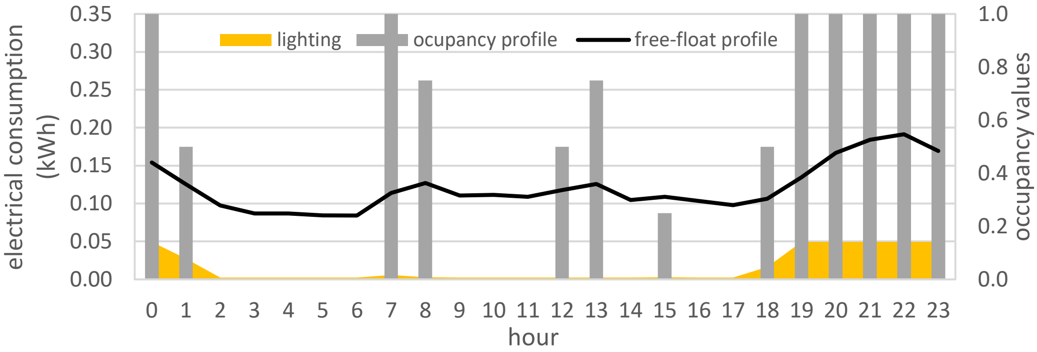


Figure 50 – free-float lighting profile for path 111 compared with total consumption and occupancy profile

Due to higher lighting needs, the winter lighting profile (Figure 51) has higher values of consumption than the remaining two lighting profiles, being the only one a with high consumption in the morning. The summer and free-float profile have similar lighting consumption profiles.

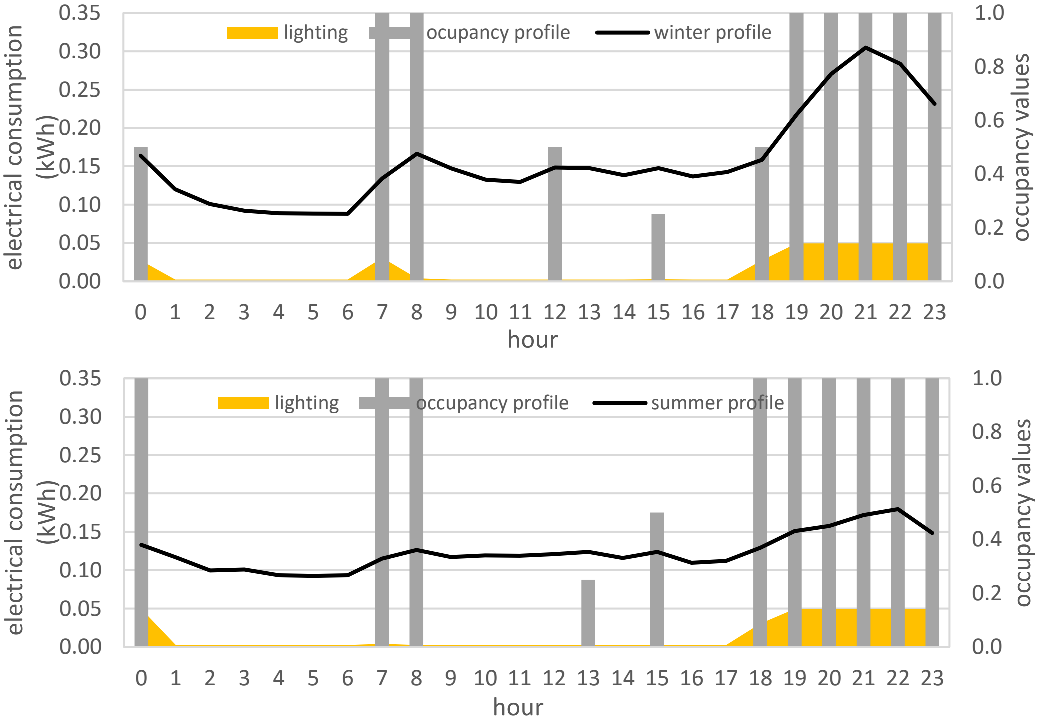


Figure 51 –light profile for path 111 compared with total consumption and occupancy profile for winter (top) and summer (bottom)

4.6.3 Baseline profile

The baseline corresponds to the minimum value of electrical consumption and does not have a direct relation to the occupancy (typically occurs during inactive occupancy periods as for example sleeping). Baseline electricity consumption has two components: the necessary constant use of appliances such as fridges and freezers, and the losses due to appliances on standby. Standby losses include computers, televisions, cookers and other equipment consuming a reduced electrical power whilst in an unused state [38]. To calculate the maximum value for baseline consumption, $\mathrm{BL\_consumption_i}$, correspondent to the minimum value of C_i, where C_i is the consumption profile (4.37). The time of minimum consumption is also retrieved, $BLtime_i$.

$$BLconsumption_i = \min C_i(BLtime_i) \quad (4.37)$$

4.6.3.1 *Fridge*

The first assumption is that the fridge has a steady consumption with an assumed 10% increase during the occupied hours due to opening and closing of the doors in active occupancy periods.

From [56] one can conclude that the majority of this equipment in Portuguese dwellings is class A followed by B and A+. For the class A equipment, the energy efficiency index (EEI) is between 42 and 55 [77]. The EEI of the refrigeration equipment for domestic use is calculated by dividing the yearly consumption of the equipment by the normalized energetic consumption per year [78]:

$$EEI = \frac{AE_c}{SAE_C} \times 100 \quad (4.38)$$

The AE_c is the annual energy consumption of the household refrigerating appliance and SAE_C the standard annual energy consumption of the household refrigerating appliance. In [79], for a certain equivalent volume of the household refrigerating appliance, the SAE_C is calculated using (4.38)(4.37), and assumed as constant. By varying the value of EEI the annual energy consumption is obtained. The hourly consumption of the fridge is calculated as follows:

$$F_c(kwh/hour) = \frac{AE_c(kwh/year)}{365 \times 24} \quad (4.39)$$

To create the fridge consumption profile, $F_i = \{F_i(t), t \in \{00:00 + 01:00k, k = 0, \dots, 23\}\}$: the hourly consumption of the fridge was combined with the occupancy profile.

$$F_i(t) = (0{,}1 \times Oa_i(t) + 1) \times F_c \quad (4.40)$$

The value of F_c was obtained by varying the values of EEI to achieve a close value of electrical consumption typically attributed to refrigeration, going from 28%, value proposed in REMODECE [68], to 32%, value proposed for fridge and freezer in ADENE [69]. That is done by comparing the total fridge consumption with the total electrical consumption, represented by the sum of the consumption profile:

$$\%\ fridgeconsumption_i = \frac{\sum F_i}{\sum C_i} \quad (4.41)$$

It was also taken into consideration that 95% of the households have this equipment with efficiency between A++ to C (from [56]) so the EEI was considered to be between 22 to 95 (from [77]), which could lead to slightly smaller values than 28%. The value of F_c is calculated for the free-float month and the same value was assumed in the summer and winter months. Three fridge consumption profiles are obtained: $F_{free-float}$, F_{winter} and F_{summer}.

4.6.3.2 *Standby*

There are three function modes for every equipment: when it is on and performing its primary function; in standby mode, when the equipment is not operating but still consuming electricity; and off mode when the equipment is not performing any function (but still have some losses as it is connected to the main electric supply) [69]. From [69], in 2004, for a sample of Portuguese dwellings, the standby consumption represented 12% of the total annual consumption and it was expected to increase drastically due to the growth of the number of electric appliances in the households. To calculate the maximum value for standby, the baseline consumption value was combined with the fridge consumption and lighting consumption at $\mathrm{BLtime_i}$:

$$\mathrm{SBconsumption_i} = \mathrm{BLconsumption_i} - \mathrm{F_i(BLtime_i)}\text{-}\mathrm{L_i(\,BLtime_i)} \qquad (4.42)$$

If the maximum value for standby obtained in summer or winter is twice the value of the free-float, the final value is reduced by 25%, due to considering the hypothesis of using heating or cooling equipment in sleeping periods, which is also the period of maximum value of standby.

To create the standby consumption profile, $\mathrm{SB_i} = \{\mathrm{SB_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \ldots, 23\}\}$: the maximum value of standby is combined with total percentage of activities related to electrical consumption $\mathrm{T\%ac_i}$, that will be developed in the next chapter.

$$\mathrm{SB_i(t)} = \frac{(100 - \mathrm{T\%eca_i(t)})}{100} \times \mathrm{SBconsumption_i} \qquad (4.43)$$

The standby profile is calculated the same way for the three months, having in the end $\mathrm{SB_{free-float}}$, $\mathrm{SB_{winter}}$ and $\mathrm{SB_{summer}}$.

4.6.3.3 *Application to the case study*

Figure 52 illustrates the fridge and standby profiles for the free-float month of path 111. It is possible to observe that the fridge slightly varies when there is occupancy, to consider the doors opening. On the other hand, the standby decreases with the occupancy and the increase of some activities, since some appliance use are associated with those activities. From 19h to 23h the consumption of standby decreases, having a slight increase at 20h due to being presumed that the household members would be having dinner, which is an activity that does not require any equipment use.

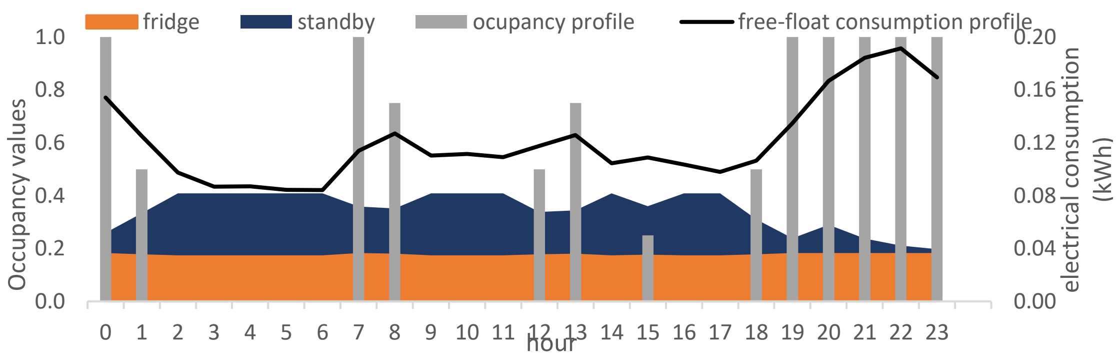


Figure 52 - fridge and standby consumption profile compared with total consumption profile and occupancy value, for the free-float month

The fridge and standby profiles for winter and summer months is obtained using the same method. The profiles are similar to the free-float month, being the only differences due to different occupancy profiles (presented alongside the remaining end-uses consumption in Figure 59).

4.6.4 Activity profiles

The objective of this step is to distribute the remaining values of electrical consumption, after removing the baseline and the lighting consumption, for the possible correspondent activity. To do so, values from a survey about the Use of time will be used [80]. Figure 53 illustrates the activities taken in different periods. Some activities do not reveal itself directly in the electric consumption of the household, like work and study, sleeping or having meals.

In this case study area, it was assumed that the hygiene was not related to the electrical consumption since gas is the typical energy used to heat up the water. This information was considered from [8], where the water heating equipment in a typical Portuguese dwelling is distributed as follows: 69% for gas heater, 12% for a gas boiler and 11%uses electric water heater. Further, the information from the energy performance certificates also supports this assumption [46].

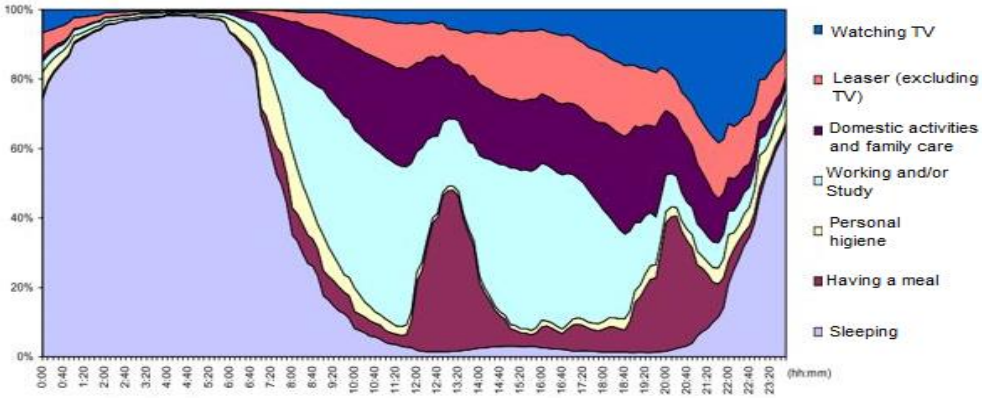


Figure 53 – Average profile of how Portuguese use their time [80]

From [80] , it is possible to withdraw the percentage values of the activities distribution that are related to the electrical consumption: leisure ($L\%_i$), watching TV ($TV\%_i$), and domestic activities ($DA\%_i$). Domestic activities will be segmented into cooking ($C\%_i$) by combining the time periods when the activity of having meals is higher than 20% and when 90% of the domestic activities are used to prepare those meals. At each hour, the total percentage of activities related to the electrical consumption, $T\%ac_i = \{T\%ac_i(t), t \in \{00:00 + 01:00k, k = 0, \dots, 23\}\}$, when the household is actively occupied, is calculated as follows:

$$T\%ac_{i_i}(t) = \begin{cases} L\%_i(t) + TV\%_i(t) + DA\%_i(t) + C\%_i, & \text{if } Oa_i(t) > 0 \\ 0, & \text{if } Oa_i(t) = 0 \end{cases} \quad (4.44)$$

For each activity, the ratio of consumption distributed for the activities is calculated as follows:

$$LR_i(t) = \begin{cases} \dfrac{L\%_i(t)}{T\%ac_i(t)}, & \text{if } Oa_i(t) > 0 \\ 0, & \text{if } Oa_i(t) = 0 \end{cases} \quad (4.45)$$

$$TVR_i(t) = \begin{cases} \dfrac{TV\%_i(t)}{T\%ac_i(t)}, & \text{if } Oa_i(t) > 0 \\ 0, & \text{if } Oa_i(t) = 0 \end{cases} \quad (4.46)$$

$$\mathrm{DAR}_i(t) = \begin{cases} \dfrac{\mathrm{DA\%}_i(t)}{\mathrm{T\%ac}_i(t)}, & \text{if } \mathrm{Oa}_i(t) > 0 \\ 0, & \text{if } \mathrm{Oa}_i(t) = 0 \end{cases} \tag{4.47}$$

$$\mathrm{CR}_i(t) = \begin{cases} \dfrac{\mathrm{C\%}_i(t)}{\mathrm{T\%ac}_i(t)}, & \text{if } \mathrm{Oa}_i(t) > 0 \\ 0, & \text{if } \mathrm{Oa}_i(t) = 0 \end{cases} \tag{4.48}$$

In the free-float month, the activity profiles, when the household is actively occupied, is directly estimated from the calculations described previously. However, for the winter months a percentage is used to report the heating use and in the summer months to report the cooling use for the corresponding heating and cooling profiles, developed in the next section. The method is first used to calculate the activities consumption profiles for the free-float month. After, it is assumed that if a household is performing the same activity, the values of consumption per activity would be the same throughout the different seasons. The method applied for the summer month is the same as the for the winter month, so the procedure will be exemplified with only one of the months.

4.6.4.1 *Free-float month*

For the free-float month, the total activity profile when the household is actively occupied will be referred as the equipment consumption profile, $\mathrm{Ec}_{\mathrm{freefloat}} = \{\mathrm{Ec}_{\mathrm{freefloat}}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$, and it is calculated as follows:

$$\mathrm{Ec}_{\mathrm{freefloat}}(t) = \\ = \begin{cases} \mathrm{C}_{\mathrm{freefloat}}(t) - \mathrm{SB}_{\mathrm{freefloat}}(t) - \mathrm{F}_{\mathrm{freefloat}}(t) - \mathrm{L}_{\mathrm{freefloat}}(t), & \text{if } \mathrm{Oa}_{\mathrm{freefloat}}(t) > 0 \\ 0, & \text{if } \mathrm{Oa}_{\mathrm{freefloat}}(t) = 0 \end{cases} \tag{4.49}$$

The activity profile when the household is not actively occupied will be referred as other consumption, $\mathrm{Otherc}_{\mathrm{freefloat}} = \{\mathrm{Otherc}_{\mathrm{freefloat}}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$, and it is calculated as follows:

$$\mathrm{Otherc}_{\mathrm{freefloat}}(t) = \\ = \begin{cases} \mathrm{C}_{\mathrm{freefloat}}(t) - \mathrm{SB}_{\mathrm{freefloat}}(t) - \mathrm{F}_{\mathrm{freefloat}}(t) - \mathrm{L}_{\mathrm{freefloat}}(t), & \text{if } \mathrm{Oa}_{\mathrm{freefloat}}(t) = 0 \\ 0, & \text{if } \mathrm{Oa}_{\mathrm{freefloat}}(t) > 0 \end{cases} \tag{4.50}$$

From the equipment profile, the profile per activity will be inferred. Firstly, for leisure consumption profile $\mathrm{Leisure}_{\mathrm{freefloat}} = \{\mathrm{Leisure}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$:

$$\mathrm{Leisure}_{\mathrm{freefloat}}(t) = \mathrm{LR}_{\mathrm{freefloat}}(t) \times \mathrm{Ec}_{\mathrm{freefloat}}(t) \tag{4.51}$$

For the TV consumption profile $\mathrm{TV}_{\mathrm{freefloat}} = \{\mathrm{TV}_{\mathrm{freefloat}}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$:

$$\mathrm{TV}_{\mathrm{freefloat}}(t) = \mathrm{TV}_{\mathrm{freefloat}}(t) \times \mathrm{Ec}_{\mathrm{freefloat}}(t) \tag{4.52}$$

For the domestic activity consumption profile $\mathrm{DA}_{\mathrm{freefloat}} = \{\mathrm{DA}_{\mathrm{freefloat}}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$:

$$\mathrm{DA}_{\mathrm{freefloat}} = \mathrm{DA}_{\mathrm{freefloat}}(t) \times \mathrm{Ec}_{\mathrm{freefloat}}(t) \tag{4.53}$$

For the cooking consumption profile $\mathrm{Cooking}_{\mathrm{freefloat}} = \{\mathrm{Cooking}_{\mathrm{freefloat}}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$:

$$\mathrm{Cooking}_{\mathrm{freefloat}} = \mathrm{Cooking}_{\mathrm{freefloat}}(t) \times \mathrm{Ec}_{\mathrm{freefloat}}(t) \tag{4.54}$$

The sleeping profile is an activity that does not imply electric consumption and it is correspondent to the inactive occupancy profile, $\mathrm{Sleeping_i} = \{\mathrm{Sleeping_i(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$ as is calculated as follows:

$$\mathrm{Sleeping_i} = \mathrm{Oina_i} \qquad (4.55)$$

The sleeping profile is calculated in the same way for the winter month and the summer month.

4.6.4.2 *Winter or summer month*

In the winter and summer, the total equipment consumption profile is divided into activity related consumption and heating or cooling consumption. First the activity related consumption is determined in two steps. The first step is the same as for the free-float month, so the equipment consumption profile, $\mathrm{Ec_{winter}} = \{\mathrm{Ec_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$, when the household is actively occupied, is calculated as follows:

$$\mathrm{Ec_{winter}(t)} = \begin{cases} \mathrm{C_{winter}(t)} - \mathrm{SB_{winter}(t)} - \mathrm{F_{winter}(t)} - \mathrm{L_{winter}(t)}, & \text{if } \mathrm{Oa_{winter}(t)} > 0 \\ 0, & \text{if } \mathrm{Oa_{winter}(t)} = 0 \end{cases} \qquad (4.56)$$

The second step is to compare the values consumed into the different activities in the free-float month with the possible values for the winter. The possible values for $\mathrm{Otherc_{winter}}$, $\mathrm{Leaser_{winter}}$, $\mathrm{TV_{winter}}$, $\mathrm{DA_{winter}}$ and $\mathrm{Cooking_{winter}}$ are all calculated as in the free-float month and compared as follows:

For the leaser consumption profile $\mathrm{FLeisure_{winter}} = \{\mathrm{FLeisure_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$:

$$\mathrm{FLeisure_{winter}(t)} = \begin{cases} \mathrm{Leisure_{winter}(t)}, & \text{if } \mathrm{Leisure_{winter}(t)} \ll \mathrm{Leisure_{freefloat}(t)} \\ \mathrm{Leisure_{freefloat}}, & \text{otherwise} \end{cases} \qquad (4.57)$$

For the TV consumption profile $\mathrm{FTV_{winter}} = \{\mathrm{FTV_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$:

$$\mathrm{FTV_{winter}(t)} = \begin{cases} \mathrm{TV_{winter}(t)}, & \text{if } \mathrm{TV_{winter}(t)} \ll \mathrm{TV_{freefloat}(t)} \\ \mathrm{TV_{freefloat}}, & \text{otherwise} \end{cases} \qquad (4.58)$$

For the domestic activity consumption profile $\mathrm{FDA_{winter}} = \{\mathrm{FDA_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$:

$$\mathrm{FDA_{winter}(t)} = \begin{cases} \mathrm{DA_{winter}(t)}, & \text{if } \mathrm{DA_{winter}(t)} \ll \mathrm{DA_{freefloat}(t)} \\ \mathrm{DA_{freefloat}}, & \text{otherwise} \end{cases} \qquad (4.59)$$

For the cooking consumption profile $\mathrm{FCooking_{winter}} = \{\mathrm{FCooking_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$:

$$\mathrm{FCooking_{winter}(t)} = \begin{cases} \mathrm{Cooking_{winter}(t)}, & \text{if } \mathrm{Cooking_{winter}(t)} \ll \mathrm{Cooking_{freefloat}(t)} \\ \mathrm{Cooking_{freefloat}}, & \text{otherwise} \end{cases} \qquad (4.60)$$

To make sure that the heating equipment use could occur when the occupants are sleeping, the other consumption profile $\mathrm{F}Otherc_{\mathrm{winter}} = \{\mathrm{Fotherc_{winter}(t)}, \mathrm{t} \in \{00{:}00 + 01{:}00\mathrm{k}, \mathrm{k} = 0, \dots, 23\}\}$ is calculated as follows:

$$\mathrm{FOtherc}_{\mathrm{winter}}(t) =$$

$$= \begin{cases} \mathrm{Otherc}_{\mathrm{freefloat}}, & \text{if } (\mathrm{Oina}_{\mathrm{winter}}(t) > 0) \wedge (\mathrm{Otherc}_{\mathrm{freefloat}} < \mathrm{Otherc}_{\mathrm{winter}}) \\ \mathrm{Otherc}_{\mathrm{winter}}, & \text{if } (\mathrm{Oina}_{\mathrm{winter}}(t) > 0) \wedge (\mathrm{Otherc}_{\mathrm{freefloat}} > \mathrm{Otherc}_{\mathrm{winter}}) \\ \mathrm{Otherc}_{\mathrm{winter}}, & \text{if } (\mathrm{Oa}_{\mathrm{winter}}(t) = 0) \wedge (\mathrm{Oina}_{\mathrm{winter}}(t) = 0) \end{cases} \qquad (4.61)$$

4.6.4.3 *Application to the case study*

The activity distribution for the free-float month for path 111 is presented in *Figure 54* and it is obtained by combining the active occupancy schedule from *Figure 45* and the time of use from *Figure 53.*

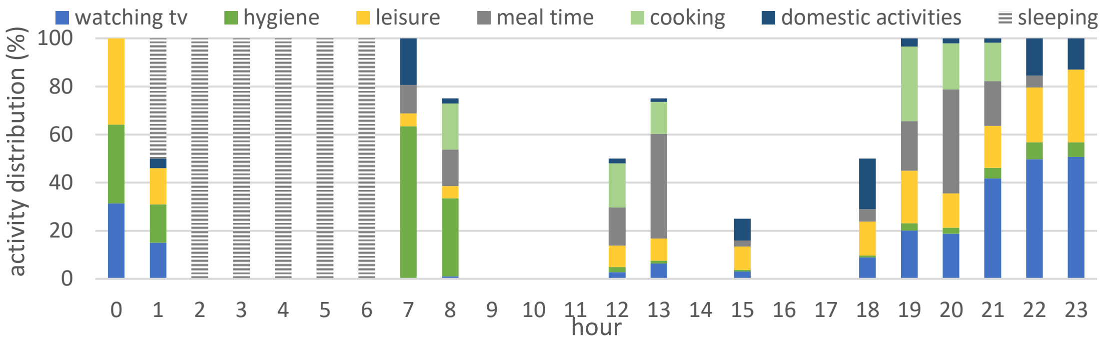


Figure 54 – free-float activity distribution in the household

In previous section, the activities from Figure 54 that reveal themselves in consumption are stipulated and the correspondent consumption is presented in Figure 55. As it is visible the cooking related consumption coincides with typical meal times, and watching TV occurs mostly from 20h to 0h, being the most consuming activity during this period.

