Title	Development of occupancy-integrated archetypes: Use of data mining clustering techniques to embed occupant behaviour profiles in archetypes
Authors(s)	Buttitta, Giuseppina, Turner, William J. N., Neu, Olivier, Finn, Donal
Publication date	2019-09-01
Publication information	Buttitta, Giuseppina, William J. N. Turner, Olivier Neu, and Donal Finn. "Development of Occupancy-Integrated Archetypes: Use of Data Mining Clustering Techniques to Embed Occupant Behaviour Profiles in Archetypes" 198 (September 1, 2019).
Publisher	Elsevier
Item record/more information	http://hdl.handle.net/10197/11565
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 (198, (22019)) https://doi.org/10.1016/j.enbuild.2019.05.056
Publisher's version (DOI)	10.1016/j.enbuild.2019.05.056

Development of Occupancy-integrated Archetypes: Use of Data Mining Clustering Techniques to embed occupant behaviour profiles in archetypes

Buttitta Giuseppina, Turner William J.N., Neu Olivier, Finn Donal P.
University College Dublin, Dublin D04 V1W8, Ireland

Abstract

Building stock modelling usually deploys representative building archetypes to obtain reliable results of annual energy heating demand and to minimise the associated computational cost. Available methodologies define archetypes considering only the physical characteristics of buildings. Uniform occupancy schedules, which correspond to national averages, are generally used in archetype energy simulations, despite evidence of occupancy schedules which can vary considerably for each building. This paper presents a new methodology to define occupancy-integrated archetypes. The novel feature of these archetype models is the integration of different occupancy schedules within the archetype itself. This allows building stock energy simulations of national population subgroups characterised by specific occupancy profiles to be undertaken. The importance of including occupant-related data in residential archetypes, which is different than the national average, is demonstrated by applying the methodology to the UK national building stock. The resultant occupancy-integrated archetypes are then modelled to obtain the annual final heating energy demand. It is shown that the relative difference between the heating demand of occupancy-integrated archetypes and uniform occupancy archetypes can be up to 30%.

Keywords: residential buildings, archetypes, stock modelling, k-mode clustering, occupancy profiles

1 Introduction

This paper introduces a new methodology that utilises data-mining clustering techniques to augment building archetype definitions with occupancy schedule information. Occupancy schedules identify the time periods during which dwellings are unoccupied or actively occupied. Dwellings are considered to be actively occupied, when at least one person is in the house and not asleep. The proposed archetypes are classified as *occupancy-integrated archetypes*. To date building archetypes have been largely characterised on the basis of building form and fabric data [1–3]. The addition of occupancy information adds an important dimension to the archetype definition, as user behaviour is known to have a significant influence on energy usage within the residential sector [4]. In particular, capturing the wide variation in user behaviour leads to more accurate energy demand simulation of building stock models, which can be used to estimate the energy consumption and to evaluate the effects of potential energy policies.

New energy policies are prompted by the increased penetration of renewable energy resources and the associated need to reduce the use of fossil fuels for space and domestic water heating loads in the residential sector.

* Corresponding author. Tel. +39 3533214489
E-mail address: giuseppina.buttitta@ucdconnect.ie

Within a European context, the promotion of renewable energy measures and the reduction of energy consumption and greenhouse gas emissions are key policy components of European short and long-term strategic goals [5-7]. Improving the energy performance of the residential building sector represents a key pathway to achieving EU energy targets, as this sector accounts for approximately 25% of energy consumption [8]. Amongst the different energy policies, the development of smart grids with integrated energy-flexible buildings, has been identified as a critical target towards this goal [9]. Energy-flexible buildings have the capability of adjusting their electricity demand to produce desirable changes in the magnitude and shape of their electricity loads, while retaining customer satisfaction and occupant comfort [10].

It is essential that building stock models provide reliable results in a transparent way in order to use them in the development of energy policies [4, 11]. The evaluation of the energy consumption of large building stock is challenging and has been approached in different ways by various researchers [12 – 16]. In these papers ([12 –16]), two distinct methodologies can be identified: 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. For this reason, a bottom-up approach is useful when the aim of the analysis is the large-scale assessment of the contribution of each building towards the aggregate energy consumption of the entire building stock. The bottom-up approach can be used to examine the effect and impact of new policies and technologies, and to assess customer engagement at an individual level. The bottom-up approach includes the archetype approach [12]. This approach is based on the classification of stock buildings within groups which are defined by similar characteristics, and are then represented by unique buildings called archetypes [16]. Archetypes are statistical composites of building category features, identified considering simplifications that depend on the intended analysis [17]. The main advantage of this technique is the reduction of the simulation time and overhead, as the number of archetypes is reduced compared to the total number of buildings in the stock.

Many methodologies to develop archetypes have been implemented in recent years [1–3]. A clear methodology is described in [1], where archetype characteristics are defined using four steps:

1. Segmentation: the number of archetypes is determined by partitioning the building stock according to pre-determined segmentation criteria, such as construction period and dwelling size/form;
2. Characterisation: each archetype is described by its technical characteristics, such as fabric U-value;
3. Quantification: the distribution of archetype buildings is determined in order to be representative of the building stock;
4. Validation: the final energy demand of the building stock is calculated using the defined archetypes and it is compared with data.

The methodology developed in [1] can be integrated with the one developed in [2], which can be used to define archetype technical characteristics which are part of the "Characterisation" step. These technical characteristics can be gathered into four main data subsets: form, envelope, operation and heating system [2]. The 'form' subset indicates the general geometry of the buildings, the 'envelope' subset considers the thermophysical properties of building envelopes, the 'system' subset includes the heating and cooling systems, as well as any HVAC and lighting systems, and the 'operation' subset indicates the parameters which affect the usage of the building and are expressed by means of a set of schedules, such as occupancy schedules. Once the building archetypes are defined, the energy consumption estimates of the archetypes can be scaled to be representative of the entire building stock, by multiplying the energy consumed by each archetype by the number of houses that it represents within the building stock [1].

The most common segmentation parameters used in literature are: shape, age, climate, heating system and primary energy [16, 18]. These parameters can define archetypes in just three (form, envelope, and system) of the four subsets which must be used according to [2]. However, archetypes must also be defined using data from the operation subset, but there is no developed methodology to classify residential archetypes considering operational schedules. From the reviewed literature, none of the available methodologies include occupant-related variables in their definition of residential archetypes, despite the observation that operation subset data should be included in the archetype definition process, as argued by [2]. The generalisation of occupancy schedules is considered one of the major shortcomings of current building stock energy models [11, 19], because energy consumption profiles are strongly correlated to occupancy schedules [20]. Additionally, several studies [3, 21] demonstrated that parameters which depend on occupancy schedules, such us heating patterns, have the strongest impact on building space heating energy demand. The analysis of measured living-room temperatures in the UK [22] showed that a large variability in internal temperature and heating pattern parameters exist between homes, challenging the assumption that one occupancy schedule fits all [23].

The use of non-differentiated occupancy schedules could be justifiable to model energy consumption at national level, because the behavioural variations amongst the individual households tend to even out when considered at large scale. However, this approach is not appropriate when it is used in national population subgroups, as the deviation of occupancy schedules from the national average becomes more significant in smaller population groups [11, 24]. The implementation of differentiated heating patterns for subgroups of the population is crucial to predict heating energy demand correctly. Thus, in order to apply the archetype approach to population subgroups, it is important to understand how to include occupant-related variables (i.e., occupancy schedules) as operational inputs to residential archetypes.

The aim of the present paper is to develop a methodology that identifies the population subgroups characterised by similar occupancy schedules. These schedules are further embedded into building archetype models to improve the estimation of annual final heating energy consumption in residential building stock, where the annual final heating energy consumption indicates the energy required to heat the building at the desired temperature, regardless of the deployed heating system. These types of archetypes, hereafter called *Occupancy-integrated* archetypes, can be used to improve building stock energy predictions when the occupancy profiles are different than the national average. The proposed methodology uses the k-modes clustering technique to achieve this aim [25]. Clustering is a common technique used for statistical data analysis, which partitions datasets into sets of well-defined groups (clusters), whose elements exhibit common similarities compared to other groups [26]. This technique has been presented in previous works of the authors [27 – 29] to group together population subgroups characterised by similar occupancy profiles. The k-mode technique has also been used in [20] to recognise behaviour patterns from the American Time Use Survey (ATUS), which was then used as input to the demographic-based probability neural networks to simulate occupancy behaviour, while hierarchical clustering has been used by other researchers [30]. The novelty introduced by the current paper is the use of the occupancy profiles obtained from clustering to characterise archetypes. In the current paper, the clustering approach is applied to datasets from national Time Use Surveys (TUS) in order to identify households characterised by similar occupancy schedules. These occupancy schedules are used to ascertain appropriate heating patterns, which are then integrated within the archetype to improve the estimation of annual final heating energy consumption in residential buildings. The heating patterns directly correlate to occupancy schedules, which are used as one of the segmentation parameters in the residential archetype definition. This proposed novel approach allows the definition of differentiated heating patterns in different archetypes, in order to capture the diversity of residential heating energy demand in building stock characterized by different percentage of households having different occupancy profiles.

The methodology developed here is applied to existing UK housing stock data. The results of the energy model using the archetypes are further compared with results obtained using the standard heating patterns for the UK [31] and against data collected in the National Energy Efficiency Data-Framework (NEED) [32]. Moreover, because the adoption of a unique representative set-point for all the archetypes is not realistic, a sensitivity analysis on the influence of the required average internal temperature on energy consumption is also undertaken.

The remainder of the paper is arranged as follows. Section 2 describes the methodology. Section 3 evaluates the developed archetypes by comparing the annual final heating energy consumption obtained from simulations of the archetypes against the NEED data. The results are obtained considering different required average internal temperature to evaluate the impact of this variable on the annual final heating energy consumption. Section 4 presents the results of the described methodology. The potential applications and limitations of the methodology are described in Section 5. Section 6 concludes the paper.

2 Methodology

This section describes the methodology by which the occupancy-integrated archetype characteristics are defined, with a special focus on occupant-related characteristics. The introduced methodology is schematised in the "Archetype definition" sub-section of Figure 1. Archetype characteristics can be defined in three steps:

i. Variable Identification: where all the relevant variables are identified;
ii. Building Stock Segmentation: where the number of archetypes is determined by partitioning the building stock according to determined segmentation criteria, such as construction period and dwelling shape;
iii. Characterisation: where each archetype is described by its technical characteristics, such as U-value.

The final archetype characteristics chosen depend on the intended use of the archetype models [17]. For the current research, the intended use is to analyse the annual final heating energy demand of residential building stock. In the current paper, archetypes are defined using the "envelope" and "form" subsets by adapting and combining the methodologies presented in [1–3]. Occupancy schedules are developed to define archetypes in the "operation" subset. These correspond to the most common schedules of the building stock, which are obtained by applying the clustering technique to the data collected by national Time Use Surveys (TUSs), as in [20, 28]. This novelty significantly enhances the usefulness of the archetype approach and allows better discrimination of energy end-use based on household groups. The operation subset is also defined by the climate zone. Once occupancy-integrated archetypes are defined, they are modelled in EnergyPlus (Version 8.5) and the annual final heating energy consumption is analysed and compared with reference data to validate the archetype models.

The terms which are used in the methodology are defined as follows: `occupant state' indicates the condition of the person who occupies the house, which can be active, non-active or absent; `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; `household state' indicates the condition of the entire household, which depends on the individual occupant state (e.g., if the occupant state of all occupants is 'non-active', the household occupant state will be 'non-active' as well); `household daily occupancy profile' indicates the sequence of the household states during one day.

