Title	A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes
Authors(s)	Buttitta, Giuseppina, Finn, Donal
Publication date	2020-01-01
Publication information	Buttitta, Giuseppina, and Donal Finn. "A High-Temporal Resolution Residential Building Occupancy Model to Generate High-Temporal Resolution Heating Load Profiles of Occupancy-Integrated Archetypes." Elsevier, January 1, 2020. https://doi.org/10.1016/j.enbuild.2019.109577.
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/11566
Publisher's statement	This is the author's version of a work that was accepted for publication in Energy and Buildings. Changes resulting from the publishing process, such as peer review, editing, corrections, structural formatting, and other quality control mechanisms may not be reflected in this document. Changes may have been made to this work since it was submitted for publication. A definitive version was subsequently published in Energy and Buildings (206, (2020)) https://doi.org/10.1016/j.enbuild.2019.109577
Publisher's version (DOI)	10.1016/j.enbuild.2019.109577

A high-temporal resolution residential building occupancy model to generate high-temporal resolution heating load profiles of occupancy-integrated archetypes

Buttitta Giuseppina, Finn Donal P.
University College Dublin, Dublin D04 V1W8, Ireland

Abstract

A strong correlation exists between occupant behaviour and space heating energy use. In particular, the occupancy status (e.g., daytime absence) is known to have a significant influence on residential heating load profiles, as well as on cumulative heating energy consumption. In the literature, many occupancy models have been utilised to predict occupancy profiles of individual dwellings as part of the larger residential building stock. However, none of the existing models consider diversity associated with occupancy-integrated archetypes to generate high-temporal resolution heating load profiles. The current paper uses Time Use Survey (TUS) data to develop a high-temporal resolution residential building occupancy model. The key feature of the proposed model, implemented using MATLAB, is the ability generate stochastic occupancy time-series data for national population subgroups characterised by specific occupancy profiles. It is shown that the results are capable of closely approximating data available from TUS. The developed model can be applied to improve the quality of modelled high-temporal resolution heating load profiles for generic building stock characterised by population subgroups represented by different occupancy-integrated archetypes. A case study is performed on a building stock sample located in London, UK. The developed occupancy model has been implemented in MATLAB and is available for download

Keywords: stochastic occupancy models; archetypes; building stock modelling; residential buildings; occupancy profiles

1. Introduction

The residential sector is a significant consumer of energy in most developed countries, and therefore a focus for energy reduction efforts [1]. Energy consumption estimation in buildings is complex as it is usually affected by building physical properties, the outdoor environment, and occupant behaviour [2]. Research on scalable residential end-use energy demand models for large scale building stocks is relatively new, if compared to the research developed to predict energy and thermal performance of individual buildings [3]. Most of the modelling techniques which have been developed to simulate the energy demand from small to large scale have been collected by several reviews [3–9]. In these papers, the modelling techniques are separated into two main categories: top-down and bottom-up. The top-down approach is based on analysis of performance data of the entire building stock and does not distinguish individual end-users. In contrast, the bottom-up approach is based on the estimation of the energy consumption of individual buildings representing the building stock. The bottom-up approach can be sub-grouped according to the data inputs [6, 10] and the methodology used [7, 9]. Data input can be on the basis of a statistical approach or a building physics approach [6]. The statistical approach is based on the analysis of historical data, while the building physics approach is based on the thermodynamic relationships that define the energy consumption of the dwellings. The thermodynamic relationships which characterise the building physics approach can be modelled through the use of equivalent Resistance-Capacitance electric circuits (RC models) [11, 12] or simulation software [13], which rely on more complex thermodynamic equations. RC models are broadly used in large scale building energy simulations because they allow the simulation of the building physics, maintaining a higher computational efficiency than simulation software. In both cases, detailed building geometry as well as physical data, such as construction elements, U values, etc., and occupancy profiles, are required as inputs of the building model. However, a key challenge associated with large scale building modelling is the collection of the data related to individual buildings. For this reason, building stock modelling usually deploys representative building archetypes to obtain reliable results, thereby minimising the associated computational cost [4]. While traditionally researchers have explored the development of archetypes by considering physical characteristics of the buildings [8, 14 – 16], studies on integration of occupant behaviour inside archetypes are more recent. Occupancy-integrated archetypes which include different deterministic occupancy profiles in the characterisation of archetypes are defined in [17]. This archetype definition overcomes the assumption that one profile fits all dwellings and allows to identify population subgroups characterised by similar occupancy profiles which can be different from the standard one. The use of occupancy-integrated archetypes produces more reliable results in the estimation of annual heating energy demand when compared to the adoption of standard occupancy schedules for residential building energy simulation provided by national guidelines (e.g., BREMEN [18] in UK and DEAP [19] in Ireland). However, the utilisation of deterministic occupancy profiles does not allow to consider the stochastic nature of occupancy profiles and the consequential diversity of the associated heating demand. The concept of diversity was introduced in [7] and it indicates the non-coincidence in energy use. The consideration of diversity is essential to generate high-temporal resolution heating load profiles. The inclusion of the diversity in large scale energy models is one of the main challenges in the energy modelling of large building stocks. In deterministic models, diversity is not modelled; in statistical models [20] the diversity is modelled through a pure mathematical procedure; in probabilistic empirical models the diversity is modelled based on the authors experience [21, 22]; and in models based on Time Use Survey (TUS) data [23 – 28] the diversity is modelled based on data available from the surveys. In a comprehensive review, Grandjean et al. observe that the most reliable probabilistic models are based on Time Use Survey (TUS) data [7].

The integration of a high-temporal resolution residential building occupancy model based on TUS data within occupancy-integrated archetypes allows occupant diversity to be included in occupancy-integrated archetype models to obtain high-temporal resolution heating load profiles. Several occupancy models based on TUS data are already available. One of the first occupancy models based on TUS data is that developed in [23]. More recent models which include diversity are those developed in [26, 28 – 30]. The model developed by Richardson et al. [26], commonly known as the CREST model, is based on a Time Use Survey (TUS) conducted in the UK in 2000. The model uses a first-order Markov Chain Monte Carlo technique to generate statistical, high time-resolution occupancy profiles. This methodology produces occupancy profiles differentiated according to the household size and the day type (weekend or weekdays). In the model developed by Widén & Wäckelgård [29], Time Use Survey data collected by Statistics Sweden in 1996 are used. Similar to the Richardson model, the Widén model also uses a first-order Markov Chain 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, or a weekday from a weekend day on the other. The model developed by Wilke et al. [30] is based on a French Time Use Survey, realised between February 1998 and February 1999. The model adopts a higher-order Markov process to address the problems associated with the Markov Chain Monte Carlo, where the current state is only dependent on the previous state. This model is based on time-dependent probabilities to activate activities and their corresponding duration distributions. Aerts et al. [28] used the data from a Belgian 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. Probabilistic occupancy profiles were obtained by applying the probability to transit from a certain state to another and the duration probability, which are both time dependent. The model is capable of producing individual occupancy profiles differentiated according to the 'typical occupancy patterns'. Additionally, the research presented by Aerts et al. has the merit of presenting a methodology to construct realistic yearly household occupancy sequences, considering consistency from day-to-day profiles.

In order to generate high-temporal resolution heating load profiles which can be integrated in occupancy-integrated archetypes, stochastic occupancy models must be able to generate differentiated occupancy profiles for population subgroups characterised by specific occupancy profiles. Most of the models which are based on TUS data differentiate households according to the number of occupants [26, 29, 31]. This differentiation has the disadvantage of producing household categories containing large occupancy numbers, which in turn consist of households characterised by significantly different occupancy profiles. Although extremely useful as a categorisation approach, the associated models cannot be readily deployed in simulation of occupancy integrated archetypes. This is because the models generate occupancy profiles that are based on an unrepresentative composite of multiple combined occupancy profiles, which cannot reproduce the specific occupancy profiles of population subgroups associated with the occupancy-integrated archetypes. 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 [26]. 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.

Just the models introduced in [23, 28] can be successfully used to capture dwellings characterised by specific household occupancy profiles, thereby specifying occupancy-integrated archetypes. The model presented in [23] is the most complete model. However, this model is based on an extensive and detailed dataset which can be difficult to collect. The model presented by Aerts et al. [28] is capable of producing profiles which can be very diversified for different households. In addition, it is also computationally more efficient than other models. Moreover, the problem of the consistency of the multi-day simulations is addressed. Nevertheless, this methodology is based on the assumption that the household individual occupants act independently, which is unlikely [28]. With this approach, the relationships between the occupancy presence of the members of the same household are not considered.

Based on this assessment and review, the arising challenge is the creation of a stochastic occupancy model which can be integrated with occupancy-integrated archetypes such that it is possible to capture the diversity provided by the stochastic behaviour of occupants to simulate high-temporal resolution heating load profiles of building stock characterised by different population subgroups. The development of this new model is thus the subject matter of the current paper.

The proposed stochastic occupancy model is: (i) scalable, as it can produce occupancy profiles for building stocks of different sizes; (ii) adaptable, as it can be used to model building stocks characterised by occupants having completely different behaviour (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); (iii) representative of diversity of occupant profiles at large-scale [5], as the stochasticity of the model allows the influence of occupant diversity on building stock energy demand to be considered. The key novelty of the model is its adaptability to building stock including population subgroups characterised by specific occupancy profiles. In addition, the model is capable of replicating similar behaviour for each household during weekdays, meaning that consistent multi-day occupancy profiles can be obtained. The model is deployable in any application where high-resolution occupancy profiles in building stock are required. An implementation of the model is available for download [32]. Because the heating demand is closely linked with the occupancy profiles of the households [2], the study of the application of such a model to obtain high-temporal resolution heating demand profiles is of particular interest. For this reason, in the current paper, the occupancy profiles are used as input in the occupancy-integrated archetype models to calculate the high-time resolution profiles of heating demand of a building stock.

The paper is organised as follows. The development of the novel stochastic model is described in Section 2. In Section 3, the developed model is used to calculate the heating demand of a hypothetical building stock of 100 identical new apartments located in London. The aim is to demonstrate and elaborate on the usability of such a model in residential building stock modelling. Initially, the deterministic occupancy profiles, typical of integrated-occupancy archetypes are used to determine the building stock heating demand. Next, the building stock heating demand is re-calculated considering the new stochastic occupancy model developed in Section 2. Additionally, the results are also compared to the ones which can be obtained using a well-established stochastic model (CREST model) [26]. The effect of use of the new stochastic occupancy model in the estimation of high-temporal resolution heating load profiles of occupancy-integrated archetypes is discussed in Section 4 of the paper. Section 5 concludes the paper.

2. Development of the stochastic occupancy model

An overview of the methodology is given in Figure 1 and is outlined in detail in this section.

Data which are used to develop the model are obtained from the Time Use Survey conducted in the UK in 2015 (TUS 2015 UK) [33]. In this model, household diaries are categorised according to the day type and the daily occupancy profiles, following the same procedure adopted in the development of occupancy-integrated archetypes to allow the integration of the model with these archetypes [17](Step 1 – 2). Three significant household states are identified to describe the occupancy profiles: (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). This choice is reasonable because it has been shown that heating demand in residential buildings is mainly influenced in particular by the succession of occupied and unoccupied periods [16, 20]. As in most of the reviewed existing models, a first-order Markov–Chain technique is selected to predict the household states, which is proved to be the most appropriate in [34].

The concept of the first-order Markov–Chain technique is that each state is dependent only on the previous state together with the probabilities of the state changing. These probabilities are held in "transition probability matrices", which must be generated for each time step, to capture the time dependence of the process (Step 3). In order to represent the different behaviour of the categories, the "transition probability matrices" which define each category are different. Once the transition probability matrices are obtained, they are used in the Markov–Chain Monte Carlo technique to obtain the household states in each time step. The sequence of household state during one day defines the household daily occupancy profile. The TUS 2015 UK data are collected with a ten-minute period resolution, and thus, in this case, the household daily occupancy profile comprises 144 integer numbers, which indicate the household states. If needed, the obtained daily occupancy profiles can also be aggregated in sequences to obtain consistent multi-day occupancy profiles (Step 4).