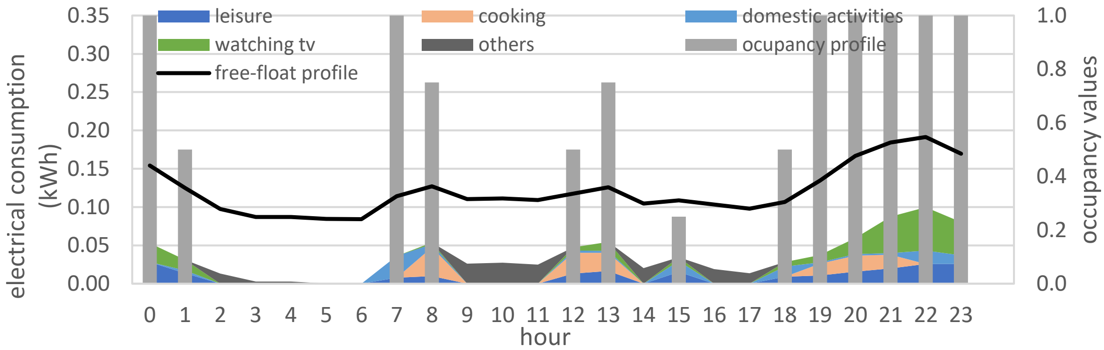
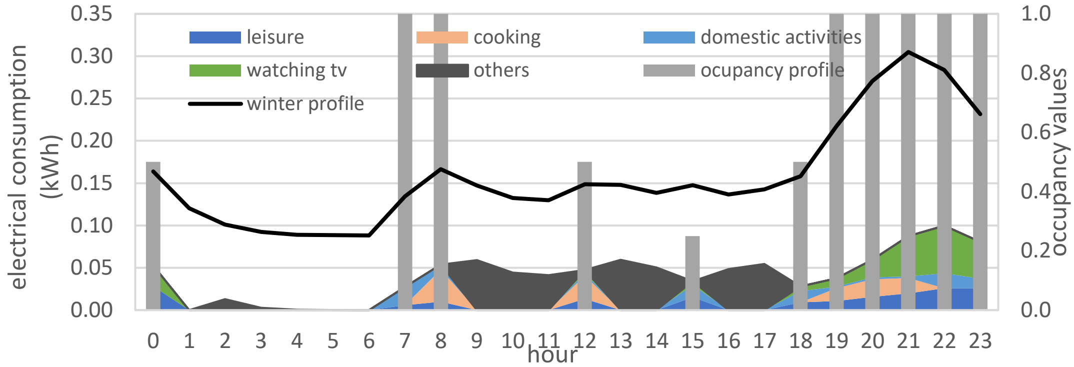


Figure 55 – Activity profiles compared with total consumption profile and occupancy values for the free-float month

The winter and the summer activity profiles are presented in Figure 56. From 19h to 0h the activity profiles are similar in the three months. Between 10 h to 17 h the activity profiles differ, especially in the winter due to the increase of other activities consumption.



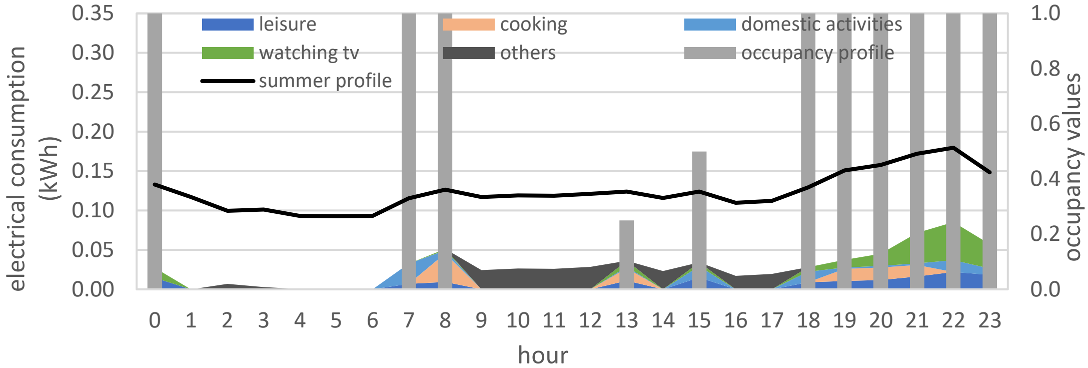


Figure 56 – Activity profiles compared with total consumption profile and occupancy values for the winter month (upper) and summer month (lower)

4.6.5 Cooling and heating profile

The method used in this thesis to determine electric space heating and cooling was based on [32]. This method consists in subtracting the average load on the mild season months from the average load on the winter month load to infer about the electric space heating. The same procedure is applied to infer about the electric space cooling. However, to obtain the heating and cooling consumption, it is not simple to compare the difference in the hourly values of the total consumption, due to the difference in the lighting needs and schedules throughout the seasons. The first assumption in this thesis, is that the free-float month does not have any electrical consumption for space heating or cooling. It was also assumed that the households would only be heated or cooled if they were occupied (active or inactive occupancy). As the method to achieve the heating profile and the cooling profile is the same, the heating profile will be developed to exemplify it.

To determine the heating consumption profile $\mathrm{Heating} = \{\mathrm{Heating}(t), t \in \{00{:}00 + 01{:}00k, k = 0, \dots, 23\}\}$:

$$\mathrm{Heating}(t) = = \mathrm{Ec}_{\mathrm{winter}}(t) - \mathrm{FLeaser}_{\mathrm{winter}}(t) - \mathrm{FTV}_{\mathrm{winter}}(t) - \mathrm{FDA}_{\mathrm{winter}}(t) - \mathrm{FCooking}_{\mathrm{winter}}(t) - \mathrm{FOtherc}_{\mathrm{winter}}(t) \quad (4.62)$$

Application to the case study

The procedure described previously was applied to the winter and cooling reference months. Figure 57 illustrate the resulting heating and cooling profiles considered in this thesis.

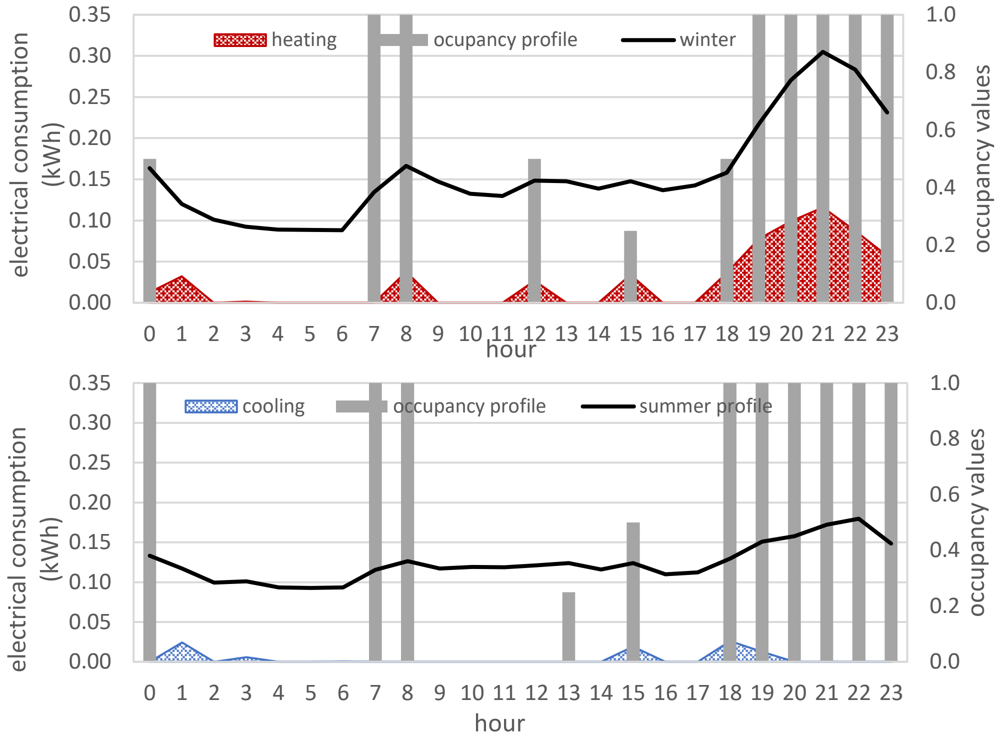


Figure 57 – Heating (upper) and Cooling (lower) profile compared with total consumption profile and occupancy values

4.6.6 Consumption profile per end-use

This section presents the resulting consumption profiles, per end-use, for path 111, taking into consideration the methodologies and resulting profiles presented in the previous sections. From Figure 58, all the activities that are related to the electrical consumption are considered, for the free-float month. The consumption profile is therefore separated into fridge consumption, standby and lighting. After subtracting those consumptions from the consumption profile, the remaining consumption will be distributed into the activities of leaser, watching TV, cooking, domestic activities and other activities.

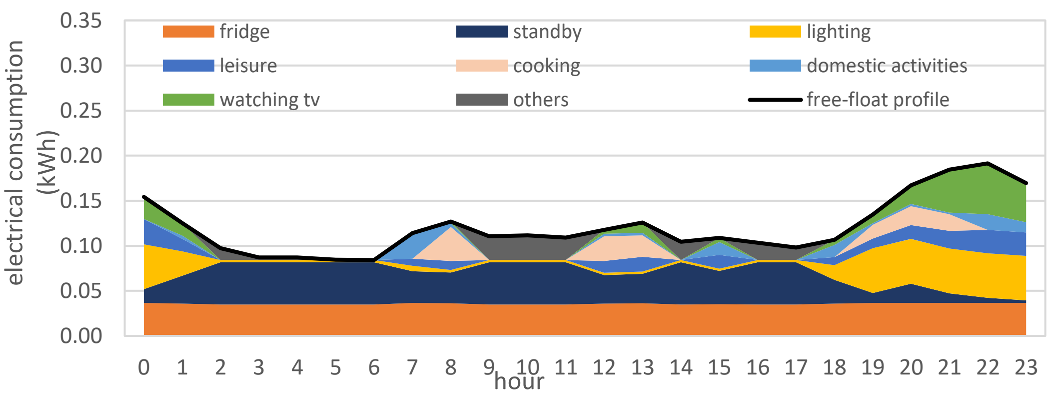
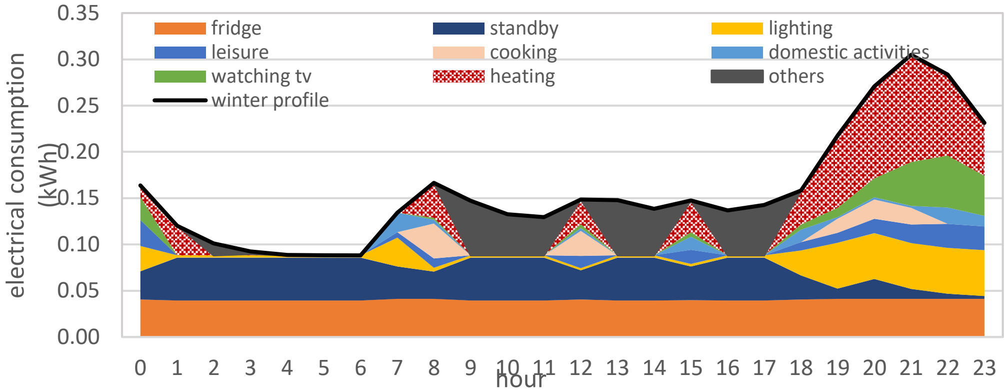


Figure 58- electrical consumption distributed by its end-use for the free-float month

The method is different when it comes to the winter month and the summer month, as it was explained previously, due to the addition of cooling or heating. From the Figure 59, it is visible the addition of heating and of cooling. From the figures presented, it is also possible to identify a big presence of other consumptions, which could be a fragility of the model since it is not related to occupancy and is not standby or fridge, it could be related to some fragility in the identification of occupancy profile or the fact that some equipment's could be left on while the household is not occupied like washing machines which was not taken into consideration by this model or that the assumption of not considering electric water heaters could not be accurate in all the scenarios.



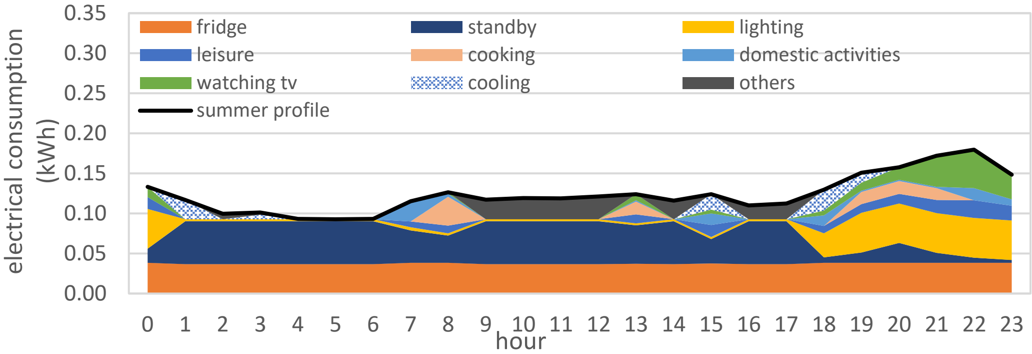


Figure 59- electrical consumption distributed by its end-use for the winter month (upper) summer month (lower)

The end-use consumption profiles for the remaining paths developed in this thesis are presented in Annex E

4.7 Parametric building energy model

Building energy simulation (BES) models are used to forecast the behavior of a complex system given a set of well-defined laws (e.g., energy balance, mass balance, conductivity, heat transfer, etc.). Building energy simulation tools allow the exhaustive calculation of the energy required to maintain specified building performance criteria, under the influence of external inputs such as weather, occupancy and infiltration. Detailed heat-balance calculations are carried out at discrete time-steps based on the physical properties of the building and mechanical systems [67]. These calculations are generally performed over the course of a full year. *EnergyPlus* is a whole building energy simulation tool used to model the energy consumption for heating, cooling, ventilation, lighting and plug and process loads [81], and it was used in this thesis as the simulation engine.

The energy simulation model is created with the input data obtained from the mathematical model developed in Section 4.6. However, other inputs such as geometry, construction characteristics, internal gains, zone airflow and HVAC systems have to be considered in order to create a complete building energy model. These input parameters used in this thesis for the case study are described in detail in the following sections.

4.7.1 Building geometry

The first step is to create a 3D model. Based on the building typology presented in the Case study section, a typical dwelling was modeled using AutoCAD [82] and Sketchup [83]. Figure 60 illustrates the 3D model used in this work.

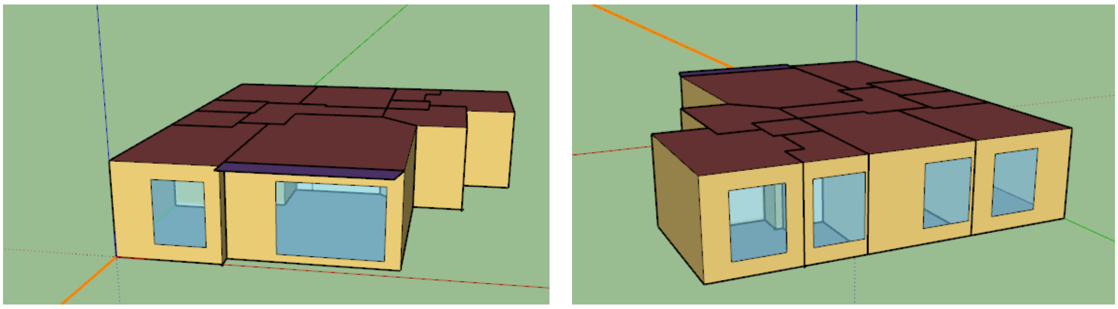


Figure 60- 3D facing north (left) and 3D facing south (south)

The *boundary* conditions are defined as adiabatic for roof and ceiling (since this is in an intermediate floor), as well as the common walls in contact to other buildings or dwellings. The shadow for the surroundings and from the upper floor balcony was modeled as showed in Figure 61.

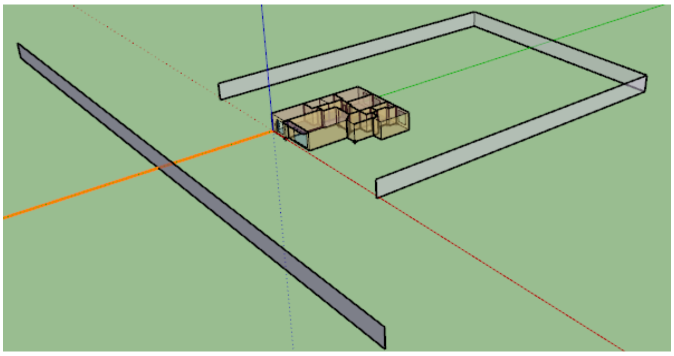


Figure 61 - shading surfaces

From there, the *OpenStudio* plugin for *Sketchup* was used to convert the geometry file into an intermediate data format (.idf) file to be used in *EnergyPLus* [81]. The floorplan area of the reference dwelling is 148.5 m^2.

4.7.2 Building envelope

In the *EnergyPLus,* the *.idf* file must be completed with information on the construction elements. The material's information was completed with the already presented in Section 3.2 for the individual materials of the exteriors and interior walls and for windows. Concerning the shading devices, a metal blind was considered, with slat width of 0.025 m and separation of 0.019 m, conductivity of 45 W/(m.ºC) and beam solar reflectance of 0.8 and 0.0 slat beam solar transmittance, from [84]. In Table 5 a comparison between the thermal transmittance values from the case study inputs and the model are presented. One can observe that the differences are insignificant.

Table 5 – Comparison between thermal transmittance from Case study and Modeled

	Case Study Total Thermal Transmittance (W/(m^2.ºC))	Modeled Total Thermal Transmittance (W/(m^2.ºC))
Exterior wall	0.54	0.539
Interior wall	1.16	1.13
Window	2.9	2.9

4.7.3 Internal gains and schedules

The internal gains are a combination of *People, Lights* and *Electric Equipment.*

The input for people is the combination of the number of people, the activity level, the occupied zones and the occupancy schedule. For the number of occupants, it was assumed that the minimum number of people per household is two and the maximum is five, based on the information of the household size. Therefore, to assume a number of occupants the maximum value of the hourly consumption profile (from the free-float month) was used and it was compared with the value obtained by dividing the average yearly consumption obtained in section 4.3 and dividing it by 2.8 which is the average dwelling occupancy for the case study area [44], using the following equation:

$$power\ per\ occupant = \frac{2474.2\ kWh/(year.dwelling)}{365 \times 24 \times 2.8\ (occupants/dweeling)} = 100.9\ W \qquad (4.63)$$

Further, the number of people dictate the number of occupied rooms as: If 2 occupants, ROOM1 is occupied; if 3 occupants, ROOM1 plus ROOM2 is occupied; if 4 or more occupants, ROOM1 plus ROOM2 plus ROOM3 is occupied. The occupancy schedules were separated into active and inactive occupancy, being the first distributed through the entire floorplan area and with an activity level of 100 W/person [65] and the second attributed to the rooms with an activity level of 40 W/person [65].

The active occupancy schedule is the combination of the previous obtained active occupancy profile. Each occupancy profile is inserted as a daily schedule, meaning that per *path* it will be 6 daily schedules of occupancy: one for the weekdays and another for the weekend, per each selected month. Then a weekly schedule was created by combining the weekdays and the weekend profiles per each month. Finally, all is combined in a yearly schedule of active occupancy.

For the active occupancy: From 1 of March to 31 of April and from 1 of September to 30 of November it was assumed a free-float occupancy profile with $Oa_{freefloath\_weekdays}$ from Monday to Friday and $Oa_{freefloath\_weekends}$ for Saturday and Sunday; From 1 of December to 28 of February it was assumed a winter occupancy profile with $Oa_{winter\_weekdays}$ from Monday to Friday and $Oa_{winter\_weekends}$ for Saturday and Sunday; From 1 of June to 31 of August it was assumed a summer occupancy profile with

$Oa_{summer\_weekdays}$ from Monday to Friday and $Oa_{summer\_weekends}$ for Saturday and Sunday. The same method was applied to inactive occupancy schedule, this time using $Oina_i$ profiles.

The installed units, $Ilight\_units_i$ calculated in section n4.6.2, were distributed throughout the entire house (Watts/m2) and the light schedule for the entire year was done with the same method as described in the occupancy schedule, using the installed units to create a ratio from lighting profile, previously developed in the methodology.

Due to the complexity of the schedules and the number of different activities it was necessary to simplify. Therefore, the fridge consumption was assumed constant, equal to Fc_i, and it was allocated to the kitchen. All the activity profiles were added into what will be called as electric equipment consumption, where the maximum value for the three months was obtained and distributed throughout the house (Watts/m^2). The maximum value for standby consumption for the three months was also obtained and distributed throughout the house (Watts/m^2). From the maximum value of equipment consumption and maximum value of standby consumption, a schedule of fraction of the peak value was created per weekdays and weekends in each month, for standby and for electric equipment. The equipment and standby schedule for the entire year was done with the same method as described in occupancy schedule.

