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

Catarina Pargana Oliveira Pires

Catarina.pargana@ist.utl.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal

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

1 Introduction

The building sector is the largest single consumer of energy in Europe, accounting for 40% of the final energy consumption in 2012 and 38% of the EU's CO_2 emissions, having 66% of the buildings' total final energy consumption being represented by residential buildings [1]. 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 [2].

To be able to measure the impact of such actions one should 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 [1]. However, it remains challenging to acquire this type of information.

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 [3].

Since then, there has been an increasing use of smart meters data in current 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 [4]–[6]. Other studies focus only on the residential load characterization [7]–[10], or on inferring about the drivers behind the residential consumption, in terms of socio-economic status, appliances stocks and dwellings characteristics [11]–[13]. Finally, electric consumption disaggregation, appliances, lighting and plug load profiles distinction [14]–[17], as well as occupancy inference and inhabitants routines are other uses of smart meter data [18]–[21].

The objective of this research is to analyze the behavior of different residential electricity consumers to identify input operation parameters and run a complete dynamic building energy simulation. The clustering *k-means* method is applied to identify different occupant's profiles. By comparing different seasons, it is possible to infer about the inhabitant consumptions habits on heating and cooling energy use. A mathematical model is developed, where hourly occupancy profiles are inferred, with disaggregated lighting, equipment, heating and cooling consumption. A dynamic energy simulation is taken and the model accuracy is evaluated against real energy data.

The final 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.

2 Methodology

The methodology proposed in this work consists of six stages: 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. Stage 2 proposes a method to distinguish residential from non-residential consumption profiles, through clustering analysis. In Stage 3, residential users are selected and characterized by using clustering analysis in 3 reference months: a free-float month, a winter month and a summer month, to identify the changes in the consumption profiles thorough the 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 and 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.

3 Case study

The building under analysis in this thesis is located in Lisbon, more precisely in Parque das Nações neighborhood. The building geometric information used in this model was provided by the Municipal Archives. A typical dwelling representative of the existing dwellings in this neighborhood was chosen to test in the energy simulation. It is a 3-room apartment with a floorplan area of 148,5m^2 and a floor-to-floor height of 3m. Information about the building's envelope, ventilation and energy systems was retrieved from 18 Energy Performance Certificates (EPC) provided by ADENE. A thermal transmittance value of 0.54 W/(m^2.ºC) was considered for exterior walls, 1.16 W/(m^2.ºC) for interior walls and 2.9 W/(m^2.ºC) for windows. The equipment considered for space heating and cooling was a multi-split with an average SCOP of 2.6 and SEER of 2.5. Moreover, 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.

4 Model development

4.1 Data cleaning and preprocessing

To create a clustering model as accurate as possible, it is necessary to identify the inconsistencies in the initial dataset, and remove them from the dataset. After removing missing-value users, zero-use residential users and outliers the size of the final dataset was 90 users. A data treatment was taken with the objective to fill in the hourly electric consumption for the users with less than 10% of missing data. To do so, hourly values from the previous and next hour were used. If this data was not available, values for the same day of the week during the same month were considered.

4.2 Typification of energy consumption

As the objective of this research is to analyze residential electrical consumption, it is necessary to identify and eliminate the non-residential users. To accomplish this goal, the 365 daily profiles representing each user were combined into an average daily consumption profile. After performing data normalization with the maximum value, a clustering analysis using *k-means* algorithm was performed to identify different user type profiles. The Mean Square Error, the Calinski Harabasz and the Davies-Bouldin validity indexes were considered efficient [8][22] and were used as adequacy measures to determine the more appropriate number of clusters. From this analysis, the number of clusters considered appropriate was 5. As a result, the final cluster centroids are presented in Figure 1. Through the visual inspection it was possible to identify the cluster centroid *c2* and *c4* as non-residential 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 by *c1, c3* and *c5.*