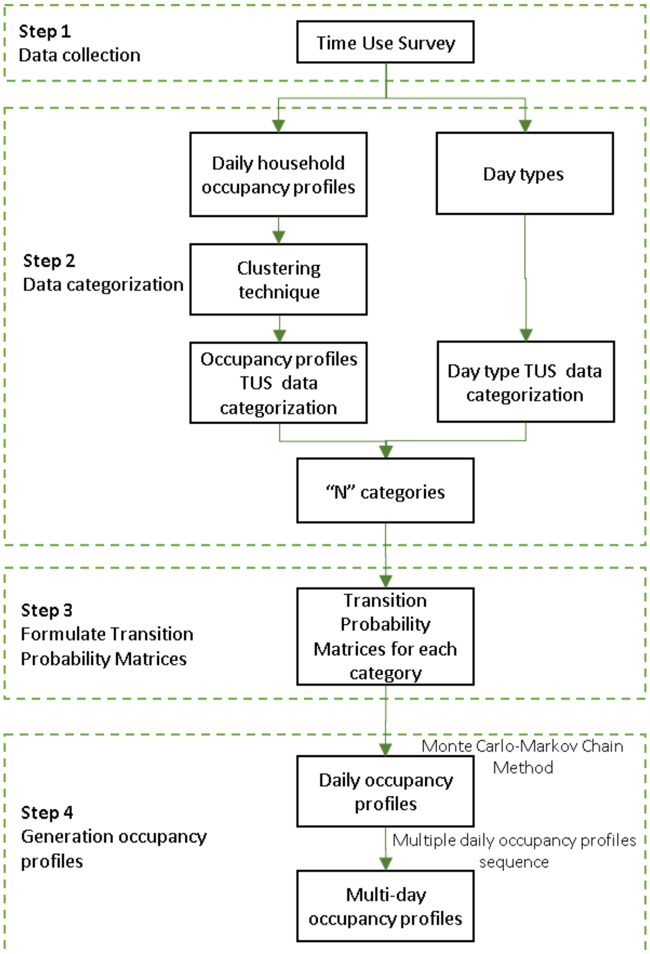

Figure 1 Model flow chart

2.1 Data Collection-Time Use Survey (Step 1)

The presented model is based on data available from the Time Use Survey 2014-2015 (UK 2015 TUS) [33]. The UK 2015 TUS recorded the everyday routines of 10,208 UK citizens belonging to 4,733 households. One household is defined as a person or group of people who have specified the accommodation as their only or main residence and share the living accommodation. The routine of survey respondents is described in detailed 24-hour diaries (household diaries), completed at ten-minute intervals. In each time slot, occupants indicated their primary and secondary activities, the location of the activity and whom the survey respondent was with. Additionally, data describing the working hours during a whole week are also available. From this data, it is possible to see that working hours related to an individual person are likely to be the same for all the working days. This leads to the assumption that it is reasonable to adopt the same daily occupancy profiles for all the working days for a determined household.

2.2 Data categorisation (Step 2)

TUS data are categorised according to the day type (weekend or weekday) and the daily occupancy profile, by the application of the k-mode clustering technique [17]. 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 [35]. The clustering is applied only on weekday occupancy profiles because it is almost impossible to recognise household having similar occupancy profiles during the weekend days. In the end, through the data categorisation, 6 different data categories are identified: 5 categories for the weekdays, represented by the modes “wd1, wd2, wd3, wd4, wd5”, and 1 category for the weekend days, represented by the mode “we”.

The result of the data categorisation is shown in Figure 2.

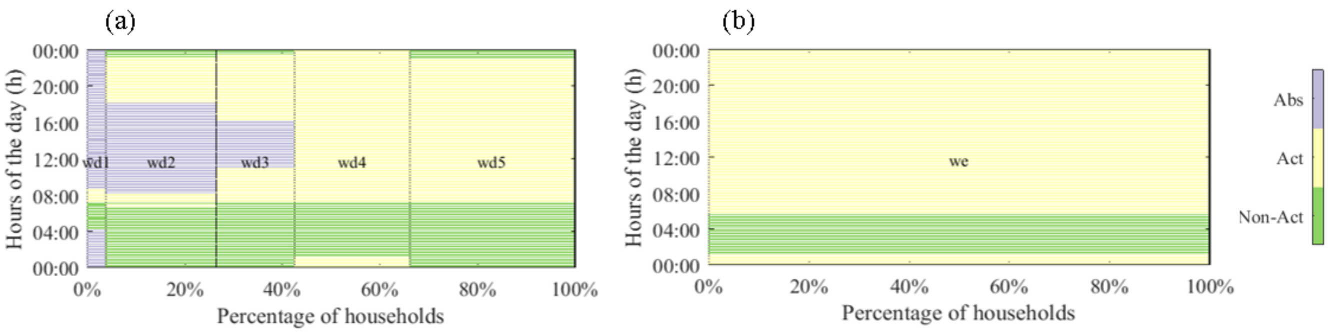

Figure 2 Occupancy categories for weekdays (a) and weekend (b)

The weekday categories correspond to the characteristic occupancy schedule (OP) of the occupancy-integrated archetypes identified in [17], which are: OP1 Daily absence, OP2 Working hours absence, OP3 Lunchtime absence, OP4 Constant presence 1, OP5 Constant presence 2, (see Table 3).

2.3 Transition probability matrices (Step 3)

The transition probability matrices are generated for each category derived from the categorisation of TUS data. In order to create daily household profiles, a transition probability matrix was created for each time step. In total, 6 x 144 transition matrixes were created, where 6 are the categories and 144 the time step in each day.

Transition probability matrices have dimensions of 3 x 3 because 3 are the possible household states (Abs, Act, Non-Act). In this model, the dimension of the transition matrixes is independent of the number of occupants in the house. This reduces the computational time compared to other models such as [26], but this also reduces the obtained information because the number of occupants in each time step is lost. Nevertheless, the heating demand is not a function of the number of occupants in the house, but it is mainly dependent to the succession of occupied and unoccupied periods, which can be accurately modelled for different occupancy-integrated archetypes as a result of the development of this model. An example of transition probability matrixes is shown in Table 1, where each of its entries is a non-negative real number representing a probability for the household to transitioning between two states from the time step "t" to time step "t+1". Probabilities of the transition probability matrixes are estimated from TUS data. The probability to pass from state *i* to state *j* is calculates as

$$p_{ij} = \frac{O_{ij}}{\sum_{k=1}^{m} O_{ik}} \tag{1}$$

O_{ij} is the number of observed transitions from state *i* to state *j*, O_{ik} is the number of observed transitions from state *i* to state *k*, *m* is the number of possible state (m = 3 in this case). This process identifies the maximum likelihood estimators of the transition matrixes [36].

Table 1 Example transition probability matrix

		Time step "t+1"		
	Household state	**Active**	**Non-Active**	**Absent**
Time step " t"	**Active**	$P_{act,act}$	$P_{act,non-act}$	$P_{act,abs}$
	Non-Active	$P_{non-act,act}$	$P_{non-act,non-act}$	$P_{non-act,abs}$
	Absent	$P_{abs,act}$	$P_{abs,non-act}$	$P_{abs,abs}$

2.4 Generation occupancy profiles (Step 4)

2.4.1 Daily occupancy profiles

The generation of daily occupancy profiles is obtained using the Markov– Chain Monte Carlo technique [37].

The first step of the technique is the generation of the start state, which is the household state at 00:00. This is randomly determined considering the probabilities found in the original TUS 2015 UK data. After the start state is established, the following household states must be determined. In order to generate the following synthetic data, a random number "r" is generated in the interval [0,1] at each time step, using a flat distribution random number generator. Comparing the number with the probabilities indicated in the transition probability matrix linked to that specific category at a given the time step. An effective way to visualise the resulting household daily occupancy profiles is through the spectrum of occupancy profiles shown in Figure 3 for each category. The x-axis of all subplots in Figure 3 indicates the households, while the y-axis corresponds to the time of the day, which is based from 00:00 to 00:00 of the successive day. The colours indicate the household state related to a single household in a defined time of the day. It is possible to see that the profiles obtained for each category are similar to the corresponding modes shown in Figure 2, though not identical. The stochastic profiles obtained for the different categories can be applied directly to the model of the associated occupancy-integrated archetypes. Thanks to the use of these stochastic profiles in the model of occupancy-integrated archetypes, it is possible to consider the diversity of occupant behaviour in the estimation of high-temporal resolution heating load profiles.

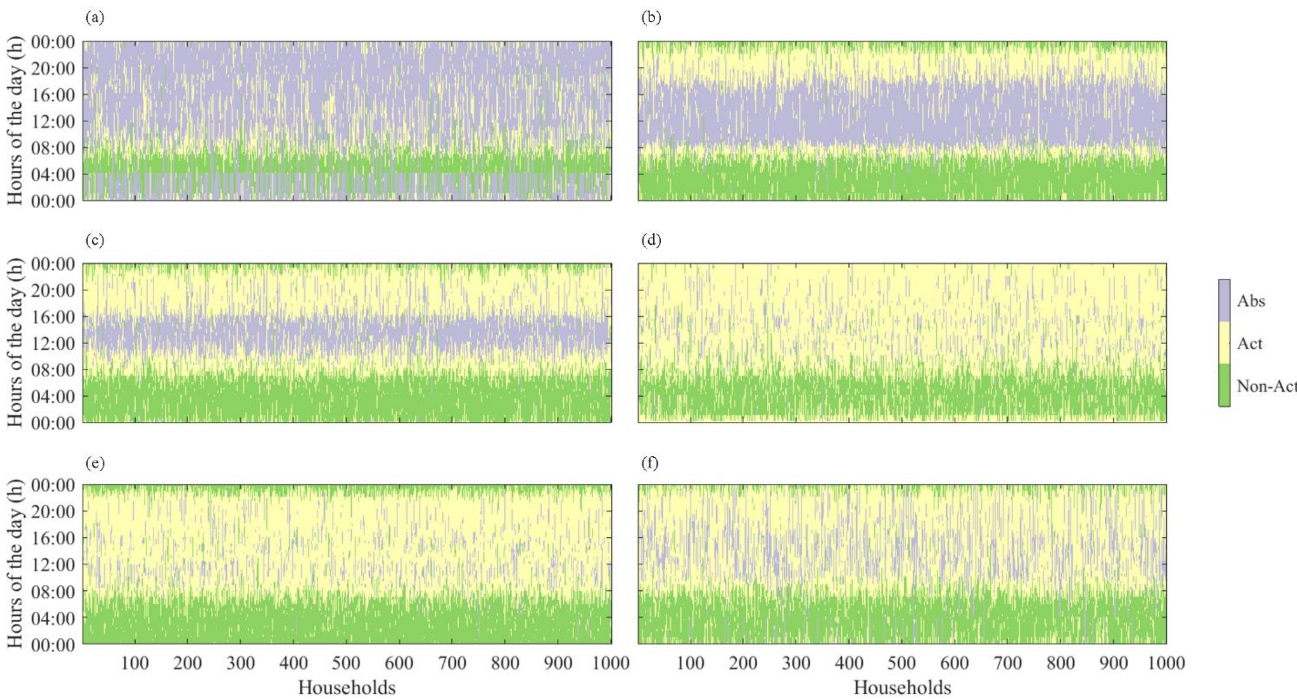

Figure 3 Spectrum of the occupancy profiles for the following categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we.

2.4.2 Multi-day occupancy profiles

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 multiday household occupancy profiles [28].

This issue is already addressed in the development of occupancy-integrated archetypes as the same deterministic occupancy profiles are used during weekdays to incorporate the likelihood that the same household behaviour is repeated during the working days because of daily working routines. This is a realistic assumption [17] because working hours are often fixed. Similar but not identical occupancy profiles during working days are obtained using the same transition matrices in the Markov-Chain process to produce the stochastic profiles during weekdays. Additionally, a different transition matrix is used for weekend days, so that the occupancy profiles obtained for non-working days are likely to be different from the one obtained for working days.

Weekly household occupancy profiles can be obtained combining 7 consecutive daily household occupancy profiles (Figure 4- Figure 8). From these figures, it is possible to see that the model works as expected: similar daily household occupancy profiles are generated for working days, and also, the household daily occupancy profiles generated by the model are very similar to characteristic modes of each category.

As the 5 categories for the weekdays correspond to the 5 occupancy schedules (OP) which are used in the identification of occupancy-integrated archetypes, it is easy to associate these weekly occupancy profiles to the occupancy integrated archetypes.

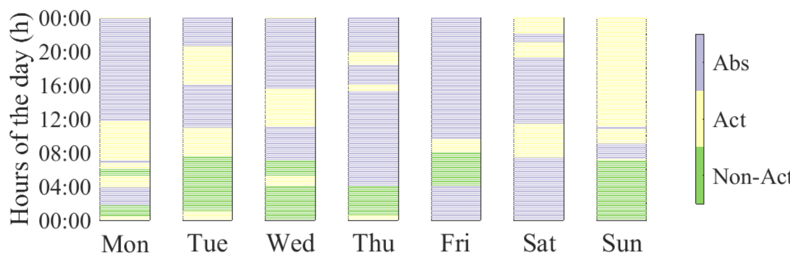

Figure 4 Weekly occupancy profiles associated with the category OP1.