As an example, the path 111 has the equipment design level of 99,8 W (0,97 W/m2) and a standby power of 53,54 W (0,52 W/m2). The equipment and the standby schedules are presented in Figure 62. Regarding the internal gains, 2 occupants were considered (0.02 pp/m2), a fridge consumption of 35,5 W (2,4 W/m2), and a maximum light level of 49,5 W (0,48 W/m2).

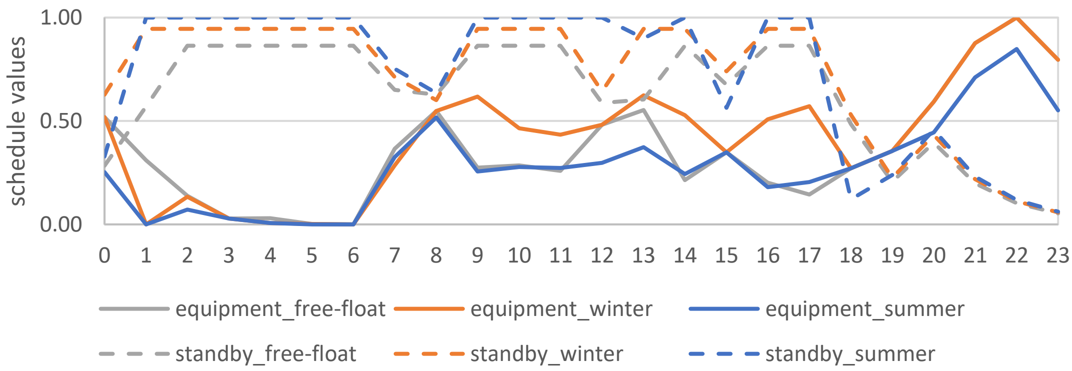


Figure 62 - equipment and standby schedule

4.7.4 Air flow and window operation

The typical value of 0.62 air changes per hour (ach) for ventilation as presented in 3.2 was considered, including infiltration and natural ventilation from opening the windows. It was assumed a value of 0.4 ach for infiltration and the remaining was modeled as natural ventilation, with the correspondent opening area of 50%. It was assumed the following rules for window interaction, if any of the rule:

- Interior temperature below 21ºC or above 25ºC the windows are closed or minimum outside temperature higher than the used setpoint for heating or maximum outside temperature bellow the cooling setpoint to open the windows;
- For the shading control, it was set that the blinds would be on if the zone air temperature is higher than 22ºC;
- Windows opening and shading devices are conditioned to the occupancy schedule;

4.7.5 Zone conditioning

The HVAC was modeled as an ideal load air system with an initial cooling setpoint of 25ºC and a heating setpoint of 19ºC. These setpoints were based on Figure 63. For an outside mean temperature in January 2017 of 8.3ºC [85] the comfort temperature can be set between 16ºC and 22ºC. In July 2016 the outside mean temperature was 24.3ºC [85], and therefore the comfort temperature can be set between 22ºC to 28ºC.

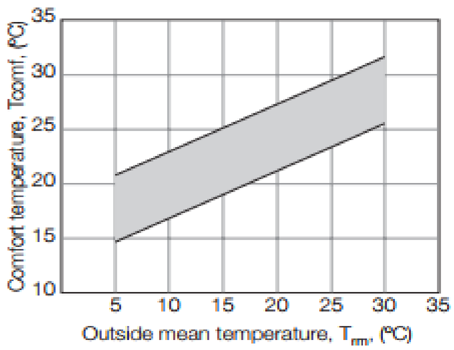


Figure 63 – Adaptive Thermal Comfort for Buildings in Portugal [86].

The areas assumed to be heated or cooled were the living room and the bedrooms. Further, the HVAC equipment only works if these rooms are occupied. The SCOP for heating and the SEER for cooling systems used are 2.6 and 2.5, respectively, as presented previously in Section 3.2

4.7.6 Climate

The weather file used in this model was obtained for Lisbon from [87]. The sun and wind exposure were also defined as no sun and no wind for all the building surfaces except outdoor walls and windows.

4.8 Model evaluation

4.8.1 Model calibration

For BES models to be used with any degree of confidence, it is necessary that the existing model closely represent the actual behavior of the building under study. This can be achieved through model calibration, where the goal is to reduce the discrepancies between BES prediction and measured building performance. During the calibration two main sets of data are needed: the simulation dataset, from the building model created, and the measured dataset, from the paths profiles.

As is showed in Figure 64, after performing the building energy simulation with the schedules determined by the mathematical model, the total consumption profiles from BES will be compared with the measured paths consumption profiles. Further, different energy systems and distinct setpoints will be tested until there is an agreement between the simulated and the measured consumptions.

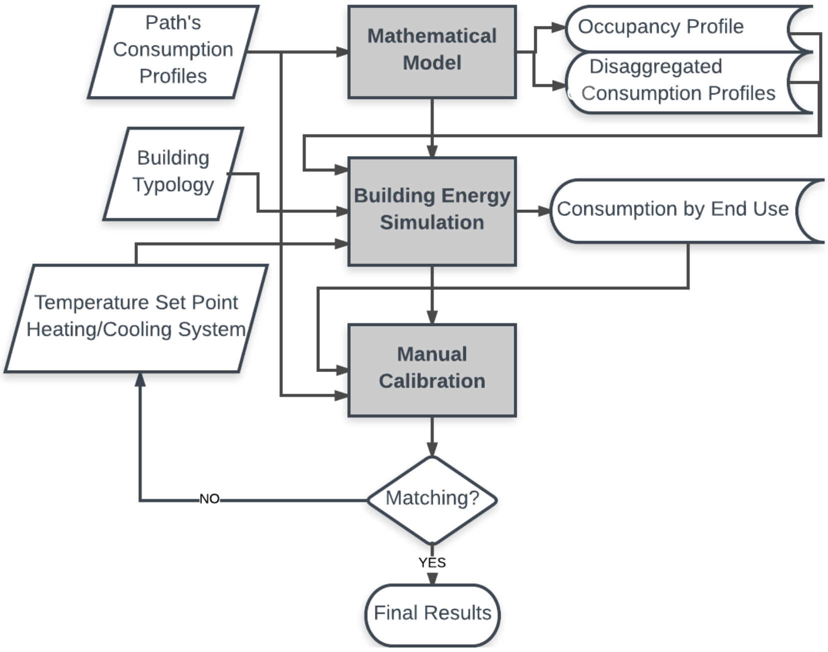


Figure 64 - Process for model validation

The validation of building energy simulation models is presently built on a model's compliance with standard criteria for Mean Bias Error (MBE) and the Coefficient of variation of the Root Mean Square Error (Cv(RMSE)). Commonly the Cv(RMSE) and MBE are the two statistical indices used. The consideration of both indices allows preventing any calibration error due to errors compensation. The MBE measures how closely simulated data is from the measured data. Due to a compensation effect (positive and negative values contribute to reduce the MBE final value), the MBE usually is not a "stand-alone" index, but it is assessed together with the Cv(RMSE). The Root Mean Squared Error (RMSE) is a measure of the sample deviation of the differences between the measured values and the values predicted by the model. The CV(RMSE) (in %) allows one to determine how well a model fits the data by capturing offsetting errors between measured and simulated data. It does not suffer from the cancellation effect. Lower Cv(RMSE) values bring to better calibration [88]. These criteria vary dependent on whether models are calibrated to monthly or hourly measured data, and are based on standard statistical indices. Currently, building energy simulation models are generally considered ‘calibrated’ if they meet the criteria set out by ASHRAE Guideline 14 [89]. These criteria propose the use of two statistical indices previously described, and are used to represent how well a mathematical model describes the variability in measured data. These indicators are calculated as follows:

Coefficient of variation of the root mean square error (CV(RMSE)):

$$CVRMSE = 100 \times \sqrt{\frac{\sum_{i=1}^{Np}(m_i - s_i)^2}{(Np-1)}} \times \frac{1}{\bar{m}} \tag{4.64}$$

Normalized mean bias error (NMBE):

$$NMBE = 100 \times \frac{\sum_{i=1}^{Np}(m_i - s_i)}{(Np-1)} \times \frac{1}{\bar{m}} \tag{4.65}$$

Where $\bar{m}$ is the average of the measured data points, $m_i$and s_i are the respective measured and simulated data points for each model instance and *Np* is the number of data points at interval *p*, and $\bar{m}$ is the average of the measured data points.

Depending on the time interval for the calibration (monthly or hourly) and in compliance with the requirements of the protocol considered, the limit threshold is subjected to slight differences, as reported in *Table 6*.

Table 6- Acceptance criteria for calibration of BES models [89]

Standard/guideline	NMBE	CV(RMSE)
Monthly criteria (%)	5	15
Hourly criteria (%)	10	30

This means that once there is reasonable agreement between measured and simulated data, the model may be considered ‘calibrated’ according to present acceptance criteria for BES models. However, the model that meets these criteria is not unique and thus there are numerous models of the same building that can be ‘calibrated’ [67].

The manual calibration was performed by assuming as a first instance for each path the SCOP, SEER and constant temperature setpoints presented in 4.7.5. Then, if the simulated profile is slightly higher or smaller than the measured in some periods, the heating and cooling setpoint is assumed to be different from the initial assumption and that the setpoint could vary throughout the day, having as minimum heating setpoint equal to 18.5°C and the maximum cooling setpoint equal to 25.5 °C. If there is still no a good match, improved systems efficiencies are tested until SCOP and SEER equal to 3.5. If after performing these sequence, the simulated profile is still higher than the measured profile, then it is assumed that there is no use of electrical equipment even though there is cooling or heating needs.

4.8.2 Model uncertainty

To address the topic of the model calibration it is important also to consider the issue of the model uncertainty, particularly for indeterminate models of complex systems. Models of complex systems are notoriously difficult to validate and have been the subject of much scientific discussion and debate in terms of quality and uncertainty, essentially due to simplifications, assumptions and simulation constraints. This is particularly true when the purpose of the model is to deliver some insight into the non-observable parts of the system, as the study taken under this thesis.

The built environment presents a complex challenge in terms of energy modelling and accurate prediction. Any given building is characterized by a multiplicity of parameters including materials properties, occupancy levels, equipment schedules, HVAC and other heating systems, and climate and weather. These represent diverse sources of model parameters uncertainty. However, this does not illustrate the entire range of potential uncertainty encapsulated by any given building model [89].

From the specific proposed method in this thesis, the major factors of uncertainty are due to the measured data from the paths, since these are determined from the cluster centroids. From this first factor, the peaks of consumption and some specific information could be lost, as cluster centroids are a representation of average daily consumption profile. From here the uncertainty increases, as the paths are used as the major input for the mathematical model, creating occupancy profiles which don not have actual measured information to validate the model. From having the occupancy profile, a *light profile, fridge* and *standby profile, activity profiles, heating* and *cooling profiles* are assumed.

Moreover, from the path analysis, another source of uncertainty identified is related to the summer month, since it is a vacation period and it is possible that in some scenarios the household was not occupied (part or the full month). Also, some occupants can be in 'inhouse' vacations which was not also considered by the mathematical model when it compares activity related consumption with the free-float month.

Other sources of uncertainty are related to the light units and the fridge power since this data was not available for the case study and therefore had to be based on assumptions from literature review.

5 Results and discussion

The profiles developed in the mathematical model presented in section 4.6 are used as input for the building energy simulation described in section 4.7, for the three reference months determined in section 4.5.1. The model evaluation and calibration is taken considering the indexes described in section 4.8 and the results are presented and discussed hereafter, for each of the five paths identified in section 4.5.4.

Two variables were used in the calibration; the type of system used for cooling and heating; and the temperature setpoint. The procedure starts with the information available in section 4.7.5 and it is calibrated in accordance with the measured data. Nonetheless, as stated before, the compliance with the thresholds can be attained with different models, as the solution is not exclusive and may not guarantee that the system assumed and temperatures setpoints are the ones actually used.

5.1 Path 111

Path 111 represents 19.75% out of the total households analyzed. The fixed inputs were presented in section 4.7.3 and end-use profiles in section 4.6.6. The calibrated results were achieved for the following inputs:

- the heating setpoint of 19.5°C for the bedroom and for the living room 19°C from 0 h to 9 h and 20°C for the remaining day;
- the cooling setpoint of 24 °C for the bedroom and for the living room 24 °C from 0h00 to 14h00 and 25°C from then on;
- A SCOP of 2.6 and SEER of 2.5.

Path 111 evaluation results, after performing the calibration, are presented in Table 7. It can be observed that the model is in compliance with the reference indexes from section 4.8.1. The free-float results reported very low values of both indexes most likely due to not having cooling or heating consumption. The winter has the highest value of CV(RMSE) but is still below the limit threshold, that can be explained by the heating consumption not having the same periods of occurrence as it was recognized in the model. The summer results have lower values of CV(RMSE) then the winter, which indicates that the deviation between the measured values and the simulated values in the summer is smaller than in the winter. The summer NMBE results reported a lower value than the winter value, since it NMBE has a compensation effect that reduces its final value due to sum of the positive and negative differences between measured and simulated values.

Table 7- Validation results for free-float, winter and summer for path 111

	CV(RMSE)(%)	NMBE(%)
Hourly limit threshold [89]	30	10
Free-float	0.8	0.1
Winter	11.7	0.8
Summer	8.3	0.1

When comparing the simulated and measured daily profiles (Figure 65) the profiles match. There is only 0,08% of difference when comparing the total daily consumption from measured and simulated, which indicates a good approximation. Since the simulation did not identified cooling or heating consumption in the month of May, the two profiles match because the consumption is a direct output of light, fridge, standby and electric equipment consumption.

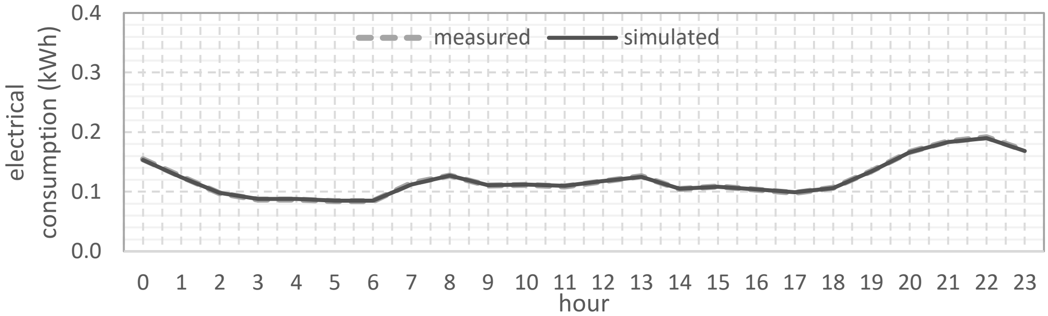


Figure 65 - Comparison between simulated (May) and measured (free-float) average daily consumption profile for path 111

Regarding the winter month, the results are presented in Figure 66. Even though the validation criteria were met, the curves do not exactly fit (35% of difference in the worst case, which occurs at 7h, being the simulated value 0.047 kW higher than the measured value). It is also important to state that the total daily consumptions from measured and simulated only differ by 0.79%. However, it is possible to observe that some difference exists, during the night period (from 19h to 22h), where the measured consumption is higher than the simulated values, the opposite occurs between 2h to 8h. Further, from the end-use disaggregation (Figure 67) it is possible to infer that this difference is most influenced by the heating consumption. As it was stated in section 4.8.1 there are numerous systems and temperature setpoints that can meet the calibration criteria. A further breakdown of the heating setpoint could improve the fit between both curves. Other variables may also be introducing uncertainty in the model, as for example building construction characteristics such as materials, infiltration, etc.), affecting the heating consumption predictions. In a future improvement of the model evaluation, a calibration model considering a wider range of variables can be considered.

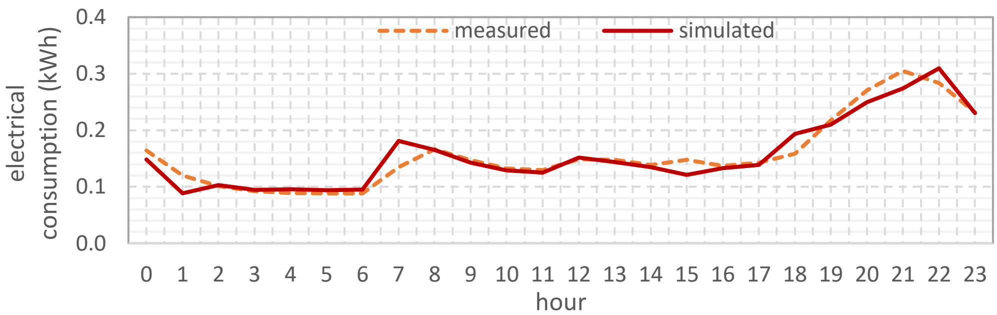


Figure 66 - comparison between simulated average daily consumption for January with winter daily consumption profile from path 111

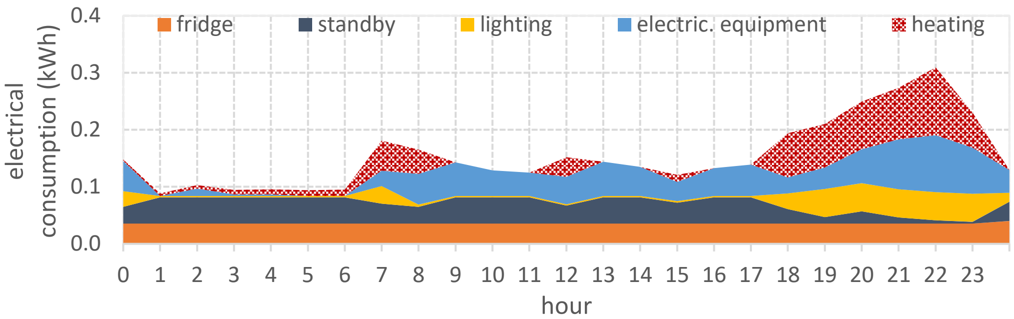


Figure 67 –End-use consumption results from simulation for January for path 111

The criteria were also met for the summer month. From Figure 68, it is possible to observe that the fitting between the simulated and the measured curve presents some diversity (26,06% of difference in the worst case, which occurs at 18h, being the simulated value 0.03 kWh higher than the measured value). From the end-use disaggregation (Figure 69) it is possible to infer that this difference is most influenced by the cooling consumption. In fact, when comparing this simulation result with the winter month, during the night period, an opposite effect is observed: for the winter month, the model underestimates the heating consumption while for the summer month it is overestimating. This opposite effect could be explained by a building insulation value assumption (in walls, windows or infiltration, for example) higher than the real one. This reinforces that, besides the comfort setpoints and systems efficiencies, the calibration should also consider other uncertain parameters related to the construction characteristics. Nonetheless, one cannot exclude the possibility that this occupant chooses to not to use the cooling or heating system, even if he/she was in a discomfort situation for a certain period.

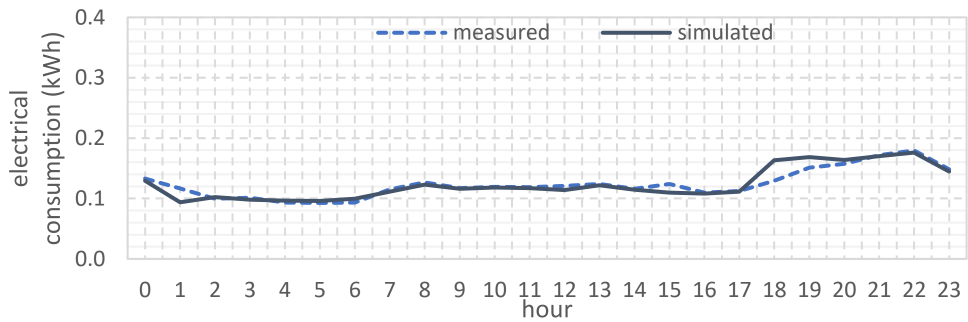


Figure 68 - Comparison between simulated average daily consumption for July with summer daily consumption profile from path 111

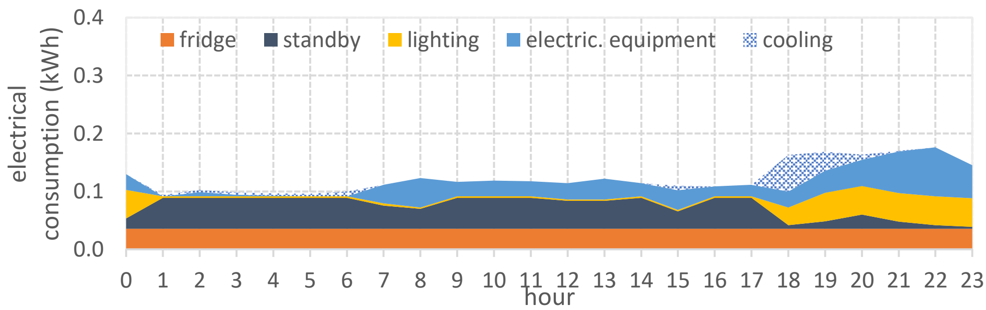


Figure 69 – End-use consumption from simulation for July for path 111

A fragility in the mathematical model related to the occupancy profile identification recognized, since ‘other activities’ refer to consumption that is not related to the household occupancy but is also not identified as fridge or standby consumption. An example using the free-float month is presented in Figure 70 where the 'other activities' consumption was not included in the total equipment consumption simulation, a difference is identified when comparing the fit with the fit obtained in Figure 65. This difference is caused since the model did not identify any active occupancy during these periods, and therefore it only considered the fridge and standby electrical consumption. Although the error is small enough to also consider the simulation valid, this highlights two improvements that must be studied in a future model:

- The active occupancy model sensitivity should be improved to identify smaller values of the threshold of turning on event and high consumption period. The current limit threshold is 0,25 for both scenarios and the daily consumption profile used as input is based on the average of all weekdays of each month. Therefore, an event that occurs once a week or even once a month would not be identified. For example, having a housekeeper coming once a week which would increase the consumption, due to the use of equipment, but when making an average of the 5 days of the week that information would be lost and that occupancy will not be identified by the mathematical model;
- The existence of equipment consumption (rather than the fridge and standby) in unoccupied periods. For example, the washing machine or dishwasher is left on while the household is not occupied.