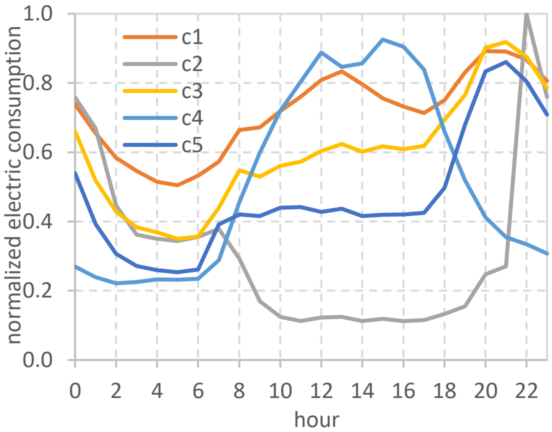

Figure 1- Final cluster centroids

4.3 Residential consumption characterization

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; a winter month, with heating needs and a summer month, with cooling needs. The selection of the appropriate months was done using the Heating Degree Days (HDD) and Cooling Degree Days (CDD) methodologies. However, other indicators as precipitation and solar irradiation were also considered. The reference months selected in this work were May for the free-float month, January for the winter month and July for the summer month. Considering these months, the dataset was filtered to perform another clustering analysis which proceeded in the same way as described in Section 4.2, with the only difference that the profiles are not normalized. In the end, 4 clusters centroids per reference month were identified and used to create different paths. These paths are the combination of the cluster centroids, representing the behavior of a certain number of users throughout the different seasons. From this analysis it was possible to distinguish: 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).

4.4 Input profiles determination

The paths identified in the previous section are used to create the inputs for the building energy simulation. A mathematical model was developed to identify the occupancy patterns, distinguish the consumption related to certain activities from the occupant's behavior, create the lighting profiles and to identify the use of electric consumption for heating and cooling (Figure 2). Electricity patterns change when occupants are present due to their interaction with the electrical loads which leads to Higher consumption periods and Turning on events. The method starts by identifying the startup and shutdown events [20] and identifying 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 then combined with a threshold of Higher consumption periods (by using the difference from hourly consumption and daily median consumption) and a threshold of Turning on events (by using the difference between the hourly consumption and its predecessor) to create an occupancy profile.

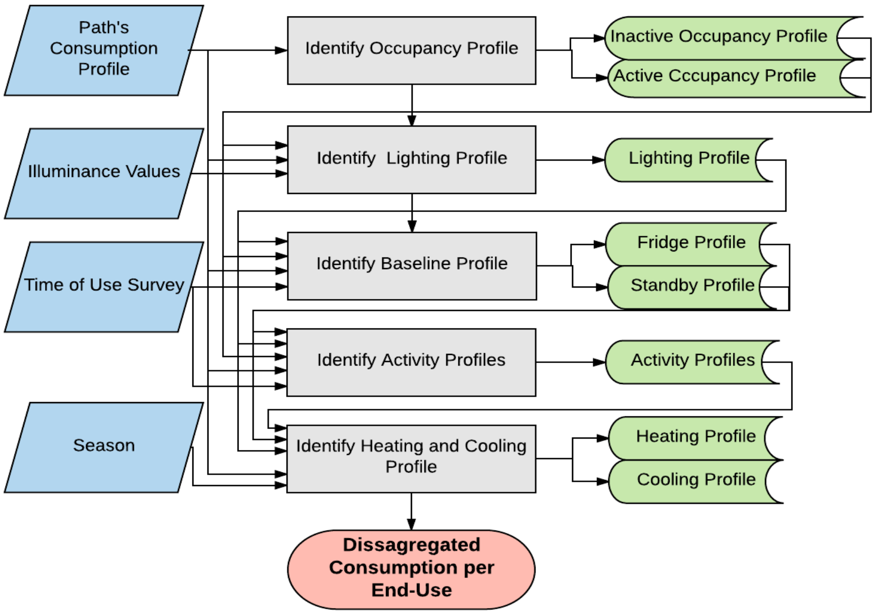

Figure 2 - Workflow of mathematical model

To determine the lighting profile, the active occupancy profile is combined with the illuminance values and the values of the installed lighting units calculated taking into consideration the Portuguese reference [23].The baseline profile corresponds to the minimum value of electrical consumption and does not have a direct relation to the occupancy. To calculate the maximum value of baseline consumption, the minimum value of the daily consumption profile is obtained. The first assumption is that the fridge has a steady consumption with an assumed 10% increase during occupied hours. Its consumption agrees with typically values attributed to refrigeration, presented in literature. To calculate the maximum value of standby, the maximum baseline consumption value previously obtained was combined with the fridge consumption and lighting consumption. To create the standby consumption profile, it was assumed that its consumption was inversely proportional to the total percentage of activities related to the electrical consumption. To do so, values from a Portuguese survey about the use of time are used [24]. Finally, the remaining values of electrical consumption, after removing the baseline and the lighting consumption, are distributed into activity, heating or cooling profiles. In the free-float month, the activity profiles are directly estimated while for the winter or summer months a percentage is used to report the heating or cooling use. This is achieved by comparing each activity consumption profile with the free-float values: if higher, the free-float values are used for winter and summer activity profiles and the remaining consumption is attributed to cooling or heating; otherwise the values are obtained directly for winter and summer.

