Title	Modelling Household Occupancy Profiles using Data Mining Clustering Techniques on Time Use Data
Authors(s)	Buttitta, Giuseppina, Neu, Olivier, Turner, William J. N., Finn, Donal
Publication date	2017-08-09
Publication information	Buttitta, Giuseppina, Olivier Neu, William J. N. Turner, and Donal Finn. “Modelling Household Occupancy Profiles Using Data Mining Clustering Techniques on Time Use Data,” August 9, 2017. https://doi.org/10.26868/25222708.2017.478.
Conference details	Building Simulation 2017, San Francisco, CA, August 7-9 2017
Item record/more information	http://hdl.handle.net/10197/9218
Publisher's version (DOI)	10.26868/25222708.2017.478

Modelling Household Occupancy Profiles using Data Mining Clustering Techniques on Time Use Data

Giuseppina Buttitta[1], Olivier Neu[2], Will Turner[3], Donal Finn[3]
[1]Electricity Research Centre, University College Dublin, Dublin, Ireland
[2]School of Electric and Electronic Engineering, University College Dublin, Dublin, Ireland
[3]School of Mechanical and Materials Engineering, University College Dublin, Dublin, Ireland

Abstract

A strong correlation exists between occupant behaviour and energy demand in residential buildings. The choice of the most suitable occupancy model to be integrated in high temporal resolution energy demand simulations is heavily influenced by the purpose of the building energy demand model and it is a trade-off between complexity and accuracy. The current paper introduces a new occupancy model that produces multi-day occupancy profiles and can be adaptable to various occupancy scenarios (e.g., at home all day, mostly absent) and scalable to different population sizes. The methodology exploits data mining clustering techniques with Time Use Survey (TUS) data to produce realistic building occupancy patterns. The overall methodology can be subdivided into two steps: 1. Identification and grouping of households with similar daily occupancy profiles, using data mining clustering techniques; 2. Creation of probabilistic occupancy profiles using 'inverse function method'. The data from the model can be used as input to residential dwelling energy models that use occupancy time-series as inputs.

Introduction

The impact of occupant behaviour and building characteristics on domestic building energy demand has been investigated by various studies such as McLoughlin et al. (2012) and Guerra Santin et al. (2009). Both papers note that total energy consumption and maximum instantaneous energy consumption are mainly influenced by the dwelling type, envelope characteristics and occupant characteristics. In particular, the occupant characteristics which strongly influence energy consumption are the age and income level of the occupants, but also the household size (i.e. the number of occupants resident within the building). Additionally, in Guerra Santin et al. (2009), it was found that one of the parameters with the highest weight in determining heating loads is the amount of time the dwelling is occupied. Famuyibo et al. (2012) showed, through a linear regression analysis, that the total energy consumption in a dwelling mainly depends on the typical weekly occupancy pattern, together with the internal air temperature, air change rate and immersion heater use frequency. Moreover, the impact of the occupant behaviour on energy consumption is predicted to increase as the energy performance of dwellings increases (De Meester et al. (2013), Neu et al. (2014)). From these studies it is clear that a large variety of occupant characteristics contribute towards the variability of energy and power demand of domestic buildings. Many authors developed their domestic total energy demand models which include some of the household characteristics which can be correlated with energy demand. Shimoda et al. (2004), for example, classified household members by gender, age and occupation. However, even occupants with the same characteristics (household size, income level, age, etc.) can exhibit completely different daily power demand profiles. This can occur even if the daily total energy consumption is the same (Grandjean et al. (2012)). According to these results, modelling occupant behaviour is one of the main challenges to be overcome in the modelling of energy load profiles of housing stocks. For this reason, presently this research topic is being addressed by a growing number of researchers, as reported in the IEA-EBC Annex 66: Definition and Simulation of Occupant Behavior in Buildings IEA-EBC Annex 66 (2016). The choice of the most suitable occupancy model is strongly determined by the modelling aim, how the model will be integrated, and the desired level of complexity and/or accuracy (Gaetani et al. (2016), Feng et al. (2015)). In the current paper a new occupancy model is described. This model has been developed such that it can be integrated with scalable bottom-up domestic energy demand model, based on the archetype approach. In the archetype approach, the housing stock is represented by 'building archetypes', which represent groups of buildings with similar properties (Reinhart and Cerezo Davila (2016)). The energy consumption estimates by modelling building archetypes can be scaled to be representative of large building stocks by multiplying the energy consumed by each archetype by the number of houses within the modelled building stock represented by each archetype. For this purpose, some occupancy model requirements are identified. First, the occupancy model must be scalable in order to produce occupancy profiles for building stocks of different sizes. Modelling different building stocks implies modelling building stocks which can be completely different, for example the dwellings which belong to the building stock can be characterised by different occupancy profiles (i.e. in some building stocks most of dwellings could be unoccupied during the day, while in other building stocks dwellings could be constantly occupied). Thus, the second characteristic of the occupancy model is the adaptability to various building stock characteristics. The third requirement of the occupancy model is the production of realistic multi-day occupancy profiles, which can be obtained by seamlessly aggregating daily sequences of occupancy profiles. This characteristic is necessary to reflect the presence of daily routines which are likely to be repeated everyday by households. Moreover, the model must be capable to capture the temporal variability of the domestic energy demand, so it needs to present high temporal resolution. The model must be suitable to be used in large building-stocks (regional or national scale). In this case, the balance between complexity and accuracy is of crucial importance. To the best of the authors' knowledge, none of the existing building occupancy models exhibit all of these required characteristics. The aim of the current paper is to present a novel occupancy model having all the aforementioned identified requirements. This model is developed to be integrated with a high temporal resolution scalable bottom-up domestic energy demand model.