In the following sections, each of the three aforementioned steps of the methodology (Variable Identification, Segmentation, Characterisation) is described in further detail.

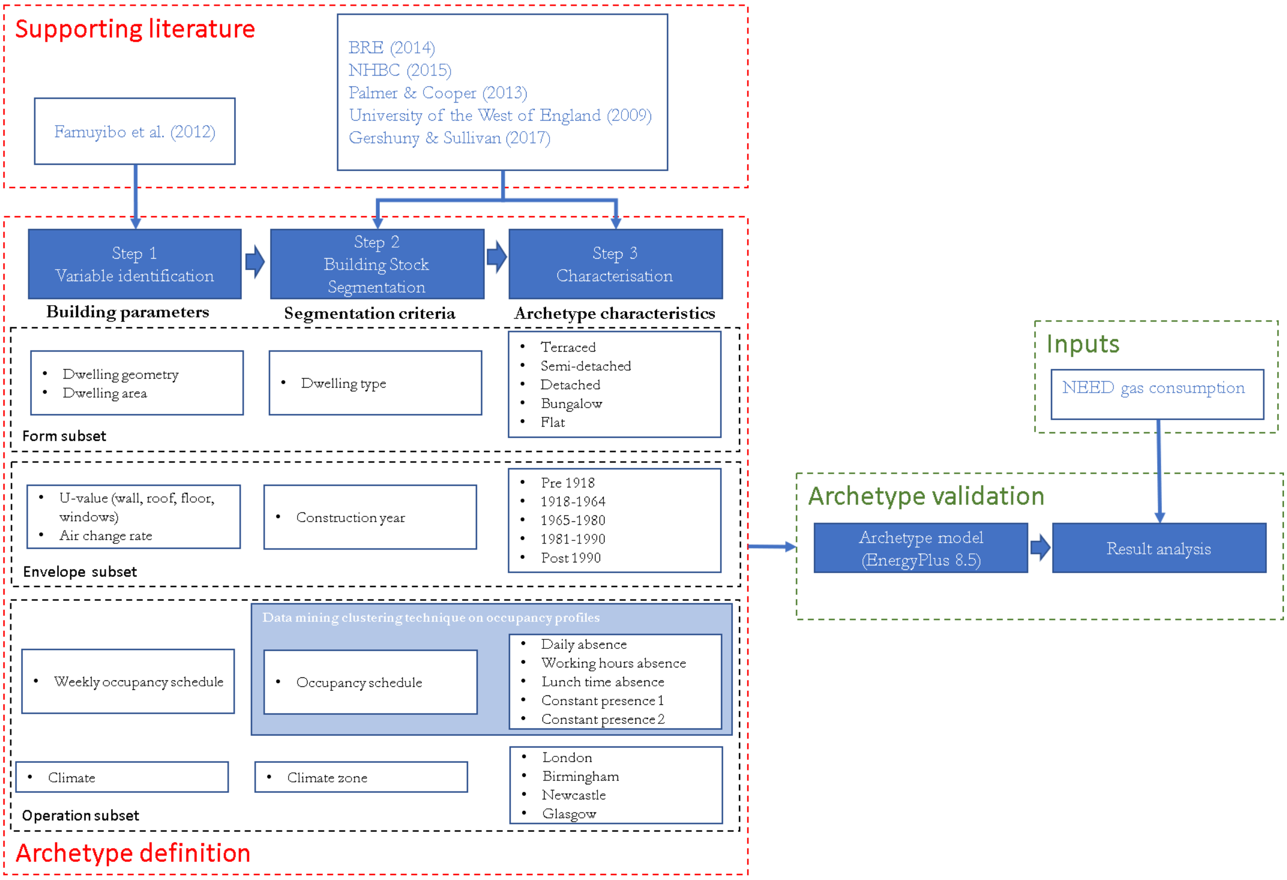

Figure 1 Overview of the proposed archetype development model

2.1 Step 1 - Variable Identification

Step 1 is concerned with the identification of the variables which predominantly influence residential heating energy consumption. The process is fundamental as the number of variables which affects the heating demand of buildings is immense and it is not possible to take into account all of them in the archetype definitions. The variable identification has been previously performed in [3]. In the paper [3], the most important variables linked to the energy consumption are selected via a literature review, and then they are ranked in terms of impact by performing a regression analysis. The results of the regression analysis show that the most relevant variables affecting heating energy consumption, from highest to lowest impact, are: internal temperature; weekly occupancy schedules (occupancy schedules during the week); air change rate; domestic hot water (DHW) use frequency. However, these variables are not sufficient to provide the parameters to define representative archetypes, thus additional variables were selected from the literature review and included: wall, ceiling, floor and window U-values; dwelling geometry; heating system; DHW system; dwelling floor area; climate.

In the current paper, the approach utilised in [3] to identify archetype variables is applied to the UK building stock. For their archetype characterisation, [3] did not include occupant-related variables. In order to improve the methodology, in the current research, weekly occupancy schedules are considered as one on the key inputs to identify occupancy-integrated archetypes. This inclusion of occupancy schedules is justified on the basis that energy load profiles are strongly influenced by household occupancy profiles, and in particular by the succession of occupied and unoccupied periods [29]. This relationship between occupancy and load, as demonstrated by previous researchers [3, 24], is due to the fact that only when occupants are home it is necessary to assure their thermal comfort in the house.

However, as in [3], the variable "internal temperature" is excluded from the characterisation of the archetype because it is not possible to define a unique internal temperature for the entire building stock. The consequences of the omission of this variable is examined in Section 3, where a sensitivity analysis is performed to analyse the impact of the required average internal temperature on archetype energy consumption for heating. As the scope of the current paper is the archetype development to assess the annual final heating energy consumption, which is independent of the heating system used, heating systems are not considered in the archetype characterisation. In addition, DHW systems are not considered, as their use does not affect the space heating demand.

Summarising, the variables identified to characterise heating demand for UK archetypes are:

1. Weekly occupancy schedule
2. Air change rate
3. Wall U-value
4. Ceiling U-value
5. Floor U-value
6. Window U-value
7. Dwelling geometry
8. Dwelling floor area
9. Climate

These nine variables allow the occupancy-related archetypes to be defined in all the main data subsets identified by Corgnati et al. in [2]: form, envelope and operation, as shown in Figure 1.

2.2 Step 2 – Building Stock Segmentation

In the segmentation process, the number of archetype buildings required to represent the building stock under consideration is outlined [1]. First, the segmentation criteria are identified. Then, the segmentation criteria are used to divide the building stock into segments. Each segment is associated with an archetype, the technical characteristics of which are defined in the characterisation step. The segmentation criteria must be identified to provide the archetype characterisation in the following data subsets: form, envelope and operation. Particular attention is given to the development of archetype characteristics in the operation subset.

The segmentation criteria used in the current research are acknowledged to define all the building parameters identified in Section 2.1. The correlation between the identified building parameters and the segmentation criteria is summarised in Figure 1. The segmentation criteria used to define archetypes in the form subset is the dwelling type, which defines the dwelling geometry and the dwelling area. Parameters of the envelope subset (such as envelope U-value and air change rate) are a function of the construction year, which is used as segmentation criteria for the envelope subset. The segmentation criteria applied to define archetypes in the operation subset are the occupancy schedule, which defines the amount of time the house is actively occupied, and the climate zone.

The knowledge of the building construction year and the dwelling type is straight forward if adequate data are available from building stock surveys. The definition of the occupancy schedules requires data on the daily activities of occupants which is not always available. Valuable data on activities are provided by TUSs conducted in most European countries, the results of which are comparable thanks to the harmonisation process supported by the 2008 Harmonised European Time Use Surveys guidelines [33].

Climate zones are defined in accordance to the climate zoning available in the building regulation codes of the country. Meteorological data from the most densely populated city in the climate zone are obtained from Meteonorrm [34], as suggested by [1].

In the UK case study, data to segment the building stock according to dwelling type and construction year are obtained from [33 – 38]. Five dwelling types (flat, bungalow, detached house, semi-detached house and terraced house) and five construction year are identified (Pre-1918, 1919-1964, 1965-1980, 1981-1990, Post-1991) (Table 1).

Occupancy schedules, used as segmentation criteria in the operation subset, are obtained from data collected by the Time Use Survey 2014-2015 (UK 2015 TUS) [39]. 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. 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.

In order to identify the most common occupancy schedules in the building stock, the clustering methodology previously developed in [27, 28] is applied on the available household daily occupancy profiles. 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, three household states are sufficient: (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. 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 using the k-mode clustering technique [25]. The k-mode technique allows for the creation of clusters of daily occupancy profiles which are similar in composition. Each cluster is characterised by a mode, which is obtained considering the most recurrent household state inside the cluster for each time step. The representative occupancy profile for each cluster is identified as the daily occupancy profile inside the cluster which is closest to the mode. The correct number of total clusters is determined using two indices: the root-mean-squared standard deviation (RMSSTD) and R-squared (RS) [26]. The RMSSTD index measures the non-homogeneity of the clusters - if its value decreases, then 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 used with categorical variables [28]. 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 a 'significant knee' [26], and it can be recognised by plotting the RMSSTD and RS indices as a function of the number of the clusters. In the present paper, the RMSSTD and RS indices dependence on the number of clusters is investigated by 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 a reference to represent the percentage variation of the two indices (Figure 2).

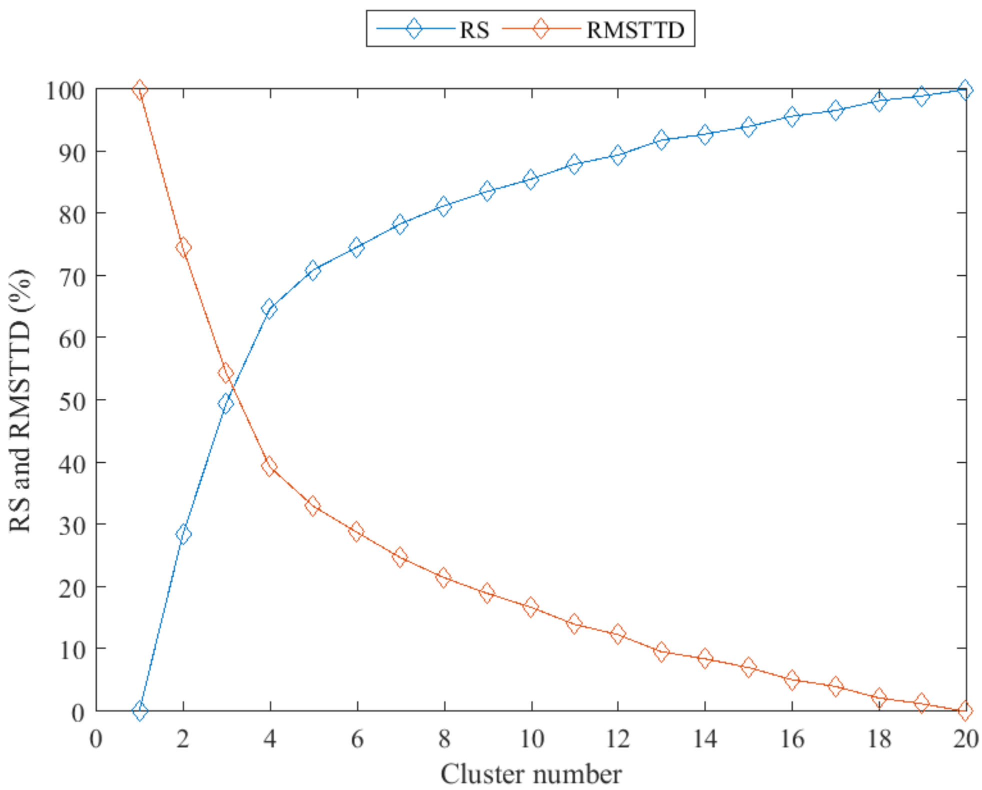

Figure 2 RS and RMSSTD percentage variation as a function of number of clusters per weekday occupancy profiles

Figure 2 indicates the percentage variation of the RS and RMSSTD indexes when the weekday occupancy profiles are clustered. In this figure, the 'significant knees' of the indexes are not easily recognizable. It has been assumed that the minimum and maximum acceptable values are 70% and 40% for RS and RMSTTD, respectively. It is possible to see that by considering five clusters, the RMSSTD index is lower than the assumed threshold and RS index is higher than 70%, indicating that acceptable levels of diversity between clusters and homogeneity within them are reached considering five clusters. Additionally, Figure 2 shows that passing from five to six clusters, the improvement of the quality of the clusters is negligible because the incremental change of both indices (considered as the difference between the two indexes at five or six clusters) is just 1%. This means that the improvement of the quality of the clusters is not sufficient to justify the increased number of final archetypes which are obtained considering more than five clusters. Thus, the appropriate number of clusters is set equal to five.