4.4.1 Application to the case study

Due to the substantial number of paths, the path 111 will be used to demonstrate the application to the case study. The obtained active occupancy profile (columns) is closely related to the consumption profile (line) as showed in Figure 3.

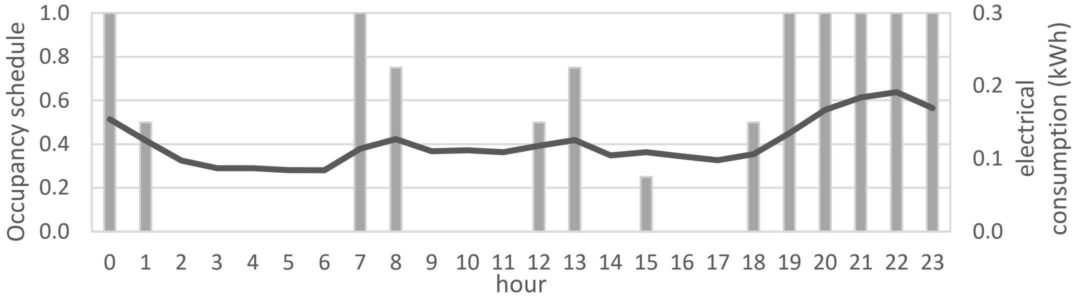

Figure 3- Free-float consumption profile and active occupancy profile

After applying the mathematical model described previously, the resulting consumption profiles, per end-use for the occupancy profile presented in Figure 3 is illustrated in Figure 4.

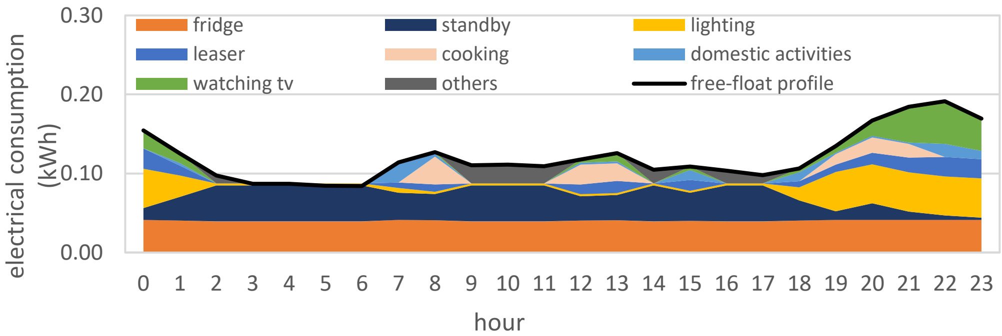

Figure 4 - Electrical consumption distributed by its end-use for the free-float month

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 and heating use. As a result, Figure 5 presents the electrical consumption distributed by its end-uses for the winter and summer months.

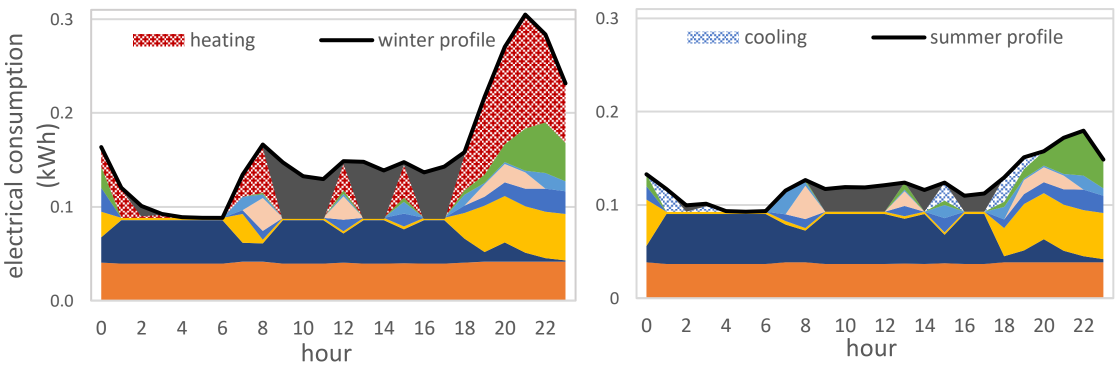

Figure 5 - electrical consumption distributed by its end-use for the winter (left) and summer (right)

4.5 Parametric building energy model

The energy simulation model is created using the *EnergyPlus* software. The profiles described previously are used as input. 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. As information about the number of occupants is not provided it was assumed a value of 100 Watts per occupant, based on the average yearly consumption per household and average members per household, this value was then compared with the maximum value of the consumption profile, from the free-float month. Peak loads from the lighting, fridge, standby and electric equipment consumption profiles are used as input and are combined into schedules (by dividing the hourly values by the peak loads). For the reference dwelling presented in the Case study, a 3D model is created (Figure 6). 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 shadows from the surroundings and from the upper floor balcony were also modeled.