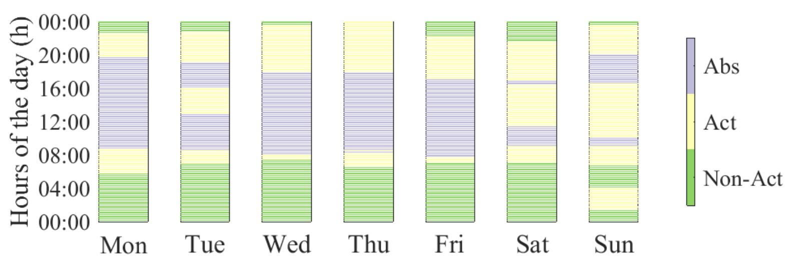

Figure 5 Weekly occupancy profiles associated with the category OP2.

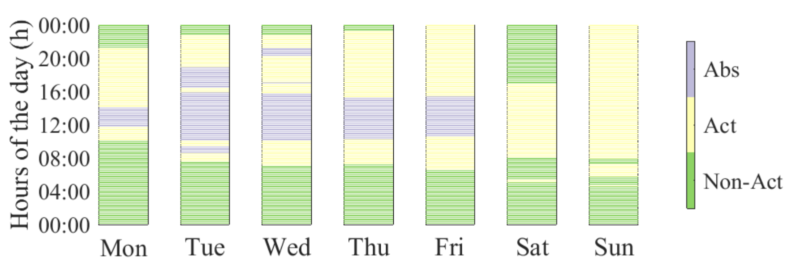

Figure 6 Weekly occupancy profiles associated with the category OP3.

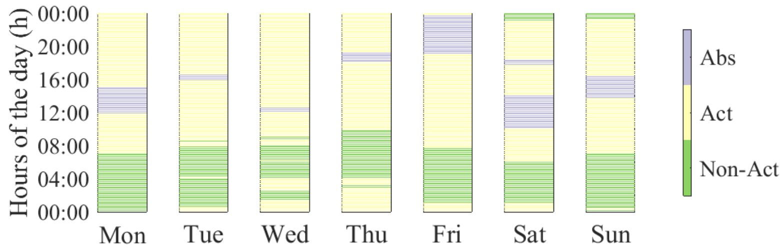

Figure 7 Weekly occupancy profiles associated with the category OP4.

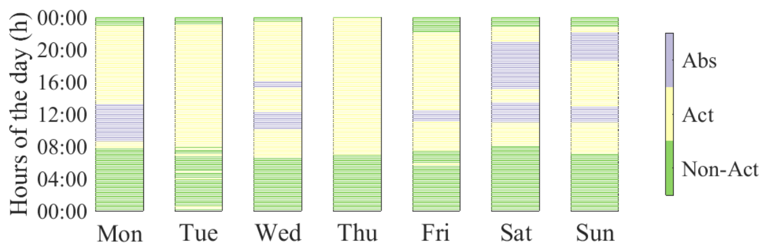

Figure 8 Weekly occupancy profiles associated with the category OP5.

2.4.3 Verification of aggregate behaviour

A comparison of the aggregated model output against TUS data for each category is shown in Figure 9. This shows the percentage of active households in each time step for each category as obtained from TUS data (real profile), compared to the average proportion of active households as generated by the model for the same category (synthetic profile). In order to obtain the synthetic profile, initially the percentage of active households in a sample of 1000 households is calculated. Then, this step is repeated 1000 times in order to obtain a reliable value of the average proportion of active households generated through the application of the presented stochastic occupancy model. The 95% confidence interval of the obtained synthetic profile is also shown in Figure 9 to indicate the reliability of the data shown for the considered sample. However, this interval is almost invisible in this case as the sample size considered is very large. Figure 10 shows the 95% confidence interval obtained when the average value of a household sample of 100 households is calculated 10 times. Comparing Figure 9 and Figure 10, it is possible to see that as expected, the confidence interval is bigger when the sample size is smaller, although it is still acceptable even when the sample is small.

The close correlation between the model output and the TUS data is quantified considering 3 metrics, namely Mean Absolute Error (MAE) (Equation 1), Root Mean Squared Error (RMSE) (Equation 2) and Coefficient of Variation (CV) (Equation 3). The MAE, the RMSE and the CV are obtained using Equation 1, 2 and 3, respectively, where $AHP_p(i)$ is the predicted active household percentage at repetition i, AHP_r is the reference active household percentage (from

TUS data), and N is the total number of repetitions, which is equal to 1000 in this case. $\overline{AHP_p}$is the average value of the AHP_p obtained for the total number of repetitions.

$$MAE = \frac{\sum_{i=1}^{N}\left|AHP_p(i) - AHP_r\right|}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(AHP_p(i) - AHP_r\right)^2}{N}} \tag{2}$$

$$CV = \frac{RMSE}{\overline{AHP_p}} \tag{3}$$

The resulting values obtained for the different time steps of the day for all the different categories are reported in Figure 11. From this figure, it is possible to see that the MAE, RMSE and CV values are always approximately equal to zero, meaning that the result produced by the presented stochastic model are very close to the data obtained from TUS. The average value of these metrics over one day is reported in Table 2. In this table, the Pearson coefficient (ϱ) is also reported. This is approximately equal to 1 for all categories, confirming that there is a close correlation between the profiles obtained from the stochastic occupancy model and the reference result obtained from the TUS data.

Table 2: Pearson coefficient, MAE, RMSE and CV error metrics for different categories

	Categories					
	Wd 1	**Wd 2**	**Wd 3**	**Wd 4**	**Wd 5**	**We**
ϱ	1	1	1	1	1	0.9996
MAE	0.00896	0.00864	0.00928	0.00908	0.00854	0.01047
RMSE	0.00781	0.00762	0.00731	0.00962	0.00790	0.00828
CV	0.00056	0.00048	0.00030	0.00026	0.00039	0.00029

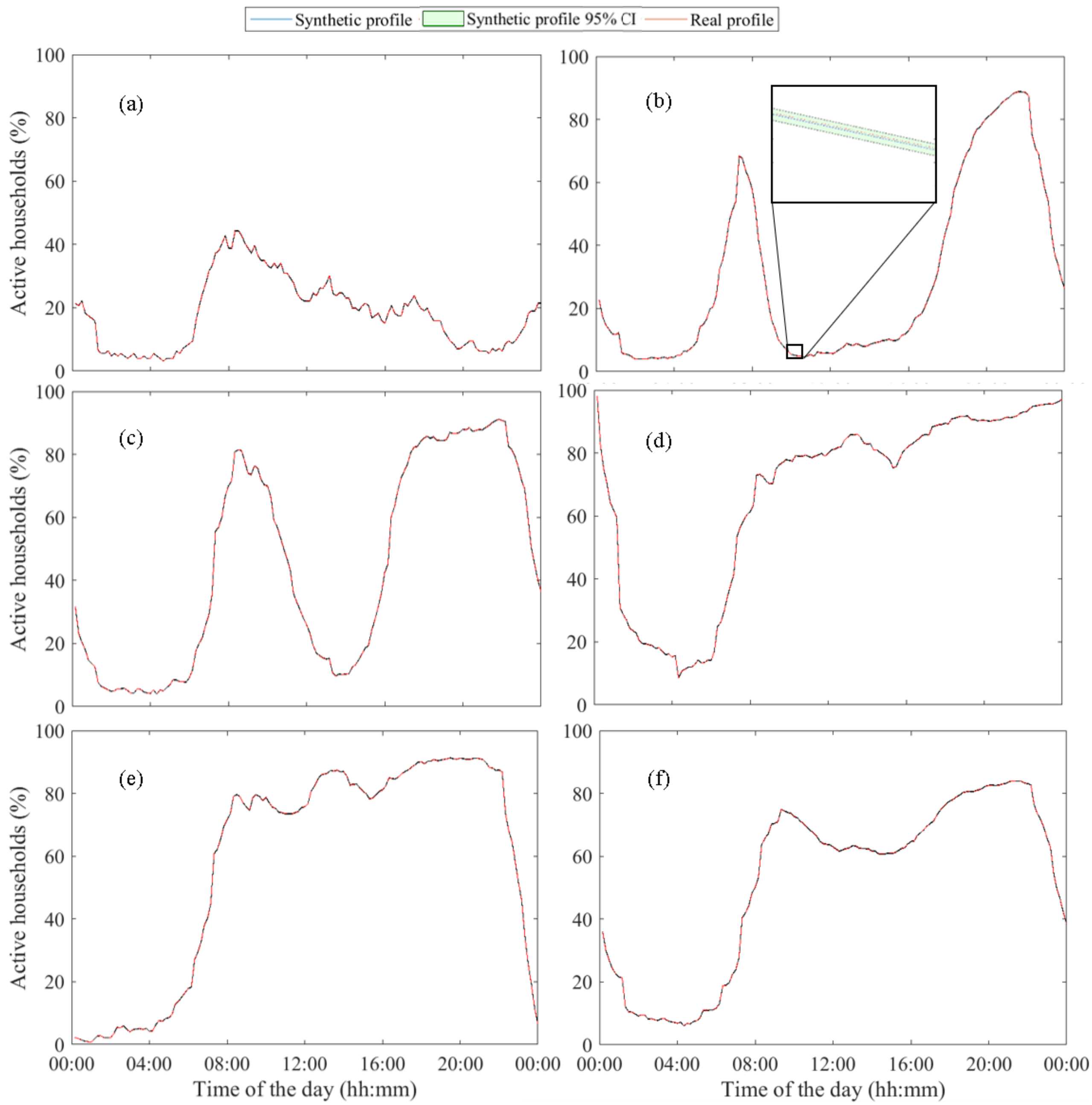

Figure 9 Comparison of real and synthetic profiles of the percentage of aggregated active occupancy for the following categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we. Sample: 1000 households. 1000 repetitions.

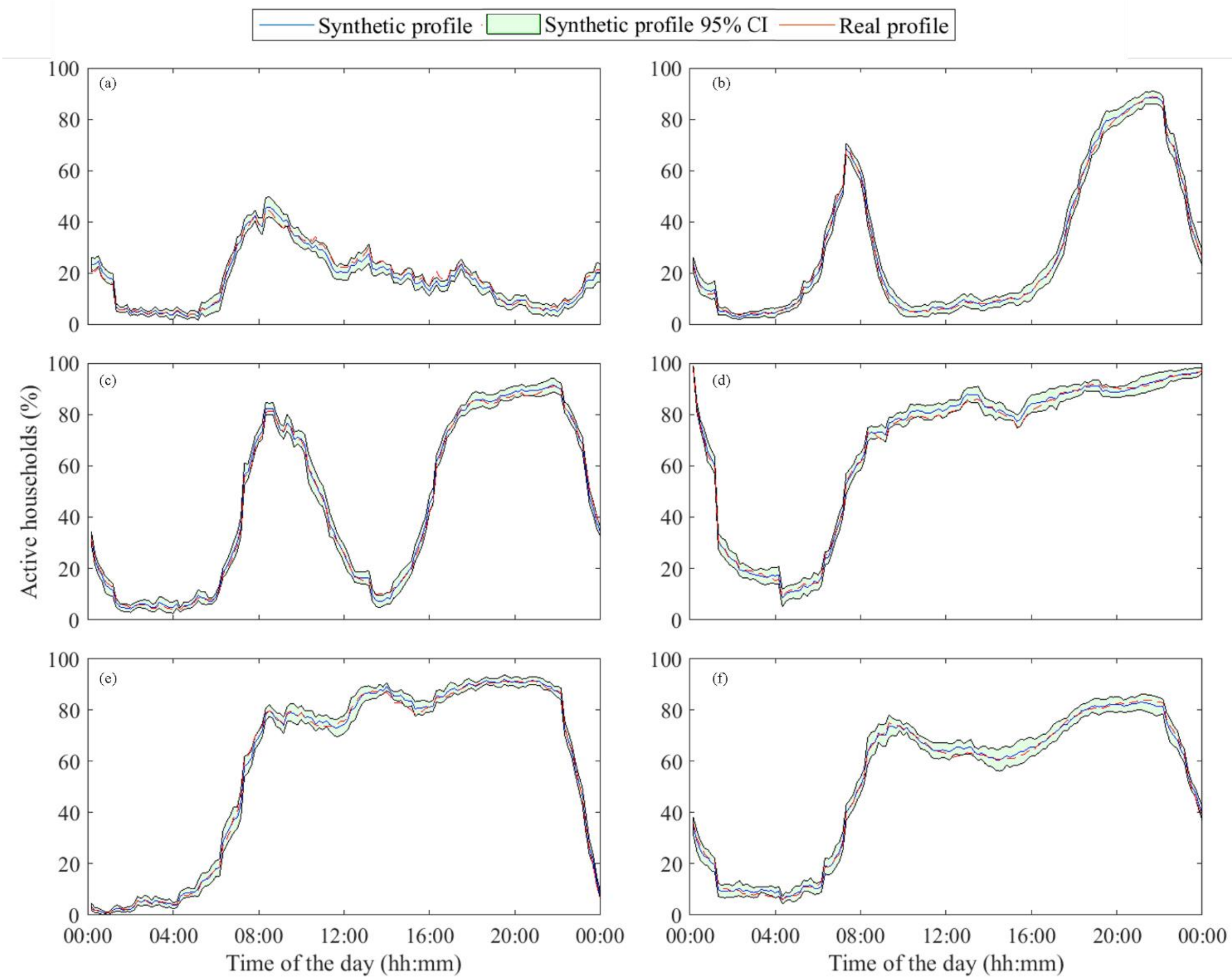

Figure 10 Comparison of real and synthetic profiles of the percentage of aggregated active occupancy for the following categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we. Sample: 100 households. 10 repetitions.