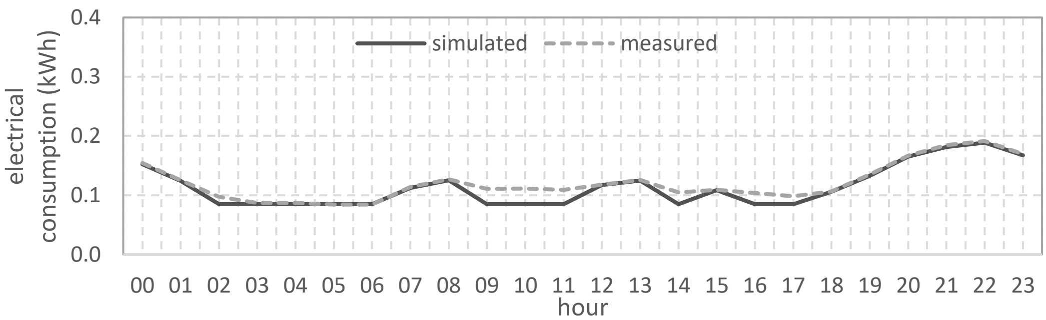


Figure 70 - Comparison between simulated (May) and measured (free-float) average daily consumption profile for path 111 (without the 'other activities' simulation)

5.2 Path 212

Path 212 represents 16.05% of the analyzed users. This household was assumed to be occupied by 4 persons, with an equipment design level of 190.8 W (1.28 W/m^2), a maximum standby power of 74.48 W (0.50 W/m^2), a fridge of 58.1 W (3.93 W/m^2) and a light level of 102.9 W (0.69 W/ m^2). The occupancy schedule and the input profiles from mathematical model are present in annex E.1. The calibrated results were achieved for the following inputs:

- Gas system for heating or no system;
- No cooling system is used.

The evaluation results after performing the calibration are presented in Table 8. It can be observed that the model is in compliance with the reference indexes presented in section 4.8.1. Similar values as the ones obtained in path 111 were achieved for the free-float month, while the winter month values of

CV(RMSE) are much lower (less 9.5%), due to not having electrical consumption related to heating. CV(RMSE) for the summer has the same value as path 111 but has higher values in terms of NMBE (2.5%).

Table 8- Validation results for free-float, winter and summer for path 212

	CV(RMSE) (%)	NMBE (%)
Hourly limit threshold [89]	30	10
Free-float	0.8	0.2
Winter	2.2	0.5
Summer	8.3	2.6

From Figure 71, it is also visible that the free-float and winter month have almost a perfect fit. For the worst case, for free-float the difference between measured and simulated is 1.64% which occurs at 7h while for the winter that difference increases to 6.20% which occurs at 20h. In the summer some differences are visible (between 10h to 12 h), having the biggest difference occurring at 11 h (23.44%), where the simulated value is 0.06 kWh less than the measured. These differences can be explained by the mathematical model having assumed 0.8% of the summer consumption for cooling (annex E.1) while no consumption for cooling was identified in the simulation after performing calibration, since the only match occurred for no use of cooling equipment (annex F.1). This overestimation can be also explained by uncertainties in the mathematical model related to uncommon or temporary occupancy patterns. For example, when comparing the measured free-float consumption profile with the measured summer profile there is a shift in the morning peak for later hours and longer period in the summer month. This can be an indicator of “inhouse” holidays, where the activity consumption shifts in that period. However, the mathematical model does not take that into consideration and assumes the increase is due to cooling use.

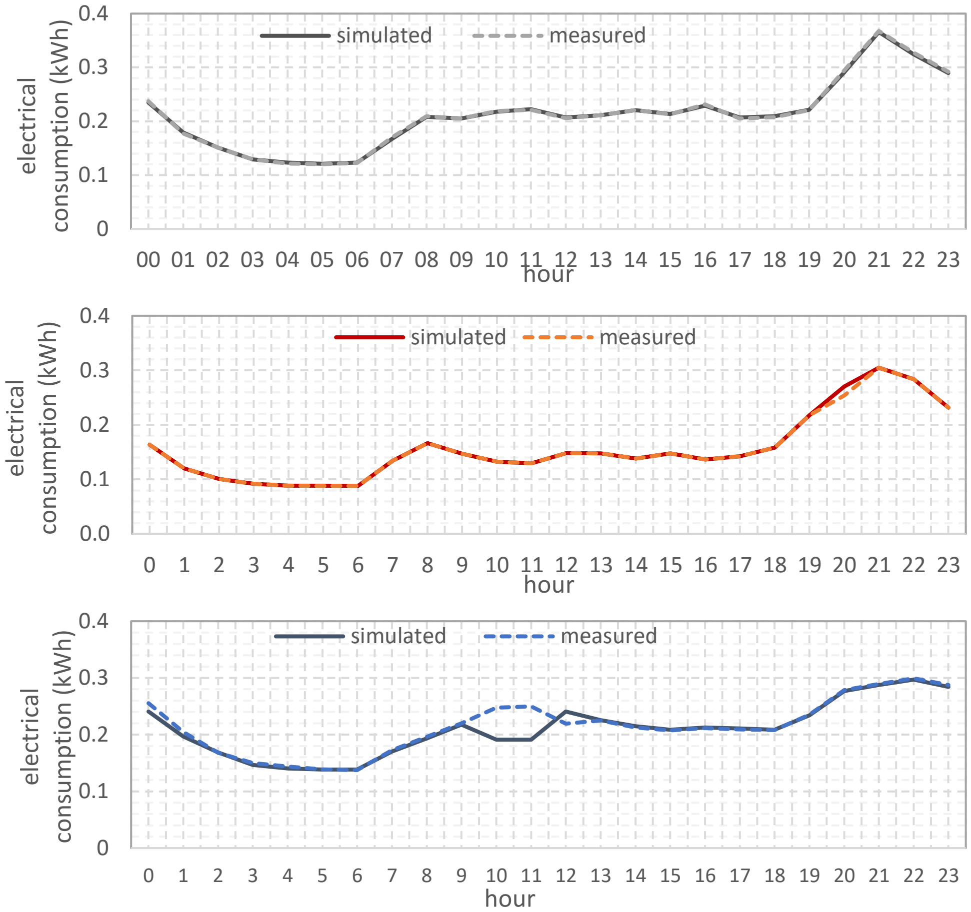


Figure 71 - Average daily consumption comparison between simulated and free-float profile (upper), winter profile (center) and summer profile (lower) for path 212

5.3 Path 222

The Path 222 represents 9.88% of the analyzed users. This household was assumed to be occupied by 4 persons, with an equipment design level of 205.13 W (1.4 W/m^2), a maximum standby power of 104.4 W (0.7 W/m^2), a fridge of 60 W (4.05 W/m^2) and light level of 106.13 W (0.71 W/m^2). The occupancy schedule and the input profiles from the mathematical model are present in annex E.2. The calibrated results were achieved for the following inputs:

- the heating setpoint of 19ºC for the bedroom and for the living room 19.5ºC from 0h to 10h and 20.5ºC for the remaining day;
- the cooling setpoint of 24ºC for the bedroom and for the living room 24ºC from 0h to 14h and 25ºC from then on;
- A SCOP of 1 and SEER of 2.5.

The evaluation results after performing the calibration are presented in Table 9. It can be observed that the model values are below the reference indexes presented section 4.8.1. Similar values as path 111 and path 212 were achieved for the free-float month, while in the winter month values for Cv(RMSE) are slightly when compared with path 111 (less 3,3%). Cv(RMSE) for the summer is also lower but the difference is not so significant as in the winter (only 0,1%), when comparing to the values obtained in path 212 for the summer similar values are achieved, except for NMBE where a smaller is here reported (less 2,4%).

Table 9- Validation results for free-float, winter and summer for path 222

	CV(RMSE)(%)	NMBE(%)
Hourly limit threshold [89]	30	10
Free-float	0.7	0.0
Winter	8.4	0.9
Summer	8.4	0.2

From Figure 72, it is also visible that the free-float, for the worst case the difference between measured and simulated is 1.67% which occurs at 0 h. From Figure 72, the overall fit between the measured and the simulated for the winter and summer months is satisfactory but not in all periods. For winter, the biggest difference occurs at 10h (27.75% of difference) where the simulated value is 0.1 kWh smaller than the measured, while for summer it occurs at 19h (26.49% of difference) where the simulated is 0.06 kWh higher than the measured.

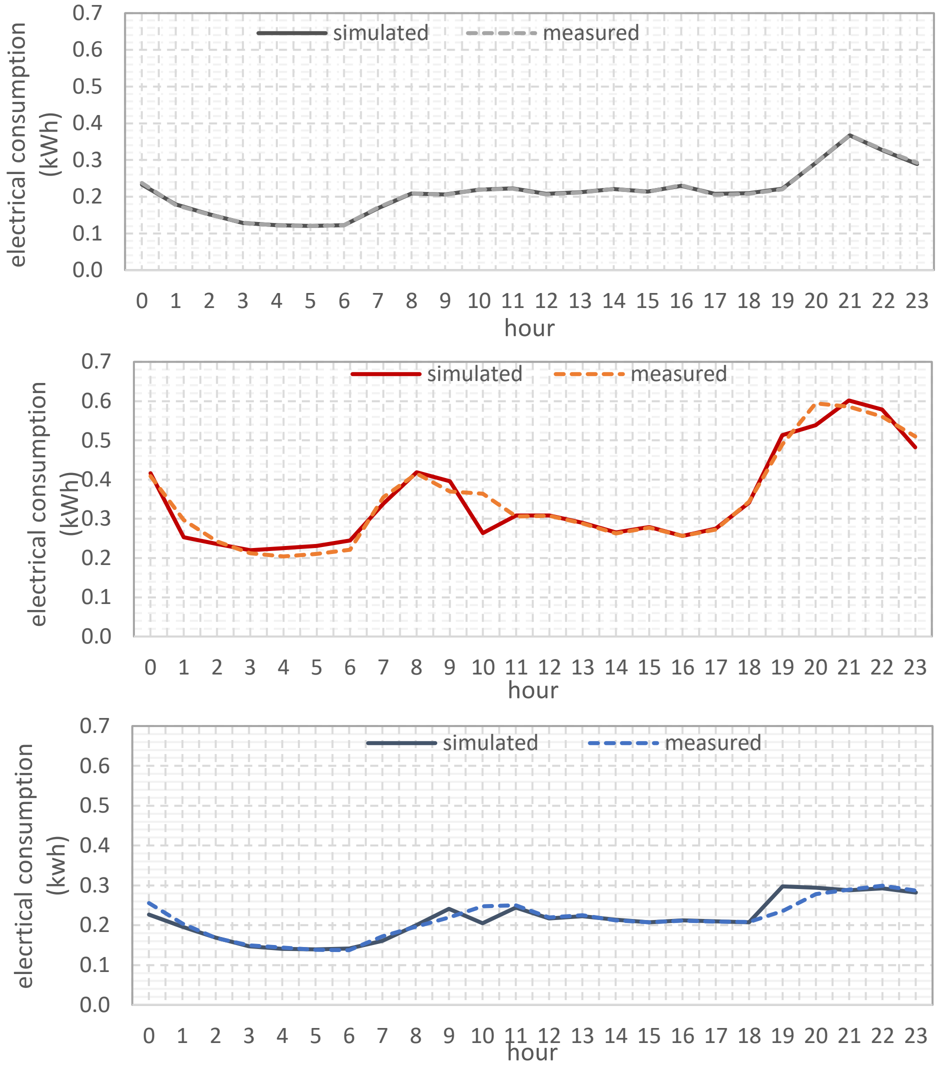


Figure 72 - Average daily consumption comparison between simulated and free-float profile (upper), winter profile (center) and summer profile (lower) for path 222

For winter, the major difference occurs at 10h (27,75% of difference) where the simulated value is 0,1 kWh smaller than the measured, while for summer it occurs at 19 h (26,49% of difference) where the simulated is 0,06 kWh higher than the measured. This difference may be explained by the heating or cooling consumption identified by the mathematical model that did not match the simulation values for this building characteristics (as developed previously in Section 5.1), visible when comparing end-use profile from model present in annex E.2 and end-use profile from simulation present in annex F.2. This equipment combination is not common, but it could be explained in two ways:

- this household could be equipped with air conditioner system with only cooling function which represents 7.2% of Portuguese households, value presented in [56].
- this household could in fact be equipped with a heat pump but choses to use other equipment for heating, in [8] it was presented that existent systems for space heating such as heat pumps, are not necessarily the most commonly used system;

5.4 Path 422

The Path 422 represents 7.4 % of the residential users. This household was assumed to be occupied by 5 persons, with an equipment design level of 351.2 W (2.36 W/m^2), a maximum standby power of 110.2 W (0.742 W/m^2), a fridge power of 93W (6.3 W/m^2) and a light level of 146.25 W (0.98 W/m^2). The occupancy schedule and the consumption profiles used to create this input are presented in annex E.3.The calibrated results were achieved for the following inputs:

- the heating setpoint of 19ºC for the bedroom and for the living room 19ºC from 0h to 12h and 20.5ºC for the remaining day;
- A SCOP of 2.6;
- No cooling system used.

The evaluation results after performing the calibration are presented in Table 10. It can be observed that the model values are below the reference indexes presented in section 4.8.1. For the free-float, the Cv(RMSE) and NMBE are similar to the values obtained in path 111. The winter month CV(RMSE) is lower in this path than path 111 by 7.8%. CV(RMSE) for the summer is the lowest value from all the paths (6.7% less than path 111), due to no use cooling equipment and the mathematical model not identifying cooling consumption as well.

Table 10 – validation results for free-float, winter and summer for 422

	CV(RMSE)(%)	NMBE(%)
Hourly limit threshold [89]	30	10
Free-float	1.0	0.1
Winter	3.9	0.9
Summer	1.6	0.4

From Figure 73, the free-float and summer month have a good fit, due to not having cooling consumption identified in the model. The winter has an overall good fit except for the early morning and between 19h to 21h, having the worst case occurring at 8h with 9% difference. The reasons behind these differences were already developed in 5.1. From the end-use consumption presented in annex F.3, it is possible to verify that coincides with the heating consumption periods.

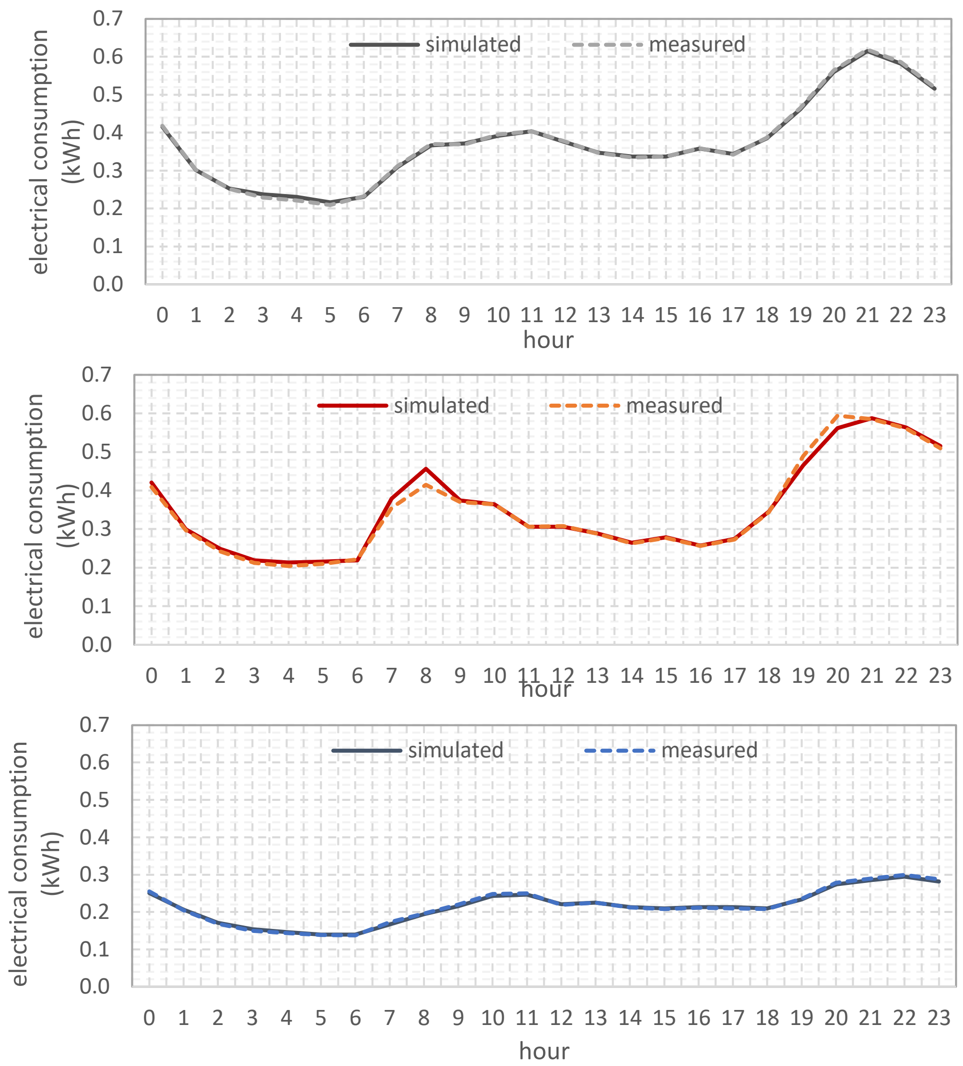


Figure 73 - Average daily consumption comparison between simulated and free-float profile (upper), winter profile (center) and summer profile (lower) path 422

In this path another fragility of the model was identified. As it was discussed in 4.5.4.4 it is likely that this household was empty at least some part of month since the summer daily consumption is 5.17 kWh/day and the free-float daily consumption is 8.93 kWh/day, and from 4h to 6h the total consumption is 0.13 kWh when the total assumed fridge consumption is 0.093 kWh, leaving just 0.04 kWh as maximum value of standby, when in the free-float month the maximum value of standby is 0.11, more than the double. The model did not identify this situation, and assumed occupied household and model the lighting consumption as it was a typically occupied household (also visible in the end-use consumption from the simulation profile in annex F.3.)

5.5 Path 444

The Path 444 represents 6.2 % of the residential users. The household was assumed to be occupied by 5 persons, with an equipment design level of 392.55 W (2.64 W/m^2), a maximum standby power of 136.07 (W/ m^2), a fridge of 93W (6.3 W/ m^2) and a light level of 152.6 W (1.03 W/m^2), the occupancy schedule and the reaming input profiles are visible in Annex E.4.

The calibrated results were achieved for the following inputs:

- the heating setpoint of 19ºC for the bedroom and for the living room 20ºC from 0h to 15h, 20.5ºC until 19h and 21.5ºC the remaining day;
- A SCOP of 1;
- No cooling system used.

The evaluation results after performing the calibration are presented in Table 11. It can be observed that the model results are below the reference indexes presented section 4.8.1 The Cv(RMSE) and NMBE values are quite similar in the three months when compared with path 111. This winter Cv(RMSE) is the highest from all the paths but NMBE was the lowest indicating that the positive differences are balanced by the negative differences between the simulated and measured values. The summer Cv(RMSE) is also the highest of all the paths (2.9% higher than path 111) and NMBE as well (3.9% higher than the value obtained in path 111).

Table 11- validation results for free-float, winter and summer for 444

	CV(RMSE)(%)	NMBE(%)
Hourly limit threshold [89]	30	10
Free-float	1.0	0.2
Winter	13.1	0.0
Summer	11.2	4.0

From Figure 74, it is possible to verify an overall good fit in the free-float month. For winter there is only a good match between 7h to 16h, when there is no heating consumption (visible in the end-use profile in annex F.4), between 1h and 7 h the simulation identified heating while the model did not, while the opposite occurs at 0h, having the worst case at 0h were the difference is 49%. Even though this hourly difference occurred the total consumption difference is below 1%. An even less efficient heating equipment with a more variable heating set point could improve the fit between both curves. For the summer, only three hours present some difference, having at 0h and 16h 33% of difference, that difference can be explained by comparing the end-use from mathematical model (annex E.4) and from the simulation (annex F.4), where at those periods the mathematical model identifies cooling the best fit occurs for no use of cooling equipment, even if there is cooling needs.

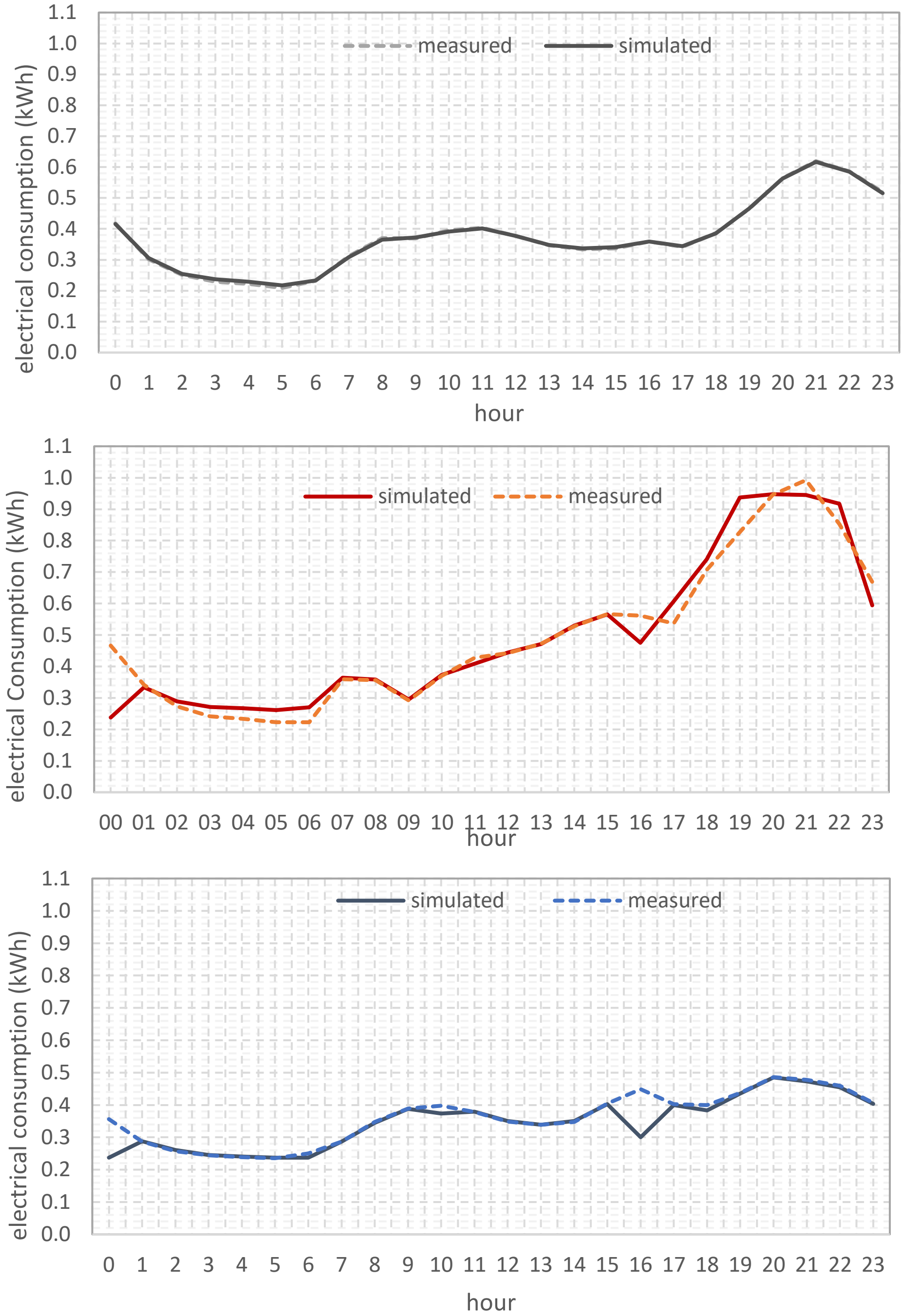


Figure 74 - Average daily consumption comparison between simulated and free-float profile (upper), winter profile (center) and summer profile (lower) for path 444

5.6 Overall end-use consumption

By comparing the simulated overall end-use consumption (Figure 75), the biggest percentage is associated with electric equipment having on average 31% of the total consumption, followed by the fridge consumption with the average of 28% and standby average value of 20% while the lighting only represents 13% and heating (average value of 6%) and cooling (average value of 1%) the remaining. When it comes to the total yearly fridge consumption, it reaches the 800 kWh/year which shows that if it was renewed by a more efficient model the percentage attributed to the fridge could be reduced. Since standby is the third responsible for the total consumption it is necessary to address measures to reduce it.