The modes associated with each cluster are shown in Figure 3, as obtained from the k-mode clustering of the weekday occupancy profiles. The modes are mainly differentiated by the period during which occupants are absent over a day, and they can be classified as follows:

- Daily absence (cluster 1): unoccupied period from 09.00 to 04:00
- Working hour absence (cluster 2): unoccupied period from 08:20 to 18:10
- Lunch time absence (cluster 3): unoccupied period from 11:10 to 16:10
- Constant presence 1 (cluster 4)
- Constant presence 2 (cluster 5)

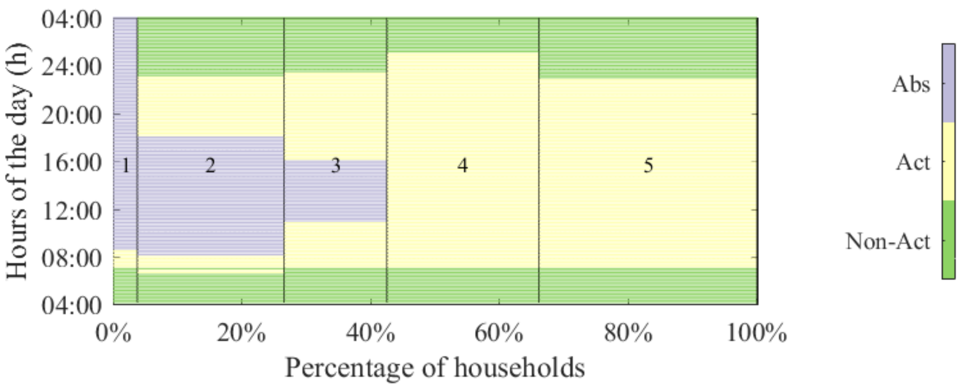

Figure 3 Weekday household daily occupancy mode profiles for each cluster (1 to 5).

Figure 3 illustrates that occupancy profiles can significantly vary amongst buildings, and this, once again, highlights the unsuitability of the use of a standard occupancy profiles in regular archetype characterisation.

Households associated to Cluster 1 have a unique behavioural schedule; they are absent from early morning until late evening. This behaviour is representative of a small percentage of households. A similar behaviour was found in [20], where 10 distinctive behaviour patterns are identified from the American Time Use Survey (ATUS).

Classification of the climate zones in the UK is done based on the climate maps presented by Meteorological Office [40].

The cities which have been chosen to represent the four climate zones identified in UK are: London, Birmingham, Newcastle and Glasgow. These cities have been selected as they have the largest population and consequently the largest number of buildings per selected climate zone [1, 41].

Table 1 summarises the segmentation categories obtained for the national UK building stock. The total number of archetypes is given by the combination of segmentation categories and it is equal to 5^4 or 625.

Table 1 Segmentation categories - UK building stock

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

2.3 Step 3 - Characterisation

In the characterisation step, archetype characteristics are defined according to the segmentation parameters, as defined in Figure 1. The archetype characteristics can be defined as the result of the statistical analysis of the national building sample, in particular, they are the most common properties statistically detected in the building categories identified in the segmentation process [42].

In the UK case study, the building data which were used to characterise archetypes in the form and envelope subsets are identifiable from national surveys [35–38, 43], and are reported in Table 2 and Table 3. From Table 2, it is possible to see, for example, that most of terraced houses built in the 1918-1964 period are characterised by uninsulated solid walls, a loft insulation greater than 150 mm and double-glazed windows. The wall type reflects the construction technique used when the dwelling was built, while the loft and window characteristics denote that most of the terraced houses built between 1918-1964 have been refurbished. These characteristics are in accordance with refurbishment trends as reported in the Energy Consumption in the United Kingdom (2015) data [32].

Table 2 Archetype envelope characteristics

Dwelling type	**Construction period**	**Envelope *characteristics***			
		External Wall [43]	Ground floor [45, 46]	Loft floor insulation (mm) [43]	Window [43]
Terraced	Pre 1918	Uninsulated solid wall pre 1918	Suspended timber	>150	Double glazing
	1918-1964	Uninsulated solid wall pre 1918	Ground bearing concrete floors 60-90	>150	Double glazing
	1965-1980	Uninsulated cavity 1965-1980	Ground bearing concrete floors 60-90	> 150	Double glazing
	1981-1990	Uninsulated cavity 1981-1990	Ground bearing concrete floors 60-90	= 150	Double glazing
	Post 1990	Insulated cavity post 1991	Ground bearing insulated floors	>150	Double glazing
Semi-detached	Pre 1918	Uninsulated solid wall pre 1918	Suspended timber	>150	Double glazing
	1918-1964	Uninsulated cavity 1919-1964	Ground bearing concrete floors 60-90	>150	Double glazing
	1965-1980	Uninsulated cavity 1965-1980	Ground bearing concrete floors 60-90	> 150	Double glazing
	1981-1990	Uninsulated cavity 1981-1990	Ground bearing concrete floors 60-90	= 150	Double glazing
	Post 1990	Insulated cavity post 1991	Ground bearing insulated floors	>150	Double glazing
Detached	Pre 1918	Uninsulated solid wall pre 1918	Suspended timber	>150	Double glazing
	1918-1964	Uninsulated cavity 1919-1964	Ground bearing concrete floors 60-90	>150	Double glazing
	1965-1980	Uninsulated cavity 1965-1980	Ground bearing concrete floors 60-90	> 150	Double glazing
	1981-1990	Uninsulated cavity 1981-1990	Ground bearing concrete floors 60-90	= 150	Double glazing
	Post 1990	Insulated cavity post 1991	Ground bearing insulated floors	>150	Double glazing
Bungalow	Pre 1918	Uninsulated solid wall pre 1918	Suspended timber	>150	Double glazing
	1918-1964	Uninsulated cavity 1919-1964	Ground bearing concrete floors 60-90	>150	Double glazing
	1965-1980	Uninsulated cavity 1965-1980	Ground bearing concrete floors 60-90	> 150	Double glazing
	1981-1990	Uninsulated cavity 1981-1990	Ground bearing concrete floors 60-90	= 150	Double glazing
	Post 1990	Insulated cavity post 1991	Ground bearing insulated floors	>150	Double glazing
Flat	Pre 1918	Uninsulated solid wall pre 1918	-	-	Single glazing
	1918-1964	Uninsulated solid wall pre 1918	-	-	Double glazing
	1965-1980	Uninsulated cavity 1965-1980	-	-	Double glazing
	1981-1990	Uninsulated cavity 1981-1990	-	-	Double glazing
	Post 1990	Insulated cavity post 1991	-	-	Double glazing

Table 3 Archetype characterisation

Dwelling type	**Construction period**	**Shape**	**Envelope *U*-value ($W/m^{-2}K^{-1}$)**				**Envelope Airtightness**
		Building area (m^2) [43]	External Wall [35, 46]	Ground floor [44, 45]	Loft floor [35]	Window [35]	ACH (1/h) [47]
Terraced	Pre 1918	80	2.1	0.8	0.16	3.1	0.56
	1918-1964	80	2.1	0.6	0.16	3.1	0.76
	1965-1980	80	1.3	0.6	0.16	3.1	0.64
	1981-1990	80	0.6	0.6	0.3	3.1	0.64
	Post 1990	80	0.45	0.3	0.16	3.1	0.51
Semi-detached	Pre 1918	90	2.1	0.8	0.16	3.1	0.56
	1918-1964	90	1.6	0.6	0.16	3.1	0.76
	1965-1980	90	1.3	0.6	0.16	3.1	0.64
	1981-1990	90	0.6	0.6	0.3	3.1	0.64
	Post 1990	90	0.45	0.3	0.16	3.1	0.51
Detached	Pre 1918	150	2.1	0.8	0.16	3.1	0.56
	1918-1964	150	1.6	0.6	0.16	3.1	0.76
	1965-1980	150	1.3	0.6	0.16	3.1	0.64
	1981-1990	150	0.6	0.6	0.3	3.1	0.64
	Post 1990	150	0.45	0.3	0.16	3.1	0.51
Bungalow	Pre 1918	73	2.1	0.8	0.16	3.1	0.56
	1918-1964	73	1.6	0.6	0.16	3.1	0.76
	1965-1980	73	1.3	0.6	0.16	3.1	0.64
	1981-1990	73	0.6	0.6	0.3	3.1	0.64
	Post 1990	73	0.45	0.3	0.16	3.1	0.51
Flat	Pre 1918	60	2.1	N/A	N/A	4.8	0.56
	1918-1964	60	2.1	N/A	N/A	3.1	0.76
	1965-1980	60	1.3	N/A	N/A	3.1	0.64
	1981-1990	60	0.6	N/A	N/A	3.1	0.64
	Post 1990	60	0.45	N/A	N/A	3.1	0.51

Weather files corresponding to the different climate zones identified in Section 2.2 are obtained from the Shiny Weather Data [48], a web service making gridded hourly weather data available in times series formats using data from Copernicus Atmosphere Monitoring Service [49]. The utilised data are related to the year 2012.

The weekly occupancy schedules, used to characterise archetypes, are shown in Figure 4. Each row indicates a different weekly occupancy profile (OP1 to OP5). Daily occupancy profiles for working days are obtained in the "building stock segmentation" step (Section 2.2). They are the same during weekdays to incorporate the likelihood that the same household behaviour is repeated during the working days because of daily working routines. 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 schedules. However, to avoid overcomplicated models, it is assumed that the daily occupancy profiles during weekend are the same for all the weekly occupancy profiles and they are equal to the mode obtained by clustering all the non-working household daily occupancy profiles. A similar approach is also used in the BREDEM [31].