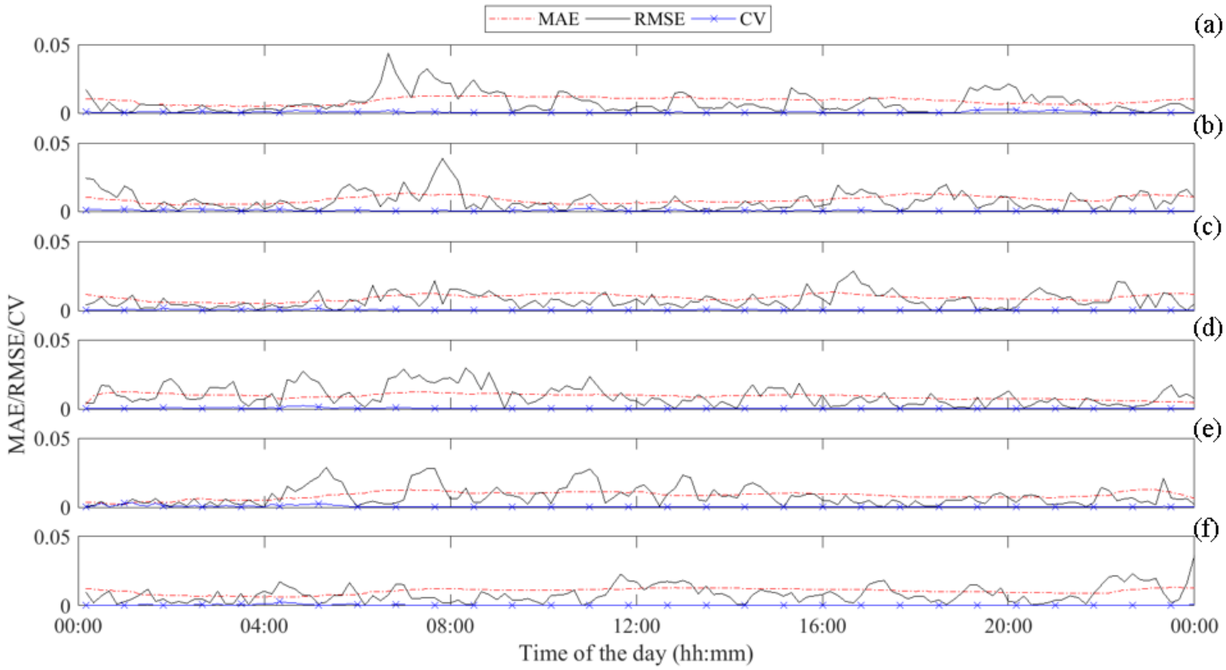

Figure 11 ME/RMSE values for the following categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we. Sample: 1000 households. 1000 repetitions.

2.4.4 Verification of transitions

The verification of transitions is obtained using a bootstrap method as suggested in [36]. This method consists of computing the maximum likelihood estimators from the samples obtained from the application of the stochastic occupancy model presented in this paper and comparing them with the maximum likelihood estimators obtained from the original TUS. The maximum likelihood estimators obtained from the original TUS correspond to the transition matrices presented in Section 2.3. The maximum likelihood estimators must be obtained for each time step and for each category. The total number is thus equal to 6 x 144. An effective way to visualise the maximum likelihood estimators is by the use of heatmaps. Figure 12 and Figure 13 show the maximum likelihood estimators obtained from TUS data and sample data, respectively. The maximum likelihood estimators obtained for different time steps are underlined by a thick vertical line. Each row in each time step indicates a household state, as well as each column. In order to maintain clarity, only a limited amount of time steps is shown. Comparing Figure 12 and Figure 13, it is possible to see that the values of maximum likelihood estimator are very similar. Their percentage difference is computed as $\frac{|p_{ij}-\widehat{p_{ij}}|}{p_{ij}} \text{ x } 100$, where $\widehat{p_{ij}}$ indicates the maximum likelihood estimator obtained from the sample. The results of the percentage difference are shown in Figure 14. In general, the values of the percentage difference between the maximum likelihood estimators are lower than 10%, except for some exceptions which are evident when p_{ij} is very small. In particular, for the time steps presented in Figure 14, the percentage difference between the maximum likelihood estimators reached

100% in the category "we" (weekend). Although this value can appear to be very high, it is justifiable by the fact that in these cases both p_{ij} and $\widehat{p_{ij}}$ are approximately equal to zero.

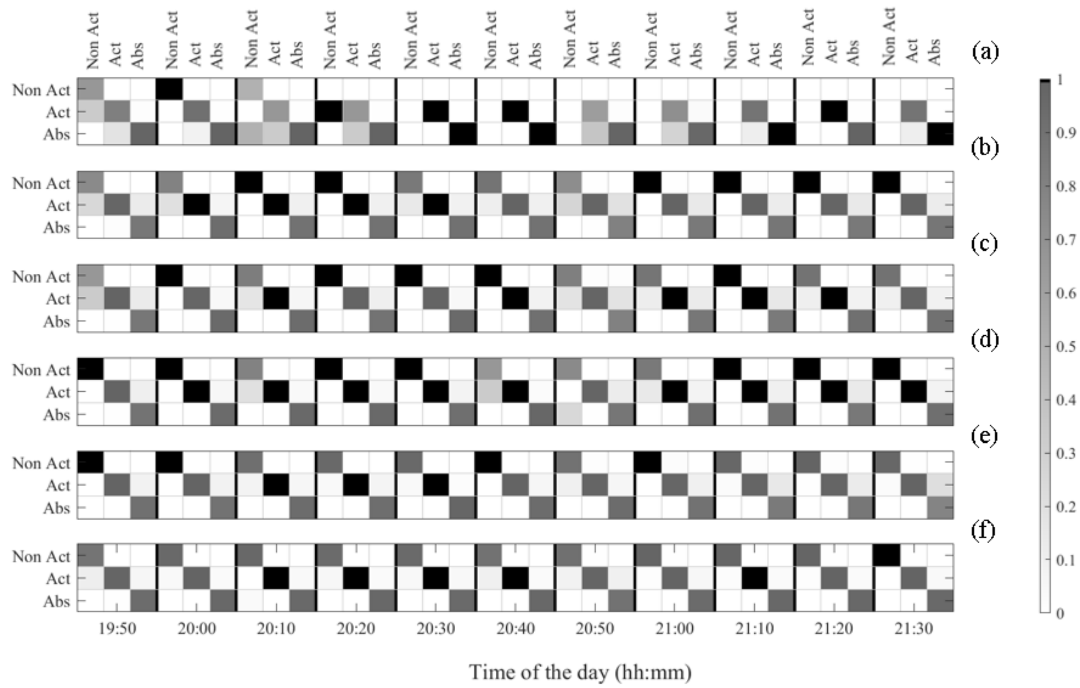

Figure 12 Maximum likelihood estimators from TUS data for categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we.

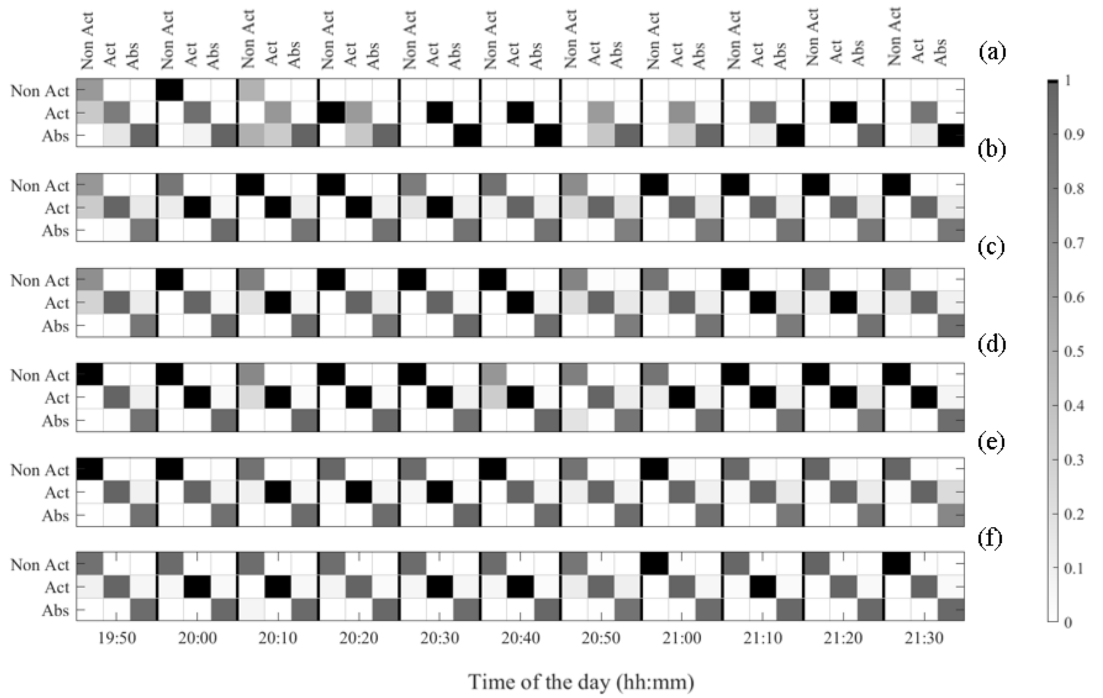

Figure 13 Maximum likelihood estimators from sample data for categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we.

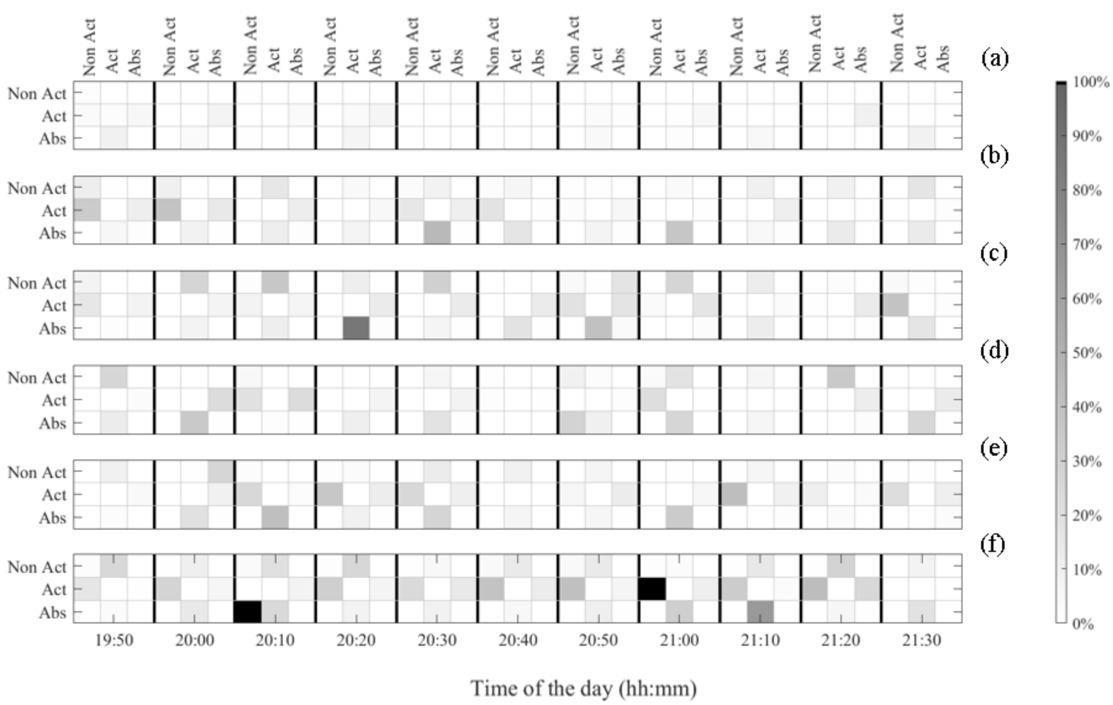

Figure 14 Percentage difference between maximum likelihood estimators from TUS and sample data for categories: (a) wd1, (b) wd2, (c) wd3, (d) wd4, (e) wd5, (f) we.