Literature review

Existing occupancy models can be classified into two main classes: deterministic models and probabilistic models. Deterministic models use fixed, a priori occupancy schedules. These models have the lowest level of complexity and they are commonly used in proprietary building energy simulation software. The deterministic approach was used by Yao and Steemers (2005). In their research, five fixed common occupancy profiles for UK households were proposed, based on the experience of the authors. The advantage of the deterministic approach is its simplicity, but its generality means that the resultant occupancy schedules do not accurately represent the diversity of occupancy presence in reality. Probabilistic models enable the consideration of the variability and stochastic nature of occupancy presence. The models are based on the probability that an event, such as an occupant returning from work, occurs or does not occur. From an exhaustive review (Grandjean et al. (2012)), it is stated that the most reliable probabilistic models are based on Time Use Survey (TUS) data. TUSs record the everyday routine of the respondents in detailed daily diaries. The key existing occupancy models which are based of TUS data are: Walker and Pokoski (1985), Richardson et al. (2008), Widén and Wäckelgård (2010), Wilke et al. (2013) and Aerts et al. (2014). The main characteristics of each model are reported below:

Walker and Pokoski (1985) This is the first model that uses the data available from Time Use Survey to construct a load profile. The households occupant probability to be home is given by the 'availability function'. This function considers the deviation of the time of the activities from typical times on a day-to-day basis. It is postulated that the deviation has a normal distribution about the typical times, with a standard deviation which depends of the activity considered and the geographical area considered. The geographical area also determines typical times when each activity is performed.

Richardson et al. (2008) This model is based on the Time Use Survey (TUS) conducted in the UK in 2000. The model uses a first-order MarkovChain Monte Carlo technique to generate statistical, high time-resolution occupancy profiles. The TUS data are categorised by the number of household residents and the day type (weekday/weekend). Thus, this methodology produces occupancy profiles differentiated according to these two characteristics.

Widén and Wäckelgård (2010) This model uses the Time Use Survey data collected by Statistics Sweden in 1996. Similar to the Richardson model, the Widen model also uses a first-order MarkovChain Monte Carlo technique to produce statistical occupancy profiles having high time resolution. This model differentiates daily occupancy profiles according to the day type (weekday/weekend) and the type of house occupied by the household (detached or apartment). This model has a good ability to differentiate occupancy profiles that characterise a detached house from an apartment on the one hand, a weekday from a weekend day on the other.

Wilke et al. (2013) This model is based on the French Time Use Survey, realised between February 1998 and February 1999. The model adopts a higher-order Markov process to bypass the problems given by the MarkovChain Monte Carlo, where the current state is only dependent on the previous state. This model is based on time-dependent probabilities to start activities and their corresponding duration distributions.

Aerts et al. (2014) This model uses the data from the Belgium Time Use Survey which was conducted in 2005. The model introduces the concept of 'typical occupancy patterns', which are identified through the application of a hierarchical clustering technique on individual occupancy profiles. The probabilistic occupancy profiles are obtained applying the probability to transit from a certain state to another and the duration probability, which are both time-dependent. The model is able to produce individual occupancy profiles differentiated according to the 'typical occupancy patterns'. Additionally, the paper has the merit of presenting a methodology to construct realistic yearly household occupancy sequences, considering consistency from day-to-day profiles.

Considering the scenario where occupancy models need to be integrated into scalable bottom-up domestic energy demand models able to simulate the domestic energy demand of large building stocks with high temporal resolution, each of the models reviewed and described previously has limitations. Most of the models which are based on TUS data differentiate households just according to the number of occupants (Richardson et al. (2008), Widén and Wäckelgård (2010) and Wilke et al. (2013)). This differentiation is not sufficient because it produces large size categories, containing households characterised by significantly different occupancy profiles. These models therefore generate occupancy profiles that are an unrepresentative composite of multiple combined occupancy profiles, which cannot reproduce specific occupancy scenarios. Thus, these models are not adaptable to building stocks characterised by diversified occupancy scenarios. Moreover, they generate occupancy profiles valid for individual days, which can be aggregated later to obtain occupancy profiles for consecutive days. However, the multi-day occupancy profiles obtained in this way do not consider patterns of consistency from day to day (Richardson et al. (2008)). Day to day profile consistency should recognise, for example, that the same household behaviour is likely to be repeated during the working days because of the daily working routine. The two models that can be successfully used to capture dwellings characterised by specific household occupancy scenarios are the ones developed by Walker and Pokoski (1985) and Aerts et al. (2014). The Walker and Pokoski (1985) model is the most complete model. However, this model is based on an extensive and detailed amount of data which can be difficult to collect. The Aerts et al. (2014) model is capable to produce profiles which can be very diversified for different households, and it is also computationally more efficient than other models. Moreover, the problem of the consistency of the multi-day simulations is faced. Unfortunately, this methodology is based on the assumption that the household individual occupants act independently, which is unlikely (Aerts et al. (2014)). With this approach, the relationships between the occupancy presence of the members of the same household are not considered. Given this background, the current research focuses on the development of a novel and simple methodology that combines the advantages of all the mentioned models.

Methodology

The aim of the presented methodology is the creation of multi-day realistic household occupancy profiles which can be integrated into scalable energy-use models of building stocks. An overview of the methodology is given in Figure 1 and is outlined in detail in this section.