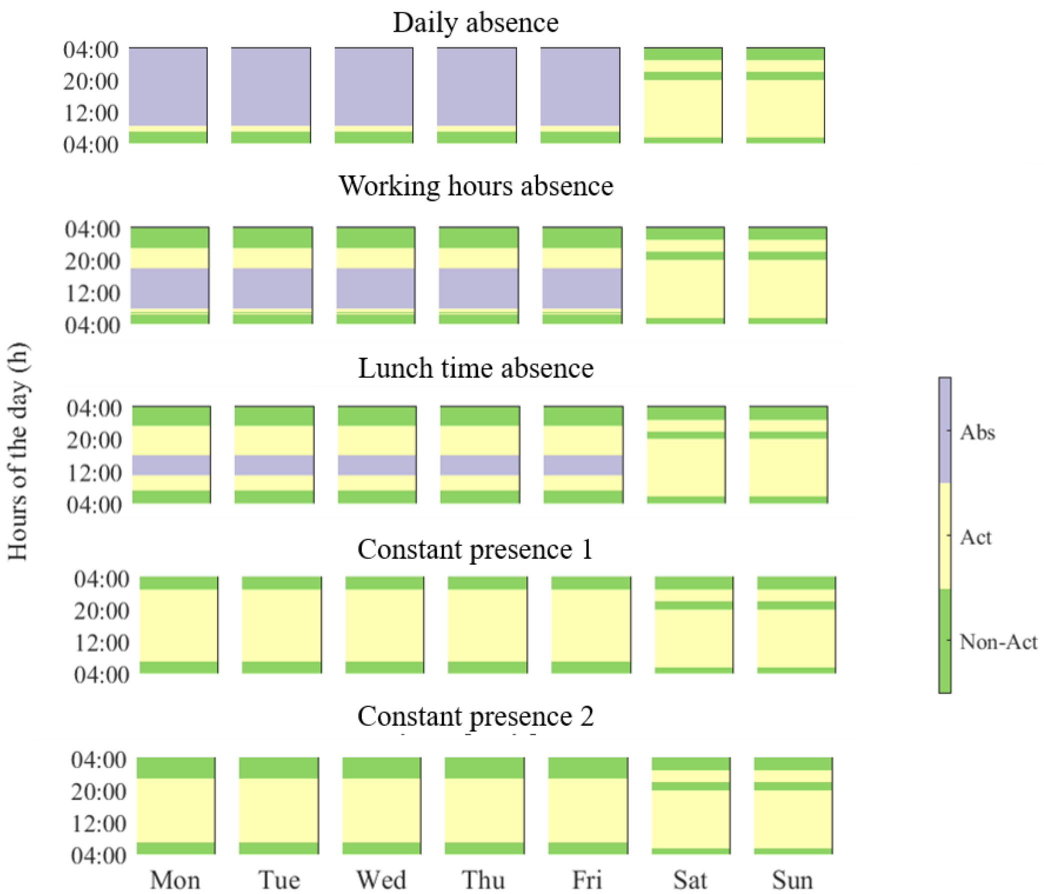

Figure 4 Weekly occupancy profiles (OP1 to OP5).

From Figure 4, it is possible to see that occupancy profiles for working and non-working days are quite different for occupancy profiles 1-3, whereas they are broadly similar for occupancy profiles 4 and 5. These last two profiles represent almost 60% of the national occupancy profiles (Figure 3). The implication of this is that for most of the population, heating patterns do not necessarily follow the general assumption of a bimodal heating pattern for weekdays, and additionally the heating patterns are quite similar for weekdays and weekends. Similar results are found from the analysis of temperature profiles in UK houses [22, 23, 50].

3 Archetype evaluation procedure

The verification of model results is challenging because the developed archetypes do not correspond to building instances, thus it is difficult to assess the accuracy of the energy models. In general, for a single building simulation, differences can be observed between the predicted and the actual energy consumption [51]. However, while it may not be possible to validate the accuracy of predictions in a building stock modelling context [4] as would be done for a specific building instance, it is still crucial to verify whether relevant general trends are captured by the model.

Evaluation of the archetypes is performed by utilising archetype energy models as outlined in Figure 5. The evaluation is based on the energy modelling of each individual archetype and the comparison of the results with available data from national surveys. The annual final heating energy consumption of archetypes is modelled using EnergyPlus (Version 8.5), a whole building energy simulation program which implements detailed building physic models. These models require additional inputs beyond Form/Envelope/Operational data, as shown in Figure 5. These include: heating control settings and internal heat gain profiles.

Additionally, EnergyPlus is based on the assumption that all conduction through the building envelope is one-dimensional, which is quite limiting because it does not consider the thermal bridge effects. For this reason, in the EnergyPlus models equivalent walls are used in order to take into account the additional heat flow per unit length and temperature difference caused by the thermal bridges. The procedure of generation of the equivalent wall is based on the distribution of bridge heat resistive components across the cross section of the wall assembly. Default values of linear thermal transmittance for the junction are obtained from SAP 2016 [35].

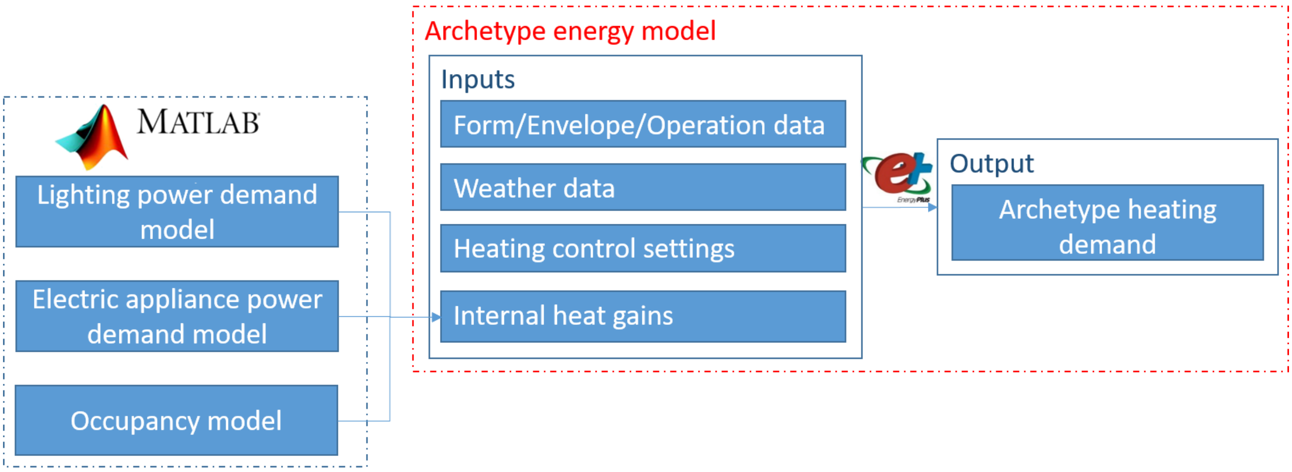

Figure 5 Architecture of archetype energy model.

3.1 Generation of archetype model inputs

3.1.1 Heating control settings

It is assumed that the overall system availability, which describes the time during which the heating system can be used, is continuous from 07:00 until 23:00 hrs However, in order to calculate a realistic annual final heating energy consumption, in the simulation, the heating system is turned on only when the building is occupied to reach the required average internal temperature. The required average internal temperature is considered as the average temperature desired inside the house by its occupants to achieve their comfort, and it is assumed to be constant on time and unvaried in space. In the sensitivity analysis, this variable is varied from 18°C to 21°C, which is the typical indoor temperature range in the UK houses [50], using incremental increases of 0.5°C. For comparison reasons, the heating demand is also calculated considering the standard heating time suggested by the BREDEM model, which is a consolidated model for calculating the energy use and fuel requirements of dwellings in the UK [31]. In this model, the heating patterns are differentiated per day type: 07:00-09:00 and 16:00-23:00 (9 hours) during weekdays and 07:00-23:00 (16 hours) during weekends.

3.1.2 Generation of internal heat gain profiles

Internal heat gains are mainly due to the presence of occupants, electric appliances, lighting and internal solar gains. Internal solar gains are estimated by EnergyPlus during the simulation, while the gains produced by other sources must be calculated in advance and inserted as inputs in the software. The methodologies used to determine the internal gains from the presence of occupants, electric appliances and lighting are outlined in the following three sub-sections.

3.1.2.1 Occupant presence

Internal heat gains correlated to occupant presence can be obtained considering the time during which the dwelling is occupied, the number of occupants in the dwelling and the performed activity. The time during which the dwelling is occupied is available from the weekly occupancy profiles which characterise the archetypes. The occupant number is defined as a function of the dwelling type. The distribution of households, characterised by a different number of occupants for each dwelling type, is available from data collected in the English Housing Survey 2015 [52]. From this data, it is possible to calculate the weighted average number of occupants for each dwelling type. The number of occupants assumed in this research is the closest integer value to the weighted average (Table 4). The same results are obtained when the formula adopted by SAP2012 [53] is used, where the number of occupants is a function of the total floor area of the dwelling.

Table 4 Occupant number according to the dwelling type

Dwelling type	Housing survey 2015	SAP 2012
Terraced	2	2
Semi-detached	3	3
Detached	3	3
Bungalow	2	2
Flat	2	2

In general, the heat generated by the human body is a function of the performed activity and it is assumed to be equal to 0.9 met, if the person is asleep and to 1.5 met if the person is active [54, 55]. These values represent typical metabolic rates per unit area of skin surface. Considering the average adult area equal to 1.8 m^2 (DuBois area), the conversion factor is assumed to be equal to 108 W/met [56]. In the current research, when the household state is “Active”, it is assumed that each household occupant generates 162 W, and 97.2 W in the case the household state is “Non-Active”.

3.1.2.2 Electric appliance internal heat gains

Internal heat gains generated by electric appliances are strictly correlated to the use of these appliances, the number of occupants and the daily occupancy profiles [57]. The internal heat gains generated by electric appliances are a function of the archetype form and operation data subsets, which respectively, determine the number of occupants and the household state during the day. The appliance use, related to the different archetypes, can be simulated after developing a model which combines already existing stochastic models [57 – 59] with the data on daily occupancy available from the archetype characterisation in the operation archetype data subset. In this case a first-order Markov-Chain model is used, as in [57].

The first step for the development of the model is the data collection about the electric appliances installed in UK dwellings. These data are available from the Household Electricity Survey - a study of domestic electrical product usage [60], which presents the electric consumption data associated to 251 households in England over the period May 2010 to July 2011. The use of some appliances is independent from the occupant activity, while other appliances can be randomly used if the occupant is active, thus they are not related to any specific activity.

For each of the appliances, the annual energy demand is measured [60]. The energy consumption correlated to some appliances (e.g., fridges) is strongly seasonal, while for other appliance groups, like audio-visual and computer equipment, or cooking devices, the annual energy consumption depends on the number of occupants. The use of the appliances can be unrelated to the occupant state (e.g., fridges), it can be dependent only on occupancy (e.g., mobile phones), or it can be related to a defined activity (e.g., oven use is linked to the activity "cooking"). Because of the strong correlation between some appliance use and occupant activities, daily activity profiles are developed for each of the weekly occupancy profiles which can characterise archetypes. Activity profiles show the probability that people perform different activities at different times of the day, as in [57]. In the current paper, the probability of a specified activity being performed takes into account the weekly occupancy profiles of the archetype. Figure 6 shows the activity probabilities related to weekly occupancy profile 2 (OP2) “Working hours absence” (Figure 6 (a)) and weekly occupancy profile 5 (OP5) “Constant presence 2” (Figure 6 (b)) over the course of a working day. These curves exhibit significantly different behavioural patterns, because they reflect population subgroups which actively occupy the house during different periods of the day. This differentiation allows the simulation of the electric demand of specific population subgroups, represented by each archetype.