It is also possible to observe that path 111 represents a low consumption household and that the fridge, standby and electric equipment have almost the same percentage of the total consumption.

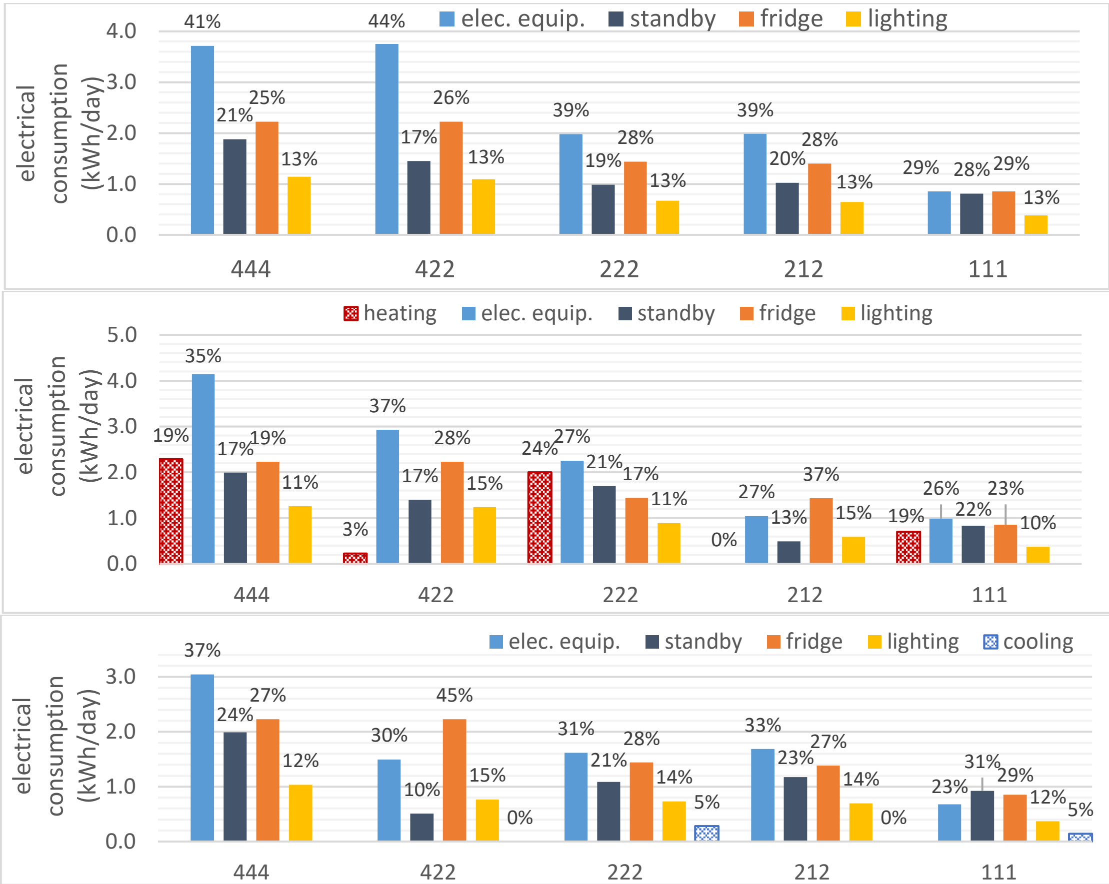


Figure 75 – end-use consumption per path for free-float (upper), winter (middle) and summer (lower)

Path 212 has in the free-float and summer month as the highest share, the consumption coming from electric equipment (39% and 33% as electric equipment) while in the winter the equipment consumption reduces to 31,5% and standby reduced 7% when compared to the free-float value while the opposite occurs to the fridge consumption (representing 27% e 28% while in the winter it rises to 37%), which could indicate that the household had some empty days that decreased the electric equipment consumption. From path 422 in the summer, it is possible to identify that the fridge consumption increases 19% when compared with free-float, and that standby decreased 7% and electric equipment 15%, which indicates that the household was empty, and when compared with winter from path 212 it is possible to conclude that in this case it was empty for a longer period. From here a fragility in the model is identified since if the household is empty the light consumption cannot be based on the values obtained from free-float, a fully occupied household. From path 111 where both systems have similar efficiency for cooling and heating it is possible to identify that in Portugal there is more consumption related to heating than cooling. As well when looking at the middle figure there are 4 paths using a kind of heating equipment while from the lower figure only 2 use cooling equipment. From path 222 with 24% and path 444 with 19% of heating consumption, both with inefficient heating systems, it is possible to test other equipment's that could reduce this consumption. For instance, when comparing path 422, which has internal loads similar to path 444 but uses a heating equipment with SCOP of 2.6, heating consumption is only has 3%.

6 Conclusions and future work

Due to the present awareness of building electrical consumption, especially in residential buildings, it is mandatory to address efficiency measures to reduce the energy consumption. To do so, these measures need to be tested through accurate models of energy consumption in buildings. However, the lack of input information necessary to create these models has been contributing to hinder this goal. The deployment of smart meters made available hourly electrical consumption information creating a fantastic opportunity to use this that data in improving the building energy model's accuracy.

The present work proposed the development of a model able to disaggregate hourly values of electrical consumption into lighting, equipment, cooling and heating consumption schedules and occupancy profiles for different households, using smart metered data. This information coupled with data on buildings geometry and construction was further used as input data in an energy simulation model.

The electrical consumption dataset included missing-value users, zero-use users and outliers which requires to perform a

Data cleaning and preprocessing procedure to improve the quality of the dataset. Only 33% of the entries of the initial dataset were compliant with the minimum requirements to perform a good analysis, corresponding to a final sample of 90 users. Nonetheless, it was still necessary to fill in missing values which increased from 1% to 3% the total final consumption in 70 users, in the remaining 20 the impact was higher than 3%. This highlights the importance of improving the accuracy of data recovery in the future since only 33% percent of the initial dataset met the minimum requirements and every user had some missing values inputted which can bring some uncertainty and reduce the accuracy of the analysis.

In order to distinguish residential from nonresidential profiles, a clustering analysis using the k-means algorithm was performed. After carrying out an analysis on the more suitable normalization method and adequacy measures, 5 cluster centroids were identified with 2 representing nonresidential profiles, corresponding to 10% of the sample.

Since the focus was on the residential consumers, a second stage of clustering analysis was performed for three reference months: free-float month (May); winter month (January) and summer month (July). For each month, a clustering analysis is performed and four cluster centroids are obtained and *paths* are created. Five paths are analyzed: path 111 (representing 20% of residential users), path 212 (16%), path 222 (10%), path 422(7%) and path 444(6%).

A mathematical model is applied to each path consumption to create input profiles (occupancy, lighting, equipment, standby and fridge) for an energy simulation suing the building characteristics presented in the case study. After performing model calibration, the best fit obtained was:

- Path 111: heating system with SCOP of 2.6 and cooling system with SEER of 2.5;
- Path 212: no use of electric system for cooling or heating;
- Path 222: heating system with SCOP equal to 1 and cooling system with SEER of 2.5;

- Path 422: heating system with SCOP of 2.6 and no use of cooling system;
- Path 444: heating system with SCOP equal to 1 and no use of cooling system.

Overall, the model proved to be effective when applied to the free-float since every path met the reference criteria and having in the worst fitting path an hour where the difference was equal to 4% (path 444), and Cv(RMSE) was always below 1% and NMBE below 0.2% (while the limit threshold values are 30% and 10%, respectably). When it comes to the winter month, all the paths met the criteria and path 212 had the best fitting results with Cv(RMSE) of 2.2% and NMBE of 0.5%, while the worst fitting path was path 444 with Cv(RMSE) of 13,1% and NMBE 0% (meaning the negative and positive differences between simulated and measured are balanced). When it comes to summer the lowest values of Cv(RMSE) were obtained for path 422 (equal to 3.9%), while path 444 obtained the highest value (equal to 11.2%) and the highest value for NMBE (equal to 4%) but still below limit values.

Path 444 was the worst fitting path for each month but the difference between total daily consumption from measured and simulated are:

- For free-float the difference was 0.22%;
- For winter the difference was below 0.01%;
- For summer the difference was below 4%.

From the worst fitting path, it is possible to concluded that even if the measured and the simulated profile consumption does not have a perfect hourly match, the total consumption does have a good approximation specially for free-float and winter. So, it is possible to conclude that the method proposed to infer the input profiles per end-use can describe the actual consumption and can be used to test energy efficiency measures in the future.

Future work includes the model optimization, especially the improvement of the occupancy profile by testing 3 different new models of combining the information from *high consumption periods* and *on event* by using the average, the maximum value or the sum of both values. Moreover, the improvement on the identification of high consumption periods and on events as a continuous evaluation proportional to the correspondent threshold instead of the actual model that is based on discrete evaluation. Future work on the mathematical model should also focus on the summer month to identify periods of inhouse vacations and unoccupied household and make appropriate changes on the end-uses consumption. Further developments on the model calibration by including an uncertainty analysis considering more variables can also improve the model predictions.

7 References

[1] European Commission, “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee, the Committee of the Regions and the European Investemnet Bank : Clean Energy For All European,” 2016.

[2] European Commission, “The Road from Paris: assessing the implications of the Paris Agreement and accompanying the proposal for a Council decision on the signing, on behalf of the European Union, of the Paris agreement adopted under the United Nations Framework Convention on Cl.” 2016.

[3] European Commission, “Towards the Paris Protocol.” [Online]. Available: https://ec.europa.eu/clima/policies/international/paris_protocol_en. [Accessed: 08-Sep-2017].

[4] European Commission, “Communication from the Commission to the European Parliament and the Council: Implementing the Energy Efficiency Directive – Commission Guidance,” 2013.

[5] European Commission, “Climate and energy priorities for Europe: the way forward,” 2014.

[6] Y. Saheb, K. Bódis, S. Szabó, H. Ossenbrick, and S. Panev, “Energy Renovation: The Trump Card for the New Start for Europe|European Commission,” 2015.

[7] European Commission, “Proposal for a Directive of the European Parliament and of the Council amending Directive 2010/31/EU on the energy performance of buildings,” 2016.

[8] M. Ferreira and M. Panão, “The role of inhabitants behaviour in the prebound effect of housing energy consumption,” in *BEHAVE 2016*, 2016, no. 4th European Conference on Behaviour and Energy Efficiency, pp. 8–9.

[9] D.-G. for E. European Commission, “Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on common rules for the internal market in electricity,” 2017.

[10] J. D. Rhodes, W. J. Cole, C. R. Upshaw, T. F. Edgar, and M. E. Webber, “Clustering analysis of residential electricity demand profiles,” *Appl. Energy*, vol. 135, pp. 461–471, 2014.

[11] A. Al-Wakeel and J. Wu, “K-means based cluster analysis of residential smart meter measurements,” *Energy Procedia*, vol. 88, pp. 754–760, 2016.

[12] M. N. Q. Macedo, J. J. M. Galo, L. A. L. Almeida, and A. C. C. Lima, “Typification of load curves for DSM in Brazil for a smart grid environment,” *Int. J. Electr. Power Energy Syst.*, vol. 67, no. May, pp. 216–221, 2015.

[13] K. Le Zhou, S. L. Yang, and C. Shen, “A review of electric load classification in smart grid environment,” *Renew. Sustain. Energy Rev.*, vol. 24, pp. 103–110, 2013.

[14] S. Ramos, J. M. Duarte, F. J. Duarte, and Z. Vale, “A data-mining-based methodology to support MV electricity customers’ characterization,” *Energy Build.*, vol. 91, 2015.

[15] G. Chicco, “Overview and performance assessment of the clustering methods for electrical load pattern grouping,” *Energy*, vol. 42, no. 1, pp. 68–80, 2012.

[16] G. J. Tsekouras, N. D. Hatziargyriou, and E. N. Dialynas, “Two-stage pattern recognition of load curves for classification of electricity customers,” *IEEE Trans. Power Syst.*, vol. 22, no. 3, pp. 1120–1128, 2007.

[17] V. Figueiredo, F. Rodrigues, Z. Vale, and J. B. Gouveia, “An electric energy consumer characterization framework based on data mining techniques,” *IEEE Trans. Power Syst.*, vol. 20, no. 2, pp. 596–602, 2005.

[18] I. Panapakidis, M. Alexiadis, and G. Papagiannis, “Evaluation of the performance of clustering algorithms for a high voltage industrial consumer,” *Eng. Appl. Artif. Intell.*, vol. 38, pp. 1–13, 2015.

[19] K. Zhou, C. Yang, and J. Shen, “Discovering residential electricity consumption patterns through smart-meter data mining: A case study from China,” *Util. Policy*, vol. 44, pp. 73–84, 2017.

[20] I. Benítez, J.-L. Díez, A. Quijano, and I. Delgado, “Dynamic clustering of residential electricity consumption time series data based on Hausdorff distance,” *Electr. Power Syst. Res.*, vol. 140, pp. 517–526, 2016.

[21] F. McLoughlin, A. Duffy, and M. Conlon, “A clustering approach to domestic electricity load profile characterisation using smart metering data,” *Appl. Energy*, vol. 141, pp. 190–199, 2015.

[22] J. L. Viegas, S. M. Vieira, R. Melício, V. M. F. Mendes, and J. M. C. Sousa, “Classification of new electricity customers based on surveys and smart metering data,” *Energy*, vol. 107, pp. 804–817, 2016.

[23] M. Almeida Duarte Pina, “Prediction of Electricity Consumers Behaviour based on Smart-metering Data,” *Mater Thesis Mech. Eng. Inst. Super. Técnico, Univ. Lisboa*, 2016.

[24] D. J. Pinto Sousa, “Análise do comportamento de consumidores de eletricidade: principais causas e variações de perfil ao longo do ano,” *Tese Mestr. em Eng. Mecânica, Inst. Super. Técnico, Univ. Lisboa*, 2017.

[25] K. Zhou, S. Yang, and Z. Shao, “Household monthly electricity consumption pattern mining: A fuzzy clustering-based model and a case study,” *J. Clean. Prod.*, vol. 141, pp. 900–908, 2017.

[26] I. Benítez, A. Quijano, J. L. Díez, and I. Delgado, “Dynamic clustering segmentation applied to load profiles of energy consumption from Spanish customers,” *Int. J. Electr. Power Energy Syst.*, vol. 55, pp. 437–448, 2014.

[27] J. L. Viegas, S. M. Vieira, J. M. C. Sousa, R. Melício, and V. M. F. Mendes, “Electricity demand profile prediction based on household characteristics,” in *International Conference on the European Energy Market, EEM*, 2015, vol. 2015–August.

[28] C. Beckel, L. Sadamori, and S. Santini, “Towards Automatic Classification of Private Households Using Electricity Consumption Data Christian,” *J. Econ. Psychol.*, vol. 3, no. 3–4, pp. 75–86, 2013.

[29] C. Beckel, L. Sadamori, T. Staake, and S. Santini, “Revealing household characteristics from smart meter data,” *Energy*, vol. 78, pp. 397–410, 2014.

[30] A. Kipping and E. Trømborg, “Modeling and disaggregating hourly electricity consumption in Norwegian dwellings based on smart meter data,” *Energy Build.*, vol. 118, pp. 350–369, 2016.

[31] L. Farinaccio and R. Zmeureanu, “Using a pattern recognition approach to disaggregate the total electricity consumption in a house into the major end-uses,” *Energy Build.*, vol. 30, no. 3, pp. 245–259, 1999.

[32] D. George and L. G. Swan, “A method for distinguishing appliance, lighting and plug load profiles from electricity ‘smart meter’ datasets,” *Energy Build.*, vol. 134, pp. 1–11, 2016.

[33] L. Stankovic, V. Stankovic, J. Liao, and C. Wilson, “Measuring the energy intensity of domestic activities from smart meter data,” *Appl. Energy*, vol. 183, pp. 1565–1580, 2016.

[34] I. Richardson, M. Thomson, D. Infield, and C. Clifford, “Domestic electricity use : A high-resolution energy demand model,” *Energy Build.*, vol. 42, no. 10, pp. 1878–1887, 2010.

[35] I. Richardson, M. Thomson, D. Infield, and A. Delahunty, “Domestic lighting : A high-resolution energy demand model,” *Energy Build.*, vol. 41, pp. 781–789, 2009.

[36] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, “Occupancy Detection from Electricity Consumption Data,” *Proc. 5th ACM Work. Embed. Syst. Energy-Efficient Build. - BuildSys’13*, pp. 1–8, 2013.

[37] I. Richardson, M. Thomson, and D. Infield, “A high-resolution domestic building occupancy model for energy demand simulations,” *Energy Build.*, vol. 40, no. 8, pp. 1560–1566, 2008.

[38] Y. G. Yohanis, J. D. Mondol, A. Wright, and B. Norton, "Real-life energy use in the UK: How occupancy and dwelling characteristics affect domestic electricity use," *Energy Build.*, vol. 40, no. 6, pp. 1053–1059, 2008.

[39] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, and P. Shenoy, "Non-Intrusive Occupancy Monitoring using Smart Meters," *5th ACM Work. Embed. Syst. Energy-Efficient Build. - BuildSys'13*, pp. 1–8, 2013.

[40] N. Costa and I. Matos, "Inferring daily routines from electricity meter data," *Energy Build.*, vol. 110, pp. 294–301, 2016.

[41] W. Kleiminger, C. Beckel, and S. Santini, "Household Occupancy Monitoring Using Electricity Meters," *Proc. 2015 ACM Int. Jt. Conf. Pervasive Ubiquitous Comput.*, 2015.

[42] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[43] U. M. Malay K. Pakhira, Sanghamitra Bandyopadhyay, "Validity index for crisp and fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 3, pp. 487–501, Mar. 2004.

[44] U. Lusófona de Humanidades e Tecnologias, "Diagnóstico Social da Freguesia do Parque das Nações," Junta de Freguesia do Parque das Nações, 2015.

[45] C. M. de Lisboa, "(nova) Freguesia do Parque das Nações." .

[46] ADENE, "Certificação Energética de Edifícios." [Online]. Available: http://www.adene.pt/certificacao-energetica-de-edificios. [Accessed: 20-Sep-2017].

[47] C. A. Pina dos Santos and L. Matias, "Coeficientes de transmissão térmica de elementos da envolvente dos edifícios," *LNEC - Informações Científicas e Técnicas*, 2006.

[48] ADENE, "Sistemas de ventilação," 2016.

[49] "Arquivo Municipal de Lisboa." [Online]. Available: http://arquivomunicipal.cm-lisboa.pt/pt/. [Accessed: 28-Sep-2017].

[50] Minitab Inc, "Minitab Statistical Software." [Online]. Available: https://www.minitab.com/en-us/. [Accessed: 04-Sep-2017].

[51] EDP, "Potência média e Consumo Mensal." [Online]. Available: https://www.edp.pt/pt/particulares/bemvindoaedp/Pages/PotenciaRecomendada.aspx. [Accessed: 29-Aug-2017].

[52] G. C. McBane, "Programs to Compute Distribution Functions and Critical Values for Extreme Value Ratios for Outlier Detection," *J. Stat. Softw.*, vol. 16, no. 3, pp. 1–9, May 2006.

[53] M. Humphries, "Missing Data & How to Deal: An overview of missing data," *Popul. Res. Cent.*, p. 45, 2013.

[54] Minitab, "How to Identify the Distribution of Your Data using Minitab | Minitab." [Online]. Available: http://blog.minitab.com/blog/adventures-in-statistics-2/how-to-identify-the-distribution-of-your-data-using-minitab. [Accessed: 05-Oct-2017].

[55] E. W. Weisstein, "Log Normal Distribution." [Online]. Available: http://mathworld.wolfram.com/LogNormalDistribution.html. [Accessed: 05-Oct-2017].

[56] INE and DGEG, "Inquérito ao Consumo de Energia no Sector Doméstico 2010." 2011.

[57] The MathWorks Inc, "Calinski-Harabasz criterion clustering evaluation object - MATLAB." [Online]. Available: https://www.mathworks.com/help/stats/clustering.evaluation.calinskiharabaszevaluation-class.html. [Accessed: 30-Aug-2017].

[58] D. van Dijk, M. Spiekmann, and L. Hoes-Van Oeffelen, "EPB standard EN ISO 52016 : Calculation

of the building ’ s energy needs for heating and cooling , internal temperatures and heating and cooling load,” *REHVA J.*, no. May, 2016.

[59] BizEE Software Limited, “Heating & Cooling Degree Days - Free Worldwide Data Calculation.” [Online]. Available: http://www.degreedays.net/. [Accessed: 12-Sep-2017].

[60] Carlos A. Santos Silva, “Energy in Buildings (2),” *Energy Manag. IST*, 2016.

[61] B. Sørensen, *Solar Energy Storage*. Elsevier Science, 2015.

[62] M. Santamouris, *Energy Performance of Residential Buildings: A Practical Guide for Energy Rating and Efficiency*. Taylor & Francis, 2010.

[63] N. A. Muhammad Usman Fahad, “Factor Affecting Short Term Load Forecasting,” *J. Clean Energy Technol.*, vol. Vol. 2, no. No. 4, 2014.

[64] K. Beckmann, “Energy Demand for Heating,” *Scotland´s Cent. Expert. Connect. Clim. Chang. Res. policy*, 2016.

[65] R. and A.-C. E. American Society of Heating, *2005 ASHRAE Handbook : Fundamentals - SI edition.* American Society of Heating Refrigerating and Air-Conditionin, 2005.

[66] Time and Date, “Climate & Weather Averages in Lisbon, Portugal.” [Online]. Available: https://www.timeanddate.com/weather/portugal/lisbon/climate. [Accessed: 24-Sep-2017].

[67] D. Coakley, P. Raftery, and M. Keane, “A review of methods to match building energy simulation models to measured data,” *Renew. Sustain. Energy Rev.*, vol. 37, pp. 123–141, 2014.

[68] A. De Almeida, P. Fonseca, B. Schlomann, and N. Feilberg, “Characterization of the household electricity consumption in the EU, potential energy savings and specific policy recommendations,” *Energy Build.*, 2011.