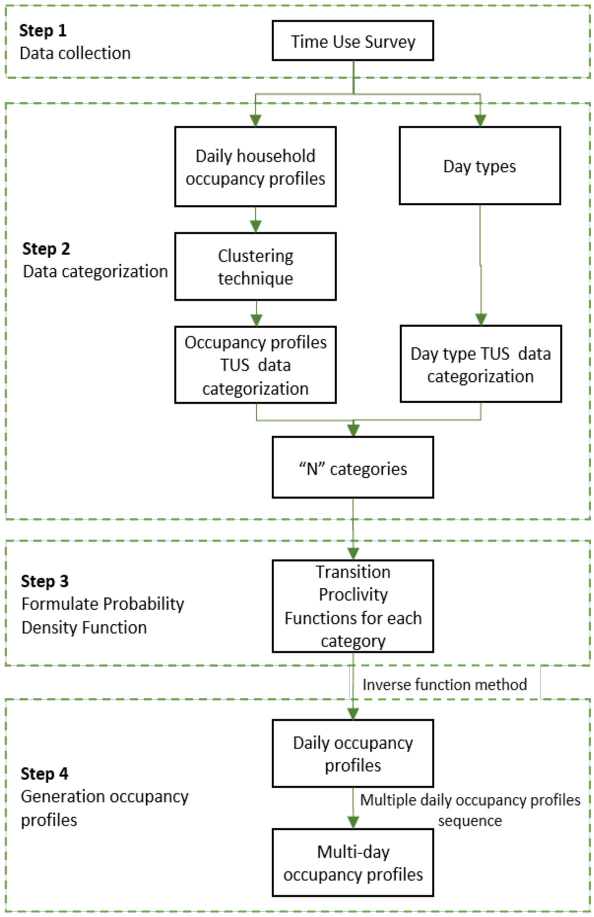

Figure 1: Flow chart of the overall methodology.

For clarity, the terms which are used in the methodology are defined as follows: 'active occupant' indicates a person who is in the house and not asleep; 'non-active occupant' defines a person who is in the house and asleep; 'absent' indicates a person who is not in the house; 'occupant state' indicates the condition of the person who occupies the house, which can be active, non-active or absent; 'household state' indicates the condition of the entire household, which depends on the individual occupant state; 'household daily occupancy profile' indicates the sequence of the household states during one day. In this methodology, the household daily occupancy profiles generation is based on the 'inverse function method', used also in previous stochastic model for the simulation of the occupancy presence (Page et al. (2008)). The inverse function method can generate a sample (in our case, the daily time series of household states) of events from a given probability distribution function (PDF). The first step of the methodology presented in the paper is the collection of the data. In particular, in this paper the data available from the UK 2000 Time Use Survey (UK TUS 2000) (Ipsos-RSL, Office for National Statistics (2003)) are used and thus they are presented in the following section. For the second stage, the UK TUS 2000 data are categorised according to the day type (weekday or weekend day) and the household daily occupancy profiles, adopting a previously developed clustering technique (Buttitta et al. (2016)). The third step of the methodology is the determination of the PDFs for each category, which are needed to apply the inverse function method. In the last step of the methodology, the PDFs are used to obtain household daily occupancy profiles. These profiles can also be aggregated in sequences to obtain multi-day occupancy profiles.

Data Collection-Time Use Survey (Step 1)

The presented methodology is based on data available from the UK 2000 Time Use Survey (UK TUS 2000) (Ipsos-RSL, Office for National Statistics (2003)). UK TUS 2000 is the first official national survey which was conducted in UK in the period between June 2000 and September 2001 and it was designed to achieve a representative sample of the population in UK. This survey records the everyday routines of 11,700 UK citizens aged above 8, belonging to 6,500 households. One household is defined as a person or group of people who have specified the accommodation as their only or main residence. The respondents were asked to complete a household questionnaire, an individual questionnaire, and two diaries - one for a weekday and another for a weekend day. In the diaries, respondent activities were recorded every ten minutes, from 04:00 to 04:00 of the successive day. In each time slot, occupants indicated their primary and secondary activities, the location of the activity and whom the survey respondent was with. In many cases not all the occupants of the same household completed the diary. For this reason, the data of some households were discarded, and just the data from 4,588 households were used. The household questionnaire allowed the determination of the number of members of each household, which defined the household size, and the type of accommodation in which they lived. From these data it is possible to understand the breakdown of different household size by type of accommodation, which can then be associated to a defined archetype. Figure 2 shows household size (in percent) by dwelling type, identified in the UK TUS 2000.

Data categorisation (Step 2)

The methodology used is based on the assumption that energy load profiles are mainly influenced by the household occupancy profiles, and in particular by the succession of occupied and unoccupied periods,

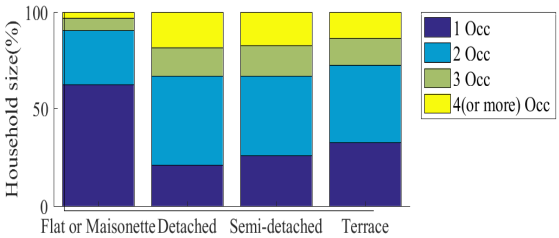

Figure 2: Distribution of different size households in each accommodation type.