2.4.5 Downloadable model

An example implementation of the model has been made available for download [32]. It is capable of creating multi-day occupancy profiles for building stock of different sizes and characterised by different shares of households belonging to different categories. The source code may be readily adapted for specific applications, with due acknowledgement to the authors.

3. Application of the model

The aim of the paper is to develop a new stochastic occupancy model which can be integrated with the occupancy-integrated archetypes described in [17], so that it can be used to accurately simulate high-temporal resolution heating load profiles of residential building stocks. The result of the integration is an archetype-based stochastic building stock model, which is schematised in Figure 15.

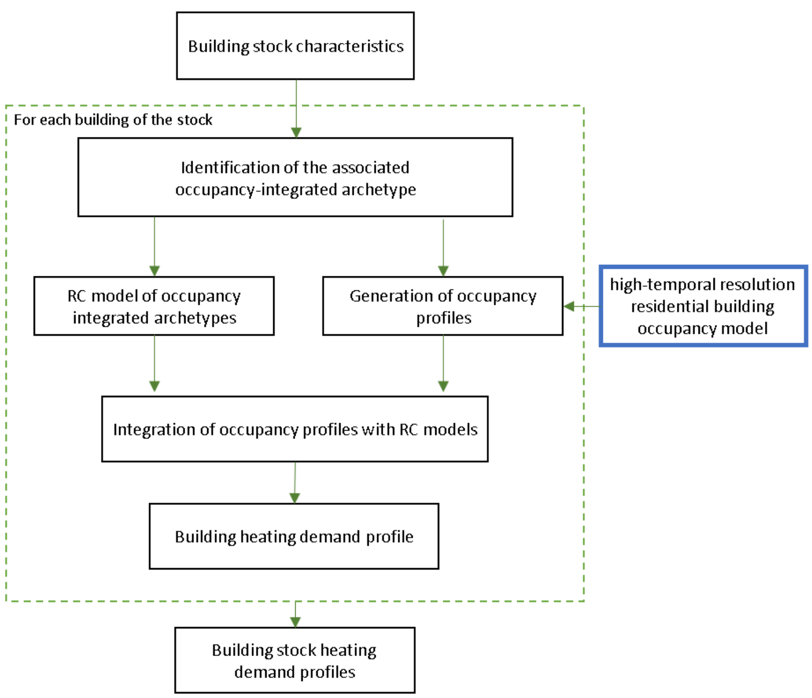

Figure 15 Scheme of the archetype-based stochastic building stock model

Given the characteristics of the building stock, each building of the stock is associated with an occupancy-integrated archetype having the most similar characteristics. The occupancy integrated archetypes in UK are classified according to the parameters presented in Table 3. The occupancy profiles correspond to those presented in Section 2.2. Once each building has been associated to an occupancy-integrated archetype, the occupancy profiles are generated by the high-temporal resolution residential building occupancy model presented in Section 2 of the paper. Such occupancy profiles are then used as inputs in the RC models of the occupancy-integrated archetypes representative of the building stock to obtain heating demand profiles. The building stock demand is obtained through the aggregation of the demand of the individual buildings.

Table 3 Archetype segmentation developed in [17]

Dwelling type	Construction year	Occupancy profiles	Climate zone
1. Flat	1. Pre-1918	1. Daily absence (OP1)	1. London
2. Bungalow	2. 1919-1964	2. Working hours absence (OP2)	2. Birmingham
3. Detached	3. 1965-1980	3. Lunch time absence (OP3)	3. Newcastle
4. Semi-detached	4. 1981-1990	4. Constant presence 1 (OP4)	4. Glasgow
5. Terrace	5. Post 1991	5. Constant presence 2 (OP5)	

In order to demonstrate the benefits of the application of the presented stochastic occupancy model to obtain high-temporal resolution heating load profiles, a hypothetical building stock of 100 identical new flats located in London is modelled. Firstly, deterministic occupancy profiles, corresponding to the modes of the occupancy categories (Figure 2), are used in the building stock model. Secondly, the stochastic profiles obtained from the high-temporal resolution residential building occupancy model developed in Section 2 are used to model the heating demand of the building stock. The difference between the two approaches is evaluated. In this case study, the scenario where the buildings are not associated with any specific occupancy profiles is also considered. In this case, it is possible to use the CREST model [26] to obtain the stochastic occupancy profiles to use as inputs in the archetype models. The comparison between the data obtained by the developed stochastic model and the CREST model is useful, as the CREST model is one of the most commonly used occupancy models in literature. Similar to the presented model, the CREST model is also based on the first-order Markov Chain Monte Carlo method, but the CREST model does not allow differentiated occupancy profiles for population subgroups characterised by different behaviours to be generated. The comparison of the results obtained using the presented stochastic models and the CREST model highlight the benefits of the application of the novel archetype-based stochastic model developed in the current research. The decision not to consider different archetypes from a structural point of view (e.g., older flats, houses) allows for isolation of the impact of occupant behaviour on building stock heating load demand to be examined.

3.1 Building stock characteristics

In the current study, a hypothetical building stock of 100 new flats located in London is modelled. It is assumed that these can be assimilated to the occupancy-integrated archetypes "Flat built after 1991" developed in [17]. The flats in the building stock can be characterised by OP 2 and 5 only. This means that the weekly days of the considered archetypes are associated to the categories represented by modes wd2 and wd5 identified in Section 2. This choice is due to the fact that the two categories together represent more than 50% of UK household occupancy profiles during weekdays (Figure 2(a)). In particular, the following cases are analysed: 100% households are associated with OP 5; 75% of households are associated with OP 5, 25% with OP 2; 50% of households are associated with OP 5, 50% with OP 2; 25% of households are associated with OP 5, 75% with OP 2; 100% households are associated with OP 2. As in Chapter 4, it is assumed that every flat is occupied by two adults.

3.2 Building model

The use of stochastic occupancy profiles obtained from the application of the presented model to the different categories allows to capture the diversity given by the stochastic behaviour of occupants. In order to take into account multiple stochastic occupancy profiles, multiple building simulations are necessary. Dynamic building equivalent Resistance-Capacitance (RC) models are used to simulate the archetypes in the current work resulting in reduced computational overhead associated with detailed simulation software models. Additionally, RC models are needed in building-to grid models [39]. In particular, the 10R7C heterogeneous RC model developed by Cabrera [38] for the apartment archetype is adopted in this paper. A calibration algorithm is used to calibrate the RC thermal network parameters [39]. Archetype energy models developed using the EnergyPlus simulation environment are used to generate reference synthetic data [17]. The RC model is calibrated by varying the values of the thermal resistances and capacitances of the RC model in order to minimise the discrepancy between the building internal temperature given by EnergyPlus and the RC model.

This yields a linear approximation of the thermal dynamics of the buildings

$$X^{t+1} = AX^{t} + BU^{t} + B_{heat}Q_{heat}^{t} \tag{4}$$

where X is the vector of state variables (i.e., temperature at the nodes of the RC model). The vectors U and Q_{heat} are the disturbance (e.g., weather) parameters and input heat gains at each node for time step t, respectively. The matrices $[A]_{n\times n}$, $[B]_{n\times m}$ and $[B_{heat}]_{n\times n}$ are state, disturbance and heat input coefficient matrices, respectively, where n is the number of nodes and m is the number of disturbance parameters. The RC network parameters used in the model are derived with the technique presented in [38].

This study is based on the assumption that the heating periods are coincident with the actively occupied periods. This assumption is adequate for modelling modern houses that are well insulated and for lightweight construction [40]. Heating desired average temperature is assumed to be 21 °C during the actively occupied hours, based on the observations presented in [41]. All the inputs required for the model (e.g., weather data, internal heat gains) are obtained as in [17]. As data from TUS 2015 UK are collected every 10 minutes, the time step used in the model is 10 minutes. In the model, it is assumed the heat produced by the heating unit is added directly to the indoor environment, such as the case with direct electric resistance heating systems. In particular, it is assumed that direct electric resistance heating systems of 3 kW are used in the modelled flats.

3.3 Resulting heating demand

3.3.1 Annual heating energy demand

Figure 16 compares the distribution of annual heating energy demand required by the building stock considering different building stock compositions. Building annual heating energy demand obtained using deterministic profiles as input in the model is observed to assume two values, depending on the associated household occupancy profiles. When stochastic profiles are used in the model, it is possible to observe that different combinations of occupancy integrated archetypes leads to different distributions of annual building stock energy heating demands, as the increase of the number of buildings characterised by OP5 leads to an increase of building characterised by a higher annual heating energy demand. Although the distribution of annual heating energy demand obtained using deterministic and stochastic profiles are significantly different, the difference between the annual average heating energy demand calculated with the stochastic profiles (E_{st}) and the deterministic profiles (E_{det}), calculated as $\frac{|E_{st} - E_{\mathrm{det}}|}{E_{st}}\ x\ 100$, is always less than 3%. Being aware that the distribution of annual heating demand across the building stock is not simulated when deterministic occupancy profiles are used, the use of the deterministic profiles for the model of the annual average heating energy demand of building stock is acceptable as the error that is introduced is negligible. The use of the CREST model does not allow different distributions of the households characterised by different occupancy profiles to be considered. Figure 16 shows that the proposed stochastic occupancy model could more accurately represent and estimate building stock annual heating energy consumption for building stocks characterised by certain occupancy profiles compared to the CREST model.

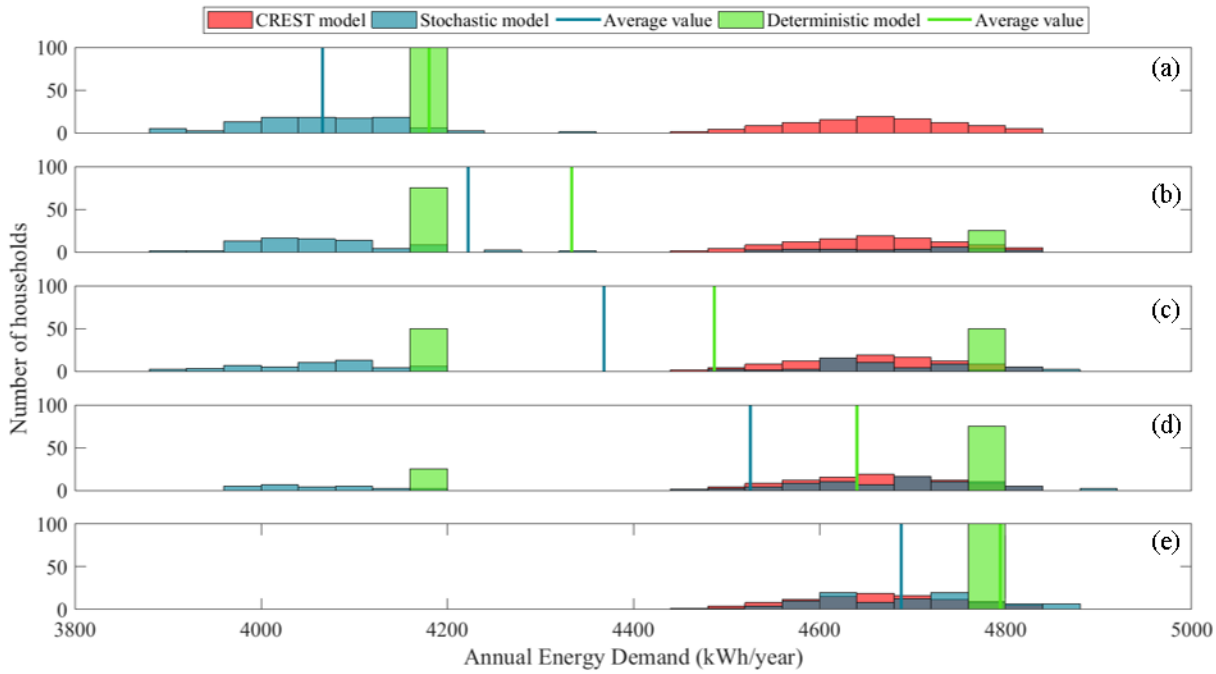

Figure 16 Distribution of annual heating energy consumption of the modelled building stock for the following cases: (a) 100% OP2; (b) 75% OP2 – 25% OP5; (c) 50% OP2 – 50% OP5; (d) 25% OP2 – 75% OP5; (e) 100% OP5.