[69] ADENE, “Eficiência energética em equipamentos e sistemas eléctricos no sector residencial.” p. 32, 2004.

[70] European Commission, “JRC’s Institute for Energy and Transport - PVGIS.” [Online]. Available: http://re.jrc.ec.europa.eu/pvgis/imaps/index.htm. [Accessed: 19-Sep-2017].

[71] Alex Ryer, “Light Measurement Handbook,” 1997. [Online]. Available: http://www.dfisica.ubi.pt/~hgil/Fotometria/HandBook/luz.html. [Accessed: 20-Sep-2017].

[72] Uc Berkeley, “Brightness vs. Distance | The Inverse Square Law.” [Online]. Available: http://bccp.berkeley.edu/o/Academy/workshop08/08 PDFs/Inv_Square_Law.pdf. [Accessed: 16-Sep-2017].

[73] IEEE GlobalSpec’s Engineering360, “Lux Meters (Light Meters) Information | Engineering360.” [Online]. Available: http://www.globalspec.com/learnmore/optics_optical_components/optoelectronics/lux_meters_light_meters. [Accessed: 21-Sep-2017].

[74] Institute of Energy and Sustainable Development. De Montfort University Leicester UK, “Daylighting and Compliance: Are current standards sufficient?”

[75] C. F. Reinhart, “Lightswitch-2002: A model for manual and automated control of electric lighting and blinds,” *Sol. Energy*, vol. 77, no. 1, pp. 15–28, 2004.

[76] H. M. Sacht, L. Bragança, M. Almeida, and R. Caram, “Estudo do desempenho lumínico para um sistema de fachada modular destinado á Portugal,” in *XIV ENTAC -Encontro Nacional de Tecnologia do Ambiente Construído*, 2012.

[77] European Commission, “Directiva 2010/30/UE do Parlamento Europeu e do Conselho no que respeita à rotulagem energética dos aparelhos de refrigeração para uso doméstico,” 2010.

[78] European Commission, “Supplementing Directive 2010/30/EU of the European Parliament and of

the Council with regard to energy labelling of household refrigerating appliances,” 2010.

[79] Topten, “Best of Europe - Selection Criteria Household Refrigerators| European Climate Foundation.” [Online]. Available: http://www.topten.eu/english/criteria/freezers_ak.html&fromid=. [Accessed: 21-Sep-2017].

[80] INE, “O uso do tempo 1999- inquérito à ocupação do tempo.” pp. 1–2, 1999.

[81] National Renewable Energy Laboratory (NREL)., “EnergyPlus | building energy simulation program.” [Online]. Available: https://energyplus.net/. [Accessed: 04-Sep-2017].

[82] AUTODESK, “AutoCAD.” [Online]. Available: https://www.autodesk.com/education/free-software/autocad. [Accessed: 25-Sep-2017].

[83] T. Inc., “3D modeling for everyone | SketchUp.” [Online]. Available: https://www.sketchup.com/. [Accessed: 04-Sep-2017].

[84] bigladder SOFTWARE, “Table of Contents: Engineering Reference — EnergyPlus 8.3.” [Online]. Available: http://bigladdersoftware.com/epx/docs/8-3/engineering-reference/. [Accessed: 25-Sep-2017].

[85] IPMA, “Bolhetim climatologico.” [Online]. Available: http://www.ipma.pt/pt/publicacoes/boletins.jsp?cmbDep=cli&cmbTema=pcl&cmbAno=2016&idDep=cli&idTema=pcl&curAno=2016. [Accessed: 06-Oct-2017].

[86] ENBRI, “Adaptive thermal comfort for buildings in Portugal.” Construction Technology in Europe, 2010.

[87] EnergyPlus, “Weather Data.” [Online]. Available: https://energyplus.net/weather. [Accessed: 25-Sep-2017].

[88] E. Fabrizio and V. Monetti, “Methodologies and advancements in the calibration of building energy models,” *Energies*, vol. 8, no. 4, pp. 2548–2574, 2015.

[89] ANSI/ASHRAE, “ASHRAE Guideline 14-2002 Measurement of Energy and Demand Savings,” *Ashrae*, vol. 8400, p. 170, 2002.

Annex A - Building geometry

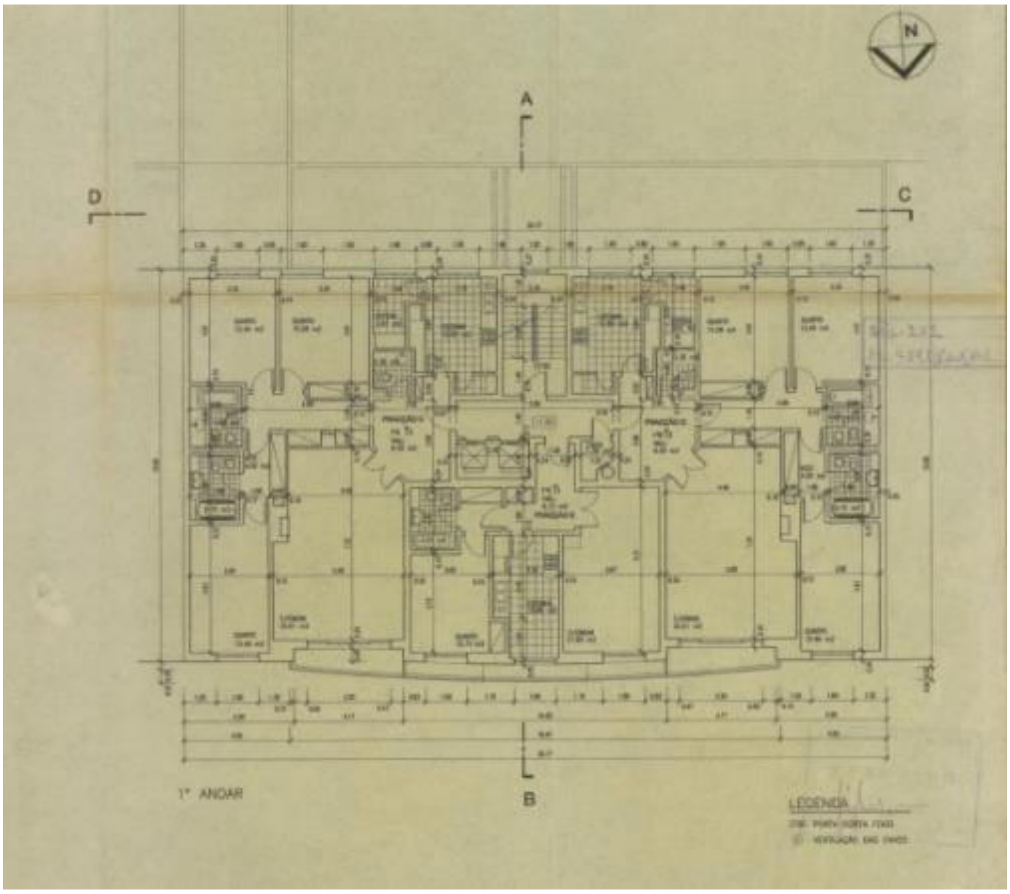


Figure 76 – 1st floor plan

Figure 77- Cut A-B and Cut C-D

Figure 78- Building elevation

Figure 79 - Block

Annex B - Cluster centroids

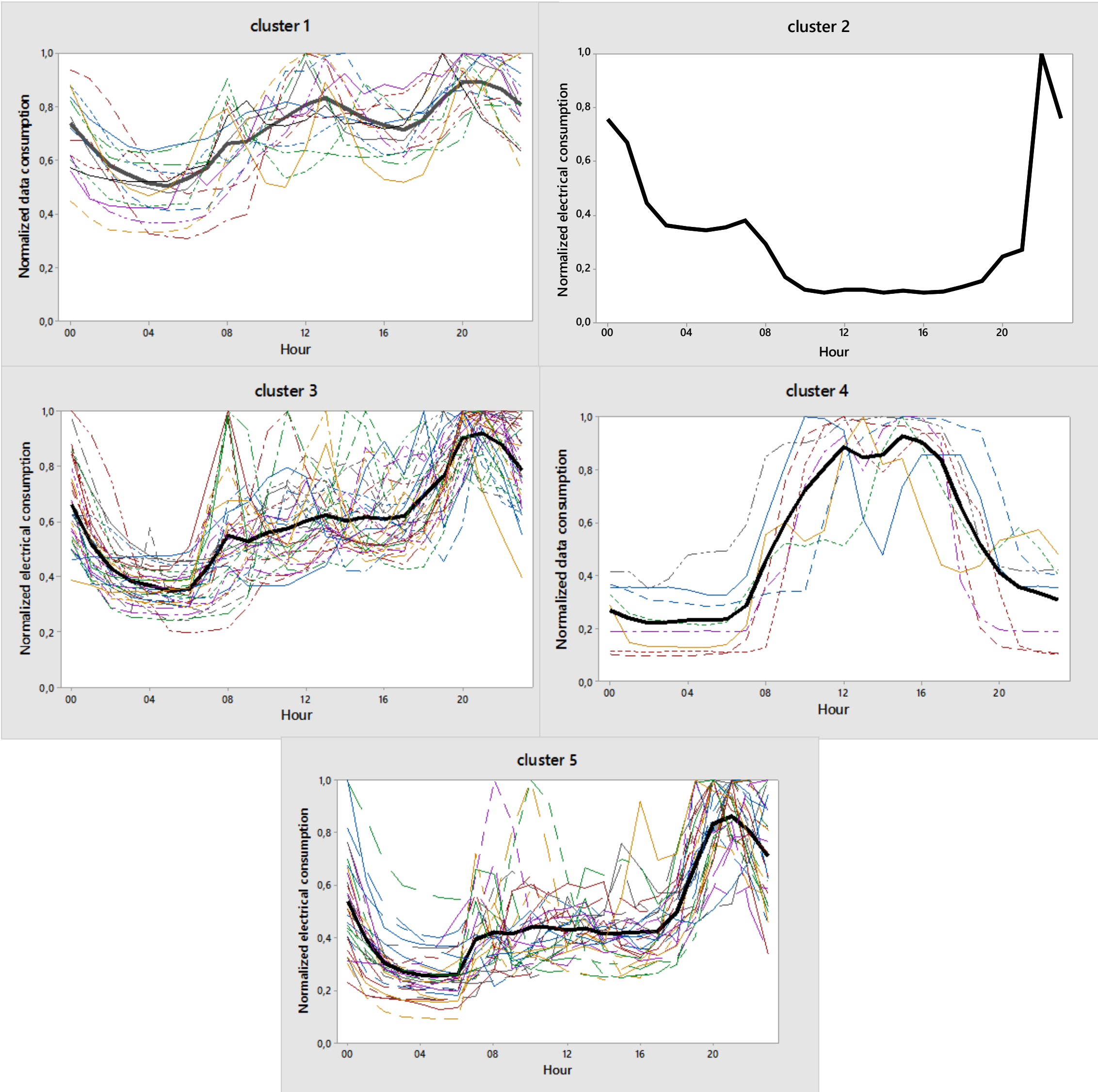


Figure 80 - Clusters for typification analysis

Annex C - Weekend paths

Table 12 -Week paths combined with weekend paths

<table>
<tr><th>week_paths</th><th>%</th><th>weekend_paths</th><th>%</th><th>week_paths</th><th>%</th><th>weekend_paths</th><th>%</th></tr>
<tr><td rowspan="7">111</td><td rowspan="7">19.8%</td><td>111</td><td>9.9%</td><td rowspan="3">222</td><td rowspan="3">9.9%</td><td>222</td><td>6.2%</td></tr>
<tr><td>121</td><td>3.7%</td><td>221</td><td>2.5%</td></tr>
<tr><td>222</td><td>1.2%</td><td>141</td><td>1.2%</td></tr>
<tr><td>122</td><td>1.2%</td><td rowspan="5">422</td><td rowspan="5">7.4%</td><td>444</td><td>2.5%</td></tr>
<tr><td>141</td><td>1.2%</td><td>222</td><td>1.2%</td></tr>
<tr><td>112</td><td>1.2%</td><td>441</td><td>1.2%</td></tr>
<tr><td>241</td><td>1.2%</td><td>221</td><td>1.2%</td></tr>
<tr><td rowspan="6">212</td><td rowspan="6">16.0%</td><td>222</td><td>4.9%</td><td>242</td><td>1.2%</td></tr>
<tr><td>211</td><td>3.7%</td><td rowspan="4">444</td><td rowspan="4">6.2%</td><td>444</td><td>2.5%</td></tr>
<tr><td>121</td><td>2.5%</td><td>144</td><td>1.2%</td></tr>
<tr><td>212</td><td>2.5%</td><td>443</td><td>1.2%</td></tr>
<tr><td>442</td><td>1.2%</td><td>343</td><td>1.2%</td></tr>
<tr><td>224</td><td>1.2%</td><td></td><td></td><td></td><td></td></tr>
</table>

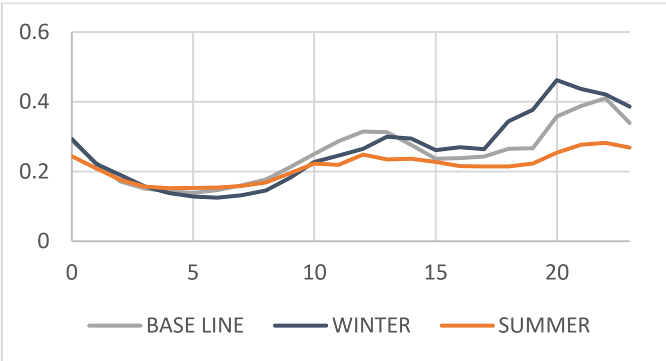


Figure 81- weekend path 222

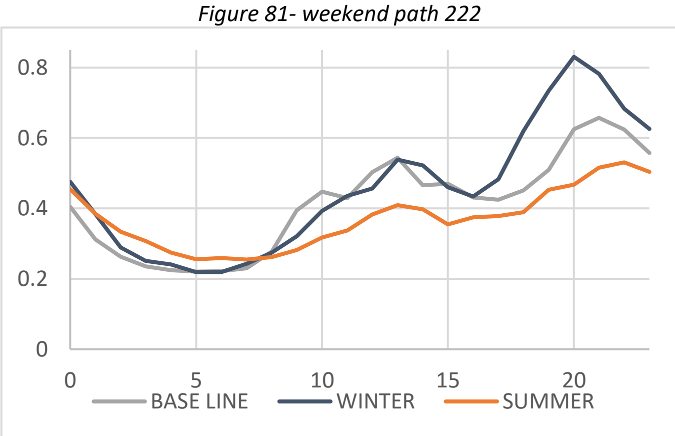


Figure 82- weekend path 444

Annex D - Light Profile

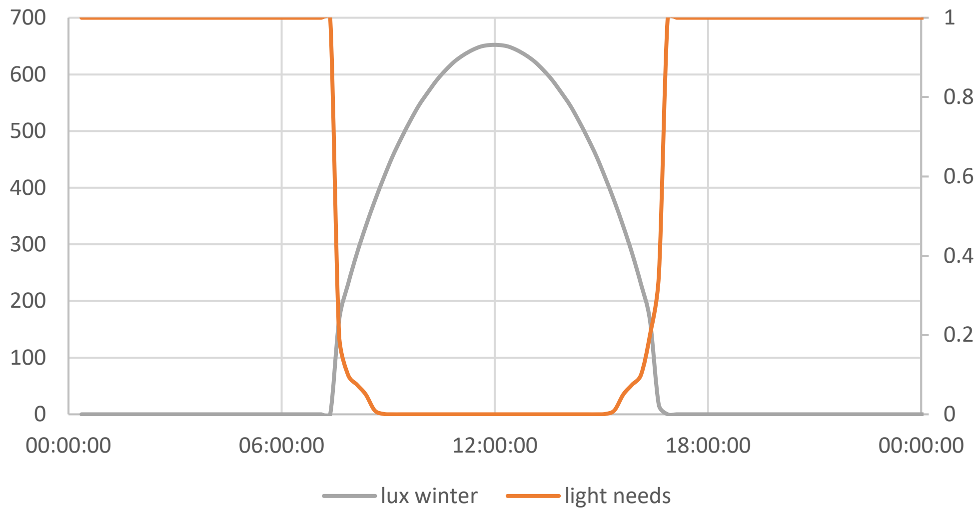


Figure 83- light needs versus lux values in winter

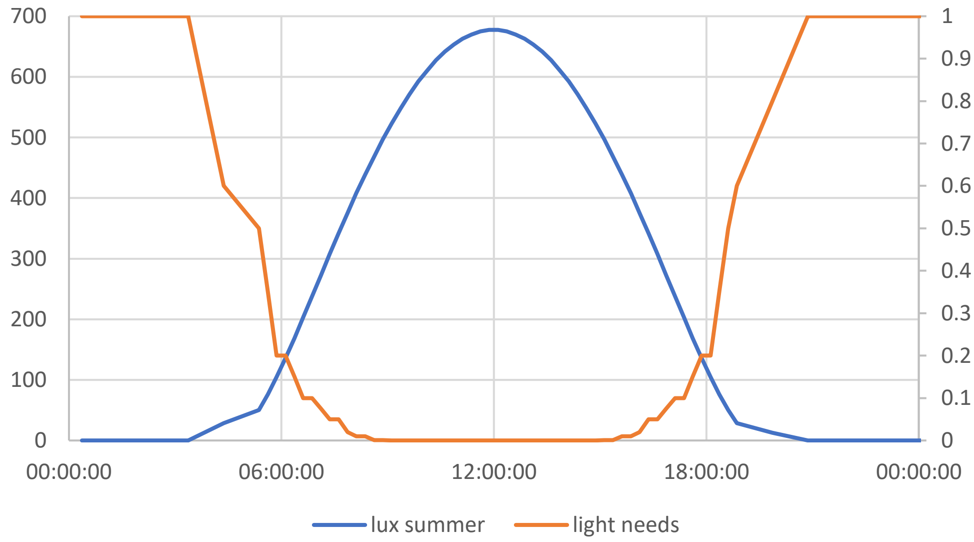


Figure 84 – light needs versus lux values in summer

Annex E - Occupancy and end-use profiles from mathematical model

E.1 Path 212

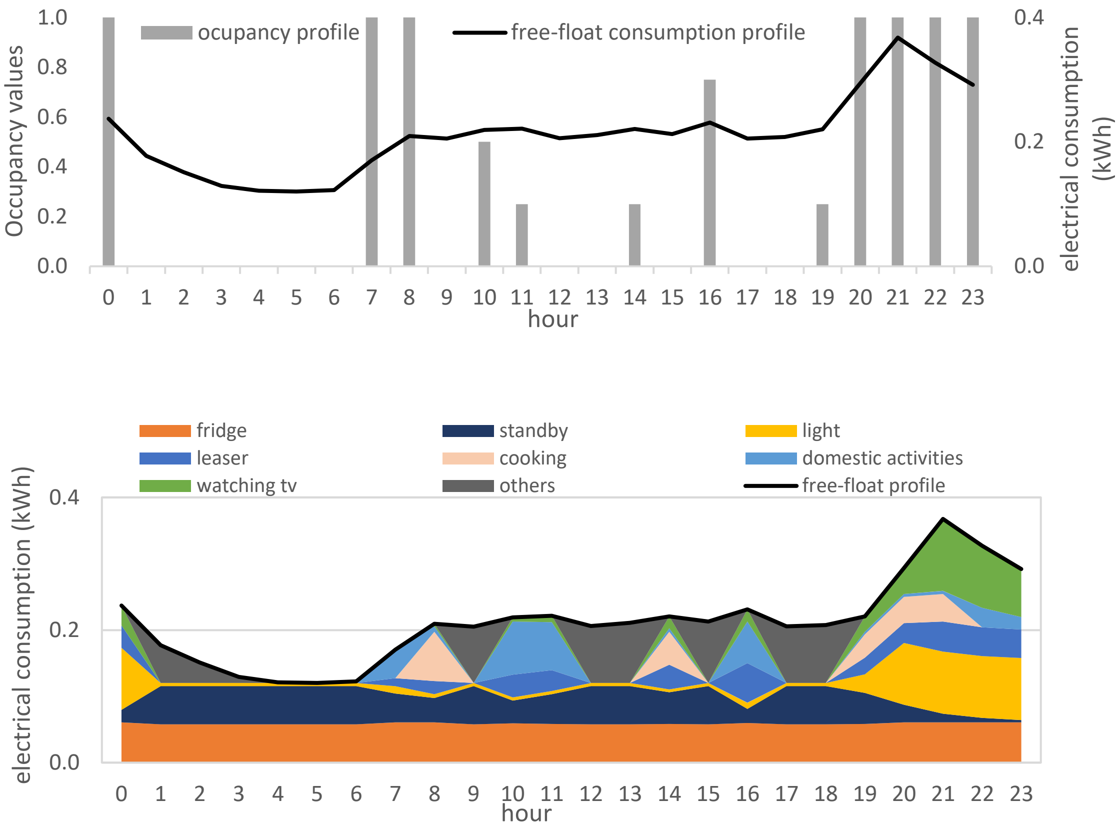


Figure 85 - Occupancy Profile (upper) and end-use profiles(lower) for free-float month of path 212

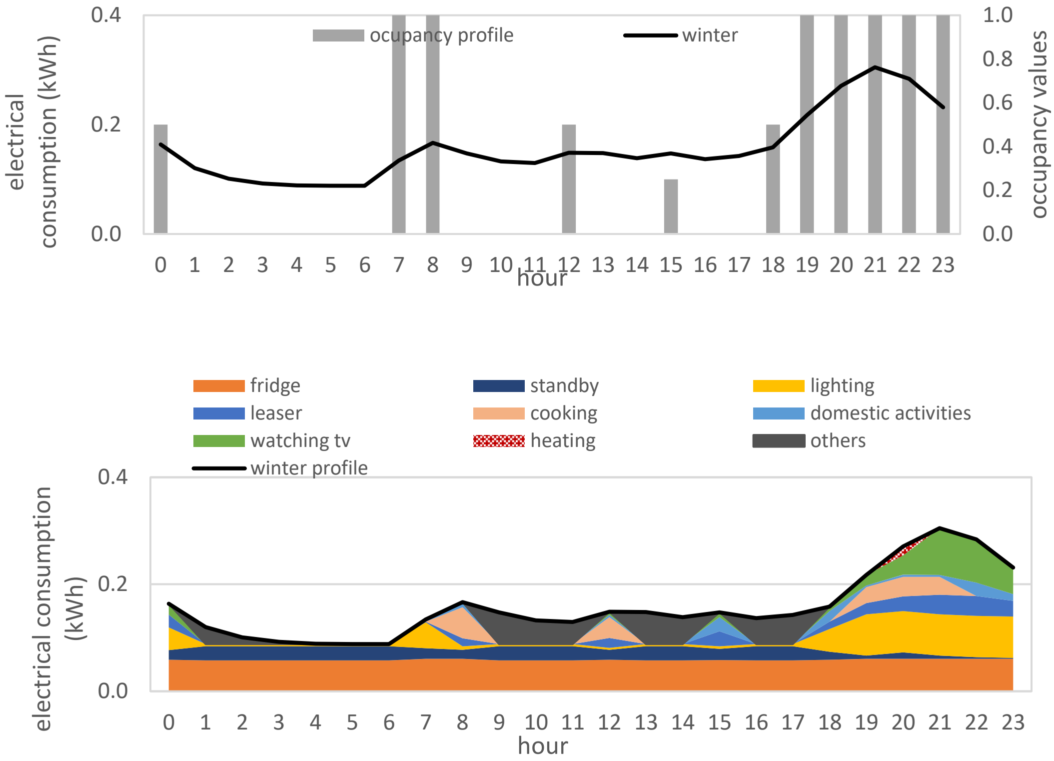


Figure 86 - Occupancy Profile (upper) and end-use profiles(lower) for winter month of path 212