and the number of household occupants. This relationship, already proved by previous researches (Yao and Steemers (2005); Famuyibo et al. (2012)), is due to the fact that only when occupants are home it is necessary to assure their thermal comfort in the house. Moreover, most of the electric appliances are used only when occupants are awake. Due to the strong impact of household occupancy profiles on energy load profiles, the household diaries are grouped according to the household daily occupancy profiles. Moreover, as a common behaviour, the occupancy profiles during the weekdays differ from the occupancy profiles during the weekend. For this reason the household diaries are grouped also according to the day type. The categorisation of household diaries according to the day type allows profiles to be differentiated for working and non-working days. The grouping of household diaries according to the number of occupants was not considered because the analysis of the diaries showed that succession of occupied and unoccupied periods is not related to the number of occupants. However, the number of occupants influences the residential dwelling energy consumption, so this information will be used when the occupancy profiles generated in this model are used to generate domestic electricity use profiles. The grouping of the household diaries according to the day type is straightforward. Indeed, the day in which the respondent routine is recorded is indicated in the TUS diaries. The grouping according to occupancy profiles is performed using a clustering methodology previously developed by the authors (Buttitta et al. (2016)). The first step of the clustering methodology is the identification of the significant household states, the sequence of which determines the household daily occupancy profiles. To identify the unoccupied and occupied periods just three household states are necessary: (i) all of the household occupants are at home and asleep (Non-Act), (ii) all of the occupants are absent (Abs), and (iii) at least one occupant is home and active (Act). The daily sequence of the household states defines the household daily occupancy profile. Figures 3 and 4 show the household daily occupancy profiles of working and non-working days, respectively, associated with the households participating in the UK TUS. The x-axis of both figures indicates the households, while the yaxis corresponds to the time of the day, which is based from 04:00 to 04:00 of successive days. The colours indicate the household state related to a single household in a defined time of the day.

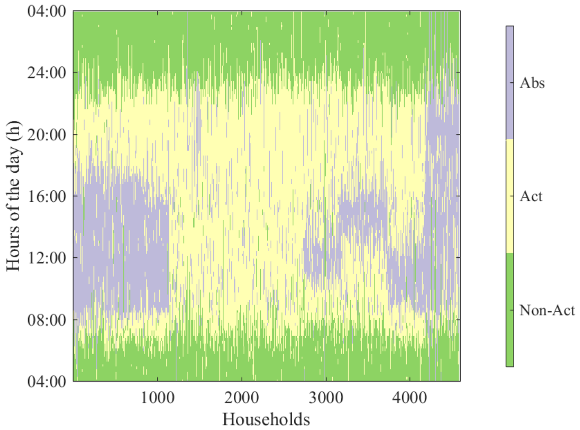

Figure 3: Weekday daily occupancy profiles of TUS households.

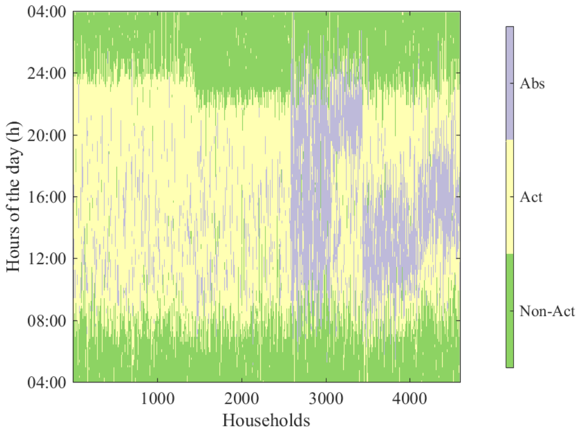

Figure 4: Weekend-day daily occupancy profiles of TUS households.

After the creation of the daily occupancy profiles for all the households considered in the TUS, the households with similar daily occupancy profiles are grouped together through the k-mode clustering technique (Jain et al. (1999)). The k-mode technique allows the creation of clusters of daily occupancy profiles which are similar in composition (e.g. time absent). Each cluster is characterised by a mode. The mode represents the centroid of the cluster and indicates the most common daily occupancy profile in that cluster. The correct number of total clusters is determined using two indices: the root-mean-squared standard deviation (RMSSTD) and R-squared (RS) uncertainty indices (Halkidi et al. (2001)). The RMSSTD index measures the non-homogeneity of the clusters: if its value decreases the obtained clusters are more homogeneous. RS is considered as the measure of the degree of difference between clusters. It is bound within the range [0,1] and it is equal to zero when there is no difference between the clusters, while a value close to one indicates that there is a significant difference between the clusters. These two indices, which were defined for quantitative data, are deployed to be used with categorical variables (Buttitta et al. (2016)). The most appropriate number of clusters corresponds to the one which determines a significant step in the value of both indices. This value is indicated as 'significant knee' (Halkidi et al. (2001)), and it can be easily recognised plotting the RMSSTD and RS indices as a function of the number of the clusters. In the presented work, the RMSSTD and RS indices dependence on the number of cluster is investigated varying the number of clusters from 1 to 20. The maximum non-homogeneity and the minimum differentiation are obtained when a single cluster is considered. This case is used as reference to represent the percentage variation of the two indices (Figure 5).

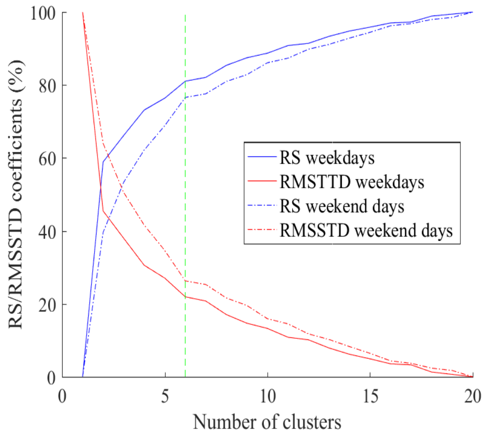

Figure 5: RS and RMSSTD indices variation according to the number of clusters.

When weekdays occupancy profiles are clustered, more than 80% of the diversity between the clusters and almost 80% of homogeneity are reached considering 6 clusters. Passing from 6 to 7 clusters, the improvement of the quality of the clusters is negligible because the incremental change of both indices is just 1%. Continuing to increase the number of clusters over 7 the quality of the clusters starts to improve again. However, in order to improve the homogeneity of each cluster and the differentiation between each cluster, some of the clusters which are obtained from the clustering technique contain just one household daily occupancy profile, which is not representative of a significant portion of UK population. A similar behaviour is observed considering the weekend occupancy profiles clustering. So, for both day types, weekdays and weekend days, the most appropriate number of clusters is found to be equal to 6. The modes associated with each cluster are shown in Figure 6 and Figure 7, as obtained from the k-mode clustering of the weekdays and weekend occupancy profiles respectively.