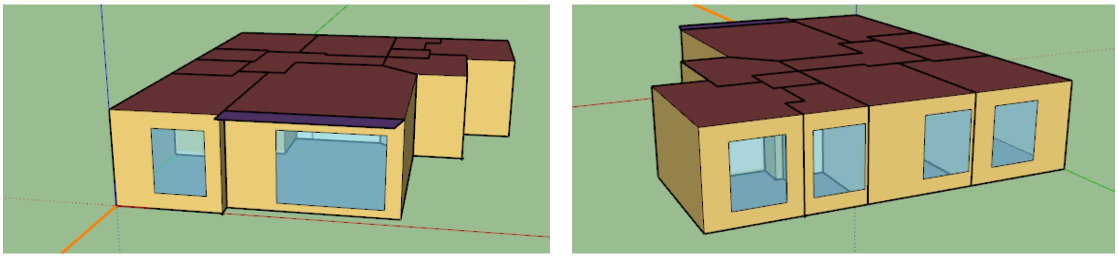

Figure 6- 3D facing north (left) and 3D facing south (south)

The value of 0.62 ach for ventilation was retrieved from the EPC and it includes 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%. The HVAC was modeled as an ideal load air system and 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.

4.6 Model evaluation

The validation of building energy simulation model is presently built on a model's compliance with the standard criteria for Normalized Mean Bias Error (NMBE) and the Coefficient of variation of the Root Mean Square Error (Cv(RMSE)). The consideration of both indices allows preventing any calibration error due to errors compensation, and the limit values were obtained from ASHRAE [25]: for the hourly criteria the maximum values are 30 % for Cv(RMSE) and NMBE of 10%. Two variables were chosen for calibration: the type of system used for cooling and heating; and the interior temperature setpoints.

5 Results and discussion

The profiles developed in the mathematical model in 4.4 are used as input for the building energy simulation described in 4.5, for the three reference months determined in 4.3. The model evaluation and calibration is taken considering the indexes described in 4.6 and the results are presented and discussed. After performing the 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.

All paths met the reference criteria for the free-float, winter and summer months (Figure 7). The free-float month had the lowest values, with the worst fitting path having only 1% of Cv(RMSE) and 0.23% of NMBE. Path 212 reported the best fitting for the winter month (Cv(RMSE) of 2.2 % and NMBE of 0.5%) due to the absence of heating consumption, while path 422 reported the best fitting for the summer month (Cv(RMSE) of 1.6 % and NMBE of 0.4%) due to the absence of cooling consumption. Path 444 reported the worst fitting, namely in the winter month (with Cv(RMSE) of 13.1% and NMBE 0%) and in the summer month (with Cv(RMSE) of 11.2% and NMBE 4%). This is due to higher heating consumption in winter and differences in the cooling use identified by the mathematical model and the simulation model.