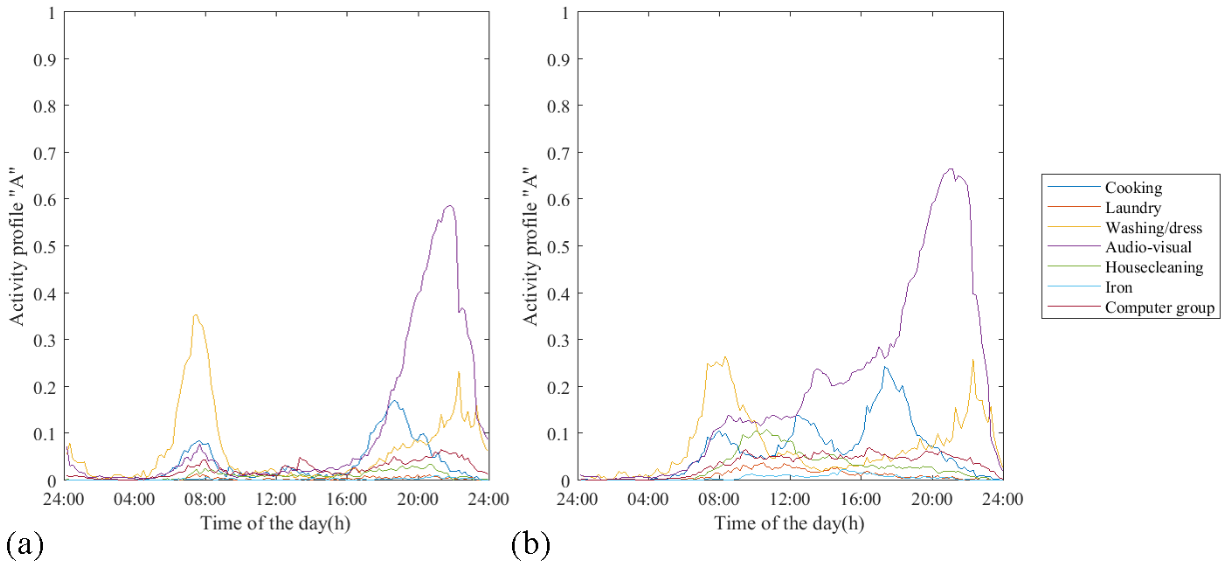

Figure 6 Activity profiles (weekdays) for (a) occupancy profile 2 (OP 2) and (b) occupancy profile 5 (OP 5).

In the current model, the switch-on probability, P_a, of any appliance, a, for any time step, t, is given by the following equation:

$$P_a(t) = S(t)\, A_i(o,t)\, C_a(m,n,o) \quad (1)$$

where: $S(t)$ is a binary variable equal to 1 if the household state is "Act"; $C_a(m,n,o)$ is the calibration scalar; and $A(o,t)$ is the activity probability itself, which is dependent on the occupancy profile, o, and time, t. The calibration scalar is used to calibrate the number of switch-on events based on appliance data. This calibration scalar is dependent on the month, m, to consider the seasonal effect, the number of occupants, n, to consider the household size effect, and it is also dependent on the occupancy profile, o. The calibration scalar is calculated such that, over a very large number of stochastic simulation runs, the mean annual consumption of the appliance will correspond to the one presented in the data available from the Household Electricity Survey [60]. For the case where the use of the appliance is not a function of a specific activity $A_i(o,t) = 1$. Additionally, if the use of the appliance is not dependent on the presence of an active occupant, $S(t)$ is also equal to 1. In this case, the final probability is equal to the calibration scalar. In order to determine if the switch-on event occurs, a random number between 0 and 1 is generated at each time step. The switch-on event occurs if the random number is smaller than the probability associated with the appliance at a specific time step. If the switch-on event occurs, the power demand of the appliance is set equal to the assigned power. The internal heat gains generated by each appliance are then calculated as a fraction of the power used by the electric appliance [61]. The total internal heat gains, generated by electric appliances in a dwelling, is the sum of the internal heat gains generated by each appliance for each time step.

This model produces different results every time it is run because of its intrinsic stochasticity, but a characteristic profile must be used in the archetype model. For this reason, the average value which can be obtained by multiple runs of this model is calculated. The minimum number of households, which was needed to be simulated, in order to obtain the average demand (for electric equipment), was found to be the one which allows a percent relative standard deviation (PRSD) [62] of the internal heat gains lower than the limit threshold to be obtained, as in [63]. The PRSD threshold is assumed to be equal to 5%. The minimum number of households is not the same for all the clusters because it is a function of activities performed by the household occupants. When clusters are more homogenous (i.e., the household performs the same activities at the same time), the number of households required to obtain the average internal heat gain is lower than when the clusters are less homogenous. For example, the minimum number is equal to 115 households for operation profile 2, whereas it is equal to 215 households for occupancy profile 5. The PRSD is calculated as per Equation (2);

$$\max\big(PRSD(i)\big) = \max\left(\frac{1}{\overline{O_{av,i}}}\sqrt{\frac{\sigma_i^2}{n \cdot 100}}\right) \tag{2}$$

where i is the time step, n is the number of households considered, which is incremented in each iteration, and σ_i^2 is the variance the timestep i. $\overline{O_{av,i}}$ is the average internal heat gain at the time step i.

3.1.2.3 Lighting internal heat gains

The use of electric lighting in the domestic sector depends mainly on the level of available natural light, coupled with the activity of the household residents. The model presented is largely based on the model developed by Richardson et al. (2009) [64]. In this model, the switch-on probability of lights in a dwelling, P_l, for any time step, t, is given by Equation (3);

$$P_l(t) = S(t)\, Irr_i(t) \tag{3}$$

where: $S(t)$ is a binary variable equal to 1 if the household state is "Act"; and $Irr_i(t)$ is a binary variable which value is the result of the natural light condition test. In the natural light condition test, first the dwelling irradiance threshold is randomly determined from a normal distribution characterised by a mean of 60 W/m² and a standard deviation of 10 W/m² [64]. Then, the irradiance threshold is compared to the current level of outdoor irradiance at each time step. If the current irradiance is below the threshold, then the resulting value of this test is 1, otherwise it is 0. Moreover, the model also allows for a five percent likelihood that a lighting unit may be used regardless of the current natural light conditions, to represent the unpredictable daytime use of lighting [64].

If a switch-on event occurs, electric power is consumed by lighting. The consumed power is a function of the installed technology. The installed technology is randomly selected considering the national share of each type of lighting for which data are obtained from the Household Electricity Survey [60]. The power consumed when a switch-on event occurs is randomly determined from a normal distribution, characterised by a mean value equal to the average installed power of the technology. In addition, the duration of the switch-on event is randomly determined from the distribution introduced by Richardson et al. in [64]. If the chosen duration is longer than the prevailing period of active occupancy in a dwelling, then the duration is truncated at the time when the active occupancy becomes zero. The consumed power by lights is then converted into internal heat gains calculated as in [61]. Finally, for this case, an average scenario must be used as input in the archetype model. The average value is obtained running the model several times, in order to obtain also, in this case, a percent relative standard deviation of the internal heat gains produced by lighting lower than 5%.

4 Results

In this section the annual final heating energy obtained as a result of the archetype modelling is presented, and it is compared with data about the energy consumption of dwellings in the relevant country to evaluate the archetypes.

Readily available data about consumed primary energy demand for heating in the UK is necessary to validate the developed archetypes. Moreover, in order to validate the individual archetypes, these data should be differentiated for different building types, construction periods, occupancy profiles and location.

The energy use and energy efficiency of domestic buildings in England and Wales are studied by National Energy Efficiency Data-Framework (NEED) [32], which collects gas and electricity consumption data on a representative sample of data. The energy consumption statistics which are available from NEED are particularly relevant because the energy consumption is classified by property attributes, household characteristics, geography and socio-demographic classifications. In the context of this research, the data which can be used to validate the archetype energy models is the annual median consumption of gas in 2012, demarcating for buildings differentiated by property type and construction age. According to statistical data available from the Energy Consumption in the United Kingdom (ECUK), the gas consumed for space heating in residential dwellings is about 85% of the total gas used is the residential sector [65], so a proportionate quantity considered from the total gas consumption available from NEED data is utilised. The reduced gas consumption, in kWh, can be converted to final energy consumption through a conversion factor equal to 1.127 [35]. The resulting quantity can be considered as the closest value to the annual final heating energy required for space heating, which can be obtained from available UK datasets. This value can be compared to the annual final heating energy consumption of archetypes characterised by different dwelling types, construction periods and occupancy profiles, considering also the variation of required average internal temperature.

In order to avoid redundant discussions, just results of the climate location "London" are presented in detail.

However, the variations of the results due to the archetype location is analysed for the archetypes identified in dwelling type "flat".

4.1 Sensitivity analysis and archetype evaluation

In general, a sensitivity analysis approach is capable of determining the effect of a building design variable on its overall performance [51]. In this work, a sensitivity analysis is used to assess the effect of the required average internal temperature on the annual final heating energy consumption. This variable is one of the parameters which has the strongest impact on building space heating energy demand [3], but it also has high variability in UK building stock [22], thus the assumption of a single internal heating temperature is not realistic. For this reason, a sensitivity analysis is performed on this parameter.

This is undertaken by a one-parameter-at-a-time (OAT) method, where the individual effect of the design parameter "average internal temperature" on the building performance is evaluated [66]. The average internal temperature is varied between 18°C and 21°C, considering incremental increases of 0.5 °C, which is the acceptable range in UK [23, 56]. To determine the design parameter sensitivity, the "sensitivity index" (SI) related to the set point temperature is calculated as;

$$SI = \frac{E_{max} - E_{min}}{E_{max}} \; 100\% \quad (4)$$

where E_{max} and E_{min} represent the maximum and minimum energy requested for heating, respectively, resulting from varying the design parameter over its entire range [66]. Values of Sensitivity Indexes obtained for the different occupancy-integrated archetypes located in London (as identified in Sect. 2) are indicated in Figure 7.

Analysing Figure 7, it is possible to see that the Sensitivity Index has a range between 35 and 60%, indicating that the required average internal temperature has a significant impact on the energy consumed for heating. This implies that the choice of a common required average internal temperature for all archetypes can bring significant errors in the estimation of energy consumption for heating, and thus it is necessary to collect accurate data about internal temperature distribution to obtain accurate estimation of energy consumed for heating in archetypes. As the required average internal temperature must be reached just when the building is occupied, the magnitude of the sensitivity index is proportionally correlated to the amount of time the household state of the building is active, which is a function of the weekly occupancy profiles.