3.3.2 Monthly heating energy demand

Since the environmental conditions change every month, the simulated heating energy demand could be affected. Comparison on a monthly basis provides a better resolution of how different building stock compositions affect building stock heating energy consumption for the various seasons. As expected, the difference between the total energy required by building stocks characterised by different occupancy profile compositions is higher for colder months (January, February, November, December), when the heating energy required is higher (Figure 17).

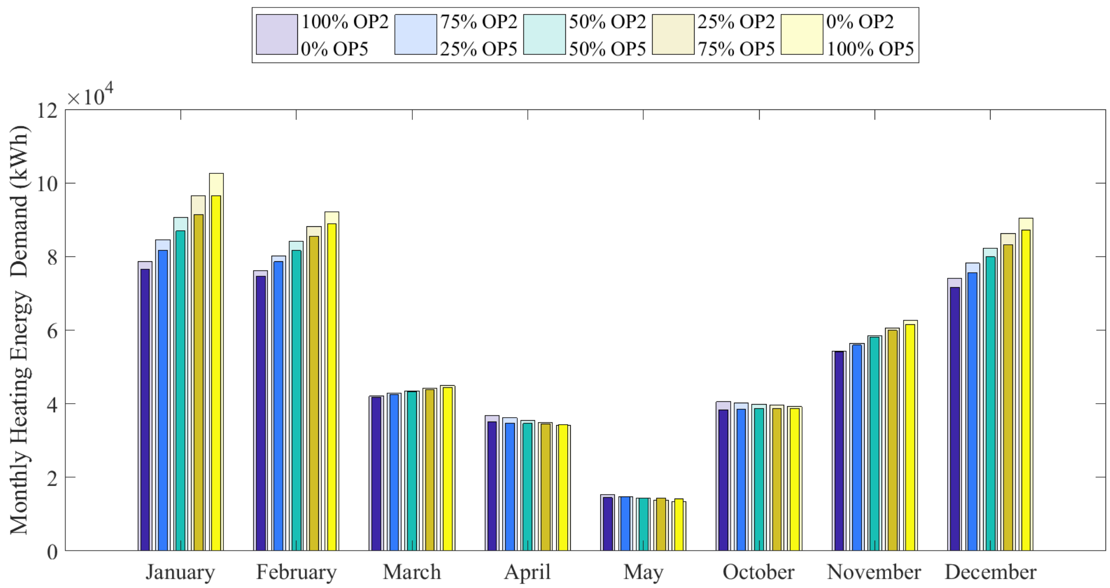

Figure 17 Building stock Heating Energy Demand for different months (Light coloured bars indicate the results obtained using deterministic profiles).

From the analysis of Figure 17, it is also possible to see that the use of deterministic profiles does not have a strong effect on the estimation of the monthly average heating energy demand (indicated by the light coloured bars). This means that also in the case in which the aim of the simulation is to calculate the monthly heating energy demand of the building stocks, the use of deterministic profile is sufficient.

The percentage difference between the monthly energy consumption obtained using the CREST model and the proposed stochastic model is calculated as $\frac{|E_{st,m}-E_{\mathrm{CREST,m}}|}{E_{st}}\ x\ 100$, where $E_{\mathrm{CREST,m}}$ is the monthly annual heating energy modelled using occupancy profiles generated by the CREST model as input (Figure 17). The differences between the results obtained by the two models are larger during colder months. Additionally, the difference between the results obtained using the presented stochastic model and the CREST model decreases when the percentage of archetype represented by OP 5 increases, as the transition matrixes used for the category wd5 are similar to the one used in the CREST model.

Figure 19 shows the distribution of monthly energy demand for the month of January, considering the different share of occupancy profile across the building stock. The behaviour is very similar to that observed for the annual heating demand distributions as shown in Figure 16, thus the same considerations are applicable as well in this case.

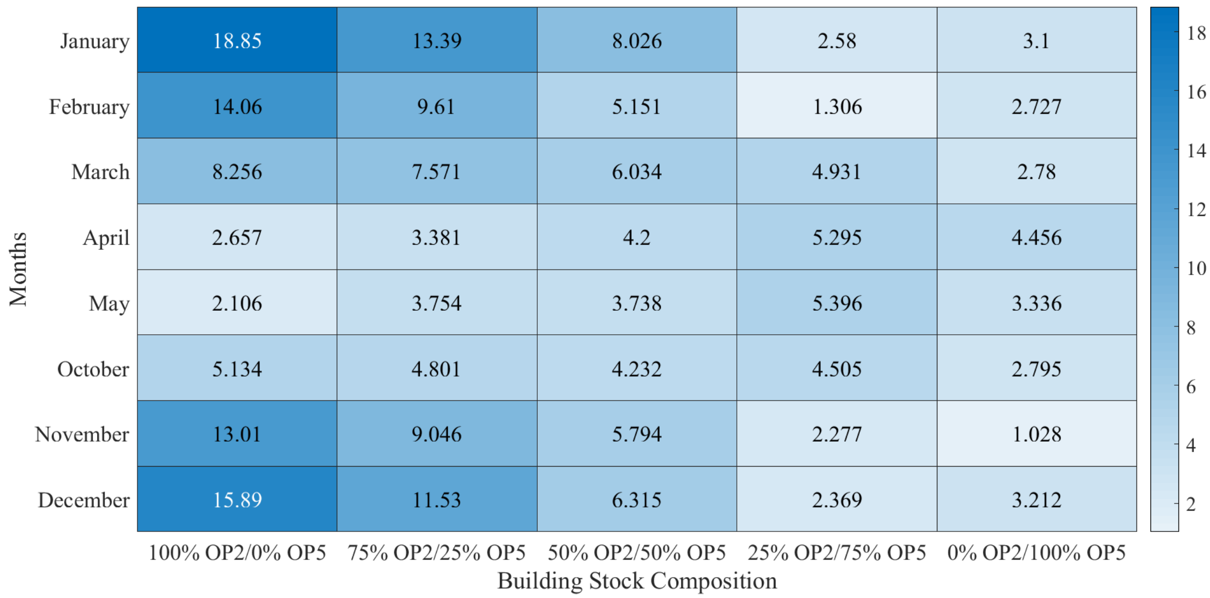

Figure 18 Percentage difference between results obtained using the proposed stochastic occupancy model and the CREST model

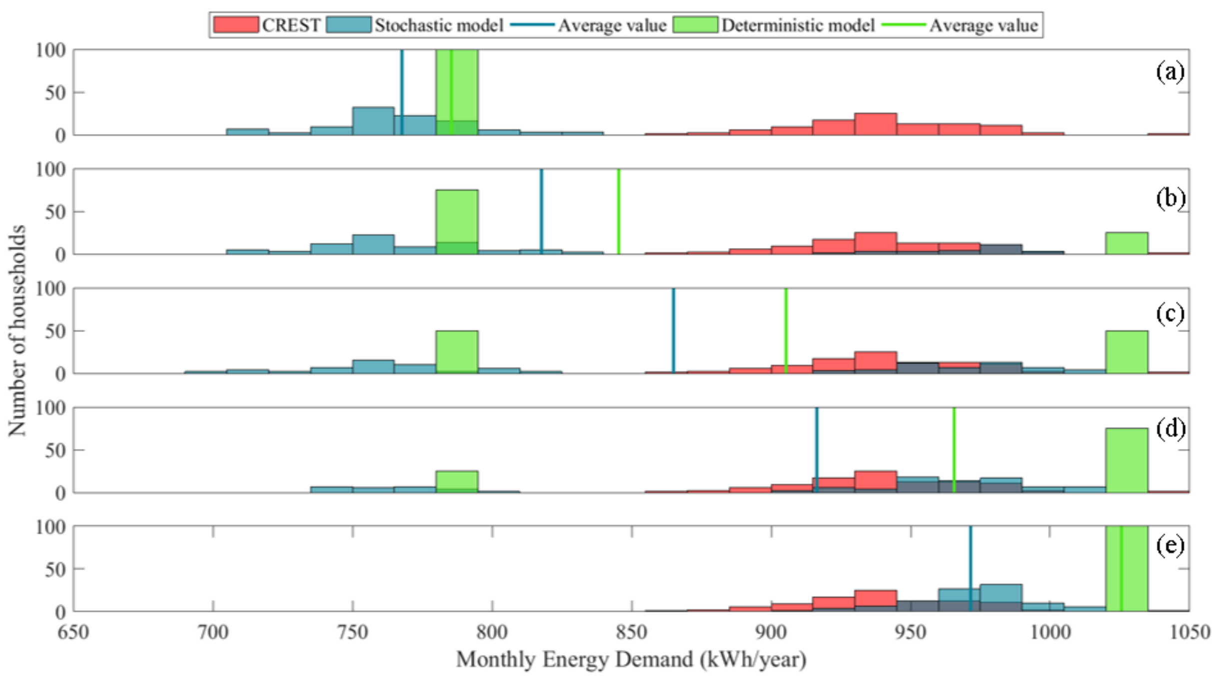

Figure 19 Distribution of monthly heating energy consumption of the modelled building stock for the month of January for the following cases: (a) 100% OP2; (b) 75% OP2 – 25% OP5; (c) 50% OP2 – 50% OP5; (d) 25% OP2 – 75% OP5; (e) 100% OP5.

3.3.3 Multi-day heating power load

One of the advantages of the proposed stochastic model is the possibility to consider patterns of consistency from day to day, as shown in Section 2.4.2. The use of multi-day occupancy profiles in the archetype models leads to the production of consistent multi-day heating power loads for a single household.

Figure 20 shows the multi-day heating power load of individual archetypes associated with different occupancy profiles. The results are shown from Jan 9th to Jan 15th 2012. This period has been selected to illustrate the different high-resolution heating load profiles in a generic winter week for archetypes characterised by different occupancy profiles. It is possible to see that for archetypes associated with OP 2 (Figure 20 (a)), the heating demand is usually required only in the morning and in the evening during weekdays, while it is required at any time of the working days for an archetype associated with OP 5 (Figure 20 (b)). This reflects that general behaviour of archetypes associated with OP 2, which are usually uninhabited during the day, while the archetypes associated with OP 5 are occupied the whole day. Figure 20 (c) also shows the multi-day heating power load of archetypes which are not associated with any particular occupancy profile. In this case, the occupancy profiles are obtained using the CREST model. It is possible to see that in this case it is not possible to recognise any consistent pattern from day to day, as already pointed out in [26].