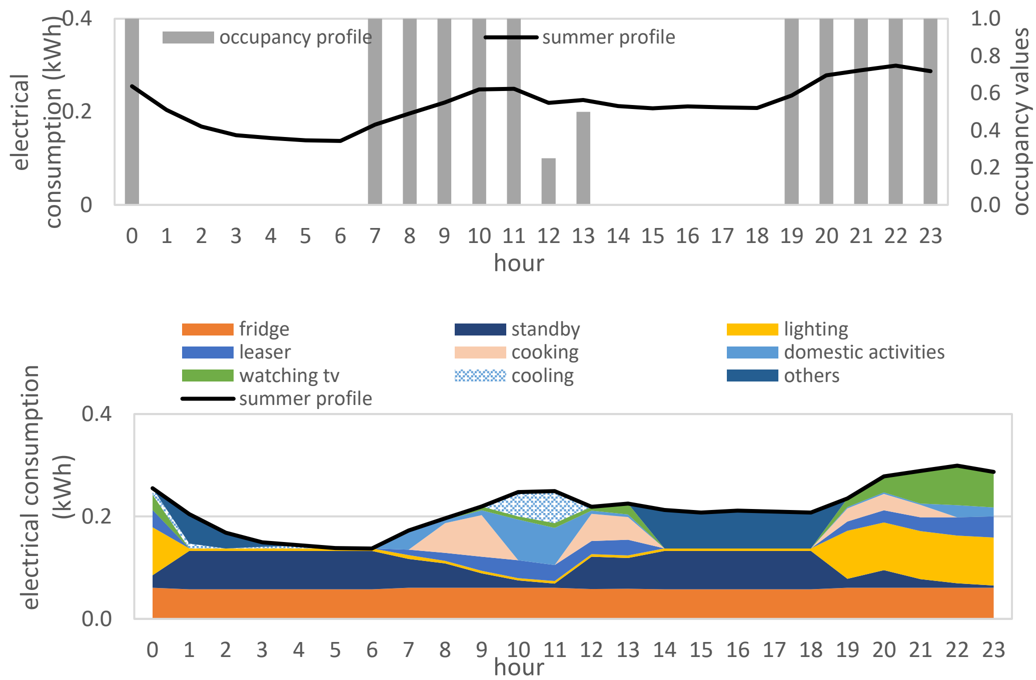


Figure 87 - Occupancy Profile (upper) and end-use profiles(lower) for summer month of path 212

E.2 Path 222

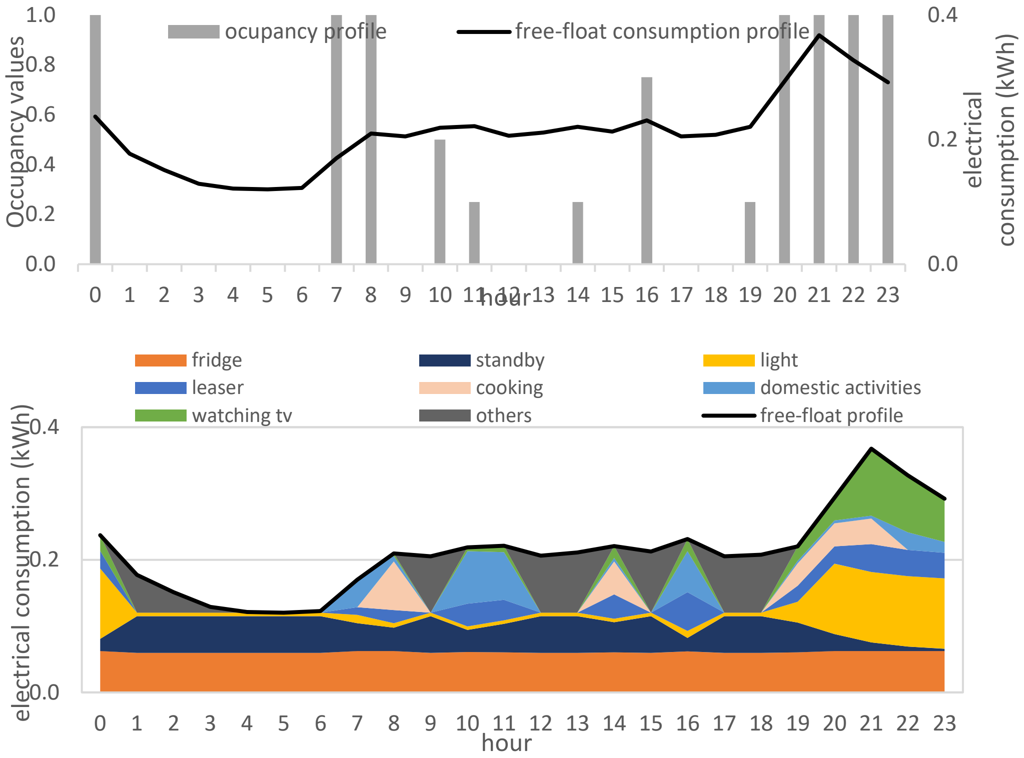


Figure 88 - Occupancy Profile (upper) and end-use profiles(lower) for free-float month of path 222

E.3

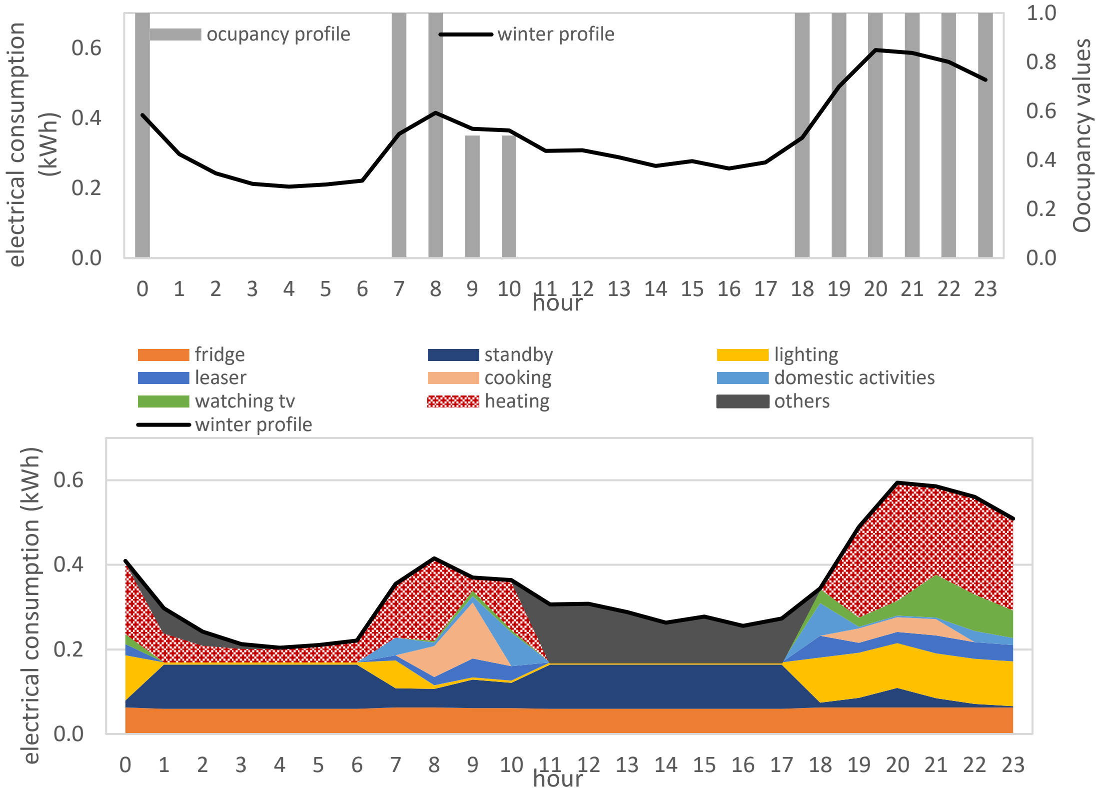


Figure 89 - Occupancy Profile (upper) and end-use profiles(lower) for summer month of path 222

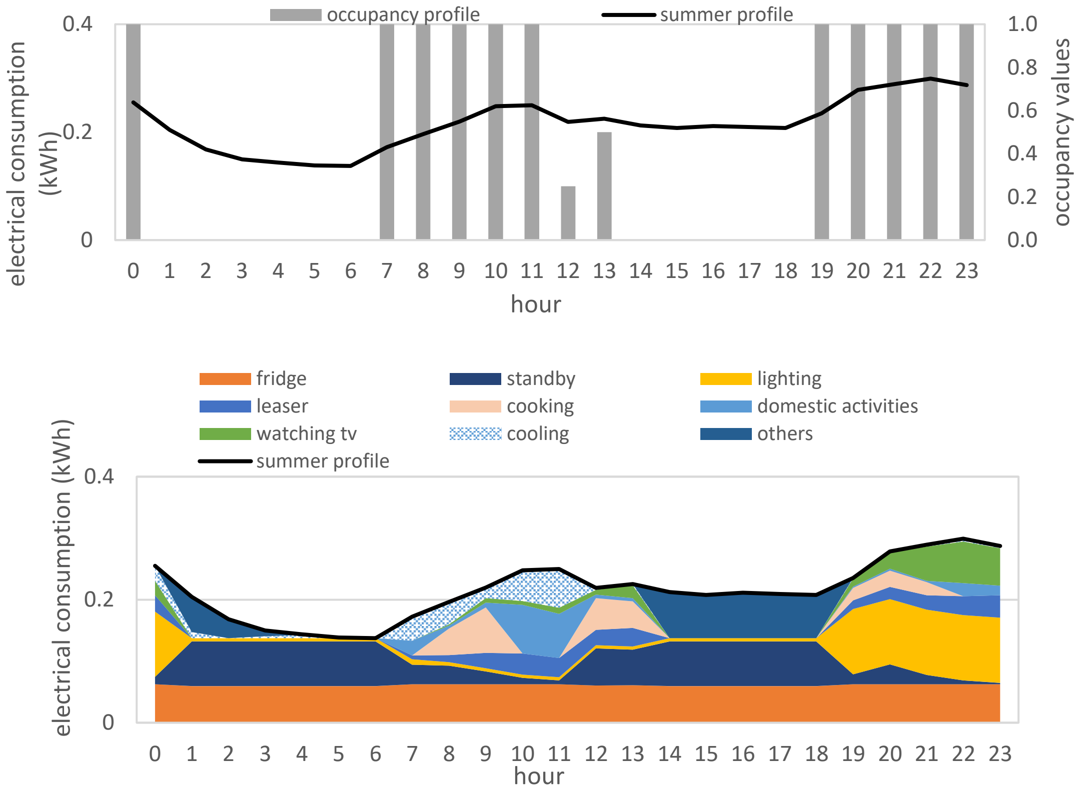


Figure 90 - Occupancy Profile (upper) and end-use profiles(lower) for summer month of path 222

E.3 Path 422

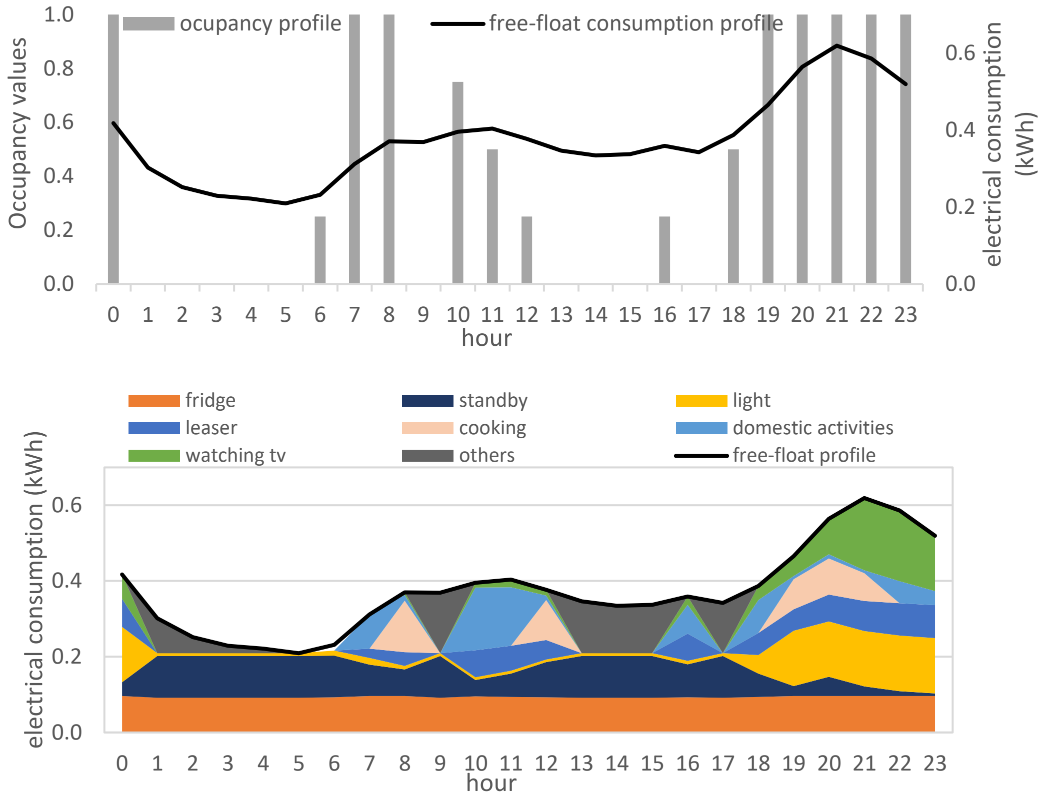


Figure 91 - Occupancy Profile (upper) and end-use profiles(lower) for free-float month of path 422

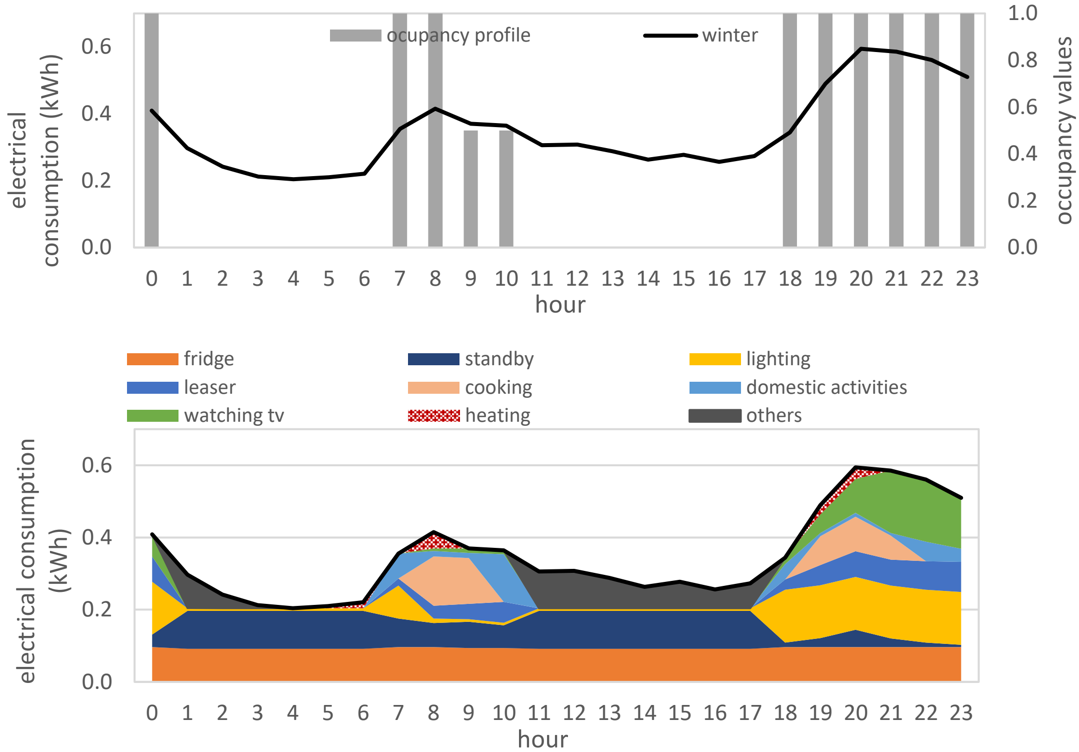


Figure 92 - Occupancy Profile (upper) and end-use profiles(lower) for winter month of path 422

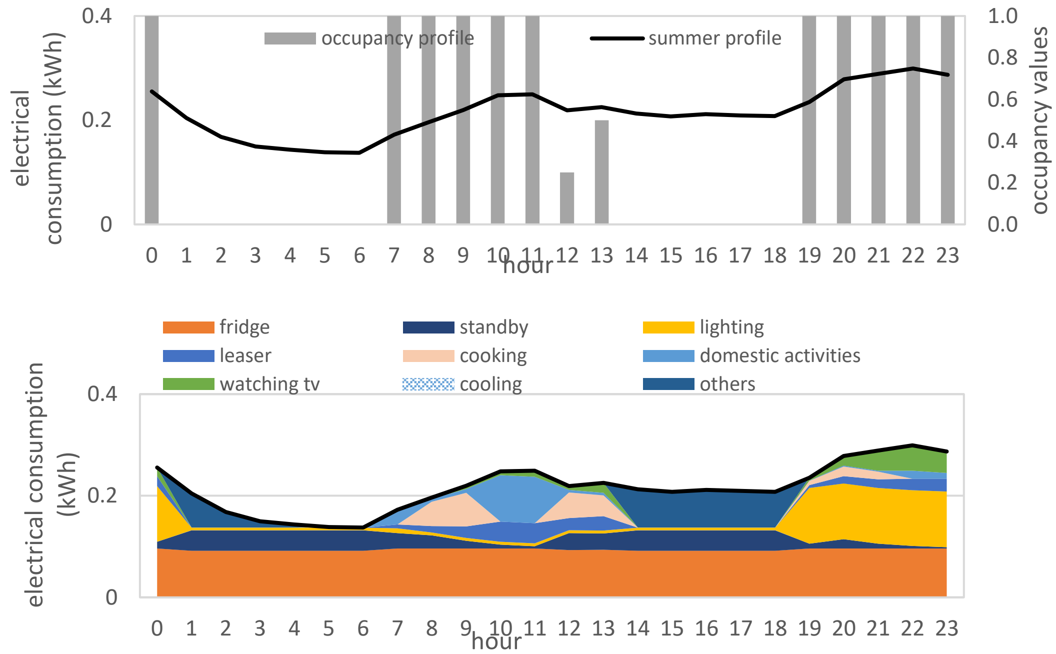


Figure 93 - Occupancy Profile (upper) and end-use profiles(lower) for summer month of path 422

E.4 Path 444

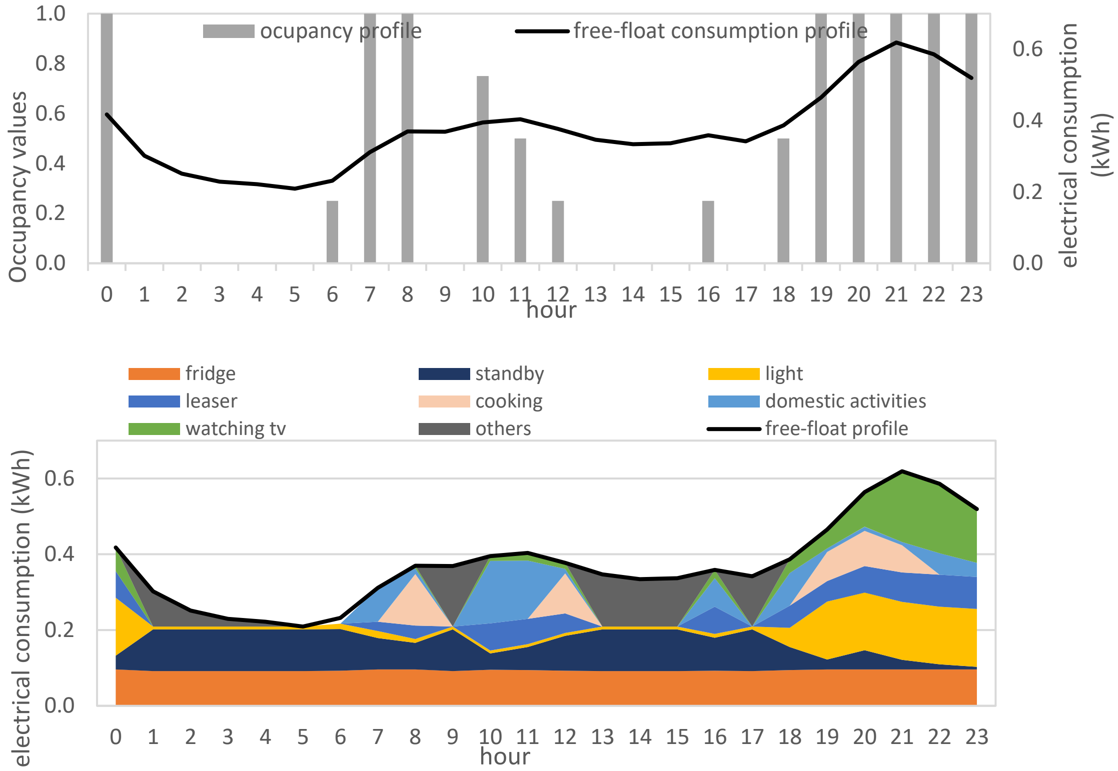


Figure 94 - Occupancy Profile (upper) and end-use profiles(lower) for free-float month of path 444