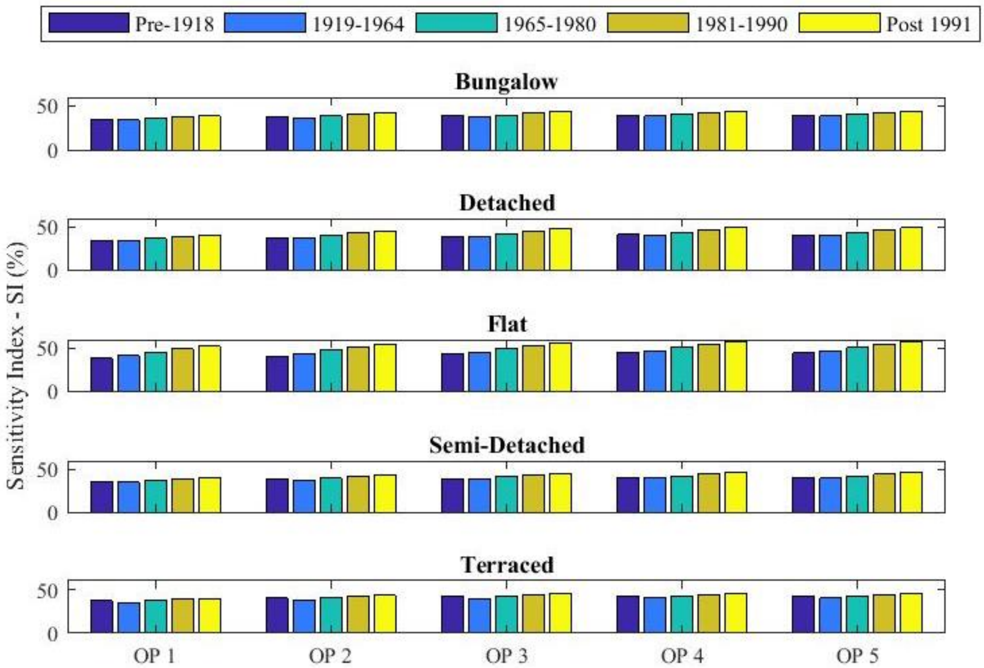

Figure 7 Sensitivity Index (%) for each occupancy-integrated archetype (OP1 to OP5) in the required average internal temperature range 18-21°C (London climate zone)

The sensitivity index is also a function of the climate zone. In Figure 8, the indexes obtained for archetypes characterized by the dwelling type "flat" are shown for different climate zones. The same trends registered in Figure 7 are visible, although the values of the index are lower in colder area such as Glasgow.

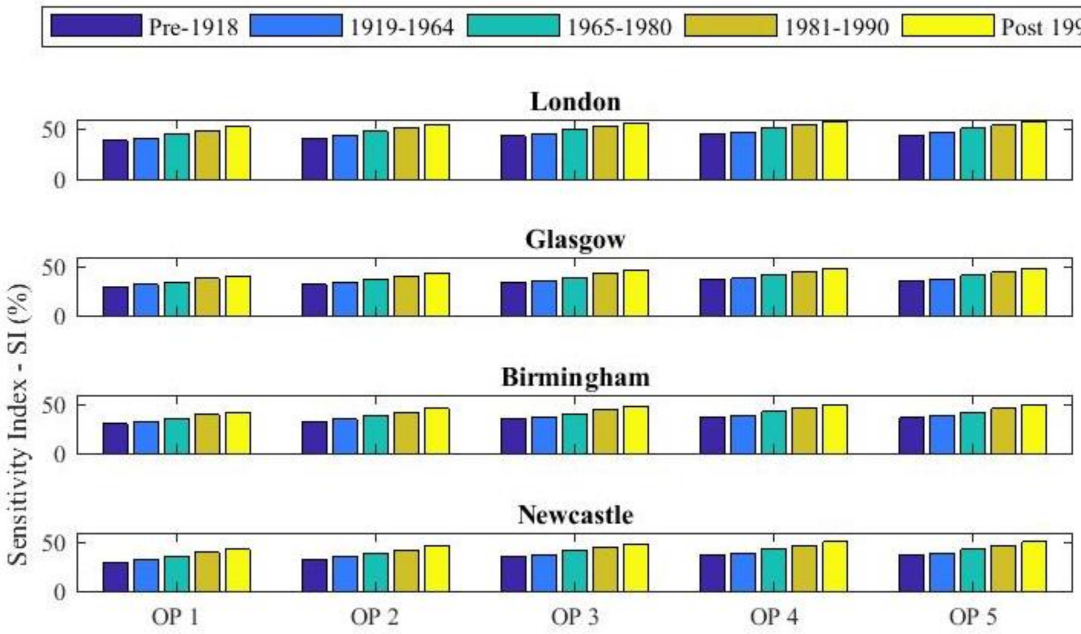

Figure 8 Sensitivity Index (%) for occupancy-integrated archetype (OP1 to OP5) characterised by dwelling type "flat" in different climate zones in the required average internal temperature range 18-21°C

The annual final heating energy consumption which is obtained from energy models of the dwelling type "flat" located in the "London climate zone" is indicated in Figure 9, as a function of the required average internal temperature. The annual final heating energy consumption is differentiated for the different weekly occupancy profiles (OP 1 to OP 5). From this figure, it is clear that the consideration of different weekly profiles has a strong impact on the final model results. The resulting final heating energy required using the standard heating hours indicated in BREDEM (E_{BREDEM}) is also indicated [31], and it is possible to observe that the use of standard heating hours leads to underestimation or overestimation of the final heating energy required, if compared to the energy required by archetypes characterised by specific occupancy profiles. E_{BREDEM} can be directly compared to the average annual final heating energy consumption, calculated as the weighted average value of the heating demand corresponding to the five weekly occupancy profiles ($E_{\overline{OP}}$). It is interesting to note that the average value is extremely close to the value which can be obtained using the standard hours indicated in BREDEM. The percentage difference (R1) is calculated using equation (5) and is never larger than 5% (Figure 11) for archetypes located in the climate zone of London. The results obtained for archetypes located in other zones are almost identical.

$$R1 = \frac{E_{\overline{OP}} - E_{BREDEM}}{E_{BREDEM}}\ x\ 100 \qquad (5)$$

This indicates that the adoption of standard heating hours is justifiable in building energy models at national level because the behavioural variations among the individual houses tend to even out when considered at scale. However, when national population subgroups are simulated, the consideration of specific occupancy profiles is critical to maximise accuracy in the estimation of annual final heating energy consumption in buildings [11, 20, 24]. Thus, in order to use the archetype approach in the context of disaggregated and differentiated population subgroups, it is important to include occupant-related variables as residential archetype operational inputs. From the analysis of Figure 9 it is possible to see that the relative difference between the heat demand calculated considered specific occupancy profiles and the one calculated considered BREDEM occupancy profiles can be up to 30% for the flat located in London. Similar results are obtained for archetypes characterised by different forms.

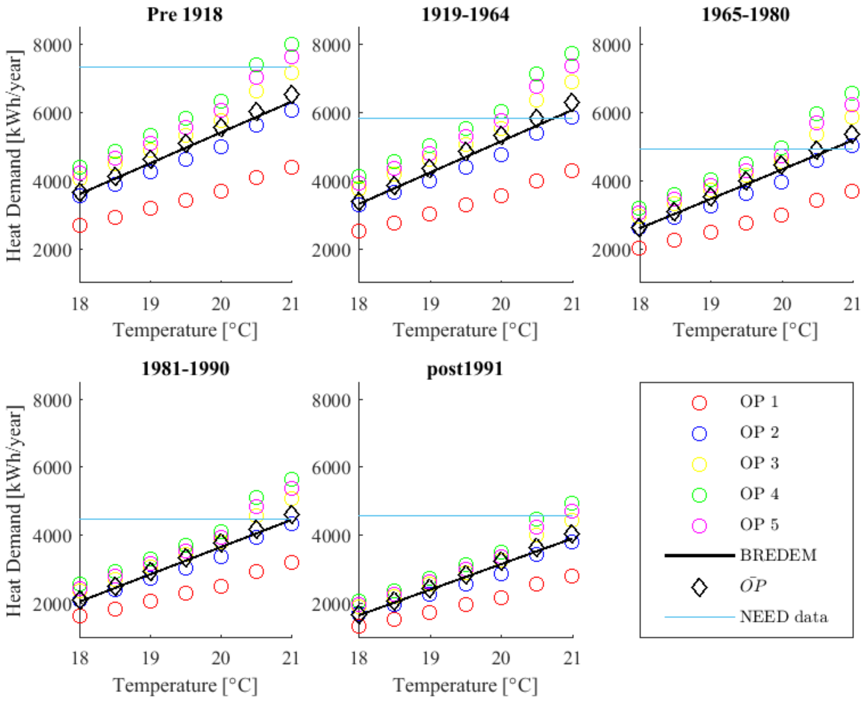

Figure 9 Annual final heating energy demand for flat archetype for different construction periods and different occupancy profiles (London climate zone)

Results obtained for different climate zone are given in Figure 10, where the annual final heating energy demand for the archetypes characterised by dwelling type "flat" and construction year "post 1991" is shown. Compared to London, the annual final heating energy demand is higher for all the other climate zones, with Glasgow being the highest.

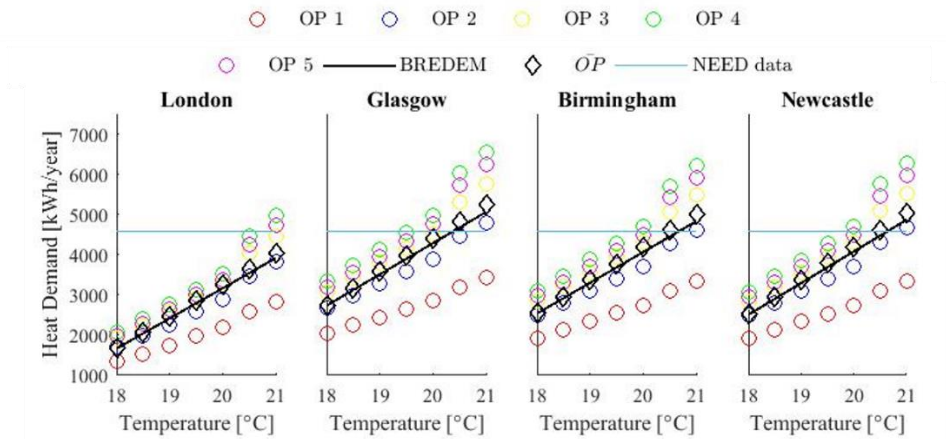

Figure 10 Annual final heating energy demand for different climate zone (archetype characterised by dwelling type "flat" and construction period "post 1991")