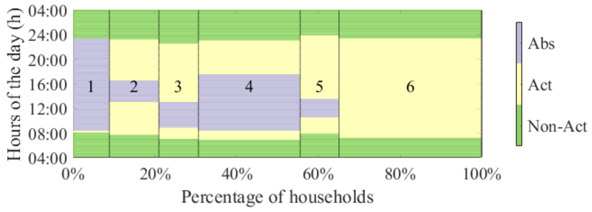

Figure 6: Weekday mode household daily occupancy profiles grouped by cluster.

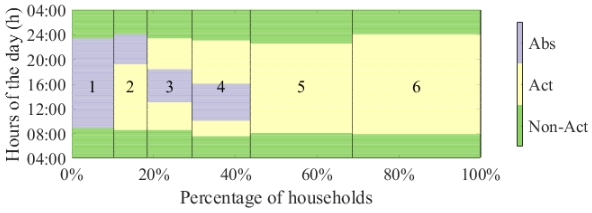

Figure 7: Weekend day mode household daily occupancy profiles grouped by cluster.

The modes are mainly differentiated by the period during which occupants are absent over a day, as depicted in Figures 6 and 7. Considering the weekday modes (Figure 6), it is possible to see that a significant percentage of houses (35%) are occupied for the entire day. Another significant percentage of houses are unoccupied from early morning to early afternoon (25%), while all the other modes represent a smaller percentage of households. During the weekend days, two modes are representative of houses which are occupied all the day. It means that during the weekend, in more than 50% of the houses, at least one occupant is at home all of the day. Moreover in the weekend modes, there is also a mode (Mode 2) which represents the houses which are unoccupied just during the evening. This is an expected behaviour during the non-working days. Summarising, through the data categorisation 12 different data categories are identified: 6 categories for the weekdays and 6 for the weekend days.

Formulate Probability Density Functions (Step 3)

At this point, in order to produce realistic profiles, it is important to include in the model the variability of the occupancy profiles. The variability is achieved by associating each mode representing the clusters to Probability Density Functions, called Transition Proclivity Functions (TPFs), which describe the probability to go from one state to another as a function of the time of the day. The number of TPFs which are obtained for each cluster depends on the number of transitions between different states present in the cluster modes. For example, if the mode which is considered is the weekday mode of Cluster 6, just two TPFs must be obtained. They indicate the distribution of the time in which the household occupancy profiles of Cluster 6 pass from the state Non-Act to Act and then again from Act to Non-Act. If the weekday mode of Cluster 4 is considered, the time of 4 state transitions must be determined (i.e. from Non-Act to Act, from Act to Non-Act, from Act to Non-Act, from Act to Non-Act). Each TPF is obtained by calculating the number of households of each cluster which change their state in each time step, and then dividing this by the number of households in each cluster. The TPFs are step functions, because transition probabilities are defined for each ten-minute time slot t and these functions are defined from 04:00 to 04:00 the next day. The cumulative distribution of the TPF over one day is equal to 1. Figure 8 shows the TPFs associated with each of the transitions state of the cluster 4 of the weekdays.

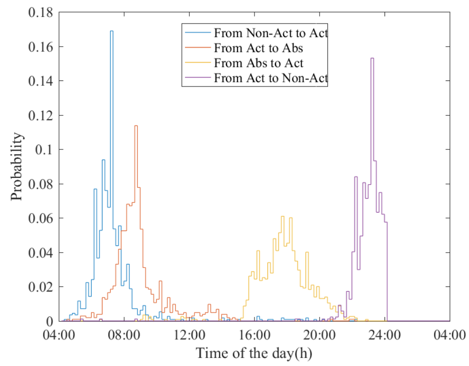

Figure 8: Transition Proclivity Function of Cluster 4 (Weekdays).

The probabilities indicated by TPFs are used to obtain the daily stochastic profiles associated with the households, as explained in the following methodology step.

Generation of occupancy profiles (Step 4)

Generation of daily occupancy profiles

The generation of daily occupancy profiles is obtained using the inverse function method, that generates daily time series of household states from the TPFs related to the different categories. Regardless of the considered category, the initial state of all the households at 04:00 is always non-active. This assumption is reasonable because the percentage of active households at 04:00 is very low (Figure 3 and Figure 4). However this assumption is not realistic for other countries such as Spain (López-Rodríguez et al. (2013)). The time of the first transition ($n = 1$) between the state non-active to the following state is obtained by generating a random number 'r' in the interval [0,1], using a flat distribution random number generator. The time in which the transition happens is the time that satisfies equation (1)

$$\sum_{t=0}^{t_{t,n}-1} TPF_n(t) \leq r < \sum_{t=0}^{t_{t,n}} TPF_n(t). \qquad (1)$$

where t_t is the time step in which there is the transition n. The time of the successive state transitions ($n > 1$) is obtained by generating other random numbers 'r' in the intervals [$\sum_{t=0}^{t_{t,n-1}} TPF_{n-1}(t)$,1] and repeating the same procedure until the transition time of all states is known. Using the TPFs of the different 12 categories found by the data categorisation, the households daily occupancy profiles related to each category can be generated.