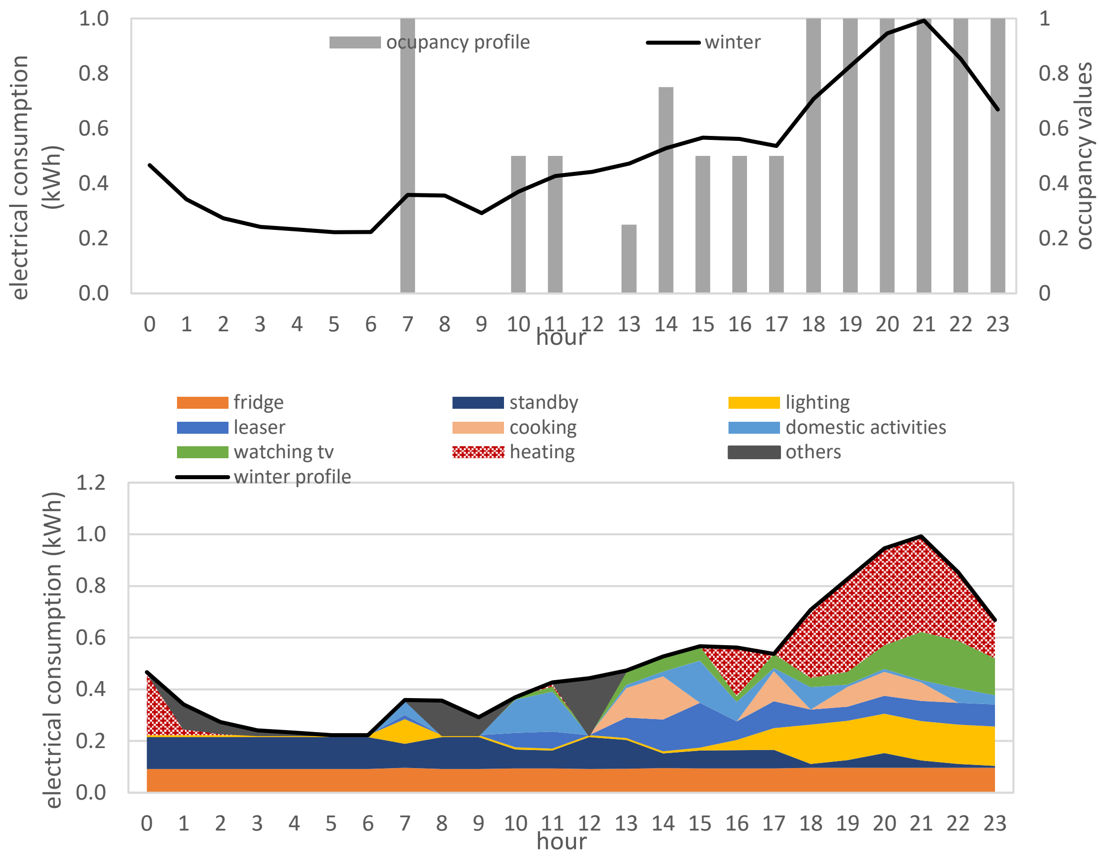


Figure 95 - Occupancy Profile (upper) and end-use profiles(lower) for winter month of path 444

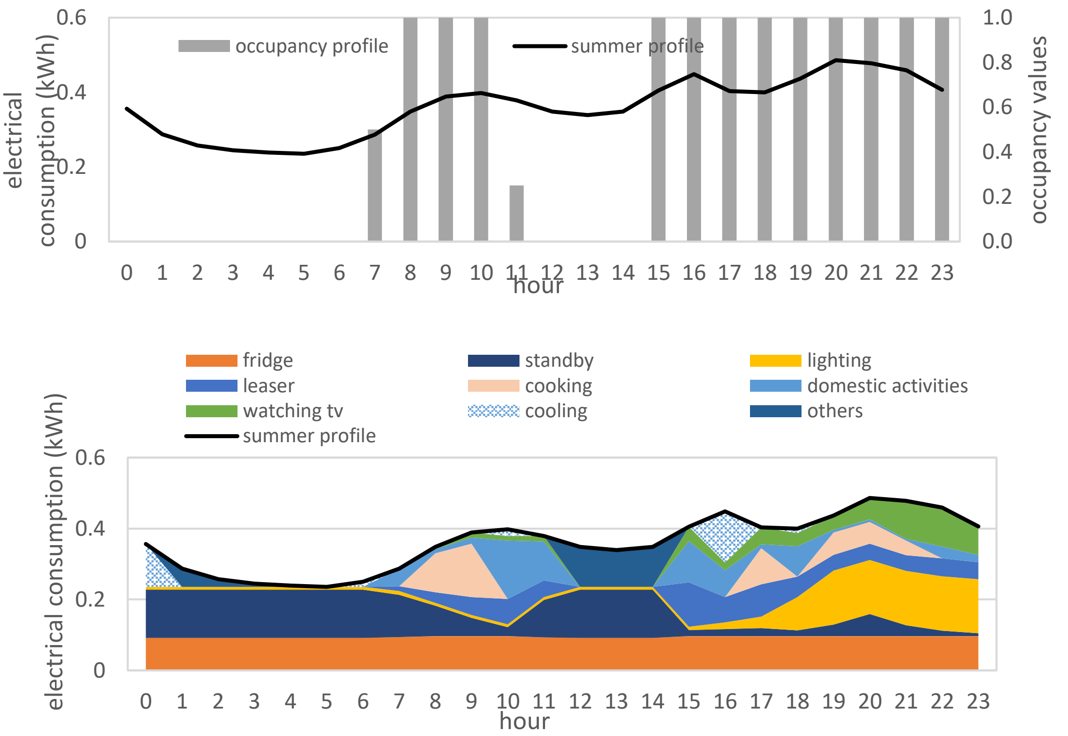


Figure 96 - Occupancy Profile (upper) and end-use profiles(lower) for summer month of path 444

Annex F - End-use consumption profiles from simulation

F.1 Path 212

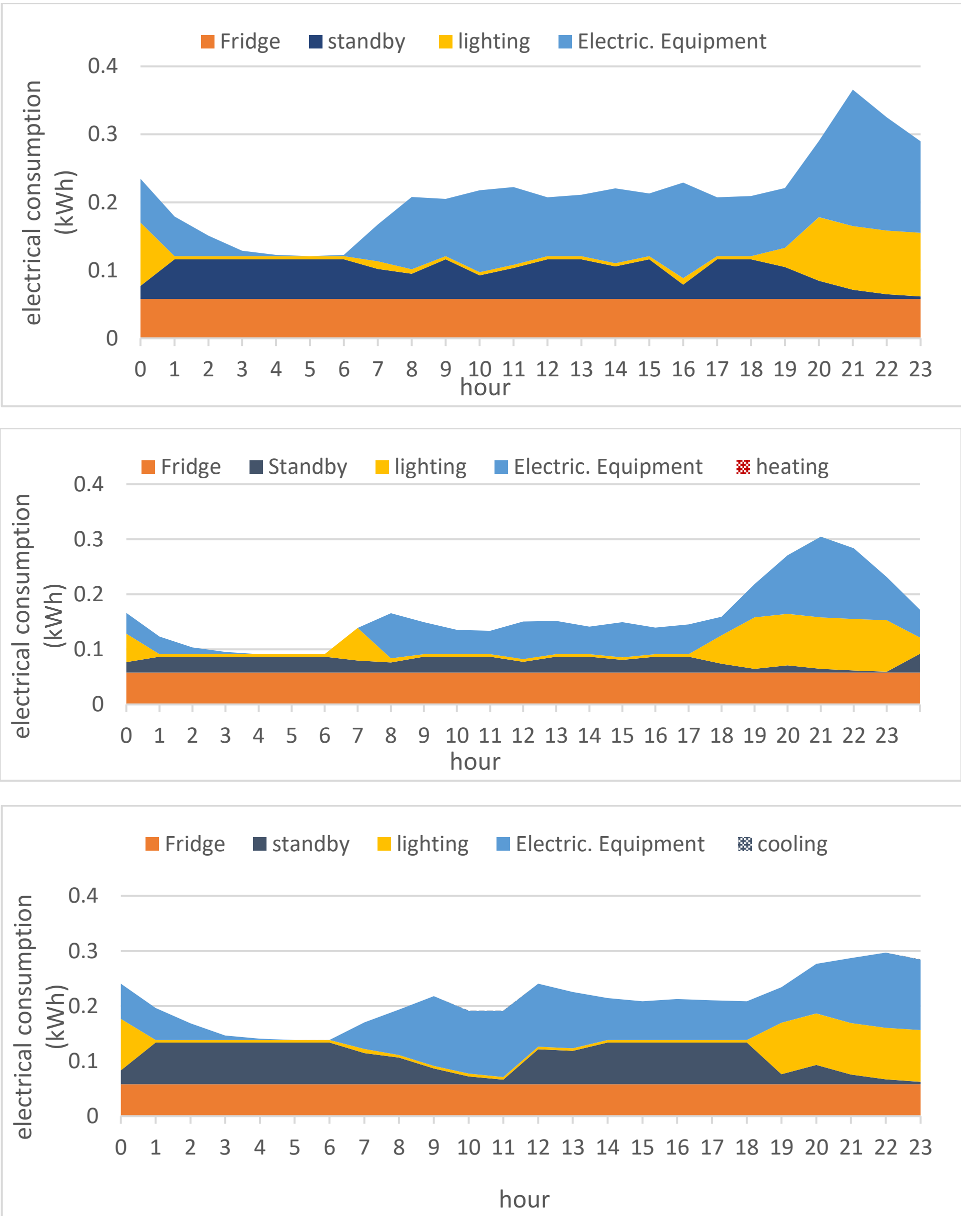
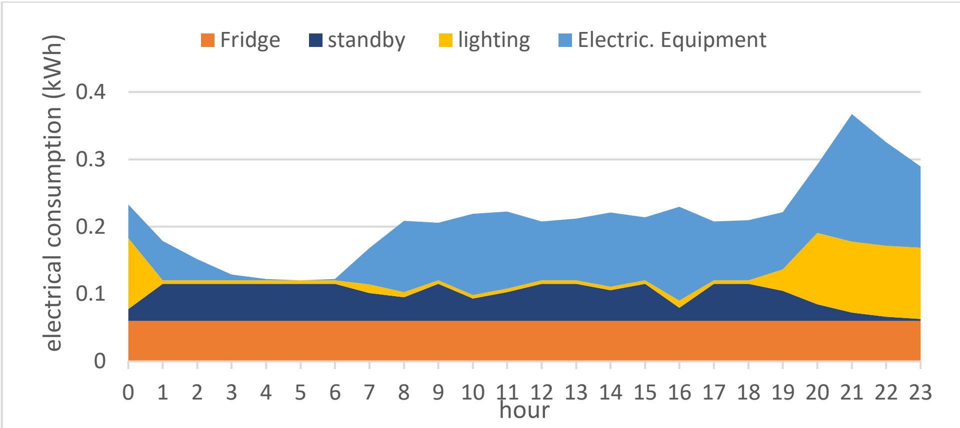
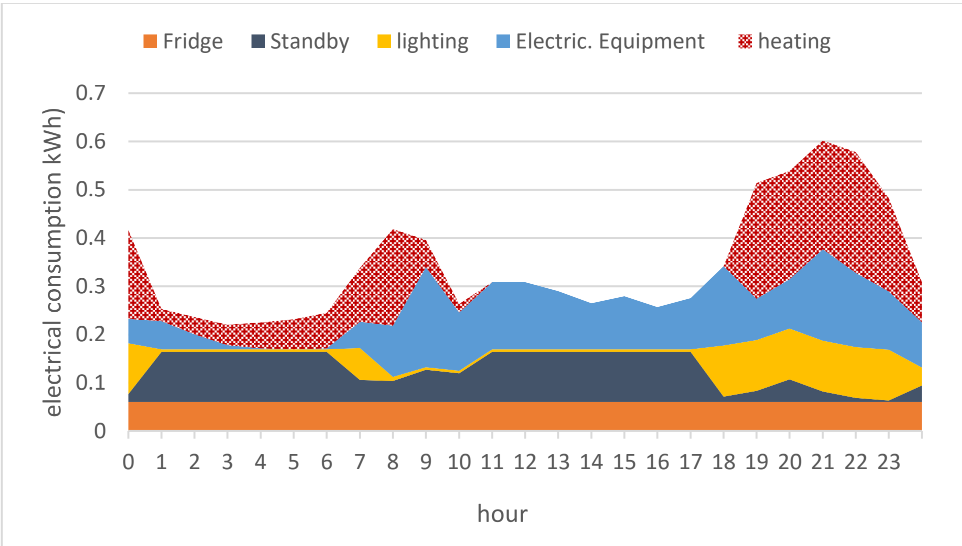


Figure 97 -End-use consumption in May (upper), January(middle) and July (lower) for path 212

F.2 Path 222





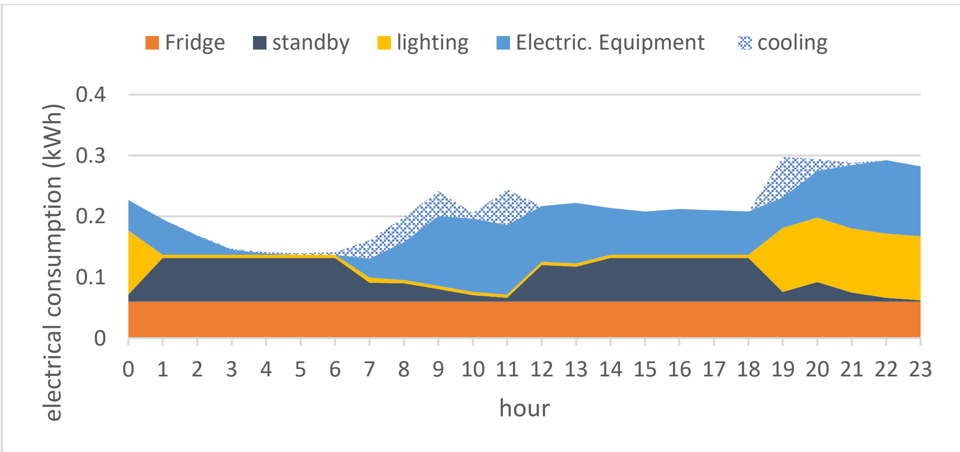
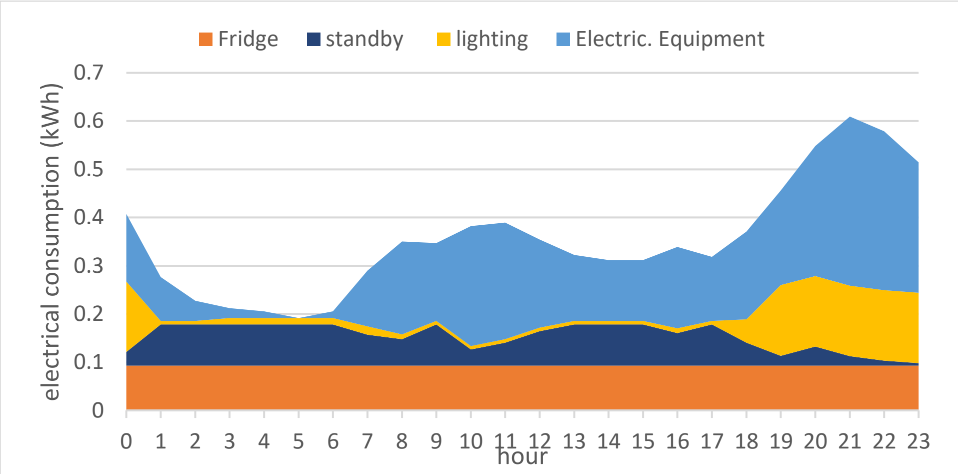
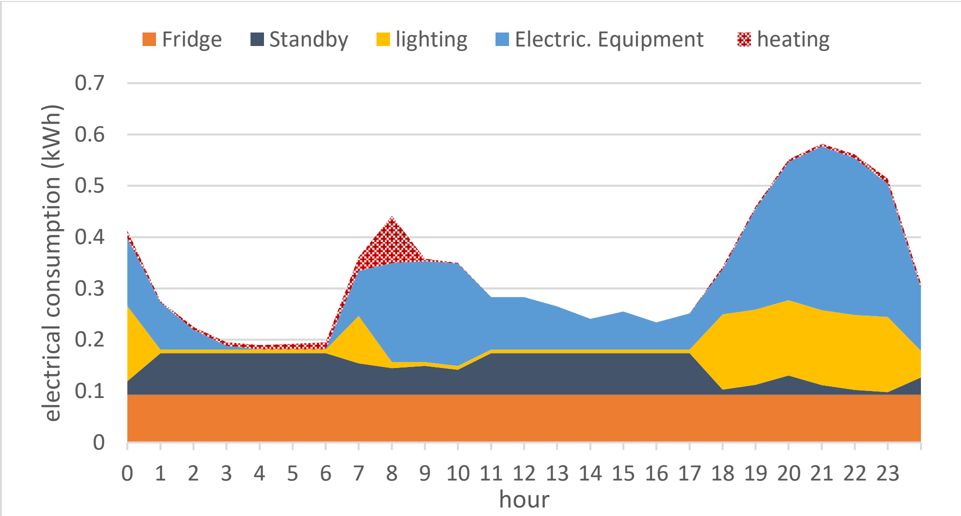


Figure 98 -End-use consumption in May (upper), January(middle) and July (lower) for path 222

F.3 Path 422





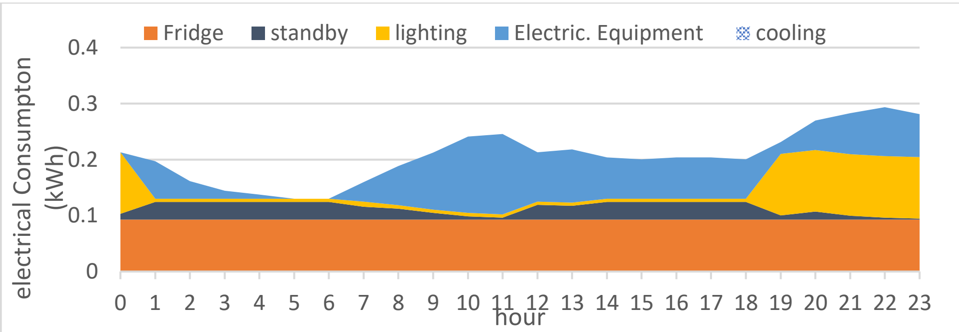
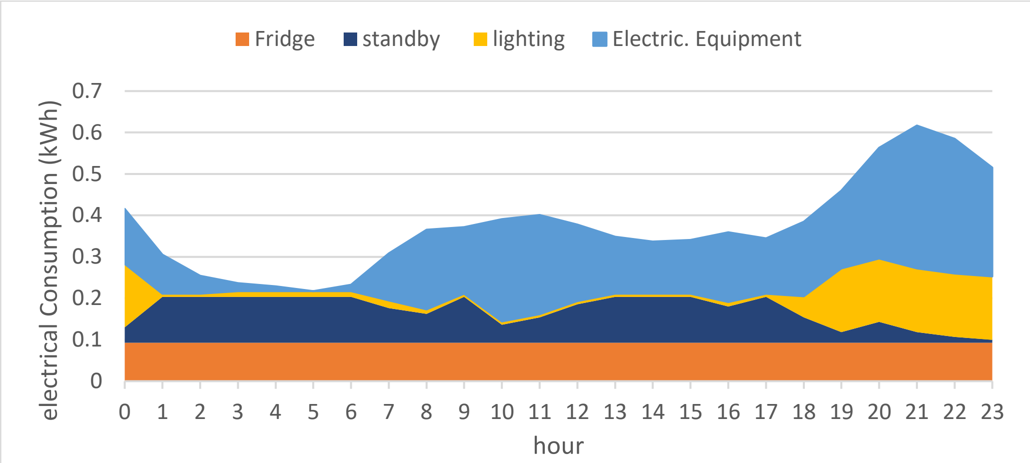
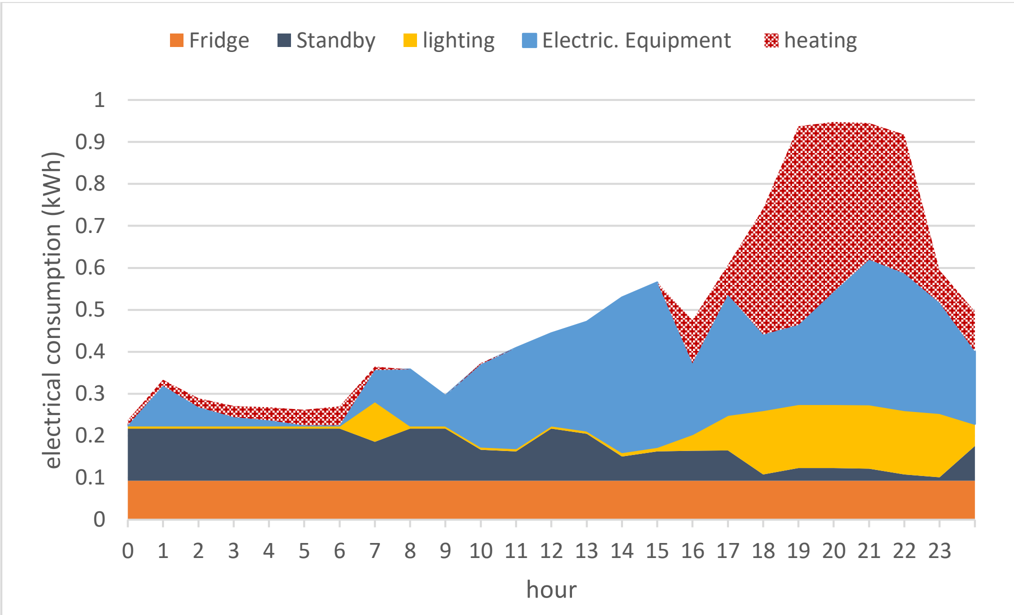


Figure 99 -End-use consumption in May (upper), January(middle) and July (lower) for path 422

F.4 Path 444





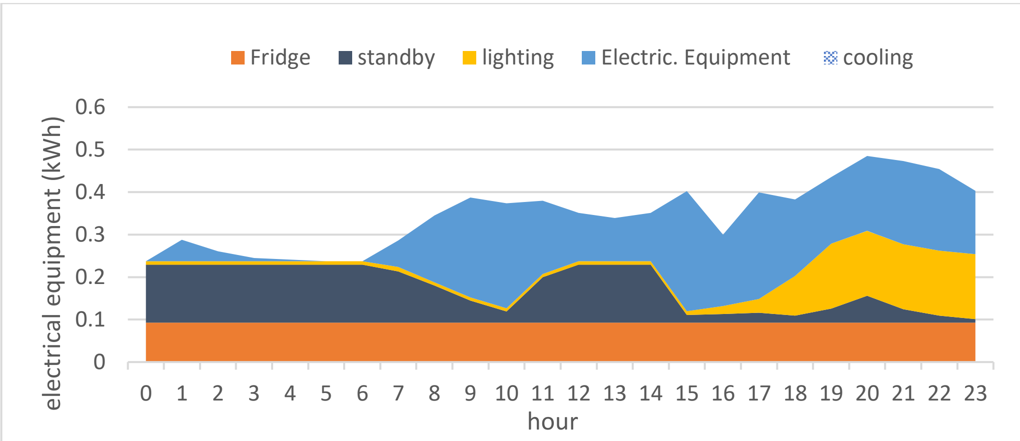


Figure 100 -End-use consumption in May (upper), January(middle) and July (lower) for path 444