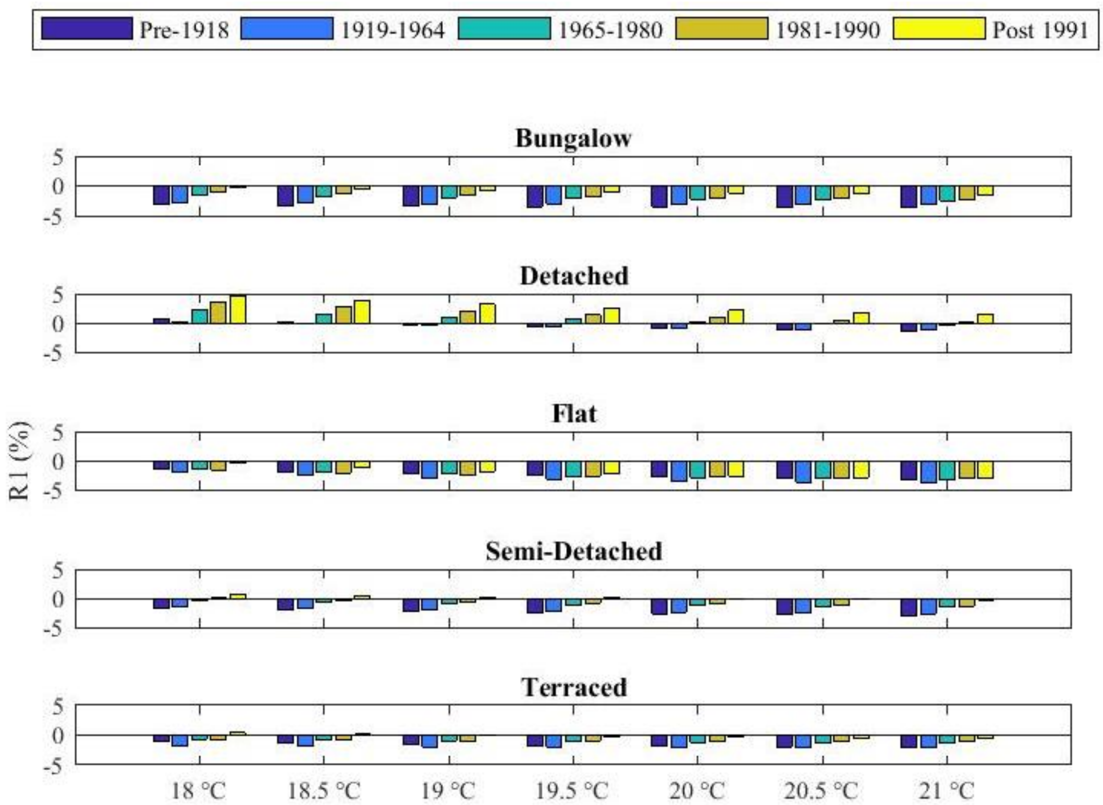

Figure 11 Percentage difference (R1) (Equation 5) between average annual final heating energy demand in occupancy-integrated archetypes ($E_{\overline{OP}}$) and annual final heating energy demand calculated using standard heating hours (E_{BREDEM}) as a function of required average internal temperature (London climate zone)

The percentage deviation (R2) of the average value of annual final heating energy consumption ($E_{\overline{OP}}$), obtained by simulating the archetypes considering the different weekly occupancy profiles, and the annual final heating energy consumption (E_{NEED}) available from NEED data [32] is calculated as:

$$R2 = \frac{E_{\overline{OP}} - E_{NEED}}{E_{NEED}} \; x \; 100 \tag{6}$$

R2 values for different average internal temperatures are indicated in Figure 12, as a function of the dwelling type and the construction year. Closer inspection shows that the R2 values are largely influenced by the internal comfort temperature. The difference between $E_{\overline{OP}}$ and E_{NEED} changes also according to the climate zone, as shown in Figure 13, where the R2 values for the archetypes characterized by the dwelling type "flat" are indicated for different locations.

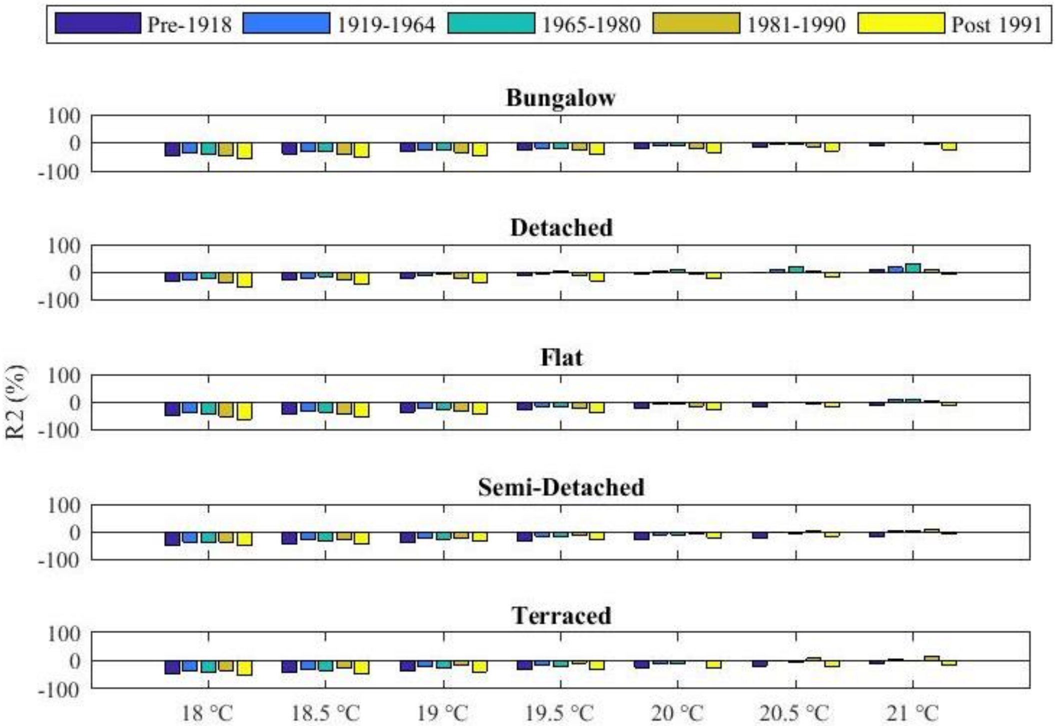

Figure 12 Percentage difference (R2) (Equation 6) between average annual final heating energy demand in occupancy-integrated archetypes ($E_{\overline{OP}}$) and annual final heating energy demand extrapolated from NEED data (E_{NEED}) as a function of required average internal temperature (London climate zone).

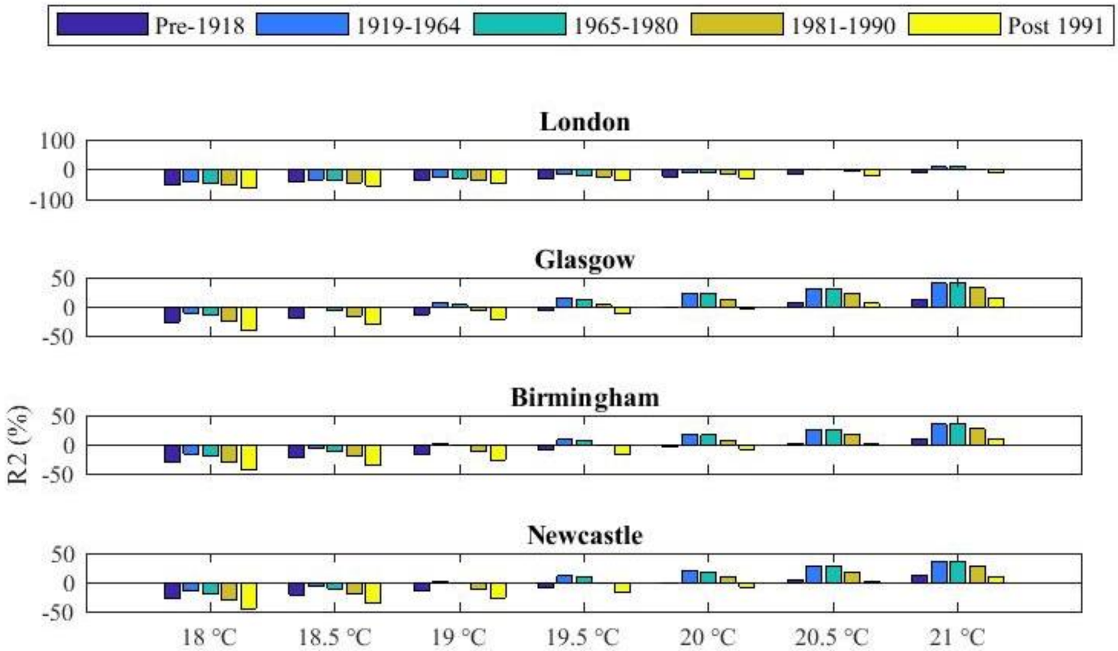

Figure 13 Percentage difference (R2) (Equation 6) between average annual final heating energy demand in occupancy-integrated archetypes ($E_{\overline{OP}}$) and annual final heating energy demand extrapolated from NEED data (E_{NEED}) as a function of required average internal temperature for different climate zones (flat dwelling type).

5 Discussion

5.1 Application for the research

The current paper presents a methodology to define occupancy-integrated building archetypes for building stock energy simulation, extending the definition of archetypes for stock modelling by including occupancy types.

The description of a building stock by means of archetypes is widely used when the collection of accurate characteristics of individual buildings is not feasible or the computational cost of modelling individual buildings is too high. The energy consumption estimates of modelled archetypes can scaled up to be representative of building stock of variable size (neighbourhood, regional or national stock) by multiplying the energy demand of the archetypes by the number of houses which fit the description of each archetype.

The use of occupancy-integrated archetypes allows the assumption that one standard pattern fits all homes to be addressed and associates houses to different archetypes by considering the occupancy profiles as well. In this way, even small national population subgroups, which are characterised by occupancy profiles which do not correspond to the national average can be modelled. In this case, the overall building stock space heating demand predictions are significantly different, if the proposed occupancy-integrated archetype models are used. The application of occupancy-integrated archetypes improves the diversification of the simulated final heating energy demand required by different household types, which cannot be achieved when a single heating profile is used in all archetypes.

The aforementioned methodology has been applied to the UK building stock, but it can be readily be replicated for other building stock. The main challenge of applying the methodology to other housing stock is the availability of necessary data on building characteristics and daily activities of occupants.

The energy model of the archetypes can contribute to the development of energy policies based on improved building energy predictions. A deeper knowledge of the segmentation of occupancy profiles within the building stock can help in the assessment of the most effective building fabric retrofits for different population subgroups.

Additionally, the presented archetypes may be used to investigate the effects of the market penetration of different space heating technologies. For example, dwellings which are almost constantly occupied over the course of the day might be more suitable for heat pumps that are most efficient when delivering constant background heat.

5.2 Limitations and future work

One of the main limitations of this methodology is that it is based on the assumption that the heating schedule perfectly matches the period of occupancy of the buildings and that the required average internal temperature is constant. This assumption is established on the expectation that occupants seek to maintain comfort by turning on the heating system whenever the indoor temperature is lower than the required average internal temperature. Although the interaction with the heating system is the most common reaction to limit the thermal discomfort, occupants can react also in other ways to adjust their thermal conditions (e.g., adaptive comfort) [68]. Additionally, multiple different internal temperatures could be required in different areas of the house, causing a spatial variation of the internal temperature. Although these variables have an impact on the final heating load of buildings, they are not included in the archetype characterisation as they are case-specific. The aim of the methodology is the development of archetypes which are representative of a wide group of buildings, and as such they cannot replicate all the characteristics of specific individual buildings. In order to overcome these limitations, additional studies need to be undertaken to improve the correlation between occupancy profiles and heating schedules.

Additionally, the archetype methodology has limitations in the use of fixed heating patterns. The use of these fixed patterns produces reliable results when the overall building stock heating energy demand is modelled, but produces unrealistic peak demand for heating when the model is used to obtain high temporal resolution energy demand profiles [27]. These profiles are necessary, for example, in the development of energy-flexible buildings capable of adjusting their electricity demand according to the needs of the power grid.

A logical extension to the outlined methodology would involve the integration of a stochastic occupancy model into the occupancy-integrated archetypes to capture the variability of human behaviour and avoid unrealistic peak demands in high-resolution energy heating models.

6 Conclusions

The current paper presents a new methodology to develop occupancy-integrated archetypes, which allows the annual final heating energy required by building stock characterised by different occupancy profiles to be modelled with better discrimination compared to the use of archetypes which do not include occupancy profiles in their characterisation. Occupancy profiles play a significant role in influencing heat demand in residential buildings. In this paper, it is shown that for UK residential archetypes, the discrepancy between the heat demand calculated using the proposed occupancy-integrated archetypes and BREDEM calculation procedures can be up to 30%. This means that the use of BREDEM occupancy profiles is not necessarily appropriate, particularly when energy profiles of disaggregated and differentiated national population subgroups is required.