Multi-day occupancy profiles

As stated before, the creation of individual household daily occupancy profiles is not sufficient when realistic monthly or annual occupancy profiles are required. In this case, patterns of consistency from day to day must be reproduced to simulate the daily routine which is likely to exist in multi-day household occupancy profiles (Aerts et al. (2014)). In the current paper, in order to obtain consistent multi-day household occupancy profiles, it is assumed that once a given household is associated with a determined weekday cluster, then the household will be associated to the same cluster for all the working days. This is a realistic assumption, because working hours are often fixed, so working day daily occupancy profiles of a given household are similar, but not identical. The day-to-day variability of the working day occupancy profiles of the same household is achieved because the individual daily stochastic profiles generated by the inverse function method are different, although they are obtained from the same TPFs. Occupancy profiles during the non-working days are more unpredictable than for working days because the daily routine is not constrained by defined times, such as working times. In this case, it is more realistic to associate a different weekend cluster with the simulated household whenever a non-working day must be simulated. The knowledge of the cluster associated to the same household for weekdays and weekend allows the definition of the number of households which are doubly associated to a defined weekday and weekend cluster. The probability that the simulated household is associated to a defined weekend occupancy cluster, given that it is associated to a weekday cluster, can be determined normalising the number of households doubly associated to a defined weekday and weekend cluster for the number of households belonging to each weekday cluster. These probabilities are given in Figure 9, where the width of each link between the weekdays and weekend clusters is proportional to the probability of transitioning between the weekday and weekend occupancy profile clusters. For example, if the dwelling is associated to the weekday cluster 6 during the working day, there is an higher probability that it is associated to the weekend cluster 6 in the non-working days. The association between the household and the weekend cluster, after that the household is associated to a defined weekday cluster, is performed generating a random number in the interval [0,1] and comparing this number with the cumulative probability to be associated to one of the weekend clusters.

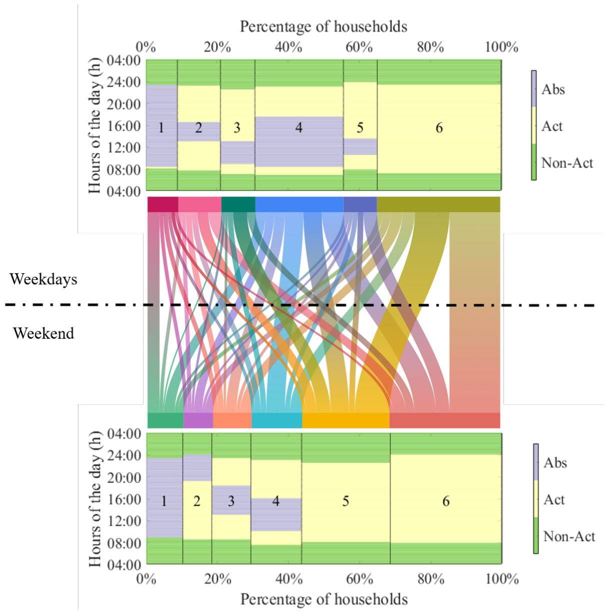

Figure 9: Links between weekday profiles and weekend profiles.

Association of the household with weekday occupancy profiles

One of the main challenges to face to use this model is to adequately associate the households with weekday clusters. The association between the households and the weekday clusters could be executed correlating the weekday clusters to the dwelling types occupied by the households. This approach gives rise to the creation of characteristic occupancy profiles differentiated according to dwelling types. This approach can be adopted if the probability to find an household belonging to a define weekday cluster as a function of the occupied dwelling type is known. Figure 10 shows weekday clusters (in percent) by dwelling type, identified in the UK TUS 2000. Since the UK TUS 2000 is a significant sample of UK households, these percentages are considered as coincident with the probabilities to find a determined weekday cluster for each dwelling type. Once these probabilities are known in the model, the association between the households and the weekday clusters is performed generating a random number between 1 and 0 and comparing this value with the probabilities related to each dwelling type.

The differences in the weekday clusters distributions according to the dwelling type is considerably differ-

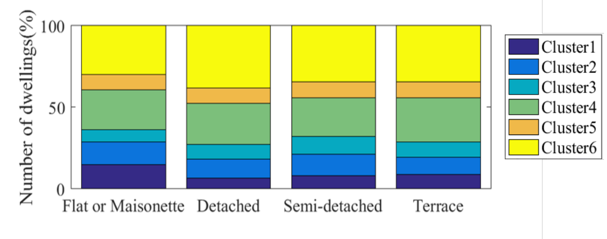

Figure 10: Distribution of different weekday clusters in each accomodation type.

ent for the flat, if it is compared to the other dwelling types. This behaviour is probably due to the fact that the flats are mainly occupied by households with one or two members, as previously seen in Figure 2. In particular the probability that the household is unoccupied during the day, represented by the weekday cluster 1, is higher, while the probability that the dwelling is occupied during all day, represented by cluster 6, is lower. However this behaviour is predictable because, when the number of occupants of the households is lower, the probability that at least one of the occupants stays at home during the day decreases. When real data about the household occupants are collected, unlike the previous case, the household weekday clusters associated to the households can be postulated according to collected data of the households. For example, if the occupants of the dwelling declare that usually the dwelling is unoccupied during the day, the households can be associated to the cluster 1. In any case the household is associated with the cluster represented by the mode which is more similar to the occupancy profiles declared by occupants. This approach is more feasible when a small group of households is modelled. The most suitable method to associate households and clusters depends on the available data of the building stock.

Results and verification of the model

Daily occupancy profiles

The daily occupancy profiles for each category are obtained applying the inverse function method using the TPFs belonging to each category, as discussed in the methodology section. This process is shown in Figure 11, where the cumulative density functions of the TPFs related to the weekdays cluster 4 are used to obtain two different daily occupancy profiles, indicated by red and blue arrows, respectively.