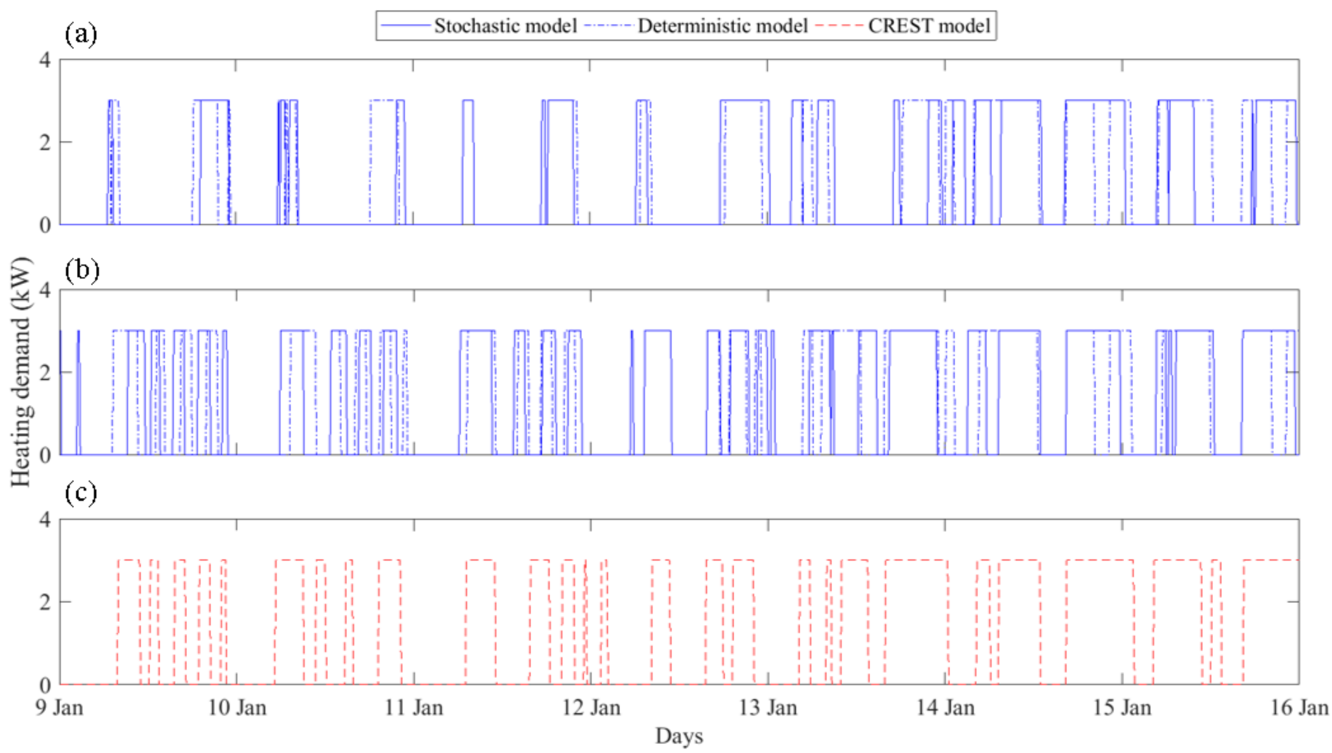

Figure 20 Multi-day heating power load of individual archetypes from Jan 9th to Jan 15th 2012 for the following cases: (a) archetype characterised by OP2; (b) archetype characterised by OP5; (c) no OP.

Figure 21 shows multi-day heating power load of the building stock from Jan 9th to Jan 15th 2012 for the following cases: (a) 100% OP 2; (b) 75% OP2 – 25% OP5; (c) 50% OP 2 – 50% OP5; (d) 25% OP2 – 75% OP5; (e) 100% OP5. Examining the results more closely, it is possible to see that building stock characterised by different occupancy compositions have different heating power demand profiles during the working days, while the power profiles during the weekend are similar for all the different cases. The results obtained using the profiles generated by the CREST model do not take into account the possible differences in composition of the occupancy profiles in the building stock. Thus, it can produce heating power loads which do not reflect the behaviour of the occupants of the building stock. This is particularly evident in Figure 21(a), where the power load in the case all the archetypes can be associated with OP 2 is shown. In this case, the heating demand is almost zero during the midday periods, and this behaviour cannot be replicated unless the data categorisation introduced in Session 2.2 is performed.

Additionally, Figure 21 also shows that when the aim of the model is to obtain high-resolution profiles of heating energy demand, the use of deterministic profiles produces unrealistic peak power demands due to the fact that all buildings with the same occupancy profile require the heating to be on at the same time. Thus, in this case, the use of stochastic occupancy profiles is recommended to simulate the diversity of occupant behaviour.

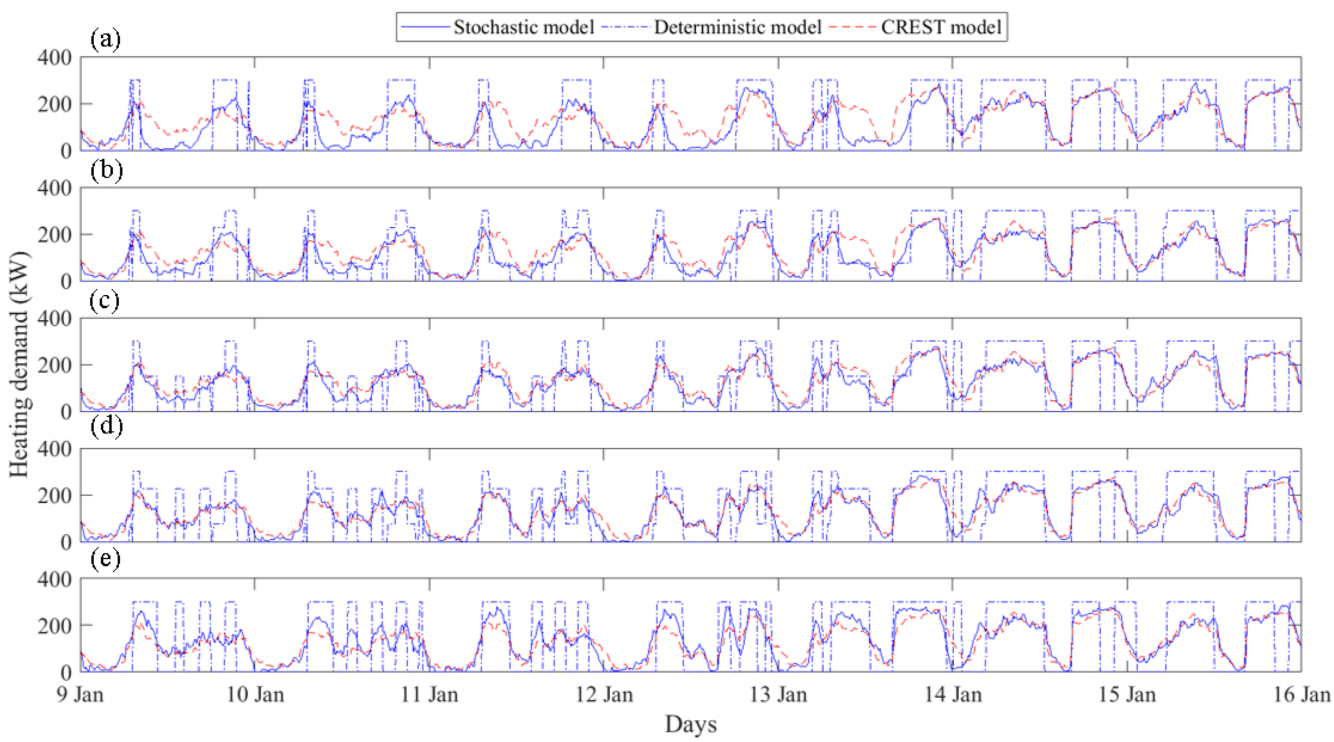

Figure 21 Multi-day heating power load of the building stock (100 archetypes) from Jan 9th to Jan 15th 2012 for the following cases: (a) 100% OP2; (b) 75% OP2 – 25% OP5; (c) 50% OP2 – 50% OP5; (d) 25% OP2 – 75% OP5; (e) 100% OP5.

3.3.4 Daily average heating power load

Figure 22 shows the average daily heating energy demand of the building stock for the month of January, considering different building stock compositions. In Figure 22 (a-e), the average heating demand obtained considering all the working days for different building stock compositions is shown, while in Figure 22 (f) the average heating demand obtained as average heating demand of all the non-working days is presented. This is not differentiated according to the different building stock compositions, as the transition matrices used to obtain the stochastic occupancy profiles of the non-working days do not change when different weekly occupancy profiles are considered.

Once again, for the working days, different building stock compositions give rise to different results, although the trend is similar. The heating energy demand is lowest during the night, followed by an initial peak at breakfast time. Then, the demand decreases during the day, until the late afternoon, when it rises towards an evening peak after which it drops again during the night. The magnitude of the required heating demand is different when different building stock occupancy compositions are considered.

It is clearly visible that when high-resolution profiles of heating energy demand are required, the use of deterministic profiles produces unrealistic power peak demand profiles, as a consequence of multiple requests of heating power at the same time. Thus, the use of deterministic occupancy profiles is not recommended when high-temporal resolution heating load profiles need to be obtained from the energy model of the building stock. The results obtained using the profiles generated the CREST model do not take into account the possible differences in composition of the occupancy profiles in the building stock. Thus, the adoption of occupancy profiles generated by the CREST model can produce heating power loads which do not reflect the behaviour of the occupants of the building stock. As anticipated by Figure 21, this difference is remarkable especially when all the archetypes can be associated with OP 2 (Figure 22 (a)). In this case, the behaviour of occupants is completely different from the one simulated using the CREST model. This shows the importance of using the proposed stochastic model in the case where the building stock is characterised by households which can be associated with occupancy profiles which are different from the national average.

While the shape of the average daily heating power loads is very similar for different months, the magnitude is different, as some months are colder than others. This is also reflected in the results of monthly heating energy demand obtained in Section 3.3.2. The results obtained using the occupancy profiles generated by the presented stochastic model are also compared to the results obtained using real occupancy profiles extracted from TUS data for each category. Because the occupancy profiles are very similar (Figure 9), the resulting high-temporal resolution heating load profiles are very similar. Also in this case, the Pearson coefficient is approximately equal to one in all cases. The 95% confidence interval of the results obtained using the occupancy profiles generated by the stochastic occupancy model presented in Section 2 is also shown in Figure 22.

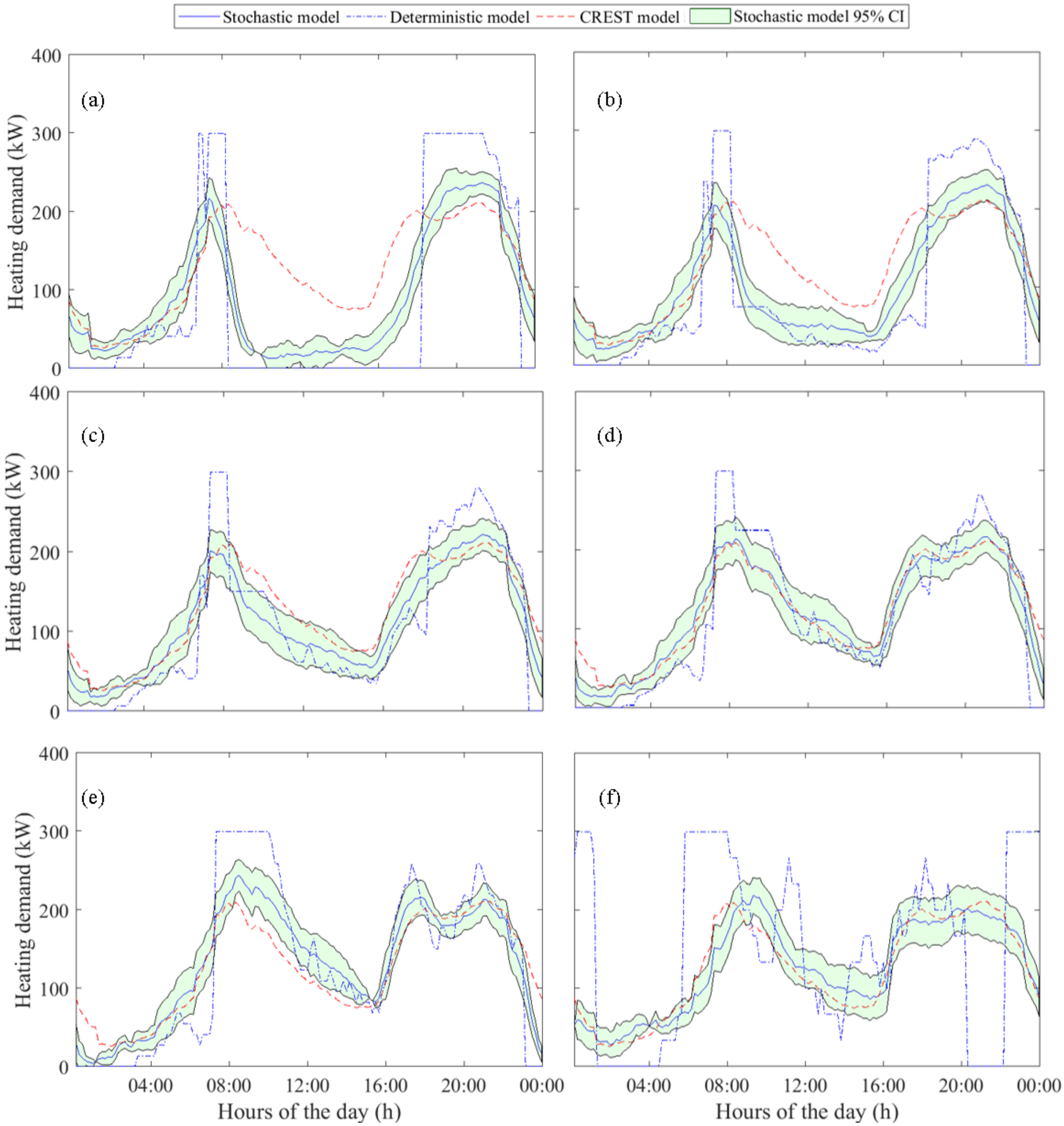

Figure 22 Building stock average daily heating energy demand in January: (a) Weekday 100% OP2, (b) Weekday 75% OP2\25%OP5, (c) Weekday 50% OP2\50%OP5, (d) Weekday 25% OP2\75%OP5, (e) Weekday 100%OP5, (f) Weekend

4 Discussion

This study proposes a new stochastic occupancy model which can be integrated with occupancy-integrated archetypes in order to produce accurate high-temporal resolution heating load profiles. The resulting model is characterised by 3 fundamental properties which are: (i) scalability; (ii) adaptability; (iii) representative of diversity.