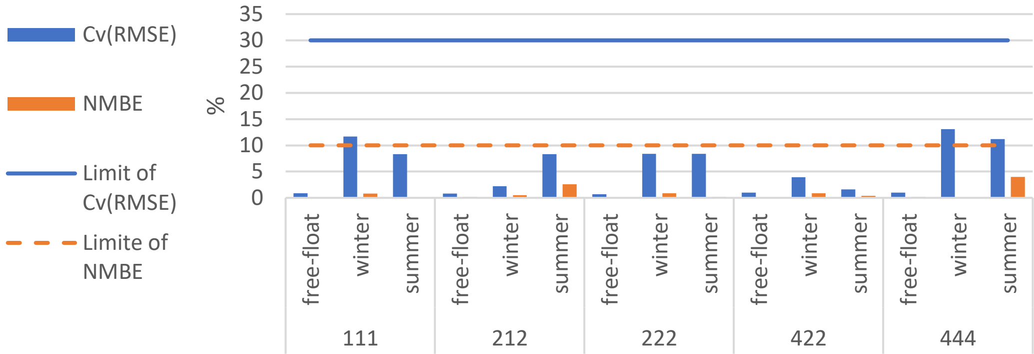

Figure 7-CVRMSE and NMBE values per path compared with Limits for each index

Figure 8, illustrates the winter and summer hourly fits for path 111, this path was chosen since it is the path with the second worst results and uses equipment's for cooling and heating. For winter, 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 and that the opposite occurs between 2h to 8h. However, the total daily consumptions from measured and simulated data only differ by 0.79%. For the summer month it is possible to observe that the fitting presents some diversity (26.06% of difference in the worst case, which occurs at 18h). Nonetheless, the total daily consumptions from measured and simulated data only differ by 0.1%. From the end-use disaggregation 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. 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.

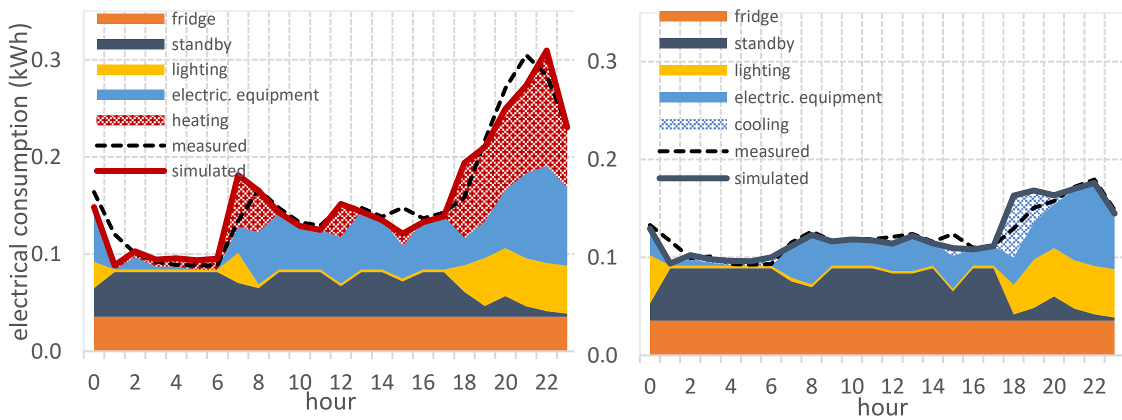

Figure 8 - Comparison between simulated and measured average daily consumption profile for path 111. Winter month (left), summer month (right)

6 Conclusions

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 is coupled with data on buildings geometry and construction as inputs for an energy simulation model.

A clustering analysis using the *k-mean* algorithm was taken with different purposes: to identify different user types and exclude non-residential users and to characterize different residential user profiles for three reference months: free-float month (May); winter month (January) and summer month (July). For each month, 4 cluster centroids were obtained and *paths* describing typical users were created. Five paths were analyzed, representing around 60% of the entire sample. The simulation results were evaluated and a calibration procedure was taken with interior temperature setpoints and heating and cooling systems efficiencies as uncertain variables.

Overall, the mathematical model proved to be effective for the simulation and modelling of the electrical consumption for different user types. All paths met the criteria for the free-float, winter and summer months. However, in the winter and the summer months the hourly fit was not as good as the one achieved in the free-float month. Nonetheless, the difference between the total daily consumption from measured and simulated data was very small: 0.22% for the free-float month, 0.01% for the winter month and below 4% for the summer month, suggesting that even if the measured and the simulated profiles do not have a perfect hourly match, the total consumption still presents a good approximation. Therefore, it is possible to conclude that the proposed method is effective in simulating the electrical consumption of different types of residential users and can be used to test energy efficiency measures in future scenarios.

Future work addresses the improvement of the estimation 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 of the identification of high consumption periods and on events as a continuous evaluation proportional to the correspondent threshold, will be considered. The identification of uncommon occupancy patterns by the mathematical model is also an improvement to consider. Further developments on the model calibration by including an uncertainty analysis considering more variables can also improve the model predictions.

7 References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] E. Delzendeh, S. Wu, A. Lee, and Y. Zhou, "The impact of occupants' behaviours on building energy analysis: A research review."

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

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

[14] 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.

[15] 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.

[16] 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.

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

[18] 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.

[19] 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.

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

[21] W. Kleiminger, C. Beckel, and S. Santini, “Household Occupancy Monitoring Using Electricity Meters,” 2015.

[22] 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.

[23] INE and DGEG, “Inquérito ao Consumo de Energia no Sector Doméstico 2010.” 2011.

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

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