From daily occupancy profiles to multi-day occupancy profiles

The first step to produce multi-day occupancy profiles is the association of the household with a determined weekday cluster. As previously stated, the association can be performed in different ways. After the association of the dwelling and the weekday clusters, the TPF functions are used to obtain the daily occupancy profiles during the working days. The association of the household with a weekend cluster is then performed using the probabilities of transitioning between the weekday and weekend occupancy profile clusters indicated in Figure 9. Using the described approach, a complete weekly occupancy profile can be obtained. Repeating the same procedure, it is possible to obtain monthly or yearly occupancy profiles. Examples of resulting weekly occupancy profiles are given in Figure 12, where the household is associated with the weekday Cluster 4, and the daily occupancy profiles related to Monday and Tuesday are the ones obtained in Figure 11. Figure 13 shows a weekly occupancy profile of a household associated to the weekday Cluster 2. In both cases, at least one on the non-working days present in the week is associated to the weekend Cluster 6, which represents the most common occupancy profile during the weekend.

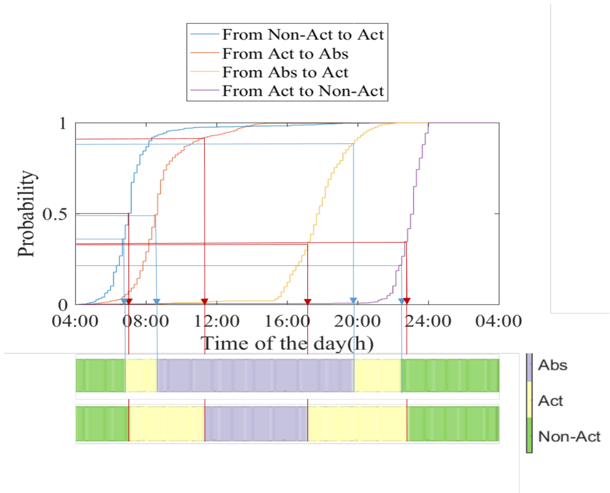

Figure 11: Generation by the inverse function method of two household daily occupancy profiles (Weekday cluster 4).

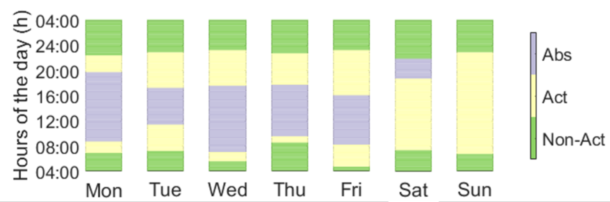

Figure 12: Example weekly occupancy profile associated to the weekday cluster 4 'Daily absence'.

Model validation

The comparison of the produced daily occupancy profiles in Figures 12 and 13 with real occupancy profiles in Figures 3 and 4 shows that the model is able to reproduce real occupancy profiles associated to each cluster. In order to validate the model, 5000 daily occupancy profiles were generated for each cluster. The comparison between simulated and real aggregated active households during weekdays shows that

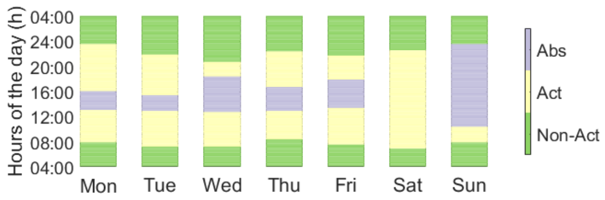

Figure 13: Example weekly occupancy profile associated to the weekday cluster 2 'Short afternoon absence'.

the model performs better when applied to some clusters, like Cluster 2 (Figures 14), giving a Pearson correlation coefficient of 0.98, but slightly worst for other clusters, like Cluster 4 (Figures 15), where the obtained Pearson correlation coefficient is 0.94.

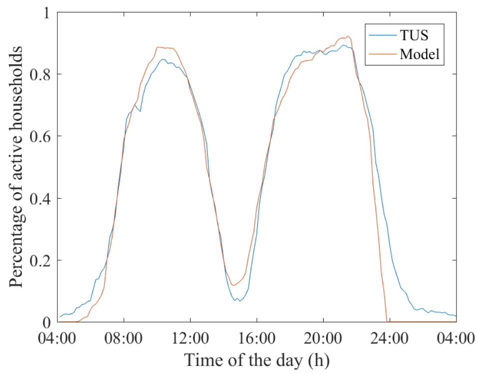

Figure 14: Comparison of simulated and TUS diaries data for weekdays.

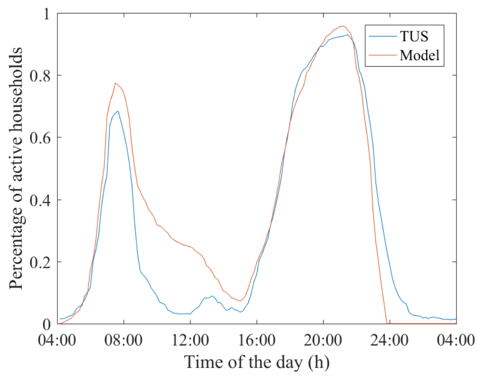

Figure 15: Comparison of simulated and TUS diaries data for weekdays.

The validation of model should include also the assessment of the capability of the model to consider day to day consistency of the generated multi-day occupancy profiles. However, occupant routine is collected just for a single day in TUS diaries, and this does not allow the validation of multiple day profiles.