The model is scalable as it can be applied to building stocks of different size, provided that each building of the considered building stock can be associated with one of the archetypes identified at national scale. Additionally, thanks to the use of occupancy-integrated archetypes, the model can be also applied to building stocks characterised by occupancy profiles which do not correspond to the national average. This assures the adaptability of the model to different building stock. These characteristics can be obtained also using occupancy-integrated archetypes with deterministic occupancy profiles. However, if deterministic profiles are used, only accurate average monthly and annual heating energy demand estimates can be obtained, as evident from the results presented in Section 3.3.1 and 3.3.2. The use of deterministic occupancy profiles does not allow the diversity of occupant behaviour to be taken into account when high-temporal resolution heating load profiles must be simulated. The use of deterministic profiles in this context produces unrealistic peak demands (Figure 22).

On the other end, the use of the well-established CREST model allows the diversity of occupant behaviour in building stock to be modelled, but it does not facilitate households to be characterised by specific occupancy profiles. Thus, if occupancy profiles generated by the CREST model are integrated into the archetypes, it is no longer possible to adapt the model to building stock having specific occupancy profile compositions, thereby lacking adaptability. The developed stochastic occupancy model allows to take into account the diversity of occupant behaviour within the model, maintaining the characteristics of scalability and adaptability already introduced thanks to the development of occupancy-integrated archetypes. It is important to emphasise that the CREST model is still valuable when the building stock is characterised by occupancy profiles which correspond to the national average. Also, the use of the deterministic occupancy profiles is still preferred in the case where the aim of the building stock model is to obtain the average daily or annual heating energy demand, as it is the simplest solution to adopt and the results are not far from the results which are obtained using the statistical model (Figure 16 and Figure 17).

Finally, the current work is based on the assumption that heating hours are coincident with occupied hours and that the heat produced by the heating emitter is added directly to the indoor environment. However, these assumptions are not always realistic when different dwelling types and heating systems are considered. Further possibilities will be explored in future work, extending the presented approach to different heating systems and different dwellings.

5. **Conclusions**

The current paper develops a new stochastic occupancy model which can be applied to occupancy-integrated archetypes to obtain high-temporal resolution heating load profiles. The proposed model is capable of modelling building stocks at different scales and characterised by different occupancy profiles, taking into account the diversity of occupant behaviour. The model, verified against data contained in the UK TUS, shows acceptable accuracy. The Mean Absolute Error, Root Mean Square Error and Coefficient of Variations calculated from the comparison of the results obtained from the stochastic model and the data available from TUS are all approximately zero. In addition, the comparison of the transitions between two different household states in each time step shows that the accuracy of the model is high.

An implementation of the model is available for download and may be adopted and adapted for any application in which the simulation of occupancy profiles for population subgroups characterised by specific occupancy profiles is required.

In this paper, by means of a case study, the stochastic model is applied to calculate the heating demand of a hypothetical building stock of 100 new flats located in London, and the results are compared to the ones that can be obtained using deterministic profiles and the CREST model. From the analysis of the results, it is possible to see how the stochasticity of the occupancy profiles allows the diversity of the occupant behaviour in the generation of high-temporal resolution heating load profiles to be considered. Compared to the CREST model, this model offers the possibility to create multiple alternative occupancy schedules based on the association of the occupancy-integrated archetype to a specific occupancy profile.

Therefore, the methodology can be used to obtain high-temporal resolution heating power loads of building stocks characterised by occupants having a different behaviour than the national average. In this way, more accurate models can be obtained, which can be used as a basis to provide more meaningful policy recommendations thanks to a more accurate estimation of potential energy and cost savings.

Acknowledgement

This work is part of the Real Value project. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 646116.

References

[1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy Build.*, vol. 128, pp. 198–213, 2016.

[2] L. Diao, Y. Sun, Z. Chen, and J. Chen, "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation," *Energy Build.*, vol. 147, pp. 47–66, 2017.

[3] H. Lim and Z. (John) Zhainwoo, "Review on stochastic modeling methods for building stock energy prediction," *Build. Simul.*, vol. 10, pp. 607–624, 2017.

[4] C. F. Reinhart and C. Cerezo Davila, "Urban building energy modeling - A review of a nascent field," *Build. Environ.*, vol. 97, pp. 196–202, 2016.

[5] G. Happle, J. A. Fonseca, and A. Schlueter, "A review on occupant behavior in urban building energy models," *Energy Build.*, vol. 174, pp. 276–292, 2018.

[6] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renew. Sustain. Energy Rev.*, vol. 13, no. 8, pp. 1819–1835, 2009.

[7] a. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," *Renew. Sustain. Energy Rev.*, vol. 16, no. 9, pp. 6539–6565, 2012.

[8] B. F. Balvedi, E. Ghisi, and R. Lamberts, "A review of occupant behaviour in residential buildings," *Energy Build.*, vol. 174, pp. 495–505, 2018.

[9] I. Gaetani, P. J. Hoes, and J. L. M. Hensen, "Occupant behavior in building energy simulation: Towards a fit-for-purpose modeling strategy," *Energy Build.*, vol. 121, pp. 188–204, 2016.

[10] M. Kavgic, A. Mavrogianni, D. Mumovic, A. Summerfield, Z. Stevanovic, and M. Djurovic-Petrovic, "A review of bottom-up building stock models for energy consumption in the residential sector," *Build. Environ.*, vol. 45, no. 7, pp. 1683–1697, 2010.

[11] N. Good, L. Zhang, A. Navarro-espinosa, and P. Mancarella, "High resolution modelling of multi-energy domestic demand profiles," *Appl. Energy*, vol. 137, pp. 193–210, 2015.

[12] A. Molina-García, A. Gabaldón-Marin, C. Álvarez, J. a Fuentes-Moreno, and E. Gómez-Lazaro, "Electrical thermal storage modeling: a tool to evaluate new opportunities and bids for residential users," in *2003 IEEE Bologna Power Tech Conference Proceedings*, 2003.

[13] D. B. Crawley, J. W. Hand, M. Kummert, and B. T. Griffith, "Contrasting the capabilities of building energy performance simulation programs," *Build. Environ.*, vol. 43, pp. 661–673, 2008.

[14] É. Mata, A. Sasic Kalagasidis, and F. Johnsson, "Building-stock aggregation through archetype buildings: France, Germany, Spain and the UK," *Build. Environ.*, vol. 81, pp. 270–282, 2014.

[15] I. Ballarini, S. P. Corgnati, and V. Corrado, "Use of reference buildings to assess the energy saving potentials of the residential building stock: The experience of TABULA project," *Energy Policy*, vol. 68, pp. 273–284, 2014.

[16] A. A. Famuyibo, A. Duffy, and P. Strachan, "Developing archetypes for domestic dwellings - An Irish case

study," *Energy Build.*, vol. 50, pp. 150–157, 2012.

[17] G. Buttitta, W. J. N. Turner, O. Neu, and D. P. Finn, "Development of occupancy-integrated archetypes : Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes," *Energy Build.*, vol. 198, pp. 84–99, 2019.

[18] BRE, "SAP 2016 The Government's Standard Assessment Procedure for Energy Rating of Dwellings," 2016.

[19] SEAI, "Dwelling Energy Assessment Procedure (DEAP) and Non- Dwelling Energy Assessment Procedure (NEAP)," *SI 243 2012*, no. June, pp. 19–20, 2012.

[20] R. Yao and K. Steemers, "A method of formulating energy load profile for domestic buildings in the UK," *Energy Build.*, vol. 37, no. 6, pp. 663–671, 2005.

[21] M. Stokes, M. Rylatt, and K. Lomas, "A simple model of domestic lighting demand," *Energy Build.*, vol. 36, no. 2, pp. 103–116, 2004.

[22] J. V. Paatero and P. D. Lund, "A model for generating household electricity load profiles," *Int. J. Energy Res.*, vol. 30, no. 5, pp. 273–290, 2006.

[23] C. F. Walker and J. L. Pokoski, "Residential Load Shape Modeling Based on Customer Behavior," *IEE Power Energy Soc.*, no. 7, pp. 1703–1711, 1985.

[24] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenzi, "Bottom-up approach to residential load modeling," *IEEE Trans. Power Syst.*, vol. 9, no. 2, pp. 957–964, 1994.

[25] J. Widén, A. M. Nilsson, and E. Wäckelgård, "A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand," *Energy Build.*, vol. 41, no. 10, pp. 1001–1012, 2009.

[26] 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.

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

[28] D. Aerts, J. Minnen, I. Glorieux, I. Wouters, and F. Descamps, "A method for the identification and modelling of realistic domestic occupancy sequences for building energy demand simulations and peer comparison," *Build. Environ.*, vol. 75, pp. 67–78, 2014.

[29] J. Widén and E. Wäckelgård, "A high-resolution stochastic model of domestic activity patterns and electricity demand," *Appl. Energy*, vol. 87, no. 6, pp. 1880–1892, Jun. 2010.

[30] U. Wilke, F. Haldi, J. L. Scartezzini, and D. Robinson, "A bottom-up stochastic model to predict building occupants' time-dependent activities," *Build. Environ.*, vol. 60, pp. 254–264, 2013.

[31] U. Wilke, "Probabilistic Bottom-up Modelling of Occupancy and Activities to Predict Electricity Demand in Residential Buildings," *École Polytech. fédérale Lausanne EPFL*, vol. 5673, 2013.

[32] G. Buttitta and D. P. Finn, "A high-temporal resolution residential building occupancy model to be used for

occupancy-integrated archetypes." Mendeley Data, 2019.

[33] J. Gershuny and O. Sullivan, "United Kingdom Time Use Survey, 2014-2015. [data collection]." UK Data Service., 2017.

[34] G. Buttitta, W. J. N. . Turner, O. Neu, and D. Finn, "Modelling residential building stock heating load demand - Comparison of occupancy models at large scale," in *2017 ASHRAE Annual Conference, June 2017, Long Beach*, 2017.

[35] G. Buttitta, O. Neu, W. Turner, and D. Finn, "Modelling Household Occupancy Profiles using Data Mining Clustering Techniques on Time Use Data School of Electric and Electronic Engineering , University College Dublin , Dublin , Ireland School of Mechanical and Materials Engineering , University Colleg," in *IBPSA Building Simulation 2017*, 2017.

[36] I. Teodorescu, "Maximum Likelihood Estimation for Markov Chains," *arXiv Prepr. arXiv0905.4131v1*, 2009.

[37] E. W. Weisstein, "Markov Chain," *MathWorld--A Wolfram Web Resource.* [Online]. Available: http://mathworld.wolfram.com/MarkovChain.html.

[38] C. H. Andrade-Cabrera, "Ensemble Calibration of Archetype Building Energy Models - PhD Thesis," University College Dublin, 2019.

[39] C. Andrade-Cabrera, W. J. N. Turner, D. Burke, O. Neu, and D. P. Finn, "Lumped Parameter Building Model Calibration using Particle Swarm Optimization," 2016.

[40] M. He, T. Lee, S. Taylor, S. K. Firth, and K. Lomas, "Coupling a stochastic occupancy model to EnergyPlus to predict hourly thermal demand of a neighbourhood," in *Proceedings of BS2015: 14th Conference of International Building Performance Simulation Association, Hyderabad, India, Dec. 7-9, 2015*, 2015.

[41] M. Shipworth, S. K. Firth, M. I. Gentry, A. J. Wright, D. T. Shipworth, and K. J. Lomas, "Central heating thermostat settings and timing: building demographics," *Build. Res. Inf.*, vol. 38, no. 1, pp. 50–69, 2010.