The sensitivity analysis on the influence of required average internal temperature on energy consumption shows that this parameter has a strong impact on the result of the building energy models, with values of Sensitivity Indexes that can reach 60%. The application of occupancy-integrated archetypes has the potential to improve the diversification of annual final heating energy demand predictions for different household types, which cannot be achieved when a single heating profile is used in all archetypes. This can be particularly useful when building stock includes population subgroups characterised by a different behaviour than the national average. Although the use of occupancy-integrated archetypes has the aforementioned advantages, further work is required to include the natural stochasticity of occupancy patterns, which would be required for energy demand profiles at high temporal resolution.

Acknowledgements

This project has received funding from the European Union Horizon 2020 Research and Innovation Programme under grant agreement No 646116 (Real Value).

References

[1] 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.

[2] S. P. Corgnati, E. Fabrizio, M. Filippi, and V. Monetti, "Reference buildings for cost optimal analysis: Method of definition and application," Appl. Energy, vol. 102, pp. 983–993, 2013.

[3] 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.

[4] M. Brøgger and K. B. Wittchen, "Estimating the energy-saving potential in national building stocks – A methodology review," Renew. Sustain. Energy Rev., vol. 82, no. 1, pp. 1489–1496, 2018.

[5] European Commission, "2050 Energy Strategy", 2018 [Online]. Available: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2050-energy-strategy. [Accessed: 12-Jul- 2018].

[6] European Commission, "2030 Energy Strategy", 2018 [Online]. Available: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2030-energy-strategy. [Accessed: 12-Jul- 2018].

[7] European Commission, “2020 Energy Strategy”, 2018 [Online]. Available: https://ec.europa.eu/energy/en/topics/energy-strategy-and-energy-union/2020-energy-strategy. [Accessed: 12-Jul- 2018].

[8] Eurostat, “Final Energy Consumption by Sector”, 2017 [Online]. Available: http://ec.europa.eu/eurostat/tgm/table.do?tab=table&init=1&plugin=1&language=en&pcode=tsdpc320. [Accessed: 12-Jul-2018].

[9] S. Ø. Jensen, M. Henrik, R. Lopes, R. G. Junker, D. Aelenei, R. Li, S. Metzger, K. B. Lindberg, A. J. Marszal, M. Kummert, B. Bayles, E. Mlecnik, R. Lollini, and W. Pasut, “Annex 67: Energy Flexible Buildings - Energy Flexibility as a key asset in a smart building future Contribution” , 2017.

[10] RealValue, “Realising value. From electricity markets with local smart electric thermal storage technology,” Tech. rep., 2015, grant agreement No 646116.

[11] V. Cheng and K. Steemers, “Modelling domestic energy consumption at district scale: A tool to support national and local energy policies,” Environ. Model. Softw., vol. 26, no. 10, pp. 1186–1198, 2011.

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] P. Caputo, G. Costa, and S. Ferrari, “A supporting method for defining energy strategies in the building sector at urban scale,” Energy Policy, vol. 55, pp. 261–270, 2013.

[18] W. J. N. Turner, O. Kinnane, and B. Basu, “Demand-side characterization of the Smart City for energy modelling,” Energy Procedia, vol. 62, pp. 160–169, 2014.

[19] D. Yan, W. O’Brien, T. Hong, X. Feng, H. Burak Gunay, F. Tahmasebi, and A. Mahdavi, “Occupant behavior modeling for building performance simulation: Current state and future challenges,” Energy Build., vol. 107, pp. 264–278, 2015.

[20] 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.

[21] S. K. Firth, K. J. Lomas, and A. J. Wright, “Targeting household energy-efficiency measures using sensitivity analysis.,” Build. Res. Inf., vol. 38, no. 1, pp. 25–41, 2010.

[22] G. M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, and A. Summerfield, "The reality of English living rooms - A comparison of internal temperatures against common model assumptions," Energy Build., vol. 66, pp. 688–696, 2013.

[23] G. M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, and A. J. Summerfield, "The shape of warmth: temperature profiles in living rooms," Build. Res. Inf., vol. 43, no. 2, pp. 185–196, 2015.

[24] 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.

[25] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," ACM Comput. Surv., vol. 31, no. 3, pp. 264–323, 1999.

[26] M. Halkidi, Y. Batistakis, and M. Vazirgiannis, "Clustering algorithms and validity measures," Proc. Thirteen. Int. Conf. Sci. Stat. Database Manag. SSDBM 2001, pp. 3–22, 2001.

[27] 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.

[28] G. Buttitta, W. J. N. Turner, and D. Finn, "Clustering of Household Occupancy Profiles for Archetype Building Models," Energy Procedia, vol. 111, pp. 161–170, 2017.

[29] G. Buttitta, O. Neu, W. J. N. Turner, and D. Finn, "Modelling Household Occupancy Profiles using Data Mining Clustering Techniques on Time Use Data" in IBPSA Building Simulation 2017, 2017.

[30] 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.

[31] B. Anderson, P. F. Chapman, N. G. Cutland, C. M. Dickson, S. M. Doran, G. Henderson, J. Henderson, L. Kosmina, and L. D. Shorrock, "BREDEM-8 Model Description 2001 update," p. 98, 2008.

[32] Department for Business Energy & Industrial Strategy, "National energy efficiency data-framework. Summary of analysis using the National Energy Efficiency Data Framework (NEED)," 2017.

[33] Eurostat, "Harmonised European time use surveys - 2008 guidelines", 2008.

[34] Meteotest, "Meteonorm version 6, global meteorological database for engineers, planners and education." Bern, Switzerland, 2009.

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

[36] NHBC, "Homes through the decades", 2015.

[37] J. Palmer and I. Cooper, "United Kingdom Housing Energy Fact File", Dep. Energy Clim. Chang., 2013.

[38] University of the West of England, "Evolution of building elements," 2011. [Online]. Available: https://fet.uwe.ac.uk/conweb/house_ages/elements/print.htm. [Accessed: 13-Jan-2017].

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

[40] Met Office, “Mean daily maximum temperature - Winter average: 1981-2010.” [Online]. Available: https://www.metoffice.gov.uk/public/weather/climate#averagesMaps. [Accessed: 23-Jan-2019].

[41] R. Arababadi, “Energy use in the EU building stock – case study: UK,” Master dissertation, Linkoping University, 2012.

[42] 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.

[43] DCLG, “English Housing Survey”, 2016.

[44] D. Johnston, “A physically-based energy and carbon dioxide emissions model of the UK housing stock”, Doctoral dissertation, Leeds Metropolitan University, 2003.

[45] HM Government, “L1A Conservation of fuel and power in new dwellings – The Building Regulation”, 2013 edition incorporating 2016 amendments.

[46] HM Government, “L1B Conservation of fuel and power in existing dwellings - The Building Regulations 2010”, 2010 edition incorporating 2010, 2011, 2013 and 2016 amendments.

[47] IP1/00 “Air Tightness in UK Dwellings”, BRE Environmental Engineering Centre.

[48] L. Lundström, “Weather data for building simulation. New actual weather files for North Europe combining observed weather and modelled solar radiation”, School of Sustainable Development of Society and Technology, Master program in Energy Optimization for Buildings, 2012.

[49] European Centre for Medium-Range Weather Forecasts, “Copernicus Atmosphere Monitoring Service (CAMS),” 2018. [Online]. Available: http://www.soda-pro.com/web-services/radiation/cams-radiation-service/info#versions. [Accessed: 23-Jan-2019].

[50] G. M. Huebner, M. McMichael, D. Shipworth, M. Shipworth, M. Durand-Daubin, and A. Summerfield, “Heating patterns in English homes: Comparing results from a national survey against common model assumptions,” Build. Environ., vol. 70, pp. 298–305, 2013.

[51] A. Ioannou and L. C. M. Itard, “Energy performance and comfort in residential buildings: Sensitivity for building parameters and occupancy,” Energy Build., vol. 92, pp. 216–233, 2015.

[52] Department for Communities and Local Government, “English Housing Survey, 2015: Housing Stock Data. [data collection].” UK Data Service. SN: 8186, http://doi.org/10.5255/UKDA-SN-8186-1, 2017.

[53] BRE, “SAP 2012 The Government ’ s Standard Assessment Procedure for Energy Rating of Dwellings,” 2012.

[54] M. Mansoubi, N. Pearson, S. A. Clemes, S. J. Biddle, D. H. Bodicoat, K. Tolfrey, C. L. Edwardson, and T. Yates, “Energy expenditure during common sitting and standing tasks: examining the 1.5 MET definition of sedentary behaviour”, BMC Public Health, vol. 15, no. 1, p. 516, 2015.

[55] B. E. Ainsworth, W. L. Haskell, M. C. Whitt, M. L. Irwin, A. M. Swartz, S. J. Strath, W. L. O’Brien, D. R. Bassett, K. H. Schmitz, P. O. Emplaincourt, D. R. Jacobs, and A. S. Leon, “Compendium of Physical Activities: an update of activity codes and MET intensities,” Med. Sci. Sport. Exerc., vol. 32, Supplement, pp. 498–516, 2000.

[56] ASHRAE (American Society of Heating Refrigerating and Air-Conditioning Engineers), "Thermal Environmental Conditions for Human Occupancy", 2013.

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

[58] G. Flett and N. Kelly, "A disaggregated, probabilistic, high resolution method for assessment of domestic occupancy and electrical demand," Energy Build., vol. 140, pp. 171–187, 2017.

[59] E. McKenna and M. Thomson, "High-resolution stochastic integrated thermal-electrical domestic demand model," Appl. Energy, vol. 165, pp. 445–461, 2016.

[60] J.-P. Zimmermann, M. Evans, J. Griggs, N. King, L. Harding, P. Roberts, and C. Evans, "Household Electricity Survey: A study of domestic electrical product usage," 2012.

[61] O. Neu, "Assessment of the electrical flexibility resource of residential building stocks using archetypes", Doctoral dissertation, University College Dublin, 2016.

[62] B. S. Everitt and A. Skrondal, "The Cambridge Dictionary of Statistics", vol. 53, no. 9. 2010.

[63] M. R. Driels and Y. S. Shin, "Determining the number of iterations for Monte Carlo simulations of weapon effectiveness," Naval Postgraduate School, 2004.

[64] I. Richardson, M. Thomson, D. Infield, and A. Delahunty, "Domestic lighting: A high-resolution energy demand model," Energy Build., vol. 41, no. 7, pp. 781–789, 2009.

[65] Department of Energy & Climate Change, "Energy Consumpion in the UK (2015) - Chapter 3: Domestic energy consumption in the UK between 1970 and 2014", 2015.

[66] P. Heiselberg, H. Brohus, A. Hesselholt, H. Rasmussen, E. Seinre, and S. Thomas, "Application of sensitivity analysis in design of sustainable buildings," Renew. Energy, vol. 34, no. 9, pp. 2030–2036, 2009.

[67] R. V. Jones, A. Fuertes, C. Boomsma, and S. Pahl, "Space heating preferences in UK social housing: A socio-technical household survey combined with building audits," Energy Build., vol. 127, pp. 382–398, 2015.

[68] C. Hanmer, M. Shipworth, D. Shipworth, and E. Carter, "How household thermal routines shape UK home heating demand patterns," Energy Effic., vol. 5, pp. 5–17, 2019.

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