Discussion and Conclusion

A new methodology that produces realistic multi-day household occupancy profiles that consider patterns of consistency from day to day is presented in this paper. The main innovation introduced is represented by the clustering of the households according to their daily occupancy profiles. This step allows the generation of differentiated occupancy profiles according to the characteristics of the considered building stock. The identified occupancy profiles, indicated by the modes of the clusters, could be used as fixed occupancy schedules in a deterministic approach, or they could be used to identify characteristic occupancy profiles in an archetype approach. In the proposed methodology, the modes of the clusters are used as inputs to produce stochastic occupancy profiles. The final probabilistic occupancy profiles are obtained considering the deviation of the time in which households change their state from the time indicated by the modes. The proposed methodology is intended to be applied in scalable residential energy load models. It allows the creation of realistic occupancy profiles at high temporal resolution, maintaining a low complexity. However, this model, as it is presented, is not suitable when modelling at high spatial resolution. In order to obtain a higher spatial resolution, the number of occupants in the dwelling should be known at each time step. In this case, the same methodology can be applied by considering different household states to characterise the occupancy profiles. For example, it could be possible to use the state initially identified in Buttitta et al. (2016). Another limitation of the methodology is that it does not consider the unexpected deviation of the occupancy profiles from the occupancy profiles described by the mode. However, further analysis on the effect of this aspect on the modelling of household energy end-use consumption has been performed in Buttitta et al. (2017).

Acknowledgement

This research has been financially supported under the auspices of the European Union Horizon 2020 research and innovation programme (Grant Agreement No 646116 Real Value).

References

Aerts, D., J. Minnen, I. Glorieux, I. Wouters, and F. Descamps (2014). A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison. *Building and Environment 75*, 67–78.

Buttitta, G., W. Turner, and D. Finn (2016). Clustering of household occupancy profiles for archetype building models. Paper presented at: Sustainability in Energy and Building, SEB-16, September 2016, Turin, Italy.

Buttitta, G., W. Turner, and D. Finn (2017). Modelling residential building stock heating load demand - Integration of occupancy models. Paper presented at: 2017 ASHRAE Annual Conference, June 2017, Long Beach.

De Meester, T., A. F. Marique, A. De Herde, and S. Reiter (2013). Impacts of occupant behaviours on residential heating consumption for detached houses in a temperate climate in the northern part of Europe. *Energy and Buildings 57*, 313–323.

Famuyibo, A. A., A. Duffy, and P. Strachan (2012). Developing archetypes for domestic dwellings - An Irish case study. *Energy and Buildings 50*, 150–157.

Feng, X., D. Yan, and T. Hong (2015). Simulation of occupancy in buildings. *Energy and Buildings 87*, 348–359.

Gaetani, I., P. J. Hoes, and J. L. M. Hensen (2016). Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy. *Energy and Buildings 121*, 188–204.

Grandjean, a., J. Adnot, and G. Binet (2012). A review and an analysis of the residential electric load curve models. *Renewable and Sustainable Energy Reviews 16*(9), 6539–6565.

Guerra Santin, O., L. Itard, and H. Visscher (2009). The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. *Energy and Buildings 41*(11), 1223–1232.

Halkidi, M., Y. Batistakis, and M. Vazirgiannis (2001). Clustering algorithms and validity measures. *Proceedings Thirteenth International Conference on Scientific and Statistical Database Management. SSDBM 2001*, 3–22.

IEA-EBC Annex 66 (2016). Occupants Behavior Research Bibliography. `http://www.annex66.org/?q=biblio`.

Ipsos-RSL, Office for National Statistics (2003). *United Kingdom Time Use Survey, 2000.* 3rd edition. UK Data Service. [data collection]. `http://dx.doi.org/10.5255/UKDA-SN-4504-1`.

Jain, a. K., M. N. Murty, and P. J. Flynn (1999). Data clustering: a review. *ACM Computing Surveys 31*(3), 264–323.

López-Rodríguez, M. A., I. Santiago, D. Trillo-Montero, J. Torriti, and A. Moreno-Munoz (2013). Analysis and modeling of active occupancy of the residential sector in Spain: An indicator of residential electricity consumption. *Energy Policy 62*, 742–751.

McLoughlin, F., A. Duffy, and M. Conlon (2012). Characterising domestic electricity consumption patterns by dwelling and occupant socio-economic variables: An Irish case study. *Energy and Buildings 48*(July 2009), 240–248.

Neu, O., V. Evon, S. Oxizidis, D. Flynn, and D. Finn (2014, May). Natural ventilation in residential building archetypes: a stochastic approach based on occupant behaviour and thermal comfort. In *Proceedings of eSim 2014: Conference of International Building Performance Simulation Association*, Ottawa, Canada.

Page, J., D. Robinson, N. Morel, and J. L. Scartezzini (2008). A generalised stochastic model for the simulation of occupant presence. *Energy and Buildings 40*(2), 83–98.

Reinhart, C. F. and C. Cerezo Davila (2016). Urban building energy modeling - A review of a nascent field. *Building and Environment 97*, 196–202.

Richardson, I., M. Thomson, and D. Infield (2008). A high-resolution domestic building occupancy model for energy demand simulations. *Energy and Buildings 40*(8), 1560–1566.

Shimoda, Y., T. Fujii, T. Morikawa, and M. Mizuno (2004). Residential end-use energy simulation at city scale. *Building and Environment 39*(8), 959–967.

Walker, C. and J. Pokoski (1985). Residential Load Shape Modeling Based on Customer Behavior. *IEE Power & Energy Society* (7), 1703–1711.

Widén, J. and E. Wäckelgård (2010, jun). A high-resolution stochastic model of domestic activity patterns and electricity demand. *Applied Energy 87*(6), 1880–1892.

Wilke, U., F. Haldi, J. L. Scartezzini, and D. Robinson (2013). A bottom-up stochastic model to predict building occupants' time-dependent activities. *Building and Environment 60*, 254–264.

Yao, R. and K. Steemers (2005). A method of formulating energy load profile for domestic buildings in the UK. *Energy and Buildings 37*(6), 663–671.